a
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

Comparison of K-Means and Hierarchical Clustering Methods in Mapping Socioeconomic Conditions of Regencies and Municipalities in Papua Provinces

Anggy Yulistya Putri [1*], and Radian Januari Situmeaang [2]

[1, 2,] Department of Statistics, Faculty of Mathematics and Natural Sciences, Cenderawasih University, Jayapura, Indonesia

[*]**Correspondance author**: gyyvat@gmail.com

ABSTRACT

This study aims to classify regencies and municipalities in Papua Province based on their socioeconomic conditions using the K-Means and Hierarchical Clustering methods, as well as to determine the clustering method that provides the best performance. The data used in this study consist of socioeconomic indicators of regencies and municipalities in Papua Province obtained from publications of Statistics Indonesia (BPS). The analysis procedures include data collection, data standardization, implementation of K-Means and Hierarchical Clustering, and evaluation of clustering results using the Silhouette Score. The results indicate that both methods are capable of grouping regions according to their socioeconomic characteristics. However, Hierarchical Clustering achieved a higher Silhouette Score of 0.418 compared to K-Means, which obtained a score of 0.331. This finding suggests that Hierarchical Clustering produces more compact clusters with clearer separation between groups, resulting in better clustering quality. The clustering results are expected to provide valuable insights for local governments in formulating more targeted development policies based on the socioeconomic characteristics of each region in Papua Province.

Keywords: ***Hierarchical Clustering, K-Means, Papua Province, Silhouette Score, Socioeconomic Conditions.***

ARTICLE INFO	
Submission received: 31 May 2026	Accepted: 30 August 2026
Revised: 13 August 2026	Published: 31 August 2026
Available on: https://doi.org/10.32493/sm.v8i2.60424	

INTRODUCTION

Papua Province is one of the regions in Indonesia characterized by substantial geographical, social, and economic diversity. The region consists of coastal areas, swamps, lowlands, and mountainous terrains that are difficult to access, resulting in disparities in development across regencies and municipalities. These disparities are reflected in various socioeconomic indicators, including Life Expectancy (LE), Human Development Index (HDI), Poverty Rate (PR), and Gini Ratio (GR).

Socioeconomic development is an important measure of regional welfare. Higher levels of development are generally associated with improved health conditions, better educational attainment, and lower poverty rates. However, socioeconomic conditions in Papua Province remain uneven, particularly between urban and remote areas. Several regencies still experience high poverty rates and relatively low levels of human development compared to other regions (BPS Papua, 2023).

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

To better understand regional socioeconomic characteristics, an analytical approach capable of grouping regions with similar characteristics is required. One of the most widely used techniques is cluster analysis. Clustering is a data mining technique that groups objects based on their similarities, such that objects within the same cluster are more similar to one another than to those in different clusters (Han et al., 2012).

Among various clustering techniques, K-Means Clustering and Hierarchical Clustering are frequently applied in socioeconomic studies. K-Means Clustering is a partition-based method that groups observations according to their distances from cluster centroids (MacQueen, 1967). This method is computationally efficient and suitable for large datasets. In contrast, Hierarchical Clustering constructs a hierarchy of clusters based on the similarity among observations (Johnson & Wichern, 2007). In this study, Agglomerative Hierarchical Clustering with Ward's linkage method was employed because it tends to produce compact and homogeneous clusters.

Previous studies have demonstrated the usefulness of clustering techniques in identifying regional socioeconomic disparities and poverty patterns. Wahyuni (2021) applied K-Means Clustering to classify underdeveloped regions in Indonesia based on poverty indicators and identified three distinct regional groups. Other studies have also shown that clustering methods can reveal socioeconomic patterns across regions and provide valuable information for policy formulation and development planning.

Therefore, this study aims to compare the performance of K-Means Clustering and Hierarchical Clustering in mapping the socioeconomic conditions of regencies and municipalities in Papua Province. The analysis is based on four socioeconomic indicators, namely Life Expectancy (LE), Human Development Index (HDI), Poverty Rate (PR), and Gini Ratio (GR). The clustering results are evaluated using the Silhouette Score to determine the method that produces the most optimal grouping structure. The findings are expected to provide insights into regional socioeconomic disparities in Papua Province and support the formulation of more targeted development policies.

MATERIAL AND METHODS

Research Data

This study employed secondary data obtained from official publications of Statistics Indonesia (BPS) and the Statistics Office of Papua Province. The data consisted of socioeconomic indicators of regencies and municipalities in Papua Province during the period 2018–2022. The data were collected from various statistical publications, including *Papua in Figures*, *Human Development Index Reports*, *Poverty Profile in Indonesia*, and other related statistical reports (BPS, 2023; BPS Papua, 2023).

The unit of analysis comprised all regencies and municipalities in Papua Province. Four variables representing regional socioeconomic conditions were used in this study, namely Life Expectancy (LE), Human Development Index (HDI), Poverty Rate (PR), and Gini Ratio (GR). These variables were selected because they represent important dimensions of socioeconomic development, including health conditions, human development quality, poverty levels, and income inequality (BPS, 2023).

Data were collected through a documentation method by accessing and downloading official statistical publications available on the Statistics Indonesia website. The collected data were subsequently selected, processed, and standardized before being used in the clustering analysis. The use of secondary data from BPS was considered appropriate due to its high validity and reliability and its extensive application in socioeconomic research in

a
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

Indonesia.

Table 1. Operational Definition of Reasearch Variabels

Variabel	Definition	Unit
Life Expectancy (LE)	Average expected years of life at birth	Years
Human Development Index (HDI)	Composite index measuring achievements in health, education, and standard of living	Index
Poverty Rate (PR)	Percentage of population living below the poverty line	Percent (%)
Gini Ratio (GR)	Indicator of income distribution inequality	Index

Research Variables

The variables used in this study represent socioeconomic indicators describing the welfare conditions of regencies and municipalities in Papua Province.

1. Life Expectancy (LE).
   Life Expectancy refers to the average number of years a newborn is expected to live under prevailing mortality conditions. This indicator reflects the overall health status of a population. Higher life expectancy values indicate better public health conditions.
2. Human Development Index (HDI).
   The Human Development Index is a composite measure used to evaluate human development achievements based on three fundamental dimensions: a long and healthy life, knowledge, and a decent standard of living. Higher HDI values indicate better human development performance.
3. Poverty Rate (PR).
   Poverty Rate refers to the percentage of the population whose average monthly per capita expenditure falls below the poverty line. Higher poverty rates indicate a greater proportion of the population living in poverty.
4. Gini Ratio (GR).
   The Gini Ratio is a statistical measure used to assess income distribution inequality within a population. Values range from 0 to 1, where values closer to 0 indicate more equal income distribution, while values closer to 1 indicate greater inequality.

Table 2. Research Variables and Codes

Variabel	Code	Operational Definition	Unit
Life Expectancy	X1	Average expected years of life at birth	Years
Human Development Index	X2	Index measuring human development achievement	Index
Poverty Rate	X3	Percentage of population below the poverty line	Percent (%)
Gini Ratio	X4	Measure of income distribution inequality	Index

a
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

Research Procedure

The analytical procedure was conducted systematically to classify regencies and municipalities in Papua Province based on their socioeconomic conditions using K-Means Clustering and Hierarchical Clustering. The research stages are described as follows.

1. Data Collection

Secondary data were collected from official publications of Statistics Indonesia (BPS) and the Statistics Office of Papua Province. The variables analyzed included Life Expectancy (LE), Human Development Index (HDI), Poverty Rate (PR), and Gini Ratio (GR) for the period 2018–2022.

2. Descriptive Statistical Analysis

Descriptive statistics were employed to provide an overview of the research data. The analysis included minimum values, maximum values, means, and standard deviations for each variable.

3. Data Preprocessing

Prior to clustering, the data underwent preprocessing, including data completeness verification and standardization. Standardization was performed using the Z-score transformation to eliminate scale differences among variables and ensure equal contributions during clustering analysis.

The Z-score standardization formula is given by:

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

where:

Z= standardized value,
X= original observation value,
μ= mean value,
σ= standard deviation.

4. Determination of the Optimal Number of Clusters

The optimal number of clusters was determined using the Elbow Method. This approach evaluates the Within-Cluster Sum of Squares (WCSS) for different cluster numbers. The optimal cluster number is identified at the elbow point, where further increases in cluster numbers no longer produce substantial reductions in WCSS.

5. K-Means Clustering Analysis

After determining the optimal number of clusters, clustering was performed using the K-Means algorithm. The procedure involves initializing cluster centroids, calculating Euclidean distances between observations and centroids, assigning observations to the nearest centroid, and updating centroid positions iteratively until convergence is achieved.

6. Hierarchical Clustering Analysis

Hierarchical clustering was conducted using the Agglomerative Hierarchical Clustering approach with Ward's linkage method. Initially, each observation was treated as an individual cluster. Clusters were then merged sequentially based on similarity until a hierarchical structure was formed. Ward's linkage was selected because it tends to produce compact and homogeneous clusters by minimizing within-cluster variance.

a
StatMat Jurnal Statistika dan Matematika
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

The clustering results were visualized using a dendrogram to illustrate the similarity relationships among regencies and municipalities.

7. Visualization of Clustering Results

The clustering results obtained from both K-Means and Hierarchical Clustering were visualized using Principal Component Analysis (PCA). PCA was employed to reduce data dimensionality and facilitate the interpretation of cluster separation patterns.

8. Cluster Evaluation

The final stage involved evaluating clustering quality using the Silhouette Score. This metric measures the degree of similarity among observations within the same cluster and their separation from observations in other clusters. Higher Silhouette Score values indicate better clustering quality, characterized by greater intra-cluster similarity and clearer inter-cluster separation.

The clustering method yielding the highest Silhouette Score was selected as the most appropriate method for mapping the socioeconomic conditions of regencies and municipalities in Papua Province.

Based on the analytical procedures described previously, the overall research workflow is illustrated in Figure 1.

Figure 1. Elbow Method

RESULTS AND DISCUSSION

Overview od Research Data

This study utilized socioeconomic data from regencies and municipalities in Papua Province. The analysis was conducted using four variables, namely Life Expectancy (LE), Human Development Index (HDI), Poverty Rate (PR), and Gini Ratio (GR). These variables were selected because they represent important dimensions of socioeconomic development, including health conditions, human development quality, poverty levels, and income inequality.

The dataset consisted of observations from all regencies and municipalities in Papua Province. Prior to clustering analysis, the data were examined descriptively to understand their general characteristics and variability. Descriptive analysis is important because it provides an initial overview of the distribution and magnitude of each variable before further statistical analysis is performed.

The selected variables exhibited varying scales and measurement units. Therefore, data standardization was required to eliminate scale differences among variables and ensure that each variable contributed equally to the clustering process. The standardized data were subsequently used in both K-Means Clustering and Hierarchical Clustering analyses.

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

Table 3. Research Variables

Variabel	Keterangan
LE	Life Expectancy
HDI	Human Development Index
PR	Poverty Rate
GR	Gini Ratio

Descrptive Statistics

Descriptive statistical analysis was conducted to summarize the characteristics of the socioeconomic variables used in this study. The analysis included the minimum value, maximum value, mean, and standard deviation for each variable.

Table 4. Descriptive Statistics of Socioeconomic Variables

Variable	Minimum	Maximum	Mean	Standard Deviation
LE	54.82	72.57	65.11	3.71
HDI	29.42	80.61	57.46	11.13
PR	10.03	43.65	28.69	9.68
GR	0.15	0.448	0.345	0.061

Based on Table 4, the Life Expectancy (LE) variable has an average value of 65.11 years, with a minimum value of 54.82 years and a maximum value of 72.57 years. The standard deviation of 3.71 indicates differences in health conditions among regencies and municipalities in Papua Province, although the level of variation is relatively low.

The Human Development Index (HDI) has an average value of 57.46, with a minimum value of 29.42 and a maximum value of 80.61. The standard deviation of 11.13 indicates substantial variation in human development levels across regions in Papua Province. This condition suggests that the quality of education, health, and living standards remains uneven among regions.

For the Poverty Rate (PR), the average value is 28.69%, with a minimum value of 10.03% and a maximum value of 43.65%. The standard deviation of 9.68 indicates considerable variation in poverty levels among regencies and municipalities. This high variability suggests that several regions continue to experience significantly higher poverty rates than others.

Meanwhile, the Gini Ratio (GR) has an average value of 0.345, ranging from 0.150 to 0.448. The standard deviation of 0.061 indicates a moderate level of variation in income inequality across regions.

Overall, the descriptive statistical analysis reveals that the socioeconomic conditions of regencies and municipalities in Papua Province remain highly diverse. The differences observed in each variable reflect the existence of regional development disparities, particularly in terms of health, human development, poverty levels, and income distribution.

Data Preprocessing

Before the clustering process was performed, the data underwent a preprocessing stage through standardization using the Z-score method. This standardization process was conducted to ensure that all research variables were measured on the same scale, thereby preventing any particular variable from dominating the cluster formation process.

a
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

The variables used in this study have different value ranges. Life Expectancy (LE) and Human Development Index (HDI) possess relatively larger scales compared to the Gini Ratio (GR). Therefore, standardization was necessary to place all variables on a comparable scale and improve the effectiveness of the clustering analysis.
Data standardization was performed using the following Z-score transformation formula:

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

Where:
Z= standardized value
X= original observation value
μ= mean
σ= standard deviation

The standardization process transformed the data into values with a mean approximately equal to zero and a standard deviation approximately equal to one. As a result, the standardized dataset was suitable for subsequent clustering analysis using both K-Means Clustering and Hierarchical Clustering methods.

Table 5. Example of Stadardized Data

No	LE	HDI	PR	GR
1	0.431	1.075	-1.882	0.706
2	-1.655	-0.058	1.034	-0.453
3	0.418	1.244	-1.581	-0.138
4	0.704	0.923	-0.365	0.358
5	1.010	0.860	-0.158	0.474

Based on the standardization results presented in Table 5, all variables have been transformed to a comparable scale, allowing them to be used more effectively in the clustering process. Positive standardized values indicate that an observation is above the mean of the corresponding variable, whereas negative values indicate that the observation is below the variable mean.

For example, in the first observation, the Human Development Index (HDI) has a standardized value of 1.075, indicating that the HDI value is higher than the overall average of the dataset. In contrast, the Poverty Rate (PR) has a standardized value of −1.882, suggesting that the poverty level of the observation is below the average of all regions included in the study.

Through the standardization process, regional clustering can be conducted more objectively because all variables contribute equally to the analysis. Consequently, the resulting clusters are expected to be more accurate and representative of the underlying socioeconomic characteristics of the regions.

Determination of the Optimal Number of Clusters

The optimal number of clusters for the K-Means Clustering method was determined using the Elbow Method. This method analyzes changes in the Within-Cluster Sum of Squares (WCSS) for different numbers of clusters. WCSS measures the degree of variation among observations within a cluster. Lower WCSS values indicate better clustering quality

a
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

because observations within the same cluster exhibit greater similarity.
The results of the Elbow Method are presented in Figure 2.

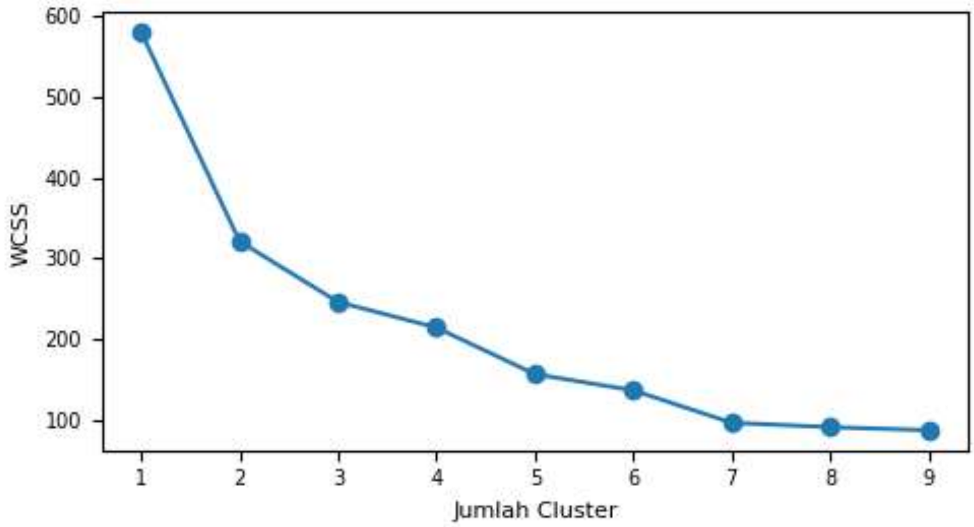

Figure 2. Elbow Method

Based on the Elbow Method graph presented in Figure 2, the WCSS value decreases substantially from one cluster to three clusters. After the number of clusters reaches three, the reduction in WCSS becomes relatively stable and no longer exhibits significant changes.

The elbow point is observed at (k = 3), indicating that three clusters represent the optimal number of clusters for this study. Therefore, it can be concluded that the most appropriate clustering solution consists of three clusters.

The selection of three clusters provides a suitable balance between similarity among observations within the same cluster and differences among clusters. Consequently, the K-Means Clustering analysis in this study was performed using three clusters.

K-Means Clustering Results

The clustering analysis using the K-Means method was conducted based on the optimal number of clusters determined by the Elbow Method.

1. Clustering Results

After standardizing the data using the Z-score method and determining the optimal number of clusters through the Elbow Method, three optimal clusters were identified. Subsequently, the K-Means Clustering algorithm was applied to the socioeconomic data of regencies and municipalities in Papua Province. K-Means Clustering groups observations based on their proximity to cluster centroids, such that observations within the same cluster exhibit a high degree of similarity (MacQueen, 1967).

The clustering results indicate that all regencies and municipalities in Papua Province can be classified into three clusters with distinct socioeconomic characteristics. These clusters reflect differences in regional socioeconomic development, as represented by Life Expectancy (LE), Human Development Index (HDI), Poverty Rate (PR), and Gini Ratio (GR).

The clustering results also show that each cluster contains a different number of members. This finding suggests that the socioeconomic conditions across regencies and municipalities in Papua Province are not homogeneous. Regions belonging to the same

a
StatMat
Jurnal Statistika dan Matematika
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

cluster share relatively similar socioeconomic characteristics, whereas regions assigned to different clusters exhibit distinct characteristics.

To further understand the characteristics of each cluster, an analysis of the average values of the variables within each cluster was conducted. The results reveal the existence of clusters characterized by relatively high socioeconomic development, indicated by high HDI and LE values combined with low poverty rates. In contrast, other clusters represent regions with relatively lower socioeconomic conditions, characterized by higher poverty rates and lower levels of human development.

These findings demonstrate that the K-Means method is capable of identifying patterns of similarity in socioeconomic conditions among regencies and municipalities in Papua Province. Therefore, the clustering results can provide useful information for policymakers in designing more targeted and effective regional development strategies (Han et al., 2012).

2. PCA Visualization of K-Means Clustering

To facilitate the interpretation of the clustering results, a visualization was performed using Principal Component Analysis (PCA). PCA is a dimensionality reduction technique that transforms multivariate data into a smaller number of principal components while preserving most of the information contained in the original dataset (Johnson & Wichern, 2007).

The PCA visualization was generated by projecting the data classified by the K-Means method onto two principal components, namely Principal Component 1 (PC1) and Principal Component 2 (PC2). The visualization results are presented in Figure 3.

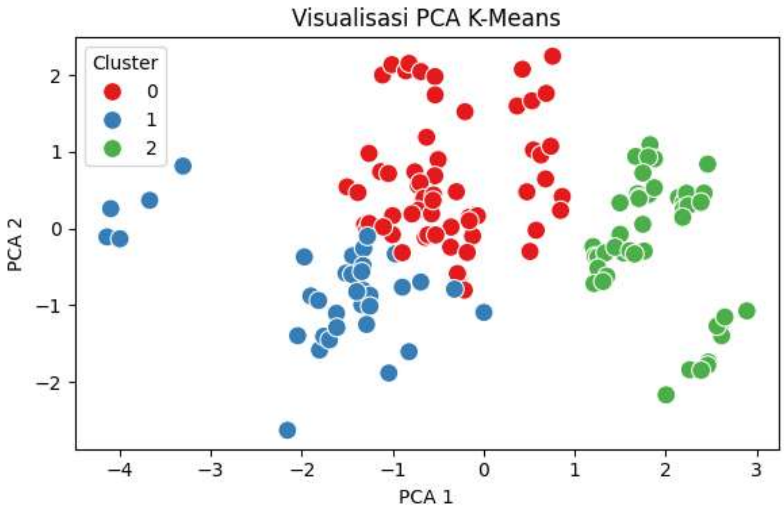

Figure 3. Visualization of K-Means Clustering.

Based on Figure 3, the clustering results obtained using the K-Means method produced three clusters, represented by different colors. Cluster 0 is displayed in red, Cluster 1 in blue, and Cluster 2 in green. Most observations within each cluster form relatively distinct groups, indicating that the K-Means method is capable of identifying patterns of similarity in the socioeconomic conditions of regencies and municipalities in Papua Province.

a
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

Cluster 2 (green) appears to be more clearly separated from the other clusters, as most of its members are concentrated on the right side of the plot. In contrast, Cluster 0 (red) and Cluster 1 (blue) exhibit several observations located close to one another in the central area of the plot. This condition suggests that some regencies and municipalities possess relatively similar socioeconomic characteristics, resulting in their positions near the boundaries between clusters.

Overall, the PCA visualization demonstrates that the K-Means Clustering method is capable of forming meaningful groups based on the socioeconomic characteristics of regencies and municipalities in Papua Province. The visualization supports the cluster interpretation process and provides a clearer understanding of the similarity relationships among regions included in this study.

1. Interpretation of K-Means Clusters

Cluster interpretation was performed based on the average values of the variables within each cluster, namely Life Expectancy (LE), Human Development Index (HDI), Poverty Rate (PR), and Gini Ratio (GR). According to Han et al. (2012), cluster interpretation aims to provide meaningful descriptions of the resulting groups so that the characteristics of each cluster can be better understood.

Based on the K-Means Clustering results, three clusters with distinct socioeconomic characteristics were identified. The average values of each variable for the three clusters are presented in Table 6.

Table 6. Average Values of socioeconomic Variables by K-Means Cluster

Cluster	AHH	IPM	PPM	RG
0	63.55	54.19	32.15	0.372
1	63.83	46.60	37.03	0.264
2	68.37	70.62	17.20	0.369

Based on Table 6, Cluster 2 represents the group of regions with the most favorable socioeconomic conditions among all clusters. This is reflected by the highest values of Life Expectancy (68.37 years) and Human Development Index (70.62). In addition, this cluster has the lowest Poverty Rate, amounting to 17.20%. These findings indicate that the regions classified in Cluster 2 generally exhibit better health conditions, educational attainment, and overall welfare levels than those in the other clusters. Although the cluster has a Gini Ratio of 0.369, indicating a moderate level of income inequality, it can generally be categorized as a group of regions with relatively advanced socioeconomic development.

Cluster 0 represents regions with moderate socioeconomic conditions. This cluster has an average Life Expectancy of 63.55 years and a Human Development Index of 54.19, which are lower than those of Cluster 2 but higher than those of Cluster 1. The Poverty Rate in this cluster reaches 32.15%, indicating that poverty remains a significant challenge. Furthermore, the Gini Ratio of 0.372 is the highest among all clusters, suggesting a relatively greater degree of income inequality. Therefore, the regions included in Cluster 0 can be classified as areas with moderate socioeconomic development that still face challenges in achieving a more equitable distribution of welfare.

Meanwhile, Cluster 1 represents regions with the least favorable socioeconomic

a
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

conditions. This is evidenced by the lowest Human Development Index value of 46.60 and the highest Poverty Rate of 37.03%. Although the Life Expectancy in this cluster is slightly higher than that of Cluster 0, the overall level of human development remains relatively low. On the other hand, the Gini Ratio of 0.264 is the lowest among all clusters, indicating a more equal distribution of income. However, this equality occurs within a context of generally low welfare levels. Therefore, the regions classified in Cluster 1 require greater attention in efforts aimed at improving human development outcomes and reducing poverty.

Overall, the clustering results reveal clear differences in socioeconomic characteristics among regencies and municipalities in Papua Province. Cluster 2 represents regions with relatively advanced socioeconomic development, Cluster 0 consists of regions with moderate socioeconomic conditions, and Cluster 1 comprises regions with relatively low levels of socioeconomic development. These findings demonstrate that the clustering approach is capable of identifying patterns of similarity in regional socioeconomic conditions and can serve as a valuable basis for formulating more targeted and effective development policies (Han et al., 2012).

2. Hierarchical Clustering Results

The Hierarchical Clustering analysis in this study was performed using the Agglomerative Hierarchical Clustering approach with Ward's linkage method. This method groups observations based on their similarity by successively merging clusters that minimize within-cluster variance, thereby producing compact and homogeneous clusters.

2.1. Dendrogram

The Hierarchical Clustering method applied in this study employed the Agglomerative Hierarchical Clustering approach with Ward's linkage criterion. According to Johnson and Wichern (2007), hierarchical clustering operates by progressively merging observations with the highest degree of similarity until all observations are grouped into a single cluster. The clustering results are visualized using a dendrogram to facilitate the interpretation of the resulting cluster structure.

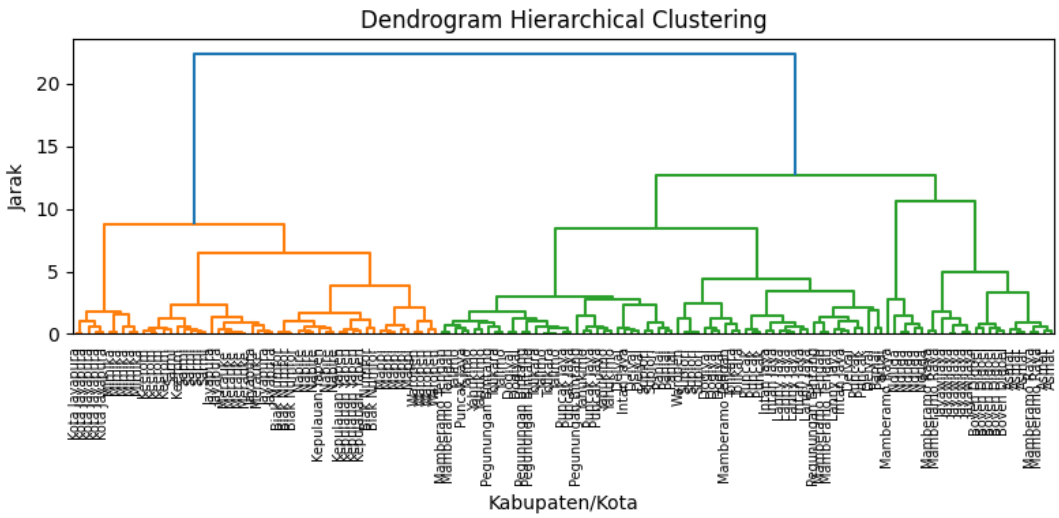

Figure 4. Hierarchical Clustering Dendrogram

Based on Figure 4, the regencies and municipalities in Papua Province are grouped into several clusters with different levels of socioeconomic similarity. The vertical axis represents the linkage distance or degree of dissimilarity between observations, while the horizontal axis represents the observations being clustered.

a
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

The dendrogram shows that the cluster merging process occurs gradually until all observations are combined into a single large cluster. A substantial increase in linkage distance can be observed at the final stage of the clustering process, occurring at approximately a distance of 22. This sharp increase indicates the existence of distinct differences among the clusters formed.

Based on the dendrogram, the clustering structure was cut before the occurrence of the large increase in linkage distance, resulting in three main clusters. The selection of three clusters was considered appropriate because it provides a clear separation among groups while maintaining a relatively high degree of homogeneity within each cluster. According to Kaufman and Rousseeuw (2009), an effective clustering solution should preserve high similarity among observations within the same cluster while maintaining clear distinctions between different clusters.

Overall, the dendrogram demonstrates that the Hierarchical Clustering method is capable of identifying patterns of similarity in the socioeconomic conditions of regencies and municipalities in Papua Province and forming relatively homogeneous regional groups based on their socioeconomic characteristics.

Clustering Results

After applying the Agglomerative Hierarchical Clustering method with Ward's linkage criterion, three clusters were identified, representing different socioeconomic conditions among regencies and municipalities in Papua Province. The clustering process was based on four variables, namely Life Expectancy (LE), Human Development Index (HDI), Poverty Rate (PR), and Gini Ratio (GR). According to Johnson and Wichern (2007), Hierarchical Clustering groups observations according to their degree of similarity, such that observations within the same cluster exhibit relatively homogeneous characteristics compared to those in other clusters.

Based on the dendrogram cut at the optimal number of three clusters, all regencies and municipalities in Papua Province were successfully classified into three groups with distinct socioeconomic characteristics. The clustering results indicate that socioeconomic conditions across regions in Papua Province remain heterogeneous, suggesting the need for different development approaches tailored to the characteristics of each regional group.

The characteristics of each cluster can be examined through the average values of the socioeconomic variables presented in Table 7.

Table 7. Average Values of Socioeconomic Variables by Hierarchical Cluster

Cluster	AHH	IPM	PPM	RG
1	67.92	69.10	18.95	0.37
2	65.47	50.43	36.17	0.32
3	58.12	50.88	29.99	0.34

Based on Table 7, Cluster 1 exhibits the highest values of Life Expectancy (LE) and Human Development Index (HDI) among all clusters, with average values of 67.92 years and 69.10, respectively. In addition, this cluster has the lowest Poverty Rate (PR), amounting to 18.95%. These findings indicate that the regions classified in Cluster 1 generally have better levels of welfare, health conditions, and human development compared to the other clusters.

a
StatMat
Jurnal Statistika dan Matematika
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

Cluster 2 is characterized by the highest Poverty Rate, reaching 36.17%, although its Life Expectancy remains higher than that of Cluster 3. The Human Development Index in this cluster is 50.43, indicating a relatively low level of human development. These results suggest that regions within Cluster 2 continue to face substantial poverty-related challenges and therefore require greater attention in efforts aimed at improving community welfare and socioeconomic conditions.

Meanwhile, Cluster 3 records the lowest Life Expectancy, with an average value of 58.12 years. The Human Development Index in this cluster is 50.88, which is only slightly higher than that of Cluster 2. However, the Poverty Rate in Cluster 3 is lower than that of Cluster 2, amounting to 29.99%. This finding indicates that the primary challenges faced by regions in Cluster 3 are more closely related to health conditions and human development outcomes rather than poverty itself.

Overall, the clustering results obtained using the Hierarchical Clustering method reveal clear differences in socioeconomic characteristics among the clusters. Cluster 1 represents regions with relatively advanced socioeconomic conditions, Cluster 2 represents regions with relatively high poverty levels, and Cluster 3 represents regions facing significant challenges in health and human development. These findings demonstrate that the Hierarchical Clustering method effectively identifies patterns of similarity in the socioeconomic conditions of regencies and municipalities in Papua Province and can therefore serve as a useful basis for formulating more targeted development policies (Kaufman & Rousseeuw, 2009).

Furthermore, the distinct characteristics observed across the clusters indicate that socioeconomic development in Papua Province remains uneven. Consequently, development strategies should be tailored to the specific needs of each regional group. Regions classified in Cluster 1 may focus on maintaining and strengthening their existing development achievements, whereas regions in Clusters 2 and 3 require more intensive interventions, particularly in poverty reduction, educational improvement, healthcare enhancement, and the equitable provision of infrastructure and public services.

3.3. PCA Visualization of Hierarchical Clustering

To facilitate the interpretation of the clustering results obtained from the Hierarchical Clustering method, a visualization was performed using Principal Component Analysis (PCA). PCA is a dimensionality reduction technique that transforms multivariate data into a smaller number of principal components while preserving most of the information contained in the original dataset (Johnson & Wichern, 2007).

In this study, PCA was employed to visualize the clustering results of regencies and municipalities in Papua Province in a two-dimensional space represented by Principal Component 1 (PC1) and Principal Component 2 (PC2). By projecting the multidimensional data onto these two principal components, the distribution of observations and the separation among clusters can be more easily observed and interpreted.

The PCA visualization of the Hierarchical Clustering results is presented in Figure 5.

a
StatMat Jurnal Statistika dan Matematika
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

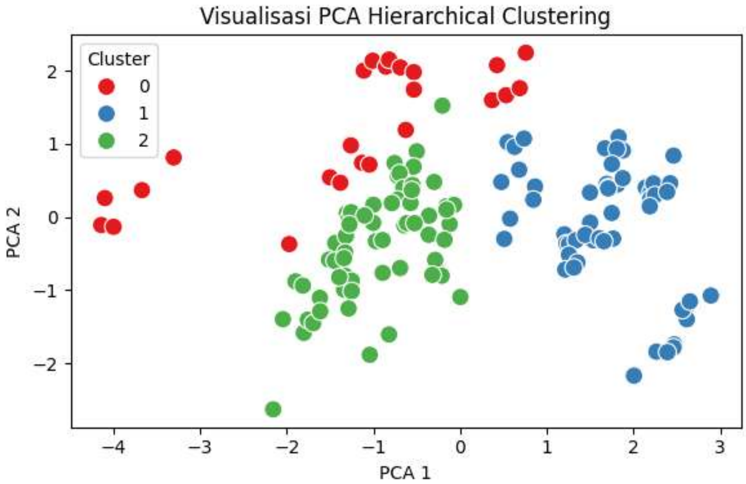

Figure 5. PCA Visualization of Hierarchical Clustering

Based on Figure 5, the Hierarchical Clustering results form three clusters, represented by different colors. Each point in the plot corresponds to a regency or municipality in Papua Province projected onto two principal components, namely Principal Component 1 (PC1) and Principal Component 2 (PC2). Overall, the three clusters exhibit relatively distinct distribution patterns, indicating differences in socioeconomic characteristics among the regional groups.

The blue cluster appears to be concentrated on the right side of the plot and is clearly separated from the other clusters. This pattern suggests that the observations within this cluster share relatively homogeneous characteristics that distinguish them from the other groups. Based on the cluster profile analysis, this cluster represents regions with relatively favorable socioeconomic conditions, characterized by higher Life Expectancy (LE) and Human Development Index (HDI) values, as well as lower poverty levels.

The green cluster is primarily located in the central to lower-left region of the plot. The relatively compact distribution of observations indicates a high degree of similarity among cluster members. This group represents regions that continue to face socioeconomic challenges, particularly in terms of poverty and relatively low levels of human development compared to the cluster with the most favorable conditions.

Meanwhile, the red cluster is positioned in the upper-left area of the plot and appears to be separated from the other two clusters. This location suggests that the regions included in this cluster possess socioeconomic characteristics that differ from those of the remaining groups. Several observations located relatively far from the cluster center also indicate the presence of variation among members within this cluster.

Overall, the PCA visualization demonstrates that the Hierarchical Clustering method produces a reasonably clear separation among clusters. Although a few observations are located in adjacent areas, most cluster members remain concentrated within their respective groups, indicating a high degree of within-cluster homogeneity and substantial between-cluster heterogeneity. According to Kaufman and Rousseeuw (2009), an effective clustering solution is characterized by high similarity among observations within the same cluster and clear differences between clusters.

These visualization results support the findings obtained from the Silhouette Score evaluation, where the Hierarchical Clustering method achieved a score of 0.418, which is higher than the score of 0.331 obtained by the K-Means method. This result indicates that

a
StatMat
Jurnal Statistika dan Matematika
STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

Hierarchical Clustering produces a superior cluster structure for grouping regencies and municipalities in Papua Province based on their socioeconomic conditions. Therefore, the PCA visualization further strengthens the conclusion that Hierarchical Clustering is the more appropriate method for this study (Kaufman & Rousseeuw, 2009).

Interpretation of Hierarchical Clusters

Cluster interpretation was conducted to understand the socioeconomic characteristics of the regencies and municipalities included in each cluster based on the average values of Life Expectancy (LE), Human Development Index (HDI), Poverty Rate (PR), and Gini Ratio (GR). According to Han et al. (2012), cluster interpretation aims to provide meaningful descriptions of the resulting groups so that the characteristics of each cluster can be explained more comprehensively.

Based on the Hierarchical Clustering results, three clusters with distinct socioeconomic characteristics were identified, as presented in Table 8.

Table 8. Average Values of socioeconomics Variavles by Hierarchical Cluster

Cluster	LE	HDI	PR	GR
1	67.92	69.10	18.95	0.37
2	65.47	50.43	36.17	0.32
3	58.12	50.88	29.99	0.34

Based on Table 8, Cluster 1 represents the group of regions with the most favorable socioeconomic conditions among all clusters. This is evidenced by the highest values of Life Expectancy (67.92 years) and Human Development Index (69.10). In addition, this cluster has the lowest Poverty Rate, amounting to 18.95%. These findings indicate that the regions classified in Cluster 1 generally exhibit better health conditions, human capital quality, and overall welfare levels compared to regions in the other clusters. Although the cluster has a Gini Ratio of 0.37, indicating a moderate level of income inequality, it can generally be categorized as a group of regions with relatively advanced socioeconomic development.

Cluster 2 represents the group of regions facing the greatest challenges related to poverty. This is reflected in its Poverty Rate of 36.17%, which is the highest among all clusters. Furthermore, the Human Development Index in this cluster is 50.43, indicating a relatively low level of human development. Although the average Life Expectancy reaches 65.47 years, the high poverty rate suggests that the welfare of communities within this cluster remains relatively low. Therefore, Cluster 2 can be categorized as a group of regions with high poverty levels that require special attention through poverty alleviation policies and programs aimed at improving living standards.

Meanwhile, Cluster 3 is characterized by the lowest Life Expectancy, averaging 58.12 years. This value indicates that health conditions in the regions belonging to this cluster remain relatively poor compared to the other clusters. Although the Poverty Rate in Cluster 3 is lower than that of Cluster 2, at 29.99%, the Human Development Index of only 50.88 suggests that human development outcomes are still far from optimal. These findings indicate that the primary challenges faced by regions in Cluster 3 are more closely related to health and human development issues than to poverty itself. Consequently, this cluster can

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

be categorized as a group of regions requiring greater attention in improving healthcare services, educational attainment, and human resource development.

Overall, the cluster interpretation reveals distinct socioeconomic characteristics across the three clusters. Cluster 1 represents regions with relatively advanced socioeconomic conditions, Cluster 2 represents regions characterized by high poverty levels, and Cluster 3 represents regions that continue to face challenges in health and human development. These differences indicate that the socioeconomic conditions of regencies and municipalities in Papua Province remain diverse, highlighting the need for development policies tailored to the specific characteristics and needs of each regional group.

The clustering results can serve as a valuable reference for local governments in formulating more targeted development strategies. Regions in Cluster 1 may focus on maintaining and further enhancing their development achievements, whereas regions in Cluster 2 require more intensive poverty reduction programs. Meanwhile, regions in Cluster 3 need greater attention in improving access to and the quality of healthcare services, education, and human resource development. Therefore, the Hierarchical Clustering results provide useful information for supporting more effective and sustainable socioeconomic development planning in Papua Province (Han et al., 2012).

3. Cluster Evaluation

Cluster evaluation was conducted using the Silhouette Score to assess the quality of the clustering results obtained from the K-Means Clustering and Hierarchical Clustering methods. This evaluation metric measures the degree of similarity among observations within the same cluster as well as the degree of separation between different clusters.

The Silhouette Score ranges from −1 to 1. A higher Silhouette Score indicates better clustering quality, suggesting that observations within the same cluster are highly similar while clusters are clearly separated from one another.

Tabel 9. Cluster Evaluation Results

Clustering Method	Silhouette score
K-Means	0.331
Hierarchical Clustering	0.418

Based on Table 9, the Hierarchical Clustering method achieved a Silhouette Score of 0.418, whereas the K-Means method obtained a score of 0.331.

These results indicate that Hierarchical Clustering produced a better clustering structure than K-Means for grouping the socioeconomic conditions of regencies and municipalities in Papua Province. The higher Silhouette Score obtained by Hierarchical Clustering suggests that the resulting clusters exhibit greater within-cluster homogeneity and clearer separation between clusters compared to those generated by the K-Means method.

On the other hand, although the K-Means method was also capable of producing reasonably well-defined clusters, the similarity among observations within the same cluster and the separation between clusters were not as optimal as those achieved by Hierarchical Clustering.

Therefore, based on the Silhouette Score evaluation, Hierarchical Clustering can be considered the more effective method for classifying the socioeconomic conditions of

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

regencies and municipalities in Papua Province using the variables of Life Expectancy (LE), Human Development Index (HDI), Poverty Rate (PR), and Gini Ratio (GR).

Overall, the findings demonstrate that the Hierarchical Clustering approach provides a more accurate representation of regional socioeconomic disparities in Papua Province. Consequently, this method is recommended for identifying groups of regions with similar socioeconomic characteristics and for supporting evidence-based regional development planning and policymaking.

CONCLUSION

This study aimed to compare the performance of K-Means Clustering and Hierarchical Clustering methods in mapping the socioeconomic conditions of regencies and municipalities in Papua Province based on four indicators: Life Expectancy (LE), Human Development Index (HDI), Poverty Rate (PR), and Gini Ratio (GR). The results showed that both methods produced three clusters representing distinct socioeconomic characteristics across regions. In the Hierarchical Clustering results, Cluster 1 consisted of regions with the most favorable socioeconomic conditions, characterized by high LE and HDI values and a low poverty rate. Cluster 2 represented regions with the highest poverty levels, while Cluster 3 comprised regions facing challenges primarily related to health conditions and human development.

Based on the Silhouette Score evaluation, Hierarchical Clustering achieved a score of 0.418, which was higher than the score of 0.331 obtained by K-Means Clustering. These findings indicate that Hierarchical Clustering produced a better clustering structure, characterized by greater within-cluster homogeneity and clearer separation between clusters. Therefore, Hierarchical Clustering can be recommended as the more suitable method for mapping the socioeconomic conditions of regencies and municipalities in Papua Province. The results of this study may also serve as a useful reference for policymakers in designing more targeted and effective regional development strategies aimed at reducing socioeconomic disparities across regions.

AUTHOR CONTRIBUTIONS

For research articles with several authors, a short paragraph specifying their individual contributions must be provided. For example:
Conceptualization: Anggy Yulsitya Putri; Data collection: Anggy Yulistya Putri; Data analysis: Anggy Yuslitya Putri; Writing original draft; Anggy Yulistya Putri; Review and editing: Anggy Yulistya Putri.

FUNDING

This research received no external funding.

ACKNOWLEDGMENTS

The author would like to express sincere gratitude to the Centeral Statistics Agency (Badan Pusat Statistik/BPS) and BPS Papua Province for providing the data used in this study. The author also thanks the lecturers and academic staff of the Statistics Study Program, Faculty of Mathematics and Natural Sciences, Universitas Cenderawasih, for their guindance, support, and valuable suggestions throughout the research process. Appreciation is also extended to all parties who have directly or indirectly contributed to the completion of this research

STATMAT (Jurnal Statistika dan Matematika)
P-ISSN: 2655-3724 E-ISSN: 2720-9881
Vol. 8, No. 2, August 2026
12th article, Pages: 335 – 352
DOI: https://doi.org/10.32493/sm.v8i2.60424

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

1. Arikunto, S. (2018). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta, hlm. 274–282.
2. Badan Pusat Statistik. (2023). *Statistik Indonesia 2023*. Jakarta: Badan Pusat Statistik.
3. Badan Pusat Statistik Provinsi Papua. (2023). *Papua Dalam Angka 2023*. Jayapura: BPS Provinsi Papua.
4. Fadilah, Y. C., Sani, A., & Andrianingsih. (2024). "Applying K-Means Clustering for Grouping Papua's Districts Based on Poverty Indicators Analysis." *Jurnal Ilmu Pengetahuan dan Teknologi Komputer*, Vol. 10, No. 1, hlm. 45–53.
5. Ghozali, I. (2021). *Aplikasi Analisis Multivariate dengan IBM SPSS*. Semarang: Badan Penerbit Universitas Diponegoro, hlm. 95–132.
6. Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Burlington: Morgan Kaufmann.
7. Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (6th ed.). Pearson Education.
8. Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley.
9. MacQueen, J. (1967). "Some Methods for Classification and Analysis of Multivariate Observations." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, hlm. 281–297.
10. Mutiah, N., et al. (2024). "Perbandingan Metode Klastering dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat." *ResearchGate Publication*, hlm. 1–12.
11. Permatasari, E. P., Iriani, L. A., & Widodo, E. (2026). "Identification of Indonesian Provinces Based on Socioeconomic Indicators in 2024 Using Hierarchical Clustering." hlm. 55–70.
12. Sugiyono. (2019). *Statistika untuk Penelitian*. Bandung: Alfabeta, hlm. 57–89.
13. Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta, hlm. 67–95.
14. Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta, hlm. 193–225.
15. Todaro, M. P., & Smith, S. C. (2015). *Economic Development* (12th ed.). Pearson Education, hlm. 241–276.
16. United Nations Development Programme. (2022). *Human Development Report 2022*. New York: UNDP.
17. Walpole, R. E. (2012). *Probability and Statistics for Engineers and Scientists* (9th ed.). Pearson, hlm. 315–352.
18. Wahyuni, R. (2021). "K-Means Clustering for Grouping Indonesia Underdeveloped Regions in 2020 Based on Poverty Indicators." *Jurnal Parameter*, Vol. 6, No. 2, hlm. 61–72.