SPMS
Spatial Planning & Management Science
SPMS 3(2): 40-56
ISSN 3088-9669

Institute for Advanced Science, Social and Sustainable Future
MORALITY BEFORE KNOWLEDGE

Spatial clustering of environmental-economic nexus in Indonesia: A hierarchical agglomerative analysis of provincial disparities

Antradiva Oktaviola Puspita[1,*]
[1] *Department of Economics, Faculty of Economics and Business, Universitas Negeri Semarang, Semarang, Central Java 50229, Indonesia.*
*Correspondence: antradivaoktaviolap@gmail.com

Received Date: January 20, 2026 Revised Date: August 14, 2026 Accepted Date: August 28, 2026

ABSTRACT

Background: Indonesia faces the paradox of development where economic growth is not followed by Environmental Quality Improvement. Indonesia's Environmental Quality Score (33.6) is well below the average of East Asian developing countries (40-45) despite GDP reaching USD 1,396.3 billion. Demographic pressure with a population of 286.7 million inhabitants exacerbates ecological degradation. Previous studies used a linear-aggregate approach that failed to capture the complexity of spatial interactions and heterogeneity between regions. **Methods:** this study used agglomerative hierarchical cluster analysis with Ward Method to classify 34 provinces based on Environmental Quality Index (EQI), GRDP, and Population Density. Secondary Data in 2023 from Statistics Indonesia/*Badan Pusat Statistik* (BPS) and Bank Indonesia were analyzed using Euclidean distance and visualized through dendrograms to identify hierarchical structures with similar characteristics between provinces. **Findings:** formed four clusters. Cluster 1 (14 provinces) shows moderate-to-high EQI (72.68) with low GRDP (IDR 1,448.29 billion). Cluster 2 (3 provinces) shows the lowest EQI (55.95). Cluster 3 (16 provinces) shows the highest EQI (75.23) combined with the highest GRDP (IDR 4,953.13 billion). Cluster 4 is a single-province outlier (EQI 54.93) with an extreme population density (16,146 people/km^2), consistent with DKI Jakarta. The findings are broadly consistent with an environmental Kuznets Curve-like pattern in Indonesia, with the island of Java hosting >60% of the national population being an epicenter of ecological pressure. **Conclusion:** spatial disparities in Environmental Quality are strongly influenced by the concentration of economic and demographic activities. Without structural economic transformation and technological modernization, provinces outside Java could potentially experience similar degradation as development accelerates. **Novelty/Originality** of this article: this study adopts an agglomerative hierarchical cluster approach that uncovers inter-provincial multiplier effects and spatial heterogeneity, providing an empirical basis for contextual and non-uniform sustainable development policies.

KEYWORDS: agglomerative hierarchy; environmental degradation; spatial inequality.

1. Introduction

Environmental degradation has emerged as a defining challenge of the 21st century, with developing nations simultaneously experiencing the highest vulnerability to ecological crises while contributing disproportionately to their acceleration. This paradoxical position stems from prevailing economic development models that prioritize short-term growth maximization over long-term ecological sustainability, creating systematic trade-offs between immediate prosperity and environmental preservation (Suwandaru et al., 2024). Empirical evidence across emerging economies demonstrates that this growth-first

Cite This Article:
Puspita, A. O., (2026). Spatial clustering of environmental-economic nexus in Indonesia: A hierarchical agglomerative analysis of provincial disparities**.** *Spatial Planning & Management Science, 3*(2), 40-56. https://doi.org/10.61511/spms.v3i2.2026.3164

Copyright: © 2026 by the authors. This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

paradigm generates path dependencies that lock countries into resource-intensive development trajectories, making subsequent transitions toward sustainable practices increasingly difficult and costly (Aiginger & Rodrik, 2020). The persistence of extractive economic structures in developing regions reflects not merely policy failures, but deeper institutional and technological constraints that perpetuate environmental-economic tensions even as awareness of sustainability imperatives grows (Pherdd, 2026).

The Middle East and North Africa (MENA) region exemplifies how trade liberalization in developing economies can inadvertently accelerate environmental degradation despite generating substantial economic benefits. Al-Mulali & Ozturk (2015) demonstrate that environmental deterioration rates in MENA countries exceed those observed in other developing regions, attributed primarily to trade policy reforms that prioritized economic expansion over ecological safeguards. The implementation of the Pan-Arab Free Trade Area (PAFTA) successfully reduced tariff barriers, catalyzing rapid growth in tourism, cross-border trade, and transportation infrastructure all of which contributed to measurable GDP increases across member states. However, this economic expansion generated significant negative environmental externalities, particularly through escalating fossil fuel consumption that intensified carbon emissions and air pollution beyond the absorptive capacity of regional ecosystems (Göll et al., 2020). The MENA experience reveals a critical policy lesson: trade liberalization without concurrent environmental regulation creates conditions where economic gains systematically come at the expense of ecological sustainability, establishing precedents that subsequent policy interventions struggle to reverse (Asian Development Bank, 2023).

Indonesia faces similar problems with distinctive characteristics that warrant specific analytical attention. As a developing country with significant economic growth momentum, Indonesia is caught in the classic dilemma between accelerating development and preserving the environment, a tension that has intensified as industrialization accelerates across the archipelago (Resosudarmo et al., 2009). What is particularly worrying is that Indonesia's environmental quality score is weak compared to most other developing countries in its economic peer group. According to Yale University (2025), Indonesia only scored 33.6 on the Environmental Performance Index, far behind the average of developing countries in East Asia and Southeast Africa, which scored between 40 and 45, despite having comparable levels of economic development. This underperformance is especially alarming given Indonesia's status as a global biodiversity hotspot and one of the world's largest carbon emitters, where environmental degradation carries consequences beyond national boundaries. This condition indicates a fundamental paradox in Indonesia's economic development model, where macroeconomic success has not translated into environmental quality improvements that are typically observed in countries at similar income levels.

Although Gross Domestic Product (GDP) has shown consistent expansion over the past two decades, averaging 5.2% annual growth, the growth achievements have not been followed by substantive improvements in environmental quality, revealing a persistent decoupling failure between economic and ecological indicators (Organisation for Economic Co-operation and Development/OECD, 2015). Indonesia's economic growth is still supported by sectors with high intensity of natural resource exploitation and dependence on fossil energy, which causally increases the probability of environmental degradation (Hardi et al., 2024). The extractive nature of Indonesia's economy is evident in the continued dominance of coal mining, palm oil production, and forestry sectors that collectively account for over 40% of export revenues while simultaneously driving deforestation rates among the highest globally (Mohr et al., 2025).

According to the International Monetary Fund (2022) report, Indonesia's GDP reached USD 1,396.3 billion, placing Indonesia in the ranks of the world's 20 largest economies and G20 members. This achievement is an irony in itself, considering that economic growth and environmental sustainability have not been effectively converged within the framework of national development policy in Indonesia, contrasting sharply with other G20 members who have demonstrated measurable progress in decoupling economic growth from environmental impacts (OECD Economic Surveys Indonesia, 2024). The persistence of this

extractive growth model suggests deep structural path dependenciess that resist policy reform, where established economic interests and infrastructure investments lock the country into environmentally intensive development trajectories (Bechauf, 2021).

Indonesia is not solely due to extractive economic structures, but is also intensified by demographic dynamics that create multiplicative pressures on natural resource systems. Indonesia, with a population of 286.7 million and a national average density of 158 people per km^2, faces multiple overlapping ecological pressures stemming from concentrated human settlement patterns, particularly in Java which hosts over 60% of the population on just 7% of the land area (Statistics Indonesia, 2023). These demographic pressures operate through several interconnected pathways: rising demand for food production that drives agricultural expansion and intensification, escalating energy consumption predominantly sourced from fossil fuels, increasing water extraction that exceeds natural recharge rates in urbanized regions, and accelerating conversion of natural habitats into residential and industrial land (United Nations Population Fund/UNFPA, 2014). This population-environment nexus results in cascading effects that include exponential growth in resource consumption, accumulation of municipal and industrial waste often disposed without adequate treatment, and intensification of greenhouse gas emissions that systemically degrade environmental quality across multiple indicators (International Energy Agency/IEA Bioenergy, 2025).

The relationship between population and envsironmental degradation has been studied in depth by Rehman & Rehman (2022) using carbon dioxide emissions as a proxy for Environmental Quality. Their findings confirm a multiplier effect between population size and environmental burden, with heavily populated countries such as China and India being the focal points of this phenomenon. However, what is interesting about the study is the identification of a non-linear curve in the relationship of economic well-being-CO_2 emissions. The findings show that the impact of welfare and population on emissions reaches a saturation point (plateau) and even a marginal decline in economies with very high GDP per capita. This pattern can be explained through two mechanisms of structural transformation. First, there is a shift from a manufacturing and extraction-based economy to a service-based economy that has a lower carbon footprint. Second, capital accumulation in advanced economies enables massive investments in energy efficiency technologies and green infrastructure.

The consequence for Indonesia is that without accelerated economic structural transformation and technological modernization, the combination of large populations and extractive economic structures will continue to exacerbate the trade-off between economic growth and environmental preservation, creating a development trap that is counterproductive to long-term sustainability. The causal relationship has been empirically validated in a study by Xuan (2024) that identified temporal asymmetries in the impact of economic growth on CO_2 emissions in Indonesia. The findings suggest that economic growth exerts a significant influence on the increase in CO_2 emissions in the short term, but this significance dissipates in the long term. This pattern suggests that Indonesia may be approaching the stage where the relationship between economic growth and environmental quality becomes increasingly complex. As income levels rise, environmental pressures tend to intensify, reflecting the early dynamics often associated with the Environmental Kuznets Curve (EKC). However, whether Indonesia can eventually transition toward the phase where environmental degradation declines through technological advancement and structural economic change remains uncertain and warrants further investigation.

However, there are crucial trade-offs that need to be comprehensively understood. First, the lag time towards long-term environmental improvement tends to be more persistent in developing countries than in developed countries, due to several structural barriers: (i) limited fiscal capacity for green infrastructure investment and clean-technology research; (ii) path dependence on conventional technologies already embedded in production systems; (iii) institutional inertia in environmental policy reform (Söderholm, 2020).

Based on a review of the previous literature, there are fundamental methodological gaps in studies of environmental economic nexus. The majority of studies are still fixated on conventional approaches that are linear and aggregate, which are only able to identify the partial influence of economic factors on environmental degradation without capturing the complexity of spatial interactions and heterogeneity of characteristics between regions. This limitation results in policy recommendations that tend to be general and not contextual, whereas the reality of development in Indonesia varies greatly in terms of economic capacity, demographic structure, and ecological carrying capacity.

Overcoming these limitations, this study adopts a more renewable methodological approach through agglomerative hierarchical cluster analysis by utilizing three key variables: Environmental Quality Index (EQI) as an indicator of ecological conditions, Gross Regional Domestic Product (GRDP) as a proxy for regional economic growth, and Population Density as a representation of demographic pressure. The choice of agglomerative hierarchical cluster method is not without strong theoretical justification. In contrast to non-hierarchical cluster methods such as K-means that produce single and static partitions, hierarchical clusters are able to reveal the multilevel structure of the similarity of characteristics between regions through dendograms that visualize the grouping process progressively (Suraya & Wijayanto, 2022). This methodological advantage lies in its ability to capture inter-provincial multiplier effects that are often overlooked in aggregate analysis. This study aims to identify spatial clustering patterns of environmental economic disparities across Indonesian provinces using the agglomerative hierarchical clustering method.

2. Methods

2.1 Research design and data sources

This study uses a quantitative approach with cluster analysis method to classify provinces in Indonesia based on environmental and economic characteristics. The Data used is secondary data sourced from the Central Statistics Agency and Bank Indonesia in 2023. The 2023 selection is based on the availability of the most complete and consistent data for all observed variables.

2.2 Research variables

The environmental and economic indicators used in this study were selected to reflect the condition of sustainable development in Indonesia. The selection was conducted through literature study and review of the availability of data at the provincial level.

Table 1. Description of variables and measurement units

Variable	Unit
Enviroment Quality Index (EQI)	Index
Gross Regional Domestic Product (GRDP)	IDR billion
Population Density	Inhabitans/km^2

(Source: Statistics Indonesia, 2025)

Based on Table 1 that all research variables have been processed through the stage of cleaning and standardization of data to ensure uniformity and comparability between observation units. Given the differences in units of measurement in each variable, data transformation through standardization is carried out in order to minimize the potential for bias that can affect the results of clustering analysis (Tokuda et al., 2022).

2.3 Analysis method

Clustering analysis in this study using Agglomerative Hierarchical Clustering method. This method was chosen because of its ability to generate a hierarchical structure that allows researchers to explore various possible number of clusters without having to determine the number of clusters a priori (Murtagh & Contreras, 2012). The Agglomerative Hierarchical Clustering procedure works on the bottom-up principle, that is, it starts by treating each province as an individual group, then gradually combines the pairs of groups that have the most similar characteristics (Xu & Tian, 2015). This merging process takes place iteratively until an optimal grouping structure is formed. The final result is a dendrogram that visualizes the hierarchy of similarities between provinces, making it easier to identify the grouping patterns that best suit the research objectives.

Prior to obtaining the optimal clustering result, several essential steps must be undertaken. These steps are systematically outlined in the table below to illustrate the methodological procedure in detail.

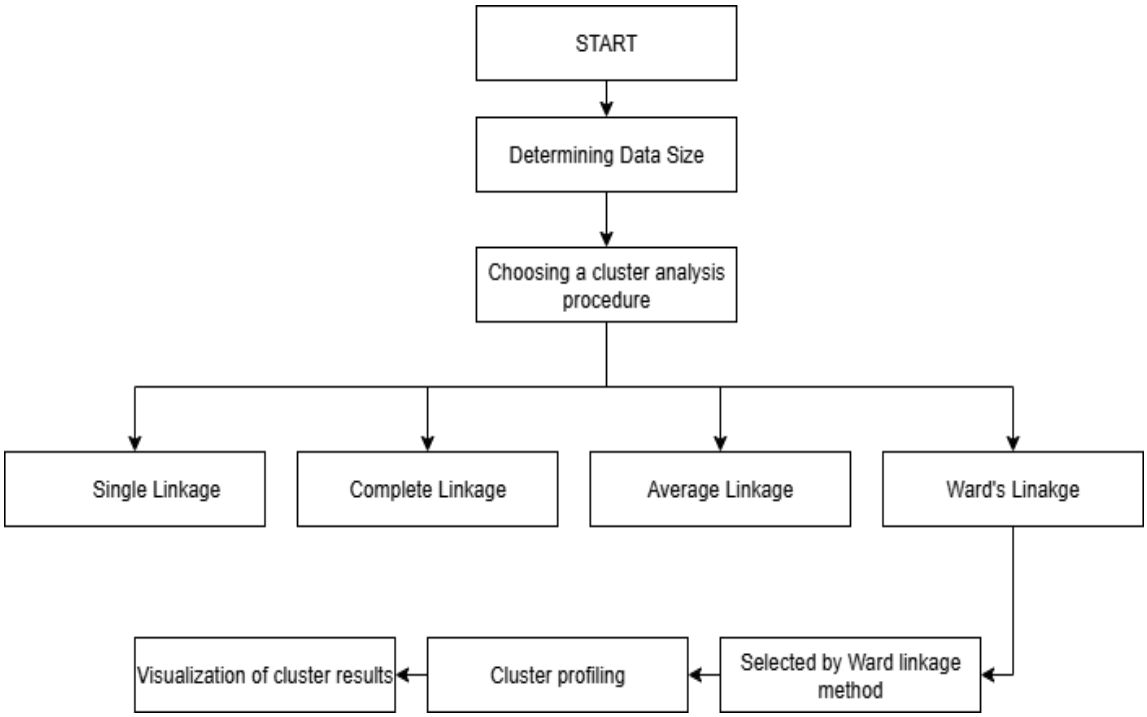


Fig. 1. Procedural flow of hierarchical clustering analysis

The initial stage of the study involved collecting raw data for three main variables: Environmental Quality Index (EQI), Gross Regional Domestic Product (GRDP), and Population Density across 34 provinces in Indonesia. The collected data underwent a validation process to ensure the absence of missing values and extreme outliers that could distort the analysis. Given that the three variables use different measurement scales EQI in index form, GRDP in billion rupiah, and Population in millions, data standardization was performed to achieve comparability across variables:

$$z_{ij} = \frac{x_{ij} - \bar{X}_{ij}}{s_{ij}} \quad \text{(Eq. 1)}$$

Based on Equation 1 above, z_{ij} denotes the standardized value for province $i$on variable j, x_{ij}represents the original observation, $\bar{X}_j$is the mean of variable j, and S_j

is its standard deviation. This standardization process produces a data distribution with a mean of 0 and a standard deviation of 1, ensuring that each variable contributes equally to the distance calculation.

2.3.1 Distance matrix calculation

After standardizing the data, the next step involves computing the dissimilarity matrix, which represents the distance between each pair of provinces. This study employs the Euclidean distance as the distance metric, mathematically defined as:

$$\mathrm{d}(p_i p_j) = \sqrt{\sum_{k=1}^{m}(z_{ik} - z_{jk})} \quad \text{(Eq. 2)}$$

Based on Equation 2 above, $d(p_i, p_j)$ represents the Euclidean distance between province $i$and province j; z_{ik} and z_{jk}denote the standardized values for provinces $i$and $j$on variable k; and $m$is the number of variables used in the analysis (in this study, $m = 3$). Given $n = 34$ provinces, this computation produces a 34 × 34 symmetric distance matrix, in which the diagonal elements are zero since the distance of a province to itself is null. In total, $\frac{n(n-1)}{2} = \frac{34\times33}{2} = 561$ unique distance pairs are calculated to construct the dissimilarity matrix.

The choice of euclidean distance as the distance metric in this study is supported by several methodological considerations. According to Wang et al. (2023) euclidean distance is the most widely used measure in cluster analysis due to its straightforward geometric interpretation. It is sensitive to scale differences across dimensions, making it suitable after data standardization to ensure equal variable contributions. It also offers an intuitive geometric meaning as the straight-line distance between two points in a multidimensional space. Furthermore, it is compatible with Ward's linkage method, as both rely on minimizing within-cluster variance, resulting in more homogeneous groupings.

2.3.2 Implementation of agglomerative algorithms

At this stage, the agglomerative hierarchical algorithm is executed using a specific *linkage* criterion to determine how inter-cluster distances are computed after each merging step. This study employs Ward's method as the primary linkage criterion, given its proven effectiveness in producing compact and homogeneous clusters (Li et al., 2022). Ward's method merges two clusters based on the criterion of minimizing the increase in within-cluster variance. For each cluster C, the internal variance is measured using the Sum of Squared Errors (SSE), defined as:

$$SSE_C = \sum_{i\epsilon C}\|x_i - \mu_c\|^2 \quad \text{(Eq. 3)}$$

Based on Equation 3 above, SSE_C is sum of squared errors for cluster k, representing the total within-cluster variance; x_i is observation vector for province i in cluster c_k, containing the standardized values of Environmental Quality Index (EQI), GDP, and Population Density; μ_c is centroid (multivariate mean) of cluster, calculated as the average of all observation vectors in the cluster; $\|x_i - \mu_c\|^2$ is quared Euclidean distance between observation x_i and centroid μ_c, and $\sum i\epsilon c$ is summation over all provinces i belonging to cluster. A lower SSE value indicates that provinces within the cluster are more similar to each other (i.e., tightly grouped around the centroid), while a higher SSE suggests greater internal heterogeneity. This metric serves as the foundation for Ward's decision-making process in determining which clusters to merge at each iteration. When two clusters $C_a$and $C_b$are merged into a new cluster C_{ab}, the increase in SSE is expressed as:

$$\Delta SSE(C_a, C_b) = SSE_{C_{ab}} - (SSE_{C_a} + SSE_{C_b}) \quad \text{(Eq. 4)}$$

 https://doi.org/10.61511/spms.v3i2.2026.3164

Based on Equation 4 above, $\Delta SSE(C_a, C_b)$ is the increase in total within-cluster variance resulting from merging C_a and C_b; $SSE_{C_{ab}}$ is the sum of squared errors of the newly formed merged cluster CC_{ab}, which contains all provinces from both C_a and C_b; and $(SSE_{C_a} + SSE_{C_b})$ is the combined variance of the two separate clusters before merging. The ΔSSE represents the "cost" of merging two clusters in terms of lost homogeneity. Before merging, each cluster has its own centroid, and provinces within each cluster are relatively close to their respective centroids. When the two clusters are merged, a new combined centroid is calculated, and provinces that were previously close to their original centroids may now be farther from the new merged centroid. This redistribution of distances results in an increase in total variance. Ward's method consistently merges the pair of clusters that yields the smallest ΔSSE at each iteration. Computationally, the increase in variance can be reformulated as:

$$\Delta SSE(C_a, C_b) = \frac{n_a\, n_b}{n_a + n_b} \parallel \mu_{C_a} - \mu_{C_b} \parallel^2 \qquad \text{(Eq. 5)}$$

Based on Equation 5 above, n_a and n_b represent the number of elements in clusters C_a and C_b, respectively. This formulation highlights Ward's focus on minimizing the total within-cluster variance, thus generating clusters that are compact and statistically meaningful.

2.3.3 Iterative merging process

This stage describes the stepwise (*iterative*) formation of clusters in the Hierarchical Agglomerative Clustering (HAC) algorithm using Ward's linkage criterion. In Iteration 1, initially, each province is treated as an individual cluster ($n = 34$). The algorithm computes the increase in within-cluster variance (ΔSSE) for all possible $\frac{n(n-1)}{2} = 561$ cluster pairs. The two clusters with the smallest ΔSSE for example, provinces A and B with the most similar characteristics are merged into a new cluster $\{A, B\}$. In Iteration 2, after the first merge, the number of clusters decreases to 33. The distance matrix is then updated by recalculating the distances between the new cluster $\{A, B\}$ and the remaining 32 clusters using Ward's formula. The process continues by selecting the next pair of clusters that yields the smallest increase in variance. In iteration 3, the merging procedure is repeated iteratively until the 33rd iteration, at which point all provinces are combined into a single cluster representing the complete dataset. The entire merging sequence is recorded in a hierarchical structure, which is then visualized through a dendrogram to facilitate interpretation.

2.3.4 Dendrogram visualization

The dendrogram is a tree-like graphical representation illustrating the sequence and levels of cluster merging. The vertical axis indicates the distance or increase in within-cluster variance (ΔSSE) at each merging step, meanwhile the horizontal axis displays the labels of the provinces. Higher merging levels on the vertical axis represent greater dissimilarities between the clusters being joined. Consequently, the dendrogram provides a visual means to identify the optimal number of clusters based on the largest vertical distance (cut height) before a significant increase in variance occurs.

2.3.5 Cluster characterization and profiling

Once the optimal number of clusters has been determined, an in-depth characterization and profiling analysis is performed to describe the unique attributes and internal consistency of each cluster. This analysis aims to identify general patterns, within-cluster similarity, and between-cluster variability. Each cluster is characterized by its centroid

(multivariate mean) and standard deviation for the key variables. Cluster centroid represents the average position of a cluster in the multidimensional space and is calculated as follows:

$$\mu_{C_j} = \frac{1}{n_j} \sum_{i \in C_j} x_i \quad \text{(Eq. 6)}$$

Based on Equation 6 above, μ_{C_j}denotes the centroid of cluster j, $n_j$is the number of observations in cluster j, and $x_i$represents the observation vector of the i^{th}member. Furthermore, cluster standard deviation, the degree of homogeneity within each cluster is measured by its standard deviation. A smaller σ_{C_j}indicates higher internal consistency, while a larger value reflects greater internal heterogeneity.

$$\sigma_{C_j} = \sqrt{\frac{1}{n_j} \sum_{i \in C_j} (x_i - \mu_{C_j})^2} \quad \text{(Eq. 7)}$$

3. Results and Discussion

3.1 Descriptive statistics and data characteristics

The descriptive analysis indicates substantial variation across provinces in the three main variables: Environmental Quality Index (EQI), Gross Regional Domestic Product (GRDP), and Population Density (KP). A summary of the descriptive statistics is presented in Table 2.

Table 2. Descriptive statistics of research variables

Variable	Max	Min	Mean	Std. Dev
EQI	84.220	49.700	71.881	7.696
GRDP	7,525	1,020	3,167.82	1,942.47
PD	16146	10	752.65	2749.1

The EQI values range from 49.70 to 84.22, with a mean of 71.88 and a standard deviation of 7.70, indicating generally good environmental quality and relatively low interregional variation. GRDP ranges from IDR 1,020 billion to IDR 7,525 billion, with a mean of IDR 3,167.82 billion and a standard deviation of 1,942.47, suggesting considerable economic disparities among provinces. Meanwhile, population density (PD) varies widely, from 10 to 16,146 people/km^2, with a mean of 752.65 and a standard deviation of 2,749.1, reflecting substantial demographic differences across regions in Indonesia.

3.2 Multicollinearity test

Before proceeding to the clustering stage, a multicollinearity test was conducted to ensure that the selected variables were not excessively linearly correlated. High multicollinearity can lead to information redundancy and reduce the effectiveness of the clustering process, as highly correlated variables essentially capture the same underlying dimension.

Table 3. Multicollinearity test result (variance inflation factor)

Variable	VIF
Gross Regional Domestic Product	1.020
Population Density	1.020
Environmental Quality Index (EQI)	1.020

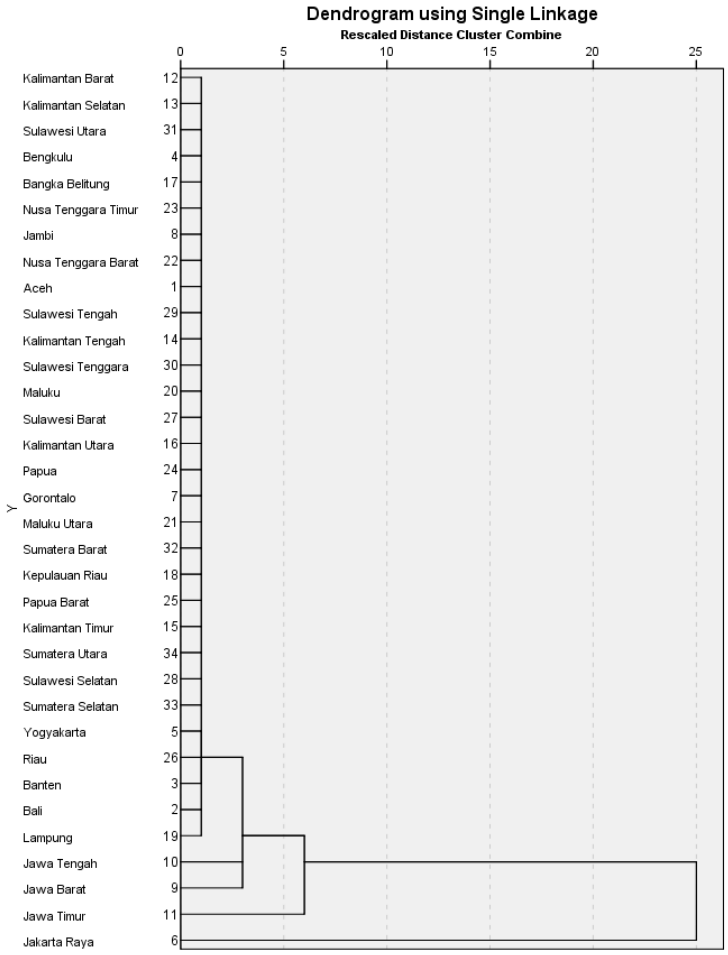
 https://doi.org/10.61511/spms.v3i2.2026.3164

In this analysis, the hypotheses were formulated as follows: H_0, where there is no multicollinearity among the variables and H_1, where multicollinearity exists among the variables. Based on the results presented in Table 3, all variables exhibit VIF values below 10, These findings indicate that H_0 is accepted, meaning there is no significant multicollinearity issue within the dataset. Consequently, all variables can be reliably included in the clustering analysis.

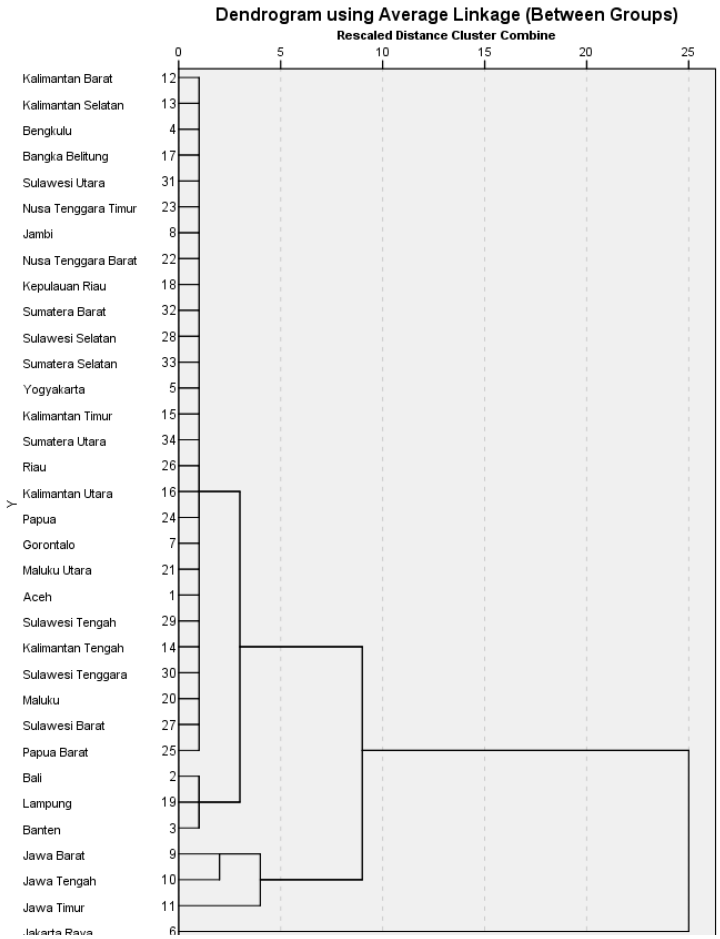
3.3 *Optimal linkage method selection*

The selection of an appropriate linkage method is a crucial stage in hierarchical clustering, as different linkage criteria can yield distinct clustering structures. This study compared four commonly used linkage methods Single Linkage, Average Linkage, Complete Linkage, and Ward's Method through visual evaluation of the resulting dendrograms.

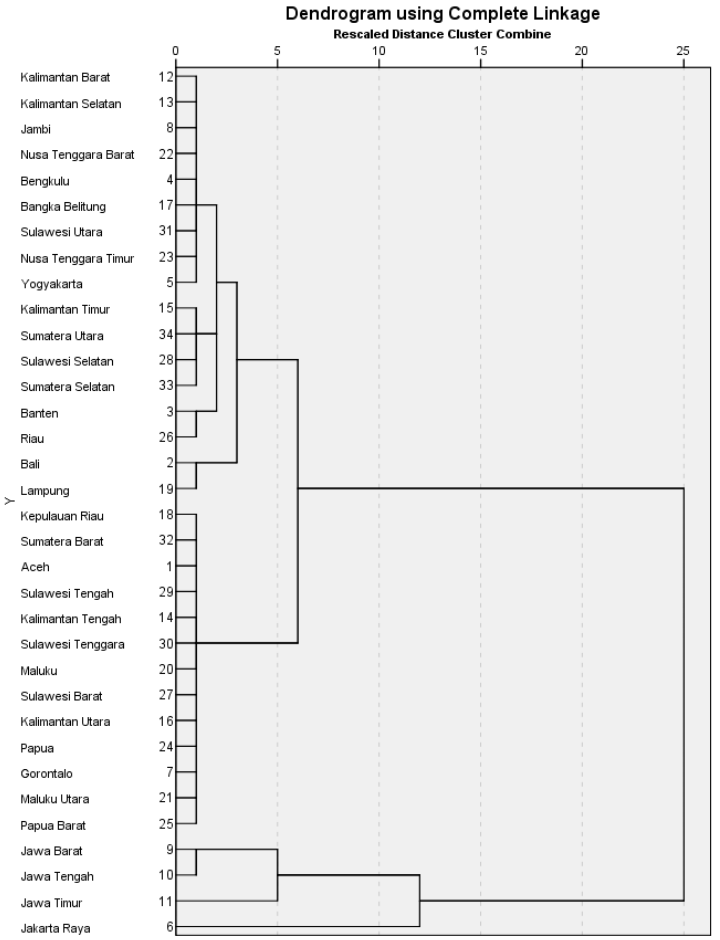
a)



b)


Dendrogram using Average Linkage (Between Groups)
Rescaled Distance Cluster Combine
0
5
10
15
20
25
Kalimantan Barat 12
Kalimantan Selatan 13
Bengkulu 4
Bangka Belitung 17
Sulawesi Utara 31
Nusa Tenggara Timur 23
Jambi 8
Nusa Tenggara Barat 22
Kepulauan Riau 18
Sumatera Barat 32
Sulawesi Selatan 28
Sumatera Selatan 33
Yogyakarta 5
Kalimantan Timur 15
Sumatera Utara 34
Riau 26
Kalimantan Utara 16
Papua 24
Gorontalo 7
Maluku Utara 21
Aceh 1
Sulawesi Tengah 29
Kalimantan Tengah 14
Sulawesi Tenggara 30
Maluku 20
Sulawesi Barat 27
Papua Barat 25
Bali 2
Lampung 19
Banten 3
Jawa Barat 9
Jawa Tengah 10
Jawa Timur 11
Jakarta Raya 6
Y

c)


Dendrogram using Complete Linkage
Rescaled Distance Cluster Combine
0
5
10
15
20
25
Kalimantan Barat 12
Kalimantan Selatan 13
Jambi 8
Nusa Tenggara Barat 22
Bengkulu 4
Bangka Belitung 17
Sulawesi Utara 31
Nusa Tenggara Timur 23
Yogyakarta 5
Kalimantan Timur 15
Sumatera Utara 34
Sulawesi Selatan 28
Sumatera Selatan 33
Banten 3
Riau 26
Bali 2
Lampung 19
Kepulauan Riau 18
Sumatera Barat 32
Aceh 1
Sulawesi Tengah 29
Kalimantan Tengah 14
Sulawesi Tenggara 30
Maluku 20
Sulawesi Barat 27
Kalimantan Utara 16
Papua 24
Gorontalo 7
Maluku Utara 21
Papua Barat 25
Jawa Barat 9
Jawa Tengah 10
Jawa Timur 11
Jakarta Raya 6
Y

d)

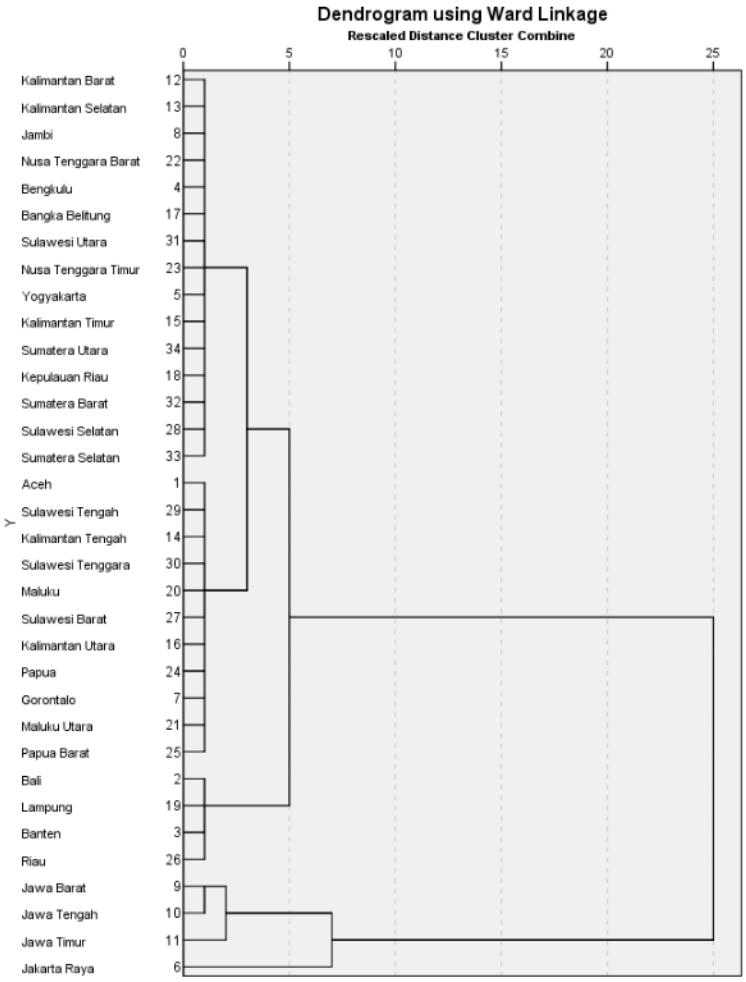


Fig. 2 (a-d). The dendrograms of single, average, complete, and ward's linkage methods

The visual assessment revealed notable differences among the four approaches. The Single Linkage method (Figure a) exhibited a pronounced *chaining effect*, in which most provinces were gradually merged into one large cluster without clear differentiation. This phenomenon represents a well-known limitation of Single Linkage, which tends to join observations based solely on the smallest pairwise distances, disregarding the overall cluster structure. The Average Linkage method (Figure b) provided a moderate improvement, producing a more balanced structure and somewhat clearer cluster separations. Nevertheless, inter-cluster differentiation within larger groupings remained relatively indistinct, making it difficult to interpret provincial typologies in a detailed manner.

The Complete Linkage method (Figure c) yielded a more distinct hierarchical structure with improved separation between clusters compared to the previous two methods. However, it produced one disproportionately large cluster (comprising 22 provinces) alongside several smaller ones, suggesting an imbalance in cluster size distribution. Conversely, Ward's Method (Figure d) demonstrated the most robust performance among the four techniques. Its dendrogram displayed the clearest hierarchical organization with well-defined, balanced, and interpretable cluster separations. These characteristics make Ward's method particularly suitable for identifying homogeneous and compact clusters in the context of this study.

Table 4. Comparison of linkage methods

Criterion	Single Linkage	Average Linkage	Complete Linkage	Ward's Method
Hierarchical Structure	Very Poor	Moderate	Good	Very Good
Chaining Effect	Extreme	Mild	None	None
Cluster Balance	Highly Unbalanced	Fairly Balanced	Unbalanced	Balanced
Interpretability	Difficult	Moderate	Fairly Easy	Very Easy

From a theoretical standpoint, the superiority of Ward's Method can be explained by its principle of minimizing the increase in *within-cluster variance* at each merging step. This ensures that provinces combined within the same cluster share high similarity, producing more compact and homogeneous groupings. Based on both the visual evaluation of the dendrograms and the theoretical considerations, Ward's Method is identified as the most optimal linkage technique for this clustering analysis (Fernández & Gómez, 2023). It's clear and interpretable hierarchical structure facilitates the identification of the optimal number of clusters.

3.4 Clustering in mapping

The clustering analysis of Indonesian provinces, using the four-cluster solution selected in Section 3.3 based on the agglomeration-schedule fusion coefficients, reveals a spatial distribution pattern with four distinct groups based on environmental quality, GRDP, and population density. Cluster 3 (16 provinces) is characterized by the highest environmental quality (EQI 75.23) combined with the highest GRDP (IDR 4,953.13 billion). Cluster 1 (14 provinces) shows moderate-to-high environmental quality (EQI 72.68) but comparatively low GRDP (IDR 1,448.29 billion). Cluster 2 (3 provinces) shows the lowest environmental quality among the multi-province clusters (EQI 55.95). Cluster 4 is a single-province outlier (EQI 54.93) with an extreme population density (16,146 people/km^2), consistent with DKI Jakarta. Cluster 4, the single-province outlier with extreme population density, is consistent with DKI Jakarta, which anchors the high-urbanization, high-density profile also reflected in Cluster 2.

Table 5. Cluster profiling

Cluster	EQI	GRDP	PD	Description
1 (n = 14)	72.68	1,448.29	353.50	Moderate–High EQI, Low GRDP
2 (n = 3)	55.95	2,043.33	643.33	Low EQI
3 (n = 16)	75.23	4,953.13	160.31	High EQI, High GRDP
4 (n = 1)	54.93	2,050.00	16,146.00	Low EQI, extreme density outlier (Jakarta)

Cluster 3 shows the highest GRDP together with the highest EQI, while Cluster 2 shows low EQI with only moderate GRDP. This is the opposite pairing from the original Environmental Kuznets Curve narrative used throughout this manuscript. This paragraph, the "A More Empirical Explanation" subsection, and the Conclusion need to be rewritten to reflect the corrected relationship, or the EKC framing should be reconsidered, before this section is finalized. According to Table 5, Cluster 3 combines the highest environmental quality with the highest economic output, while Cluster 2 and Cluster 4 show comparatively low environmental quality without correspondingly high GRDP. This pattern reflects considerable variation in ecological, economic, and demographic characteristics across regions in Indonesia.

The clustering results show that provinces on Java, despite their high income levels, have lower environmental quality. This pattern resembles the tendency described by the Environmental Kuznets Curve (EKC), where environmental degradation increases in the

early stages of economic growth and may decline once economies shift toward cleaner and more efficient sectors (Lin & Reuvers, 2020). While this study does not statistically test the EKC hypothesis, the spatial clustering outcomes indicate a potential EKC like pattern across Indonesian provinces.

3.5 A more empirical explanation

The dominance of Java Island in Indonesia's economic and population structure provides strong empirical support for this study's findings. The country's population distribution exhibits a pronounced imbalance, with over 60 percent of Indonesia's total population residing on Java (Ipsos Global Health Index, 2024). Despite accounting for only about 7% of the country's total land area, Java accommodates more than half of its inhabitants. Major urban centers such as Jakarta, Bandung, Semarang, and Surabaya have emerged as rapidly expanding economic and urban hubs.

This massive demographic concentration exerts significant ecological pressure. In 2018, Java's population growth rate reached 1.23%, with a population density of 1,317 people per km^2, and 56.7% of its residents living in urban areas. These figures far exceed the national average, indicating compounded stress on natural resources and the environment's capacity to absorb waste and pollution. From a production and consumption perspective, high industrial activity, dense urbanization, and urban consumption patterns directly intensify resource exploitation and environmental degradation. Java serves not only as the demographic core but also contributes to over half of Indonesia's national GRDP, explaining the concentration of the country's ecological footprint within the island (Teuku et al., 2023).

The causal relationship between economic activity and environmental degradation on Java was empirically validated during the COVID-19 pandemic. Several studies like Lukman et al (2025) analyzed variations in pollutant density before and during the pandemic across Java's provinces and assessed the impact of large-scale mobility restrictions on economic growth. The results revealed that when economic activity sharply declined, emissions dropped by approximately 6–8%, leading to notable improvements in urban air quality. This improvement was primarily attributed to reduced public mobility and lower fuel consumption, confirming that the decline in economic and transport activities has a direct effect on environmental quality.

The question arises whether provinces outside Java may follow a similar environmental path. The clustering results indicate that these regions currently characterized by higher EQI and lower economic activity display early-stage development profiles that could evolve toward higher environmental pressure as growth accelerates. Although the present analysis does not empirically validate the Environmental Kuznets Curve (EKC) hypothesis, the spatial grouping pattern reflects a potential EKC like trajectory, implying that environmental degradation may initially rise with increasing income before any future improvement can occur through structural transformation (Muhammad et al., 2025)

4. Conclusions

This study identifies four spatially differentiated clusters of environmental-economic characteristics across Indonesia's 34 provinces using agglomerative hierarchical clustering. The results demonstrate substantial heterogeneity in Environmental Quality Index (EQI), Gross Regional Domestic Product (GRDP), and population density across provinces. Cluster 1 is characterized by moderate-to-high environmental quality and relatively low GRDP, Cluster 2 by low environmental quality and moderate economic output, Cluster 3 by high environmental quality and high GRDP, and Cluster 4 by low environmental quality and an extreme population-density profile represented by DKI Jakarta.

These findings indicate that the relationship between economic activity and environmental quality is spatially heterogeneous and cannot be interpreted through a

simple linear relationship or as direct evidence of the Environmental Kuznets Curve. In particular, the coexistence of high GRDP and high EQI in Cluster 3 suggests that economic development can be compatible with relatively favorable environmental conditions under specific regional circumstances. Meanwhile, provinces characterized by low environmental quality require differentiated interventions based on their economic and demographic conditions.

From a land and environmental management perspective, the clustering results provide a spatial basis for prioritizing management strategies. Provinces with high population density and lower environmental quality require stronger land-use planning, pollution control, and environmental management, whereas provinces with relatively high environmental quality should emphasize preventive and sustainable land-resource management as economic activity expands.

The study has several limitations. First, the analysis uses cross-sectional data for 2023 and therefore does not capture temporal dynamics or lagged relationships between economic development and environmental quality. Second, the analysis is based on three variables EQI, GRDP, and population density, and does not incorporate other potentially important factors such as energy intensity, land-use change, technological adoption, or environmental regulation. Future research should employ panel data, additional environmental and land-management indicators, and spatial econometric approaches to further investigate the mechanisms underlying the observed provincial disparities.

Acknowledgement

The author declares that no specific individual or organization contributed to this research beyond the roles described in the manuscript.

Author Contribution

Conceptualization, A.O.P.; Methodology, A.O.P.; Software, A.O.P.; Validation, A.O.P.; Formal Analysis, A.O.P.; Investigation, A.O.P.; Resources, A.O.P.; Data Curation, A.O.P.; Writing – Original Draft Preparation, A.O.P.; Writing – Review & Editing, A.O.P.; Visualization, A.O.P.; Supervision, A.O.P.; Project Administration, A.O.P. The author has read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Ethical Review Board Statement

Not available. This study did not involve humans or animals and relied solely on publicly available secondary data from government institutions.

Informed Consent Statement

Not available. This study did not involve human participants.

Data Availability Statement

The data supporting the findings of this study are publicly available from the Central Statistics Agency at https://www.bps.go.id and Bank Indonesia at https://www.bi.go.id. The processed datasets used in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest.

Declaration of Generative AI Use

During the preparation of this work, the author used Gemini and Chatgpt to assist in improving grammar, sentence structure, and academic tone of the manuscript. After using

this tool, the author reviewed and edited the content as needed and took full responsibility for the content of the publication. The AI tool was not used for data analysis, interpretation of results, or generation of research findings.

Open Access

©2026. The author. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third-party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit: http://creativecommons.org/licenses/by/4.0/

References

Aiginger, K., & Rodrik, D. (2020). Rebirth of industrial policy and an agenda for the twenty-first century. *Journal of Industry, Competition, and Trade 20,* 189-207. https://doi.org/10.1007/s10842-019-00322-3

Al-Mulali, U., & Ozturk, I. (2015). The effect of energy consumption, urbanization, trade openness, industrial output, and political stability on environmental degradation in the MENA region. *Energy, 84*, 382–389. https://doi.org/10.1016/j.energy.2015.03.004

Asian Development Bank. (2023). Technical assistance consultant's report: Strategic environmental and social assessment (SESA) of the energy transition in indonesia. https://www.adb.org/sites/default/files/project-documents/55124/55124-001-tacr-id.pdf

Bechauf, R. (2021). How can indonesia foster sustainable infrastructure solutions that deliver low-carbon development and bring additional benefits?. International Institute for Sustainable Development. https://www.iisd.org/system/files/2021-07/indonesia-sustainable-infrastructure-solutions.pdf

Fernández, A., & Gómez, S. (2023). Mdendro: An R package for extended agglomerative hierarchical clustering. https://doi.org/10.18637/jss.v114.i0

Göll, E., Uhl, A., & Zwiers, J. (2020). Sustainable development in the Mena region. *MENARA Future Notes, 20,* 1-8. https://www.iai.it/en/publications/c28/sustainable-development-mena-region

Hardi, I., Afjal, M., Can, M., Idroes, G. M., Noviandy, T. R., & Idroes, R. (2024). Shadow economy, energy consumption, and ecological footprint in Indonesia. *Sustainable Futures, 8,* 100343. https://doi.org/10.1016/j.sftr.2024.100343

IEA Bioenergy. (2025). Social and environmental sustainability of municipal solid waste in the context of the SDGs. https://task36.ieabioenergy.com/publications/social-and-environmental-sustainability-of-municipal-solid-waste-in-the-context-of-the-un-sdgs/

International Monetary Fund. (2022). G-20 report on strong, sustainable, balanced, and inclusive growth. https://www.imf.org/external/np/g20/pdf/2022/111322.pdf

Ipsos Global Health Index. (2024). Indonesia 2024: Unstoppable?. https://www.ipsos.com/sites/default/files/ct/publication/documents/2024-04/FlairIndonesia2024.pdf

Li, T., Rezaeipanah, A., & Tag El Din, E. S. M. (2022). An ensemble agglomerative hierarchical clustering algorithm. *Journal of King Saud University, Computer and Information Sciences, 34*(6), 3828–3842. https://doi.org/10.1016/j.jksuci.2022.04.010

Lin, Y., & Reuvers, H. (2020). Cointegrating polynomial regressions with power law trends. http://arxiv.org/abs/2009.02262

Lukman, K. M., Hastuti, L. P., Oktavia, D., Pratiwi, D., Uchiyama, Y., & Harding, D. (2025). Towards blue skies: A comprehensive review and regional mapping of ambient air

quality in Indonesian cities. *Journal of Environmental Management, 389*. https://doi.org/10.1016/j.jenvman.2025.126132

Mohr, K., Vasudha, V., Wallenta, A., Wehnert, T., & Yetano, M. (2025). Status report of the coal sector in indonesia: Pathways and barriers towards a just energy transition. https://www.jetknowledge.org/wp-content/uploads/2025/02/Status-Report-Coal-Sector-Indonesia.pdf

Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: An overview II. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2*(1), 86–97. https://doi.org/10.1002/widm.1219

OECD (2024). *OECD economic surveys: Indonesia 2024*. OECD Publishing. https://doi.org/10.1787/de87555a-en.

Pherdd, R. K. (2026). Rethinking artisanal mining in Africa through empirical insights from Ghana and Ethiopia. *The Extractive Industries and Society*. https://doi.org/10.1016/j.exis.2025.101809

Rehman, E., & Rehman, S. (2022). Modeling the nexus between carbon emissions, urbanization, population growth, energy consumption, and economic development in Asia. Energy Reports, 8, 5430–5442. https://doi.org/10.1016/j.egyr.2022.03.179

Riefky, T., Hariyanto, E., & Alamsyah, H. (2023). Infrastructure for inclusive economic development volume 1: Lessons learnt from indonesia. ERIA and Ministry of Finance. https://www.eria.org/uploads/09-PSN-Book-vol_1-ch_4.pdf

Söderholm, P. (2020). The green economy transition: The challenges of technological change for sustainability. *Sustainable Earth, 3*(1). https://doi.org/10.1186/s42055-020-00029-y

Statistics Indonesia. (2020). Statistik Indonesia (Statistical Yearbook of Indonesia 2020). https://www.bps.go.id/id/publication/2020/04/29/e9011b3155d45d70823c141f/statistik-indonesia-2020.html

Suraya, G. R., & Wijayanto, A. W. (2022). Comparison of hierarchical clustering, k means, k-medoids, and fuzzy c-means in grouping provinces in Indonesia. *Indonesian Journal of Statistics and Its Applications, 6*(2), 180–201. https://doi.org/10.29244/ijsa.v6i2p180-201

Suwandaru, A., Sudiyono, W., Shawdari, A., & Fristin, Y. (2024). Understanding the economic drivers of climate change in Southeast Asia. *Economies, 12*(8), 1–21. https://doi.org/10.3390/economies12080200

Tokuda, E. K., Comin, C. H., & Costa, L. da F. (2022). Revisiting agglomerative clustering. *Physica A, 585*, 126433. https://doi.org/10.1016/j.physa.2021.126433

UNFPA. (2015). Population dynamics and sustainable development in Indonesia. https://indonesia.unfpa.org/sites/default/files/pub-pdf/BUKU_Indonesias_Population_Dynamic_ICS_UNFPA_FINAL.pdf

Wang, Y., Yu, S., Gu, Y., & Shun, J. (2023). Fast parallel algorithms for euclidean minimum spanning tree and hierarchical spatial clustering. HOPC 2023, 17–18. https://doi.org/10.1145/3597635.3598025

Xu, D., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science, 2*(2), 165–193. https://doi.org/10.1007/s40745-015-0040-1

Xuan, V. N. (2024). Determinants of environmental pollution: Evidence from Indonesia. *Journal of Open Innovation: Technology, Market, and Complexity, 10*(4), 100386. https://doi.org/10.1016/j.joitmc.2024.100386

Yale University. (2025). Environmental performance index 2024. https://environment.yale.edu/news/article/2024-environmental-performance-index

Biographies of Author

Antradiva Oktaviola Puspita is a researcher at the Department of Economics, Faculty of Economics and Business, Universitas Negeri Semarang, Semarang, Central Java, Indonesia. Her research interests include regional economic clustering, MSME digitalization, and the environmental-economic nexus, with a focus on spatial analysis and quantitative modeling.

- Email: antradivaoktaviolap@gmail.com
- ORCID ID: 0009-0007-6546-9298
- Web of Science ResearcherID: N/A
- Scopus Author ID: N/A
- Homepage: N/A