Jurnal Varian
Vol. 9, No. 1, February 2026, pp 39–52
e-ISSN: 2581-2017
DOI: https://doi.org/10.30812/varian.v9i1.5933

Comparing SOM, DBSCAN, and K-Affinity Propagation in Labor Economic Patterns

Wiwit Pura Nurmayanti[1], Desi Yuniarti[1], Meiliyani Siringoringo[1], Ika Purnamasari[1], Desi Febriani Putri[1], Siti Hadijah Hasanah[2]
[1]Universitas Mulawarman, Samarinda, Indonesia
[2]Universitas Terbuka, Tangerang Selatan, Indonesia

Article Info

Article history:

Received : 11-21-2025
Revised : 02-27-2026
Accepted : 03-10-2026

Keywords:

Density-Based Spatial Clustering of Applications with Noise (DBSCAN);
Economic Growth;
K-Affinity Propagation (K-AP);
Labor Force;
Self-Organizing Maps (SOM).

ABSTRACT

The objective of this research is to identify the most effective clustering method for grouping Indonesian provinces by labor–economic indicators to support more precise, data-driven policy formulation. Regional disparities in Indonesia's economic growth, driven by unequal labor characteristics, remain a significant obstacle to achieving inclusive development. An analytical approach capable of grouping provinces by labor and economic indicators is therefore essential. This study applies a comparative clustering analysis using three unsupervised algorithms: Self-Organizing Maps (SOM), Density-Based Spatial Clustering of Applications with Noise (DBSCAN), and K-Affinity Propagation (K-AP). The dataset consists of five key indicators, namely economic growth, total population, labor force, employment rate, and average wage level obtained from Statistics Indonesia (BPS) for the year 2024. The clustering performance is evaluated using internal validation criteria based on the ratio of within-cluster variation (Sw) to between-cluster variation (Sb), where a smaller ratio indicates more compact, well-separated clusters. The results show that each method produces different clustering structures. SOM and DBSCAN generate three clusters with varying provincial distributions, whereas K-AP produces five clusters with more balanced, representative groupings. The evaluation results indicate ratios of 3.1906 for SOM, 0.2000 for DBSCAN, and 0.1779 for K-AP, indicating that K-AP provides the most optimal clustering performance. These findings confirm that K-Affinity Propagation is the most effective and stable method for classifying Indonesian provinces by labor and economic characteristics. The outcomes of this study provide empirical insights and analytical references for labor-driven economic policy formulation and data-driven regional development planning in Indonesia.

Corresponding Author:

Wiwit Pura Nurmayanti,
Statistics Study Program, Universitas Mulawarman,
Email: wiwit.adiwinata3@gmail.com

How to Cite:
Nurmayanti, W. P., Yuniarti, D., Siringorino, M., Purnamasari, I., Putri, D. F., & Hasanah, S. H. (2026). Comparing SOM, DBSCAN, and K-Affinity Propagation in Labor Economic Patterns. *Jurnal Varian*, *9*(1), 39–52.

A. INTRODUCTION

In the era of digital transformation, the rapid growth of big data has significantly influenced how information is utilized to support decision-making, policy formulation, and future forecasting. The increasing demand for efficient and accessible data management has positioned data analysis as an essential tool across various fields, including economics and demography (Dwitiyanti et al., 2024; Nisrina et al., 2022). Among numerous analytical techniques, clustering plays a crucial role in pattern recognition and data organization, particularly within the framework of unsupervised learning (Mason et al., 2023; Oyewole & Thopil, 2023). The main goal of clustering is to categorize data objects according to their similarity, forming groups that exhibit strong internal uniformity while remaining clearly distinct from one another (Lopez et al., 2018).

The application of clustering techniques in labor economics has become increasingly relevant in developing countries such as Indonesia. With a labor force of over 52 million (BPS, 2024), Indonesia faces structural challenges, including unequal labor distribution, regional wage disparities, and differences in economic productivity across provinces. Conversely, the ongoing demographic bonus presents a strategic opportunity to accelerate economic growth if managed through data-driven and well-targeted policies (Hosan et al., 2022). Therefore, this study employs several labor and economic indicators, namely the open unemployment rate (TPT), labor force participation rate (TPAK), total labor force, and average net income of workers, as primary variables in the clustering process (Zhang et al., 2022). Clustering based on these variables is expected to reveal spatial patterns in economic growth and labor inequality across provinces, thereby supporting the development of more targeted, inclusive, and sustainable economic policies.

Several previous studies have applied clustering algorithms in socio-economic and regional analysis. Wati et al. (2024) aimed to identify spatial earthquake patterns by comparing DBSCAN and SOM, applied density-based and topology-based clustering approaches, and found that while DBSCAN effectively captured spatial density patterns, its performance was highly sensitive to parameter selection, whereas SOM provided clearer spatial mapping. Hastuti et al. (2023) aimed to classify agricultural economic indicators using SOM and K-Affinity Propagation, employed exemplar-based clustering, and reported that K-AP generated more stable and representative clusters due to its automatic exemplar selection mechanism. Nurmayanti et al. (2022) aimed to group heterogeneous health insurance participant data using DBSCAN and demonstrated that density-based clustering effectively handled socio-demographic variability. Although these studies confirm the usefulness of clustering techniques in socio-economic contexts, they are generally limited to single-method applications or comparisons between only two algorithms. They are not specifically designed to analyze provincial economic growth patterns driven by labor indicators.

The research gap between this study and previous works is that prior research has not conducted a comprehensive comparative evaluation of three fundamentally different clustering paradigms: topology-based (SOM), density-based (DBSCAN), and exemplar-based (K-AP), within the context of labor-driven economic growth at the provincial level (Jaeger & Banks, 2023; Oyewole & Thopil, 2023) Existing studies also lack a quantitative evaluation framework that objectively determines the most appropriate clustering method for regional economic analysis (Hassan et al., 2024). The difference between this study and previous ones lies in the integration of these three clustering approaches within a single analytical framework, combined with an internal validation criterion based on the ratio of intra-cluster and inter-cluster variation (Sw/Sb). This allows an objective and systematic comparison of clustering performance when applied to macroeconomic and labor indicators simultaneously (Diao et al., 2019; Iammarino et al., 2019). This novelty aligns with recent trends in data science that emphasize multi-algorithm benchmarking and objective internal validation to ensure robustness and reproducibility of clustering results in complex socio-economic datasets (Ikotun et al., 2025; Jaeger & Banks, 2023; Oyewole & Thopil, 2023).

The contribution of this study is the provision of an empirical and methodological comparison of SOM, DBSCAN, and K-Affinity Propagation for classifying Indonesian provinces based on labor–economic characteristics, accompanied by a quantitative evaluation scheme to identify the most optimal clustering structure (Hastuti et al., 2023; Wati et al., 2024). By situating this analysis within recent clustering literature, this study underscores the novelty of applying multi-method clustering to labor-based economic classification at the national scale and offers practical implications for data-driven regional policy formulation (Agustino et al., 2023; Iammarino et al., 2019; Zhang et al., 2022).

This research employs three clustering techniques: Self Organizing Maps (SOM), Density Based Spatial Clustering of Applications with Noise (DBSCAN), and K-Affinity Propagation (K-AP) to examine Indonesia's economic growth in relation to labor-driven indicators. SOM, originally proposed by Teuvo Kohonen in 1981, is an unsupervised neural network that projects multidimensional data onto a lower-dimensional map through an adaptive learning process (Melin et al., 2020). DBSCAN, in contrast, forms clusters by evaluating data density using two essential parameters: ε (radius of the neighborhood) and minPts (minimum required points) (Agustino et al., 2023; Sorkhi et al., 2024). Meanwhile, K-Affinity Propagation (K-AP), an enhancement of the standard Affinity Propagation algorithm, is designed to automatically determine the ideal number of exemplars and improve the robustness of clustering results (Selmi et al., 2025).

This research seeks to characterize the clustering patterns of Indonesian provinces in 2024 using economic and labor indicators, identify the optimal cluster count for each algorithm, and assess their performance using ratio indices, intra-cluster variation (Sw), and inter-cluster variation (Sb) to determine the most effective clustering approach.

B. RESEARCH METHOD

This study employed a quantitative cross-sectional research design, using provincial-level socioeconomic data from Statistics Indonesia (Badan Pusat Statistik/BPS) for 2024 (Tenaga Kerja - Tabel Statistik - Badan Pusat Statistik Indonesia). The dataset consists of annual aggregated data representing provincial labor and economic conditions rather than time-series, daily, or monthly observations. The unit of analysis comprises 38 provinces in Indonesia. The clustering of economic growth based on labor indicators was conducted using several key variables: economic growth rate, total population, labor force, Labor Force Participation Rate (TPAK), Open Unemployment Rate (TPT), and average net income of workers.

The stages of the comparative data analysis using the SOM, DBSCAN, and K-AP methods to cluster the economic growth of Indonesian provinces based on labor indicators are as follows:

1. Data Collection
   The dataset was sourced from the official platform of Indonesia's Badan Pusat Statistik (BPS).
2. Data Pre-processing
   This stage aimed to clean the dataset by removing incomplete, duplicate, inconsistent, outlier, or missing values to ensure data quality and accuracy.
3. Descriptive Analysis
   Descriptive analysis was performed using graphs, tables, and plots to provide an overview of the dataset characteristics.
4. Data Standardization
   Data standardization was performed to normalize variables with different measurement units using the Z-score method.
5. Euclidean Distance Calculation
   To assess the degree of similarity among data objects, Euclidean distance was employed and computed using the formula below:

$$d_{Euclid} = \sqrt{\sum_{i=1}^{n} (x_{ik} - x_{jk})^2} \tag{1}$$

   Where d_{Euclid} refresents the Euclidean distance between data objects x_i and xj, n denotes the total number of variables, x_{ij} indicates the value of object i on variable k, and x_{jk} represent the value of object j on variable k.
6. SOM Clustering Analysis
   a. Initialize random weight values W_{ij}
   b. Set cluster parameters (m) and learning rate (α)
   c. For each input vector x, compute the Euclidean distance

$$D_{(j)} = \sum_{ij}^{n} (w_{ij} - x_i)^2 \tag{2}$$

   Where $D_{(j)}$ represent the distance for each j, w_{ij} denotes the weight connecting input node i and output node j, x_i is the input value at node i, and n indicates the total number of input-layer nodes.
   d. Identify the node j that yields the minimum $D_{(j)}$
   e. Update the weights using:

$$W_{ij(new)} = w_{ij(old)} + \alpha(x_i - w_{ij(old)}) \tag{3}$$

   f. Update all units within the neighborhood of j according to the neighborhood function.
7. DBSCAN Clustering Analysis
   a. Initialize the parameters ε (epsilon) and minimum points (minPts)
   b. The optimal epsilon parameter was determined using the knee of the curve method
   c. Several combinations of epsilon and minPts were tested, and the best values were selected based on the highest Silhouette Coefficient score.
8. K-AP Clustering Analysis
   a. Compute the similarity matrix:

$$s\,(i,k) = -Xi - Xk2 \tag{4}$$

   b. Initialize the availability matrix to zero:

$$r\,(i,k) \leftarrow s\,(i,k) - max_{k\prime \in k}\,\{a\,(i,k') + s(i,k')\} \tag{5}$$

DOI: https://doi.org/10.30812/varian.v9i1.5933

c. Compute the responsibility value:

$$a(i,k) \leftarrow min\left\{0, r\left(k,k\right) + \sum_{i\prime \notin \{i,k\}} max\left\{0, r(i',k\right\}\right\} \tag{6}$$

d. Update availability:

$$\eta^{in}\left(i\right) = a\left(i,i\right) - max_{k':k\prime \notin i}\left\{a\left(i,k'\right) + s(i,k')\right\} \tag{7}$$

e. Update confidence:

$$\eta^{in}\left(i\right) = -R\left(\left\{\eta^{in}\left(j\right), j \neq i\right\}\right) \tag{8}$$

f. Combine availability and responsibility values to obtain cluster assignment:

$$c\left(i,k\right) = arg\ max_j\left\{a\left(i,k\right) + r(i,k)\right\} \tag{9}$$

9. Cluster Evaluation
   To determine the optimal clustering method for grouping economic growth by labor indicators, each algorithm was evaluated using ratio analysis, the intra-cluster standard deviation (Sw), and the inter-cluster standard deviation (Sb).
10. Conclusion.

The analysis was conducted using Microsoft Excel and R Studio Version 3.6.3. The overall workflow of the research analysis process is summarized in Figure 1 (Flowchart).

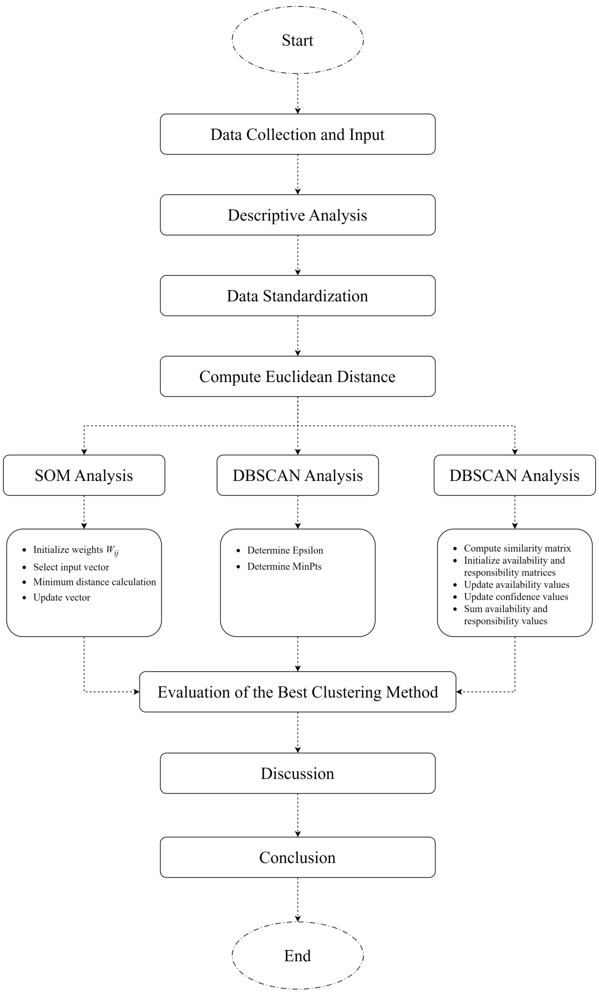

Figure 1. Flowchart of the Data Analysis Process

C. RESULT AND DISCUSSION

1. Descriptive Analysis

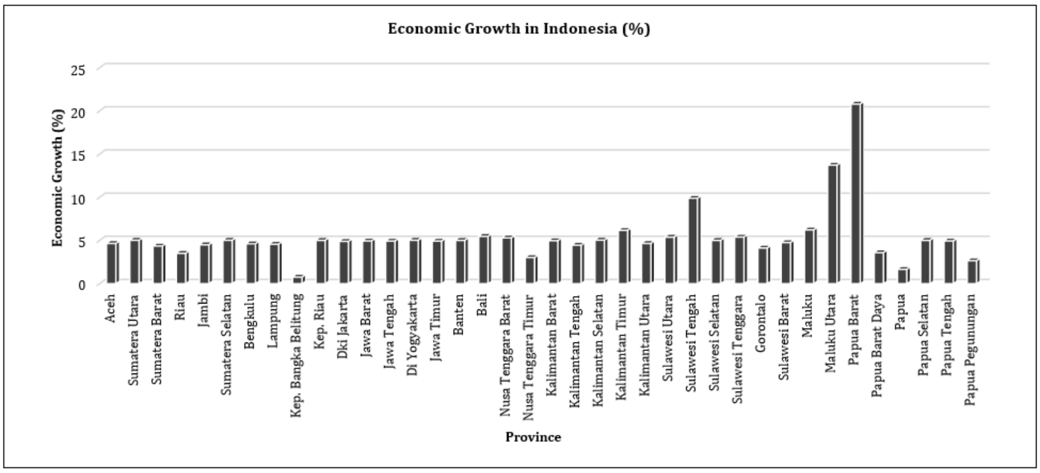

Figure 2. Economic Growth in Indonesia

Based on Figure 2, West Papua records the highest economic growth in Indonesia at 20.8%, far exceeding the national average of 5.03%. From the perspective of regional development theory, this pattern reflects the characteristics of regions in the early phase of economic expansion, where growth acceleration is typically driven by intensive capital investment and resource exploitation rather than balanced income distribution, as explained in the Growth Pole Theory and the Core–Periphery framework. Such rapid growth is often associated with extractive industries and limited labor absorption, suggesting that high economic growth does not necessarily translate into improvements in labor welfare or employment quality. This condition highlights the structural imbalance between economic output and labor-based development (Iammarino et al., 2019). In contrast, Bangka Belitung Islands Province demonstrates the lowest economic growth at 0.77%, indicating a slowdown compared to the previous year. The province's economic performance is mainly driven by tin mining and agriculture, sectors that are vulnerable to productivity stagnation and limited technological advancement. From the viewpoint of labor economics, this condition aligns with the Lewis Dual Sector Model, where regions remain dependent on traditional sectors with low labor productivity and limited wage growth (De Vries et al., 2015; Diao et al., 2019). These contrasting conditions imply different policy priorities across clusters. Provinces such as West Papua require policies that focus on improving labor absorption, workforce skills, and income distribution to ensure that rapid economic growth translates into inclusive labor welfare. Meanwhile, provinces such as Bangka Belitung need structural transformation policies to increase labor productivity, promote sectoral diversification, and develop human capital to escape low-growth traps. Therefore, the clustering results not only describe economic patterns but also provide a basis for differentiated, labor-oriented regional development policies.

2. Self-Organizing Map (SOM) Analysis

Self-Organizing Maps (SOMs) are employed to perform data clustering and organize variables into groups exhibiting similar patterns through a visual mapping process. The selection of the most appropriate cluster configuration is assessed using internal validation metrics. Three internal validation methods, Dunn Index, Silhouette, and Connectivity, are employed to establish the appropriate number of clusters for the SOM method. A Dunn Index value approaching 1, the highest Silhouette value, and the lowest Connectivity value are used as criteria for selecting the optimal number of clusters. A comparison of these three internal validation measures is presented in the following Table 1.

Table 1. Internal Cluster Validation

		3	**4**	**5**
SOM	Connectivity	10.4607	16.6571	19.2405
	Dunn	0.3123	0.0611	0.0611
	Silhouette	0.6171	0.4788	0.4753

Based on Table 1, the internal cluster validation using the three methods indicates that cluster solutions with 3, 4, and 5

DOI: https://doi.org/10.30812/varian.v9i1.5933

clusters should be evaluated to determine the optimal structure. The results show that the optimal number of clusters for the SOM analysis is three, as this configuration yields the lowest connectivity value and the highest Dunn and Silhouette indices compared to the other cluster solutions.

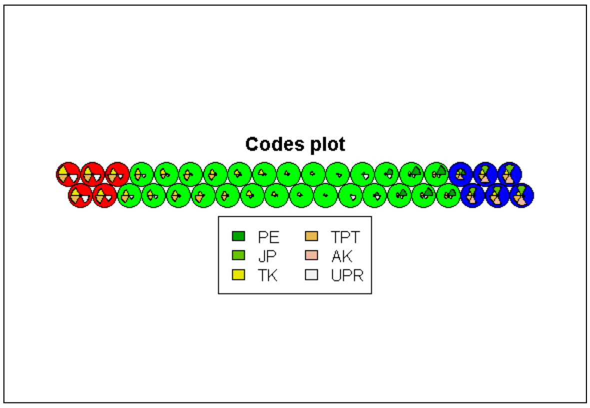

Figure 3. SOM Clustering Topology

Figure 3 shows the Self-Organizing Maps (SOM) algorithm as a code plot. In this SOM topology, each node (neuron) reflects a prototype vector that represents the characteristics of the input variables. At the same time, different colors and boundary separations indicate the cluster structure formed during training. The variation in color intensity within each node illustrates the relative magnitude of each variable (PE, JP, TK, TPT, AK, and UPR), allowing for a comparative interpretation of variable dominance across clusters. The spatial arrangement of the nodes demonstrates how similar provinces are mapped closer together, while dissimilar provinces are positioned farther apart. Therefore, the SOM visualization not only highlights cluster membership but also reveals underlying patterns and relationships among economic and labor indicators across provinces. The cluster groups are presented in Table 2:

Table 2. Members of the SOM Clustering

Cluster	Members
Cluster 1	Aceh, North Sumatra, West Kalimantan, Central Kalimantan, South Kalimantan
Cluster 2	West Sumatera, Riau, Jambi, South Sumatera, Bengkulu, Lampung, Kep. Bangka Belitung, Kep.Riau, DKI Jakarta, West Java, Central Java, DI Yogyakarta, East Java, Banten, North Sulawesi, Central Sulawesi, South Sulawesi, Sulawesi Tenggara, Gorontalo, West Sulawesi, East Kalimantan, North Kalimantan, Maluku, North Maluku, West Papua, Southwest Papua, Papua
Cluster 3	Bali, West Nusa Tenggara, East Nusa Tenggara, South Papua, Central Papua, Papua Pegunungan

Based on Table 2, which presents the members of each cluster, the mean value for each cluster was calculated to identify characteristic profiles of labor-related factors influencing economic growth in Indonesia.

Table 3. Profile of Cluster Groups

	Cluster 1	Cluster 2	Cluster 3
Economic Growth (%)	4.83	5.69	4.41
Population Size (Millions)	6.75	8.52	3.18
Employment Rate (%)	44.71	43	25.97
Open Unemployment Rate (%)	4.88	4.67	2.61
Total Labor Force (in persons)	3507120.2	4573628.9	1847336.8
Average Wage of Workers (Rp)	2982719.6	3341215.4	3489832.8

Table 3 presents the profiling results for the three clusters formed through the SOM algorithm. Cluster 2 shows the highest average economic growth and population size, whereas Cluster 3 exhibits the lowest values for both indicators. In contrast, labor absorption reaches its highest level in Cluster 1, which simultaneously records the highest open unemployment rate; both indicators are lowest in Cluster 3. The labor force size is greatest in Cluster 2 and smallest in Cluster 3. However, Cluster 3 stands out with the highest average worker wage, while Cluster 1 records the lowest.

3. Density-Based Spatial Clustering of Application with Noise (DBSCAN)

The DBSCAN algorithm was further employed to classify provinces based on density characteristics. Before clustering, outlier detection was performed to identify extreme observations that may affect cluster formation. Figure 4 indicates that outliers are present across five of the six variables, except for the Open Unemployment Rate (OUR), which shows a relatively uniform distribution. The presence of outliers highlights the heterogeneity of provincial labor and economic indicators.

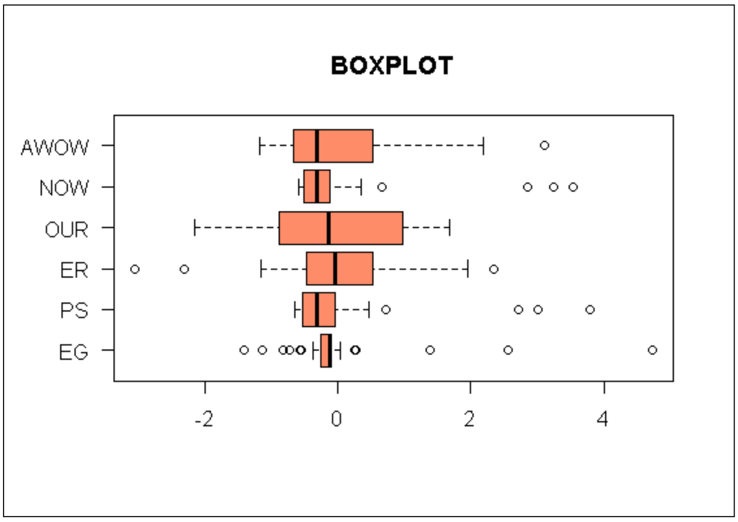

Figure 4. Boxplot of Variables

a. Determination of Epsilon and MinPts

Selecting appropriate values for epsilon and MinPts is crucial for obtaining well-formed clusters. Several combinations were tested to identify the optimal parameters, as the epsilon value directly influences core point identification, and MinPts determines neighborhood density. The k-distance plot with k=2 was used to determine the knee point, as shown in Figure 5.

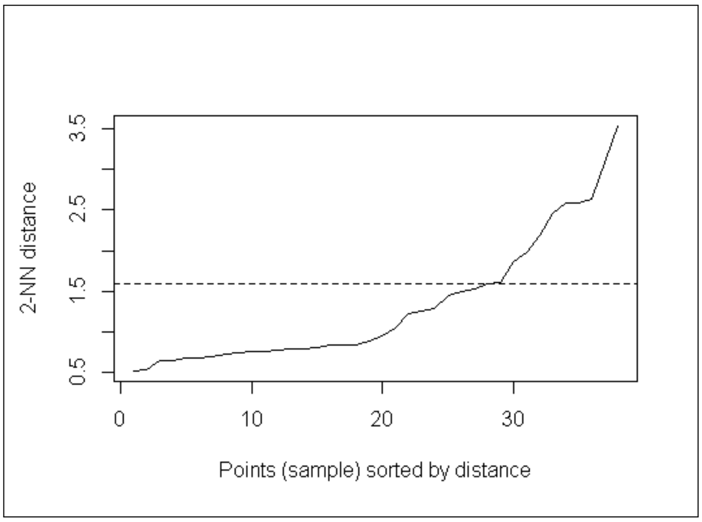

Figure 5. Epsilon Determination Plot

Based on Figure 5, using the knee-of-curve method, the optimal epsilon value is approximately 1.67, corresponding to the point where the curve transitions from a gradual increase to a sharper upward slope. This “elbow” point indicates the distance threshold where data points begin to be separated as noise rather than belonging to a dense cluster. Selecting epsilon at this value helps ensure that clusters remain sufficiently compact while minimizing misclassifying border points as noise. Therefore, the chosen epsilon strikes a balance between sensitivity to cluster density and robustness in identifying meaningful group structures within the dataset.

b. Evaluation of Optimal Cluster Numbers

Cluster validity was assessed using the Silhouette Index, with the optimal number of clusters corresponding to the configuration with the highest silhouette value. Experimental results are summarized in Table 4.

Table 4. Evaluation of Optimal Clusters

Epsilon	MinPts	Cluster	Noise	Silhouette Index
1.65	3	2	7	0.04364
1.65	2	3	5	0.03649
1.67	3	2	7	0.01167
1.67	2	3	5	0. 07844
1.70	2	2	4	0.01162

Based on Table 4, the combination of epsilon = 1.67 and MinPts = 2 yields the highest silhouette score (0.07844), indicating the most coherent DBSCAN configuration for this dataset. The silhouette metric ranges from -1 to 1, with higher values indicating more clearly formed, well-structured clusters.

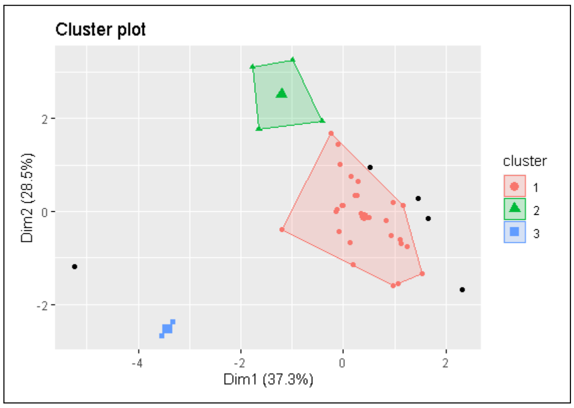

Figure 6. Cluster Visualization

Based on Figure 6, the spatial distribution of the clusters reveals noticeable structural differentiation among provinces. Cluster 1 forms a dense, centralized grouping in the reduced-dimensional space, indicating relatively homogeneous economic and labor characteristics across most provinces. In contrast, Cluster 2 and Cluster 3 appear more isolated, suggesting distinct structural profiles that deviate from the dominant pattern. The identification of noise points (black dots) indicates provinces with atypical or extreme indicator values, highlighting DBSCAN's robustness in distinguishing core clusters from peripheral observations. The detailed provincial distribution across clusters is presented in Table 5.

Table 5. Members Cluster

Cluster	Members
Noise	West Java, North Kalimantan, West Papua, Central Papua, Papua Pegunungan
Cluster 1	Aceh, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Kep. Bangka Belitung, DI Yogyakarta, Bali, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, Central Kalimantan, South Kalimantan, North Sulawesi, Central Sulawesi, South Sulawesi, Sulawesi Tenggara, Gorontalo, West Sulawesi, Maluku, North Maluku, Southwest Papua, Papua, South Papua
Cluster 2	Kep. Riau, DKI Jakarta, Banten, East Kalimantan
Cluster 3	Central Java, East Java

Based on Table 5, the DBSCAN results indicate the formation of three distinct clusters with varying provincial compositions. Cluster 1 comprises 27 provinces and represents the dominant group, while Clusters 2 and 3 include 4 and 2 provinces, respectively. Additionally, six provinces were classified as noise, implying that their economic and labor characteristics did not satisfy the minimum density requirements defined by the selected epsilon and MinPts parameters. This numerical distribution further confirms the structural grouping observed in the cluster visualization. The clustering results in this study indicate distinct characteristics among the clusters. The detailed cluster profiling analysis is presented in Table 6.

Table 6. Profile of Cluster Groups

	Cluster 1	**Cluster 2**	**Cluster 3**
Economic Growth (%)	5.01	5.27	4.93
Population Size (Juta Jiwa)	4.40	7.51	39.99
Employment Rate (%)	39.57	60.90	39.49
Open Unemployment Rate (%)	4.23	6.10	4.48
Total Labor Force (in persons)	2305565	3715643	23145303
Average Wage of Workers (Rp)	3013254	4877879	2559415

Table 6 shows that Cluster 2 has the highest economic growth and population size, while Cluster 1 has the lowest. Employment rate and open unemployment rates are highest in Cluster 2 and lowest in Cluster 1. The largest labor force size is observed in Cluster 2, whereas Cluster 1 has the smallest. Average worker wages peak in Cluster 2 and are lowest in Cluster 3.

4. K-Affinity Propagation (K-AP) Analysis

K-Affinity Propagation (K-AP) was applied as an exemplar-based clustering approach to group Indonesian provinces using the same labor and economic indicators analyzed in the previous methods. The K-AP algorithm identifies exemplars within the dataset and forms clusters around them. Cluster validity was assessed using several internal indices to determine the most representative cluster structure, presented in Table 7.

Table 7. Cluster Membership and Exemplars

Cluster	Cluster Province	Exemplar
Cluster 1	Kep. Riau, DKI Jakarta, Banten, East Kalimantan	Kep. Riau
Cluster 2	West Java, Central Java, East Java	East Java
Cluster 3	Aceh, North Sumatra, West Sumatra, Kep. Bangka Belitung, West Kalimantan, North Kalimantan, North Sulawesi, Maluku, Southwest Papua, Papua	North Sulawesi
Cluster 4	Riau, Jambi, South Sumatera, Bengkulu, Lampung, DI Yogyakarta, Bali, West Nusa Tenggara, East Nusa Tenggara, Central Kalimantan, South Kalimantan, South Sulawesi, Sulawesi Tenggara, Gorontalo, West Sulawesi, South Papua, Central Sulawesi	Riau
Cluster 5	North Maluku, West Papua	North Maluku
Cluster 6	Central Papua, Papua Pegunungan	Central Papua

Based on Table 7 , the K-AP clustering results show the distribution of provinces and their respective exemplars across six clusters. The results indicate that Cluster 1 (4 provinces) has Kep. Riau as its exemplar; East Java represents cluster 2 (3 provinces); North Sulawesi exemplifies cluster 3 (10 provinces); Cluster 4 (17 provinces) has Riau as its exemplar; North Maluku represents cluster 5 (2 provinces); and Central Papua exemplifies Cluster 6 (2 provinces). In the K-AP method, an exemplar refers to the most representative data object within a cluster, selected based on its highest similarity to other members, thereby reflecting the core characteristics of the cluster.

Table 8. Cluster Profiling

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**
Economic Growth (%)	5.27	4.93	4.13	4.98	17.26	3.80
Population Size (Millions)	7.51	43.49	4.10	4.53	0.68	1.41
Employment Rate (%)	60.90	41.49	42.76	38.55	36.18	8.61
Open Unemployment Rate (%)	6.10	5.24	5.54	3.45	4.08	2.03
Total Labor Force (in persons)	3715643	24158647	2131496	2390574	495911	911273
Average Wage of Workers (Rp)	4877879	2964443	3234460	2900317	3424772	4580517

Based on Table 8, cluster profiling reveals notable heterogeneity across groups. Cluster 5 exhibits the highest economic growth, whereas Cluster 6 records the lowest. In terms of population size, Cluster 1 has the largest, while Cluster 6 has the smallest. The employment rate and unemployment rate both peak in Cluster 1 and reach their lowest levels in Cluster 6. The labor force size is also greatest in Cluster 1 but smallest in Cluster 5. Regarding income indicators, average wages are highest in Cluster 1 and lowest in Cluster 4.

5. Evaluation of the Best Clustering Method

The optimal clustering method was determined using the ratio of within-cluster to between-cluster standard deviations (Sw/Sb), where a smaller ratio indicates better cluster compactness and separation. This evaluation was applied consistently to the clustering results produced by SOM, DBSCAN, and K-Affinity Propagation (K-AP) to enable an objective comparison across methods. The resulting Sw, Sb, and ratio values for each method are summarized in Table 9.

Table 9. Evaluation of Clustering Methods

Method	Sw	Sb	Ratio
SOM	5600641.5	175716.2	3.1906
DBSCAN	298023.4	1489784	0.2000
K-AP	240009.8	1340129	0.1779

As shown in Table 9, the K-AP method achieved the lowest ratio (0.1779), demonstrating superior clustering performance compared to SOM and DBSCAN. This advantage is attributed to K-AP's adaptive exemplar selection, lower memory requirements, and reduced computational overhead (Jaeger & Banks, 2023). The spatial distribution of the resulting clusters is illustrated in Figure 7.

These results are in line with and supported by previous studies. Hastuti et al. (2023) reported that K-Affinity Propagation produced more stable and representative clusters compared to SOM when applied to socio-economic data. Similarly, Selmi et al. (2025) emphasized that exemplar-based clustering provides better structural representation of heterogeneous data without requiring prior specification of the number of clusters. The findings of this study are that K-Affinity Propagation provides the most effective clustering structure for classifying Indonesian provinces based on labor and economic indicators. The comparative evaluation demonstrates that topology-based and density-based approaches are less capable of capturing the complex heterogeneity of provincial socio-economic characteristics. Therefore, exemplar-based clustering is more suitable for regional economic analysis and can serve as a reliable analytical foundation for labor-oriented regional development policy formulation.

Figure 7. Cluster Map

Cluster 1 (Kep. Riau, DKI Jakarta, Banten, East Kalimantan) is marked by relatively high wage levels, strong employment absorption, and a large labor force. Such characteristics are typically associated with advanced and highly dynamic labor markets. In this context, policy intervention should not merely aim at job creation but rather at improving labor market efficiency through workforce upskilling, technological adoption, and digital economic expansion to maintain productivity growth while minimizing labor market frictions (Aleca & Mihai, 2025).

Cluster 2 (West Java, Central Java, East Java) represents provinces with very large population and labor supply but moderate income and growth levels. This profile indicates the role of these provinces as national labor buffers. Therefore, policies need to transform labor abundance into higher economic value through vocational education alignment with industry needs, productivity-oriented training programs, and the strengthening of labor-intensive sectors capable of absorbing large numbers of workers (Beber et al., 2025; Canton, 2021; McKenzie, 2017).

Cluster 3 (including Bangka Belitung, North Sulawesi, West Kalimantan, Maluku, and Papua regions) reflects transitional regional economies with balanced but not yet optimal labor–economic performance. The appropriate strategy for this cluster lies

DOI: https://doi.org/10.30812/varian.v9i1.5933

in strengthening local leading sectors, encouraging regional investment, and fostering entrepreneurship to accelerate structural economic upgrading (Putra et al., 2025).

Cluster 4 (including Riau, Bali, NTB, NTT, Sulawesi, and Kalimantan regions) shows that economic activities have not been fully translated into improved wage outcomes. This suggests the presence of value chains with limited value-added for workers. Hence, policies should emphasize wage–productivity alignment, MSME empowerment, and downstream industrial development to enhance income distribution (Aleca & Mihai, 2025).

Cluster 5 (North Maluku, West Papua) demonstrates very high economic growth accompanied by very small labor absorption, a pattern commonly observed in capital-intensive extractive economies. Policy orientation in this cluster should focus on inclusive growth by increasing local labor participation and promoting community-based economic activities to ensure a more equitable distribution of welfare (Stromquist, 2019).

Cluster 6 (Central Papua, Papua Highlands) represents structurally lagging regions characterized by low labor participation and weak economic performance. In such cases, foundational policies related to infrastructure provision, access to education, health services, and basic skills development are prerequisites before more advanced economic strategies can be effectively implemented (Stromquist, 2019).

D. CONCLUSION AND SUGGESTION

The results show that clustering Indonesian provinces by labor economic indicators yields different structures across methods. SOM and DBSCAN each generated three clusters, whereas K-Affinity Propagation produced five. Based on the Sw/Sb ratio index, K-AP achieved the best performance (0.1779), outperforming DBSCAN (0.2000) and SOM (3.1906), indicating that K-AP provides the most effective clustering structure for labor-based economic growth patterns in Indonesia.

The novelty of this study lies in the integration of three different clustering paradigms: topology-based (SOM), density-based (DBSCAN), and exemplar-based (K-Affinity Propagation), within a unified framework, combined with quantitative internal validation using the Sw/Sb ratio for objective performance comparison.

This study is limited by the use of a restricted set of indicators, sensitivity to parameter selection, and the use of cross-sectional data for a single year. Future research should incorporate additional socio-economic variables and longitudinal data to improve the robustness of clustering.

ACKNOWLEDGEMENT

The authors sincerely acknowledge the support of all parties who contributed to the completion of this study. Special thanks to Statistics Indonesia (BPS) for providing the essential datasets that formed the basis of the analysis. The authors also express their appreciation to the Faculty of Mathematics and Natural Sciences, Mulawarman University, for its financial assistance provided through the PNBP funding program.

DECLARATIONS

AUTHOR CONTRIBUTION

All authors contributed equally to all aspects of this study, including conceptualization, data collection, data analysis, and manuscript preparation.

FUNDING STATEMENT

The Internal PNBP Grant funded this research for Fiscal Year 2025 from the Faculty of Mathematics and Natural Sciences, Universitas Mulawarman.

COMPETING INTEREST

The authors declare that they have no financial or personal conflicts of interest related to this research.

REFERENCES

Agustino, D. P., Budaya, I. G. B. A., Harsemadi, I. G., Dharmendra, I. K., & Pande, I. M. S. A. (2023). Comparison of the DBSCAN Algorithm and Affinity Propagation on Business Incubator Tenant Customer Segmentation. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, *12*(2), 315–321. https://doi.org/10.32736/sisfokom.v12i2.1682

Aleca, O. E., & Mihai, F. (2025). The Role of Digital Infrastructure and Skills in Enhancing Labor Productivity: Insights from Industry 4.0 in the European Union. *Systems*, *13*(2), 113. https://doi.org/10.3390/systems13020113

Beber, B., Lakemann, T., Schnars, R., & Lay, J. (2025). Employment Effects of Skills Trainings in Sub-Saharan Africa: A Systematic Review of Recent Randomized Controlled Trials. *De Economist*, *173*(1), 87–120. https://doi.org/10.1007/s10645-024-09442-6

Canton, H. (2021, July 8). *The Europa Directory of International Organizations 2021* (23rd ed.). Routledge. https://doi.org/10.4324/9781003179900

De Vries, G., Timmer, M., & De Vries, K. (2015). Structural Transformation in Africa: Static Gains, Dynamic Losses. *The Journal of Development Studies*, *51*(6), 674–688. https://doi.org/10.1080/00220388.2014.997222

Diao, X., McMillan, M., & Rodrik, D. (2019). The Recent Growth Boom in Developing Economies: A Structural-Change Perspective. In M. Nissanke & J. A. Ocampo (Eds.), *The Palgrave Handbook of Development Economics* (pp. 281–334). Springer International Publishing. https://doi.org/10.1007/978-3-030-14000-7_9

Dwitiyanti, N., Siti Ayu Kumala, & Shinta Dwi Handayani. (2024). Comparative Study of Earthquake Clustering in Indonesia Using K-Medoids, K-Means, DBSCAN, Fuzzy C-Means and K-AP Algorithms. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *8*(6), 768–778. https://doi.org/10.29207/resti.v8i6.5514

Hassan, B. A., Tayfor, N. B., Hassan, A. A., Ahmed, A. M., Rashid, T. A., & Abdalla, N. N. (2024). From A-to-Z review of clustering validation indices. *Neurocomputing*, *601*, 128198. https://doi.org/10.1016/j.neucom.2024.128198

Hastuti, S. H., Nurmayanti, W. P., & Saputri, A. A. (2023). Penerapan Metode Clustering Self Organizing Maps (som) dan K-Affinity Propagation (k-Ap) dalam Mengelompokkan Nilai Tukar Petani di Indonesia 2022. *VARIANCE: Journal of Statistics and Its Applications*, *5*(1), 79–88. https://doi.org/10.30598/variancevol5iss1page79-88

Hosan, S., Karmaker, S. C., Rahman, M. M., Chapman, A. J., & Saha, B. B. (2022). Dynamic links among the demographic dividend, digitalization, energy intensity and sustainable economic growth: Empirical evidence from emerging economies. *Journal of Cleaner Production*, *330*, 129858. https://doi.org/10.1016/j.jclepro.2021.129858

Iammarino, S., Rodriguez-Pose, A., & Storper, M. (2019). Regional inequality in Europe: Evidence, theory and policy implications. *Journal of Economic Geography*, *19*(2), 273–298. https://doi.org/10.1093/jeg/lby021

Ikotun, A. M., Habyarimana, F., & Ezugwu, A. E. (2025). Cluster validity indices for automatic clustering: A comprehensive review. *Heliyon*, *11*(2), e41953. https://doi.org/10.1016/j.heliyon.2025.e41953

Jaeger, A., & Banks, D. (2023). Cluster analysis: A modern statistical review. *WIREs Computational Statistics*, *15*(3), e1597. https://doi.org/10.1002/wics.1597

Lopez, C., Tucker, S., Salameh, T., & Tucker, C. (2018). An unsupervised machine learning method for discovering patient clusters based on genetic signatures. *Journal of Biomedical Informatics*, *85*, 30–39. https://doi.org/10.1016/j.jbi.2018.07.004

Mason, L., Hicks, B., & Almeida, J. S. (2023). EpiVECS: Exploring spatiotemporal epidemiological data using cluster embedding and interactive visualization. *Scientific Reports*, *13*(1), 21193. https://doi.org/10.1038/s41598-023-48484-9

McKenzie, D. (2017). How Effective Are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence. *The World Bank Research Observer*, *32*(2), 127–154. https://doi.org/10.1093/wbro/lkx001

Melin, P., Monica, J. C., Sanchez, D., & Castillo, O. (2020). Analysis of Spatial Spread Relationships of Coronavirus (COVID-19) Pandemic in the World using Self Organizing Maps. *Chaos, Solitons & Fractals*, *138*, 109917. https://doi.org/10.1016/j.chaos.2020.109917

Nisrina, S., Nurmayanti, W. P., Basirun, Kertanah, & Muhammad Gazali. (2022). Penerapan Metode Clustering SOM dan DBSCAN dalam Mengelompokkan Unmet Need Keluarga Berencana di Nusa Tenggara Barat. *J Statistika: Jurnal Ilmiah Teori dan Aplikasi Statistika*, *15*(2), 237–244. https://doi.org/10.36456/jstat.vol15.no2.a5549

Nurmayanti, W. P., Ratnaningsih, D. J., Nisrina, S., Rahim, A., Malthuf, M., & Kusuma, W. (2022). Clustrering of BPJS National Health Insurance Participant Using DBSCAN Algorithm. *Jurnal Varian*, *6*(1), 25–34. https://doi.org/10.30812/varian.v6i1.1886

Oyewole, G. J., & Thopil, G. A. (2023). Data clustering: Application and trends. *Artificial Intelligence Review*, *56*(7), 6439–6475. https://doi.org/10.1007/s10462-022-10325-y

Putra, I. K., Sukoco, B. M., & Wu, W.-Y. (2025). Landscape and opportunity of rural entrepreneurship: Systematic literature review. *Journal of Innovation and Entrepreneurship*, *14*(1), 122. https://doi.org/10.1186/s13731-025-00580-y

Selmi, A. T. E., Zerarka, M. F., & Cheriet, A. (2025). Efficient technique utilizing an embedding hierarchical clustering-based representation into crossed cubes for TSP optimization. *Cluster Computing*, *28*(2), 92. https://doi.org/10.1007/s10586-024-04741-8

Sorkhi, M. A., Rabbani, M., Akbari, E., & Motameni, H. (2024). A density-based clustering method with calculating the Eps parameter. *International Journal of Nonlinear Analysis and Applications*. https://doi.org/10.22075/ijnaa.2024.33345.4962

Stromquist, N. P. (2019). World Development Report 2019: The changing nature of work: By the World Bank. Washington, DC, World Bank, 2019, 151 pp. ISBN 978-1-4648-1342-9 (hbk). ISBN 978-1-4648-1328-3 (pbk), ISBN 978-1-4648-1356-6 (eBook), DOI:10.1596/978-1-4648-1328-3. *International Review of Education*, *65*(2), 321–329. https://doi.org/10.1007/s11159-019-09762-9

Wati, R. K., Pratiwi, H., & Winita Sulandari. (2024). Perbandingan Algoritma Density-Based Spatial Clustering Algorithm with Noise (DBSCAN) dan Self-Organizing Map (SOM) untuk Clustering Data Gempa Bumi. *Statistika*, *24*(2). https://doi.org/10.29313/statistika.v24i2.3645

Zhang, X., Lin, F., Wang, Y., & Wang, M. (2022). The impact of digital economy on employment polarization: An analysis based on Chinese provincial panel data. *Labor History*, *63*(5), 636–651. https://doi.org/10.1080/0023656X.2022.2133101

[This page is intentionally left blank.]