Economic Structure-Based Clustering and Development Strategy Formulation for Indonesian Provinces Using K-Means and Entropy Weight-TOPSIS

Rizka Britania[1*], Amir Tjolleng[2]
[1,2] Industrial Engineering Department, Faculty of Engineering, Bina Nusantara University, Jakarta

Email: [1]rizka.britania@binus.ac.id, [2]amir.tjolleng@binus.ac.id

Received: 2026-0820 Received in revised from 2026-08-26 Accepted:2026-09-25

Abstract

The development gap among provinces in Indonesia highlights the need for an approach that consider the differences in the economic structure of each province. This study aims to group Indonesian provinces based on their economic structures and assess their development performance to identify priority areas for intervention within each economic cluster. It further identifies the key indicators contributing to performance gaps, explores their potential root causes, and formulates specific improvement strategies to support more targeted and differentiated regional development planning. K-Means clustering is used to identify provinces with similar economic structures, Entropy Weight – TOPSIS is used to evaluate provincial performance, and weighted gap analysis is used to identify indicators that significantly contribute to the development performance gap. Fishbone analysis is used to explore potential root causes that support the formulation of improvement strategies. The analysis resulted in five clusters of Indonesian provinces: manufacturing, agriculture, mining, metropolitan services, and tourism and public services. Metropolitan services cluster has the best performance based on the TOPSIS Score (0.693), followed by manufacturing (0.427), mining (0.159), agriculture (0.113), and tourism and public services (0.092). *Pembentukan Modal Tetap Bruto* (PMTB) is the main contributor to the performance gap in the manufacturing cluster (45.49%) and mining (50.65%), internet penetration in the agriculture cluster (38.42%), and poverty rate in the tourism and public services cluster (52.27%). Root cause analysis highlights issues related to investment capacity, digital connectivity, infrastructure accessibility, productivity, employment, and poverty. The findings provide a basis for prioritizing interventions and formulating differentiated development strategies according to the economic characteristics and performance gaps of each cluster.

Keywords: *Economic Structure; Entropy Weight-TOPSIS; K-Means Clustering; Regional Development; Weighted Gap Analysis*

1. Introduction

Provinces in Indonesia have diverse economic characteristics, reflected in differences in their dominant economic activities. These differences are influenced by several factors, such as resource availability, the development of economic sectors, investment, labor force, infrastructure, and the level of technology adoption in each region. The economic structure of a region can be observed from the contribution of various business sectors to its Gross Regional Domestic Product (GRDP). Understanding these differences is crucial because provinces with distinct economic structures may face different development opportunities and constraints. Consequently, a one-size-fits-all policy approach may not be suitable for addressing region-specific issues. Identifying groups of provinces with similar economic structures can therefore provide a basis for comparing their development performance and determining more targeted development priorities according to their economic characteristics.

In regional development, a single policy approach does not always resolve the problems in every area. For instance, provinces relying on the agricultural sector face issues different from those in areas dominated by industry, mining, trade, or services. Provinces with similar economic structures can sometimes exhibit varying levels of development. Differences in investment, poverty, unemployment, labor productivity, digital connectivity, and infrastructure can lead to disparities in a region's growth potential compared to others. Therefore, the government requires information that not only emphasizes the economic characteristics of a region but also reflects the performance and development gaps among regions with similar economic profiles.

The importance of considering regional characteristics in development planning is also evident in Indonesia's planning. Through Regional Profile and Analysis, Bappenas emphasizes that identifying regional issues is one of the keys in regional development. Nevertheless, development gaps are still noticeable in various areas such as infrastructure, investment, and access to basic services. For example, a study in West Java shows gaps in electricity and telecommunications infrastructure between the northern and southern regions, alongside high investment concentration in the Bekasi and Karawang areas. This condition indicates that strong economic performance at the top level does not always have balanced performance within it. Therefore, an analytical approach is required that not only identifies performance gaps between provinces but also is able to link these gaps to characteristics based on the economic structure and determine which development indicators should be prioritized in development.

Several studies have employed cluster analysis to identify groups of regions based on similarities in economic or socio-economic characteristics. This approach is useful for capturing patterns and heterogeneity across regions. [1] applied K-Means clustering to group Indonesian provinces according to poverty levels, while [2] used the same method to classify provinces based on economic vulnerability indicators. [3] employed Fuzzy Geographic Weighted Clustering to identify three groups of regencies and cities in Central Sulawesi characterized by low, moderate, and high levels of economic development. [4] similarly identified three clusters in East Java using an ensemble clustering approach based on 12 economic variables. Meanwhile, [5] identified five clusters of regencies and cities in Central Java based on human development, infrastructure, and economic performance, revealing that regional performance was not necessarily consistent across these dimensions. Another study was conducted by [6], who applied clustering to economic disparities, poverty, economic vulnerability, welfare, and employment, using spatial clustering to group regions within each province in Indonesia based on characteristics of economic inequality. [7] used K-Means and K-Medoids clustering—to group poverty levels in Banten Province. However, these studies generally focused on cluster formation based on similarities in selected indicators and did not explicitly assess whether regions within the same economic group exhibit comparable levels of development performance. [8] employed clustering analysis to identify superior industrial clusters in Indonesia based on several economic indicators such as production growth, productivity, and contribution to GDP,

On the other hand, studies employing Multi-Criteria Decision-Making (MCDM) methods to evaluate regional performance generally rank regions based on multiple indicators but do not explicitly identify the sources of performance gaps among regions with similar economic characteristics. [9] used TOPSIS to rank the priorities of 17 economic sectors in Indonesia. [10] used Multi-Criteria Analysis (MCA) to evaluate the performance and rationale of development clusters in East Java. The findings revealed differences in performance among regions within the same cluster. However, existing studies remain limited in their ability to simultaneously examine similarities in economic structure and variations in development performance within these clusters. [11] similarly combined Fuzzy C-Means clustering and TOPSIS but reported rankings rather than gap attribution and specifically only for districts located in East Java.

Another gap lies in identifying development priorities. Rankings or performance scores can show provinces with high and low performance, but they cannot explain which indicators play a crucial role in creating these performance disparities. Therefore, gap analysis is necessary because it can consider the relative contribution of each economic indicator to the overall existing gap. These economic indicators can then be analyzed to determine the main causes, which can subsequently serve as the basis for recommending improvement strategies.

This study aims to (1) classify Indonesian provinces based on their economic structure, (2) evaluate provincial development performance, (3) identify the indicators that most contribute to performance gaps, and (4) analyze the root causes of performance gaps and formulate strategic recommendations appropriate to regional characteristics and development needs. The study proposes an integrated framework comprising K-Means Clustering, Entropy Weight-TOPSIS, Weighted Gap Analysis, and Fishbone Analysis. K-Means Clustering is used to classify 38 provinces in Indonesia based on the contribution composition of 17 GRDP sectors in 2025. Entropy Weight-TOPSIS is employed to assess provincial performance based on development indicators, including GDP per capita, HDI, poverty rate, unemployment rate, labor productivity, internet penetration, and road infrastructure. Once performance rankings are obtained, a Weighted Gap Analysis is conducted to compare groups of provinces with relatively high and low performance in each cluster, as well as to identify the indicators that play the most significant role in these differences. The results are then used in a Fishbone Analysis to explore the root causes of the priority indicators, followed by an analysis of specific solutions. This framework is expected to provide a more robust analysis for formulating development policies that align with the economic characteristics of each group.

2. Method

This study was conducted in several stages, as shown in Figure 1. The first stage involved collecting data of the Gross Regional Domestic Product (GRDP) for each province across 17 economic sectors and development indicators data. After data collection, data preprocessing is conducted, consisting of checking for missing data, Exploratory Data Analysis (EDA), and data standardization, with the aim of understanding the characteristics of the data and its distribution, as well as correlation analysis to identify relationships between variables and reduce the potential for redundant information before the data processing is carried out. The next step involves data processing through several stages (1) grouping provinces based on the similarity of their economic structures using K-Means Clustering, (2) analyzing development performance for each province with combination of the Entropy Weight Method and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), (3) implementing weighted gap analysis to compare groups of provinces with the highest (Top 25%) and lowest (Bottom 25%) TOPSIS scores (as in [12] and [13]) within each cluster to identify the indicators having significant contribution to the development performance gap within each cluster, (4) analyzing root causes for the performance gaps with fishbone diagram, and (5) developing specific development strategies based for each development indicators.

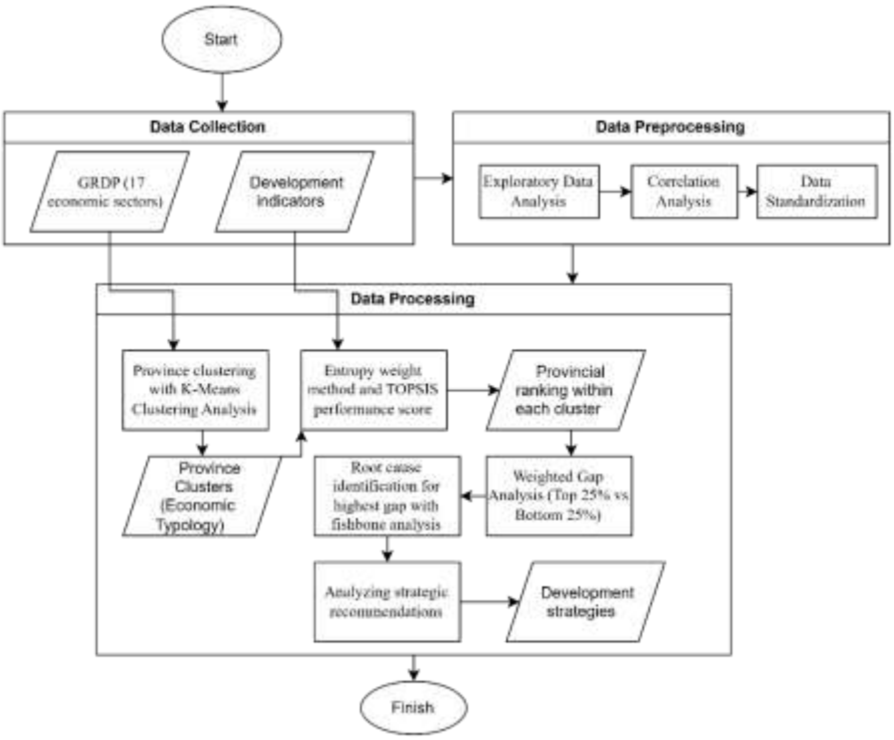

Figure 1. Research flowchart

2.1 Data Collection

This study uses secondary data obtained from the Central Statistics Agency (BPS) website, consisting of two sets of data: provincial GRDP data across 17 economic sectors and development indicator 2025 data for 38 provinces in Indonesia. The first dataset consists of GRDP data for each province across 17 economic sectors, which can be downloaded at (https://www.bps.go.id/id/statistics-table/2/MjI2NyMy/-seri-2010--quarterly-GRDP-at-constant-prices-by-economic-sector-in-all-provinces-of-Indonesia--billion-rupiah-.html). This data covers 17 economic sectors according to the industry classification established by BPS, namely: (1) Agriculture, Forestry, and Fisheries; (2) Mining and Quarrying; (3) Manufacturing; (4) Electricity and Gas Supply; (5) Water Supply, Waste Management, Wastewater Treatment, and Recycling; (6) Construction; (7) Wholesale and Retail Trade, and Repair of Motor Vehicles and Motorcycles; (8) Transportation and Warehousing; (9) Accommodation and Food Services; (10) Information and Communication; (11) Financial and Insurance Services; (12) Real Estate; (13) Business Services; (14) Public Administration, Defense, and Mandatory Social Security; (15) Educational Services; (16) Health Care and Social Services; and (17) Other Services. This data is used as the basis for grouping provinces according to their economic structure characteristics using the K-Means Clustering method.

The second set of data consists of development indicators used to evaluate the performance of each province after the clustering process was completed. Table 1 lists the seven indicators along with their sources. These seven indicators were selected based on their ability to represent the aspects of investment, human development, social welfare, economic productivity, digital transformation, and infrastructure—all of which are important dimensions of regional development.

Table 1. Data sources of development indicators

No	Data - unit	Source	Notes
1	*Pembentukan Modal Tetap Bruto – PMTB* – (billion rupiah)	https://www.bps.go.id/id/statistics-table/2/NTM0IzI=/-seri-2010--1--pdrb-atas-dasar-harga-berlaku-menurut-pengeluaran--milyar-rupiah-.html	Investment indicator
2	Human development index - index	https://www.bps.go.id/id/statistics-table/3/V25GaFNHaExaMnhITm1sWmRrUlJZelJzYUc1SGR6MDkjMyMwMDAw/indeks-pembangunan-manusia-menurut-provinsi.html?year=2025	Human development quality indicator
3	Poverty rate - %	https://www.bps.go.id/id/statistics-table/2/NTA0IzI=/indeks-keparahan-kemiskinan--p2--menurut-provinsi-dan-daerah--persen-.html	Social welfare indicator
4	Unemployment rate - %	https://www.bps.go.id/id/statistics-table/2/MjU2MiMy/tingkat-pengangguran-terbuka-menurut-provinsi--triwulanan---persen-.html	Social welfare indicator
5	Labor productivity – million rupiah / Rupiah/number of worker	(GRDP/Number of workers) https://www.bps.go.id/id/statistics-table/2/MjMzMyMy/penduduk-berumur-15-tahun-keatas-yang-bekerja-selama-seminggu-terakhir-menurut-provinsi-dan-lapangan-usaha--orang-.html	Economic productivity indicators
6	Internet signal reception - %	https://www.bps.go.id/id/statistics-table/3/TDJsTlFXOXhXVW81VEZoa1EwUk1iRUphZEVkRFVUMDkjMyMwMDAw/banyaknya-desa-kelurahan-menurut-provinsi-dan-penerimaan-sinyal-internet-telepon-seluler.html?year=2025	Digital transformation indicators
7	Road length - km	https://www.bps.go.id/id/statistics-table/3/U0VOeFZEZFNiVnByUkdGMlNrOTFVVGRHY1ZkVGR6MDkjMyMwMDAw/panjang-jalan-menurut-provinsi-dan-tingkat-kewenangan-pemerintahan--km-.html?year=2025	Infrastructure indicator

2.2 Data Pre-Processing

Data preprocessing was conducted to ensure that the dataset is suitable for analysis and to improve the reliability of the results. Python programming language was employed to conduct data preprocessing. The missing-value check confirmed that the dataset contained no missing observations;

therefore, no data imputation was required. The data represent the percentage contribution of each economic sector to the total regional value. The results of the exploratory data analysis are shown in Figure 2, 17 boxplots illustrating the distribution of data across 17 business sectors for all provinces in Indonesia.

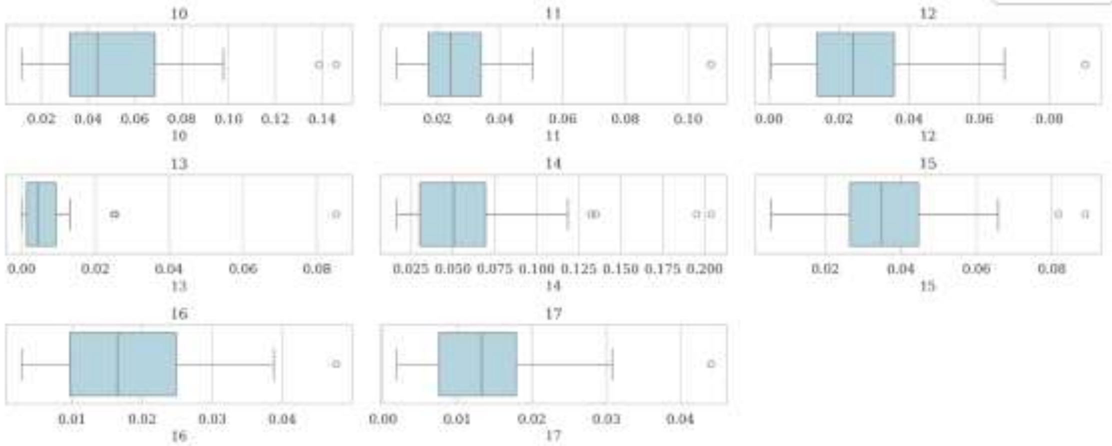

Figure 2. Boxplot of the data distribution for Indonesia's 17 economic sectors

It is shown that most business sectors have varying value distributions. Some sectors have a wide interquartile range (IQR) such as Agriculture, Forestry, and Fisheries (Sector 1), Mining and Quarrying (Sector 2), Manufacturing (Sector 3), Wholesale and Retail Trade (Sector 7), and Transportation and Warehousing (Sector 8). This means that these sectors have significantly different contributions to GRDP across provinces, indicating differences in economic specialization in each region. Some sectors, such as Electricity and Gas Supply (Sector 4), Water Supply, Waste Management, and Recycling (Sector 5), Financial and Insurance Services (Sector 11), Real Estate (Sector 12), and Health and Social Services (Sector 13) have relatively low median values with narrow distribution ranges. This means that the contributions of these sectors to GRDP are relatively small and their values are relatively similar in each province. Several sectors show the presence of outlier values, such as in Sector 2 (Mining and Quarrying), Sector 4 (Electricity and Gas Supply), Sector 6 (Construction), Sector 9 (Accommodation and Food Services), Sector 16 (Health and Social Services), and Sector 14 (Public Administration, Defense, and Mandatory Social Security). The presence of this outlier indicates that there are several provinces with sector contribution values significantly higher than those of other provinces. As the aim of this study is to identify the economic characteristics of each province, these outlier values are not removed, as they are important for subsequent clustering analysis.

Figure 3 shows the correlation plot among 17 GRDP business sectors. The correlation coefficients range from -0.62 to 0.79, indicating relationships that vary from negative to positive. Some pairs of variables show relatively low to medium correlations ($|r|<0.70$), suggesting that these sectors provide relatively distinct information on provincial economic structures. Some pairs exhibit strong positive correlations, such as sector 10 vs sector 17 ($r = 0.65$), indicating that provinces with high contributions from one sector usually also have high values in its paired sector. Conversely, negative correlations are observed in Sector 2 and Sector 7 ($r = -0.62$), Sector 3 and Sector 16 ($r = -0.61$), and Sector 3 and Sector 15 ($r = -0.52$), suggesting that greater reliance on one sector is associated with lower contributions from its paired sector.

No pair of variables has a very strong correlation ($|r| \geq 0.90$), indicating that there is no multicollinearity, thus no information redundancy. Therefore, all variables are used in further analysis to understand the economic structure of each province. The existing correlation values indicate diverse relationships between business sectors, reflecting heterogeneity in the economic structure of each province. This diversity supports the implementation of K-Means clustering, where each variable provides relevant information in distinguishing the economic characteristics of each province.

The next step is to perform data standardization before the clustering analysis. Standardization is carried out to ensure that the contribution of each variable, measured on different scales, can be compared in the clustering process. Figure 4 shows the snapshot of standardized values for first five provinces.

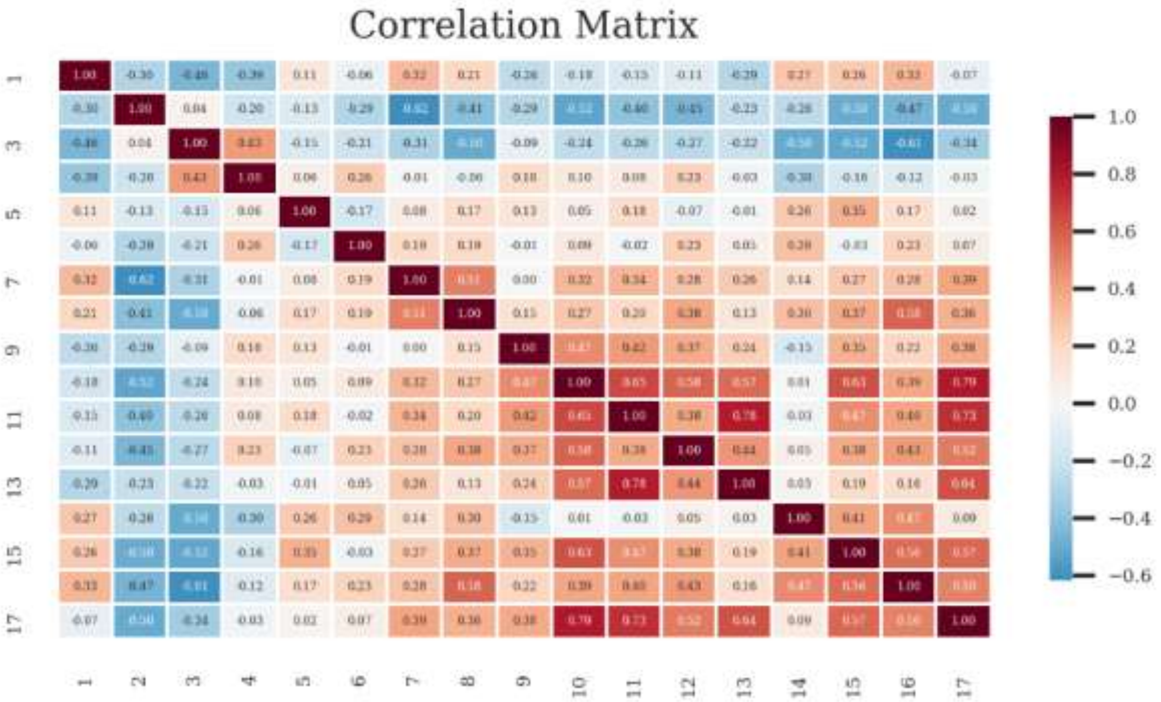

Figure 3. Correlation matrix for 17 GRDP business sectors

First Five Rows

	Provinsi	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	Aceh	0.276	0.073	0.043	0.002	0.000	0.090	0.157	0.076	0.015	0.043	0.015	0.042	0.006	0.086	0.027	0.033	0.016
1	Sumatera Utara	0.252	0.012	0.169	0.001	0.001	0.121	0.186	0.049	0.025	0.035	0.028	0.044	0.010	0.029	0.021	0.010	0.005
2	Sumatera Barat	0.215	0.038	0.091	0.001	0.001	0.088	0.168	0.103	0.012	0.092	0.031	0.021	0.005	0.055	0.041	0.018	0.020
3	Riau	0.268	0.136	0.325	0.001	0.000	0.091	0.104	0.008	0.005	0.012	0.010	0.010	0.000	0.018	0.005	0.003	0.005
4	Jambi	0.271	0.218	0.100	0.001	0.001	0.073	0.106	0.036	0.012	0.045	0.022	0.015	0.013	0.029	0.033	0.014	0.011

Figure 4. Snapshot of standardized data

3. Results and Discussion

3.1 Clustering Indonesian Provinces

The first step in K-Means clustering is to determine the optimal number of clusters. Three evaluation techniques were employed for this purpose: the Elbow Method, the Silhouette Score, and the Davies–Bouldin Index (DBI), as each provides a different perspective. The results of these three techniques are presented in Figure 5.

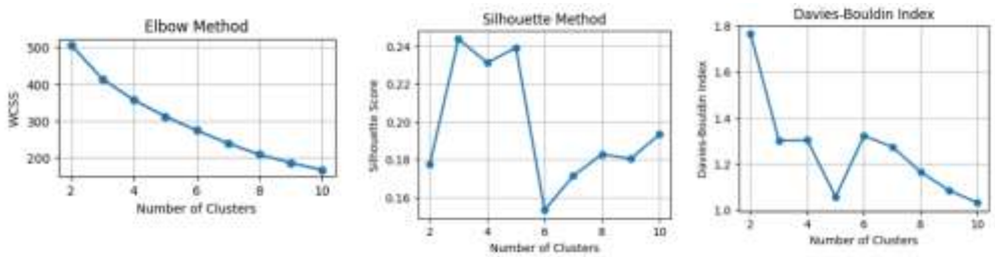

Figure 5. Result of evaluation techniques to determine the optimal number of clusters

Based on the Elbow Method graph, the WCSS (Within-Cluster Sum of Squares) value decreases as the number of clusters increases. The rate of decrease begins to level off after about 4–5 clusters, indicating that adding 4–5 clusters results in increased homogeneity within smaller clusters. Thus, the results of the Elbow Method indicate that the optimal number of clusters is 4–5. In Silhoutte method, a higher silhouette value indicates that objects within a cluster tend to be closer to one another and more distinct from other clusters. The result shows that the highest silhouette value is obtained at k = 3, with a value of approximately 0.244. The DBI value indicates that the best result is obtained at k = 10, with a value of approximately 1.03.

The three evaluation techniques provide different results regarding the optimal number of clusters. This difference indicates that there is no single k value that can simultaneously satisfy all three evaluation criteria. At k = 3, provinces with very different economic structures may be grouped together, reducing the ability to distinguish important economic characteristics. Conversely, increasing the number of clusters to k = 10 would result in many groups, which might have only a few provinces each, thus providing limited additional substantive differentiation for regional development analysis. This study selects k = 5 because it represents a balance between cluster separation and economic interpretability. The choice of k = 5 is not only related to statistical validity but also to economic interpretability and the substantive usefulness of the clustering results.

The application of the K-Means algorithm to standardized GRDP data across 17 sectors resulted in five groups of provinces with varying cluster sizes. Figure 6 illustrates this clustering. The differences in cluster size indicate that the characteristics of Indonesia's provincial economic structures are not evenly distributed, with most provinces having relatively similar economic patterns, while a few provinces have more specific economic structures.

Figure 6. Provincial Cluster in Indonesia

To interpret the characteristics of each cluster, the average contribution of each GRDP sector to each cluster was used, as shown in Figure 7. Table 2 shows the labels of each cluster as well as the number of provinces included (n).

Cluster	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	0.070	0.040	0.360	0.010	0.000	0.120	0.140	0.040	0.040	0.060	0.030	0.030	0.010	0.020	0.030	0.010	0.020
1	0.220	0.080	0.100	0.000	0.000	0.110	0.140	0.060	0.020	0.050	0.030	0.030	0.010	0.080	0.040	0.020	0.010
2	0.120	0.310	0.280	0.000	0.000	0.070	0.080	0.020	0.010	0.020	0.010	0.010	0.000	0.040	0.010	0.010	0.000
3	0.000	0.000	0.110	0.000	0.000	0.110	0.160	0.040	0.050	0.140	0.110	0.060	0.090	0.030	0.040	0.020	0.040
4	0.100	0.010	0.090	0.000	0.000	0.100	0.090	0.060	0.150	0.110	0.040	0.060	0.010	0.060	0.070	0.030	0.020

Figure 7. Average contribution of each GRDP sector for each cluster

Table 2. Labels for each cluster

Cluster	n	Dominant Sector	Economic Typology
0	5	Manufacturing, Trade, Construction	Manufacturing-Oriented
1	23	Agriculture, Transportation, Trade	Agriculture-Dominated
2	7	Mining, Manufacturing	Mining–Manufacturing
3	1	Trade, ICT, Finance, Real Estate	Metropolitan & Service-Oriented
4	2	Accommodation & Food Services, ICT, Services	Tourism & Service-Oriented

3.2 Entropy Weight Analysis

The next stage is to evaluate the development performance of each province using seven development indicators; Gross Fixed Capital Formation (GFCF), Human Development Index (HDI), poverty rate, unemployment rate, labor productivity, internet penetration, and road length. The entropy weight method is used to determine the relative weight of each indicator based on the degree of variation of information among provinces contained in it. To ensure consistency in interpretation, all indicators are oriented in the same direction, so that the higher the indicator value, the better the performance. Therefore, the poverty and unemployment rate indicators (cost criteria) are transformed, while the other indicators are considered as benefit criteria. A probability matrix is created to calculate the proportion of relative values of each province to the total provinces for each indicator. This matrix is used to calculate the entropy value, which reflects the uniformity of information distribution across provinces. An entropy value close to 1 indicates a uniform distribution, thus having low discriminating power. The diversification value is calculated based on “1 - the entropy value”. Indicator weights are obtained by normalizing the diversification values so that the total weight equals one. A summary of entropy values, diversification values, and indicator weights can be seen in Table 3.

Table 3. Entropy values for each indicator

Rank	Indicator	Entropy	Diversification	Weight
1	PMTB	0.802	0.198	38.83%
2	Internet Penetration	0.883	0.117	23.06%
3	Road Length	0.940	0.060	11.85%
4	Labor Productivity	0.940	0.060	11.81%
5	Poverty Rate	0.947	0.053	10.5%
6	Unemployment Rate	0.981	0.019	3.81%
7	HDI	0.999	0.000	0.13%

The entropy weight shows that PMTB has the highest value (38.83%), followed by internet penetration (23.06%). Entropy weights should not be interpreted as measures of the substantive importance of an indicator for regional development; rather, it represents its relative discriminatory power within the dataset. The higher the entropy weight, the greater the cross-provincial variation, thus the greater the information content for distinguishing provinces. Therefore, the high weights of PMTB and internet penetration indicate that these two indicators have greater variability across provinces and contribute more strongly to differentiating provincial performance in the TOPSIS analysis.

3.3 Provincial Performance Evaluation Using TOPSIS

The next stage is to evaluate the development performance of each province using the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method. This method is used to determine the relative position of each province based on its proximity to the positive ideal solution and its distance from the negative ideal solution. The seven development indicators are used, along with the weights obtained from the entropy weight method, to create a weighted decision matrix. Table 4 shows the resulted matrix.

Table 4. Weighted normalized matrix in TOPSIS

Cluster	PMTB	HDI	Poverty Rate	Unemployment Rate	Labor Productivity	Internet penetration	Road length
0	0.015	0.000	0.008	0.004	0.009	0.077	0.026
1	0.063	0.000	0.017	0.004	0.012	0.071	0.044
2	0.019	0.000	0.021	0.004	0.010	0.016	0.024
3	0.069	0.000	0.022	0.005	0.028	0.023	0.026
4	0.013	0.000	0.022	0.005	0.015	0.019	0.013

Next, the closeness coefficient is calculated; a higher value indicates that a province is closer to the positive ideal condition and farther from the negative ideal condition. Therefore, provinces with higher TOPSIS scores relatively have better performance. Table 5 shows the TOPSIS Score for each province.

The result shows a significant variation in development performance among provinces, which can be seen from the TOPSIS values ranging from 0.0358 to 0.6929. Provinces within the same economic cluster do not always have the same development performance. The province with the highest performance value is Jakarta (0.693), which is in Cluster 3, followed by East Java (0.687) and West Java (0.552), which are in Cluster 0. The provinces with the lowest values are West Sulawesi (0.036) (Cluster 1), Gorontalo (0.037) (Cluster 1), and Highland Papua (0.039) (Cluster 1).

Table 5. TOPSIS Score for each province

Cluster 0		Cluster 1						Cluster 2		Cluster 3		Cluster 4	
Province	TOPSIS Score	Province	TOPSIS Score	Province	TOPSIS Score	Province	TOPSIS Score	Province	TOPSIS Score	Province	TOPSIS Score	Province	TOPSIS Score
JATIM	0.687	SUMUT	0.338	SUMBAR	0.121	PAPUA	0.061	RIAU	0.281	DKI Jakarta	0.693	BALI	0.124
JABAR	0.552	ACEH	0.242	KALUT	0.117	PAPUA BARAT DAYA	0.048	KALTIM	0.258			DIY	0.059
JATENG	0.511	SULSEL	0.238	JAMBI	0.110	MALUKU	0.048	SULSEL	0.199				
BANTEN	0.225	LAMPUNG	0.157	SULTENGG	0.107	PAPUA SELATAN	0.043	SULTENG	0.155				
KEP. RIAU	0.163	NTT	0.141	SULUT	0.103	PAPUA PEGUNUNGAN	0.039	MALUT	0.089				
		KALSEL	0.130	KEP. BABEL	0.091	GORONTALO	0.037	PAPUA BARAT	0.085				
		KALBAR	0.129	BENGKULU	0.071	SULBAR	0.036	PAPUA TENGAH	0.049				
		KALTENG	0.123	NTB	0.071								
Average													
0.427		0.113						0.159		0.693		0.092	

To evaluate the performance gap between economic clusters, the TOPSIS Score is averaged by cluster. Cluster 3 has the highest TOPSIS value, but because it only consists of one province, it is not possible to calculate the variation in performance between provinces in this cluster. Cluster 0 is the cluster with the next highest average score, 0.427, with a range of 0.163 to 0.687. This value states that there is a significant diversity in the development performance between provinces in this cluster, even though they have similar economic structures. Cluster 2 has an average score of 0.159, slightly higher than cluster 1 (0.113) which has the most provincial members, 23 provinces. Cluster 4 has the lowest average score (0.092) with a range of 0.059 – 0.124, but this cluster only consists of 2 provinces, so there is a possibility of insufficient data to capture performance variations in this cluster.

3.4 Weighted Gap Analysis

After TOPSIS generates performance scores for each province, the next step is to identify the economic indicators that contribute most to the performance gaps among provinces within the same cluster. Each cluster is divided into two groups based on TOPSIS scores: the top 25% and the bottom 25%. The average score for each indicator is then compared between these two groups. Table 6 shows the results of the weighted gap analysis for the seven economic indicators in each cluster formed. (Note: Cluster 3 consists of only one province, DKI Jakarta, so a gap analysis could not be performed).

Cluster 0 and Cluster 1 have same top three indicators. The PMTB indicator in each province within these two clusters has a significant impact, as do access to digital infrastructure (internet penetration) and physical infrastructure (road length). Cluster 2 has different top three indicators. The PMTB gap in this cluster is highly dominant, meaning that differences in investment/capital are the most significant factor distinguishing high- and low-performing provinces within the mining-based cluster. Other key indicators include the poverty rate and labor productivity, which indicate that issues within the mining cluster are not only related to investment and infrastructure but also to the distribution of economic benefits, socioeconomic conditions, and labor productivity. Cluster 4 has top 3 indicators that are very different from those of the other clusters. The primary indicator is the poverty rate, which accounts for 52.27% of the total absolute gap. The second indicator is the unemployment rate (19.14%). Together, these two indicators account for more than 70% of the gap, meaning that differences in development performance within Cluster 4 are primarily related to poverty and employment opportunities.

Table 6. Weighted gap for each indicator

Indicator	Cluster 0			Cluster 1			Cluster 2			Cluster 4		
	Priority	Weighted Gap	Contribution	Priority	Weighted Gap	Contribution	Priority	Weighted Gap	Contribution	Priority	Weighted Gap	Contribution
PMTB	**1**	0.111	45.49%	**2**	0.028	28.65%	**1**	0.061	50.65%	5	0.003	6.88%
Internet Penetration	**2**	0.082	33.57%	**1**	0.037	38.42%	5	0.011	8.87%	4	0.004	9.49%
Road Length	**3**	0.033	13.63%	**3**	0.020	20.26%	4	0.014	11.95%	**3**	0.004	10.15%
Labor Productivity	4	0.011	4.66%	5	0.002	1.57%	**3**	0.015	12.82%	6	0.001	2.06%
Poverty Rate	5	0.005	2.12%	4	0.009	9.59%	**2**	0.018	14.96%	**1**	0.022	52.27%
Unemployment Rate	6	0.001	0.54%	6	0.001	1.49%	6	-0.001	0.72%	**2**	0.008	19.14%
HDI	7	0.000	0.00%	7	0.000	0.02%	7	0.000	0.03%	7	0.000	0.02%

3.5 Root Cause Analysis for the Performance Gaps

The next step is to analyze the root causes of the gaps in economic indicators. There are six economic indicators that rank among the top three with the largest gaps: PMTB, internet penetration, road length, poverty rate, labor productivity, and unemployment rate. Figure 8 shows the analysis of the causes of these gaps for the six indicators.

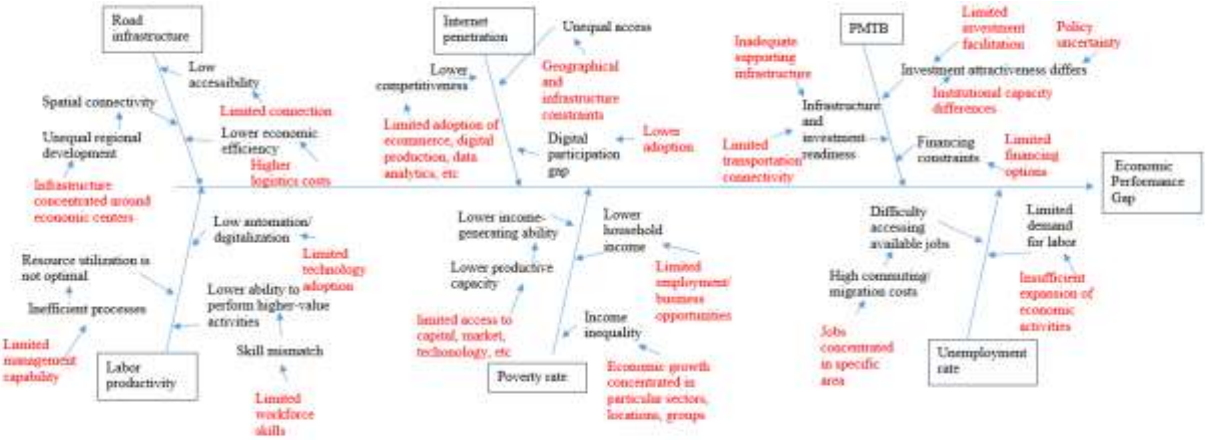

Figure 8. Fishbone diagram to analyze performance gaps in each indicator

The Fishbone Analysis suggests that gaps in regional economic performance may arise from the interrelationships among investment (PMTB), poverty, unemployment, labor productivity, internet penetration, and road infrastructure. Investment limitations may hinder the creation of capital and jobs. Meanwhile, deficiencies in skill, technology, and productive capital can reduce productivity and household income. Limitations in physical and digital connections may restrict access to markets, information, technology, and business opportunities. These findings suggest that improving regional economic performance requires an integrated strategy, not just a strategy for a single indicator.

3.6 Strategy Formulation

Because each province has a different economic structure, the strategy recommendations given do not focus on single, uniform policies, but on the emphasis of intervention priorities aligned with the economic characteristics of each region. For PMTB, local governments should strengthen investment attractiveness and readiness by simplifying licensing procedures, increasing business certainty, providing facilities for economic activities, and expanding access to funding. Investment should not only be directed to increase the capacity of existing sectors, but also to diversify the economy and valueadded activities, thereby producing a broader multiplier effect for local economic activities. For internet penetration, the strategy should extend beyond the expansion of internet infrastructure to include improvements in digital literacy, service affordability, and technology adoption among businesses. Digital technologies should be applied in ways that align with regional economic activities, including the digitization of production processes, product marketing, e-commerce, digital financial services, and data-driven approaches to improving business efficiency. For road infrastructure, development is not only about increasing the length of roads, but also about quality, network connectivity, and accessibility between regions. Better infrastructure is expected to lower logistics costs, make markets more accessible, increase labor mobility, and make investments more attractive. For labor productivity, the competence of the labor needs to be enhanced to align with the region's economic needs. Vocational training and education programs should be adapted to the needs of the main economic sectors and the continuously evolving technology. Besides skill development, productivity can also be improved through the adoption of technology, increased productive capital, and better work processes, thus productivity growth does not rely solely on the abilities of the labor. For poverty rate, enhancing the capacity of communities to participate in productive economic activities. Local governments can improve access to employment, financing, markets, skills development, and entrepreneurial opportunities. In this way, increased investment and economic activity can generate broader economic benefits and help reduce gaps in the distribution of development outcomes. Lastly, for unemployement rate, expanding productive job opportunities and reducing skill mismatches. Investment policies should be aligned with local job creation, while training programs should focus on competencies that match the real labor market needs. Economic diversification is also important to reduce dependence on certain sectors and strengthen the resilience of job opportunities against economic changes.

4. Conclussion

This study groups 38 Indonesian provinces based on their economic structure using K-Means Clustering, which resulted in five clusters: manufacturing, agriculture, mining, metropolitan services, and tourism and public services. Furthermore, the Entropy Weight-TOPSIS method reveals differences in development performance among provinces, with PMTB serving as the indicator with the highest objective weight, while the Weighted Gap Analysis identifies the indicators that most significantly contribute to performance gaps in each cluster. Fishbone analysis shows that these gaps are potentially related to limitations in investment, digital and physical connectivity, labor productivity, employment opportunities, and the fair distribution of economic benefits. The methodological contribution of this study lies in integrating economic-structure-based clustering with objective performance evaluation, performance-gap decomposition, and exploratory root-cause analysis to provide a systematic basis for developing differentiated regional strategies. Based on these findings, strategic recommendations are proposed that emphasize strengthening investment, infrastructure, digital transformation, labor productivity, job creation, and economic inclusion. However, the unequal cluster sizes, particularly the very small metropolitan services and tourism and public services clusters, should be considered when interpreting within-cluster comparisons. In addition, the Fishbone Analysis is exploratory and represents potential rather than empirically validated causal relationships. Future studies could incorporate socioeconomic and environmental indicators and validate the identified root causes and proposed strategies through stakeholder or expert assessment.

Reference

[1] A. Bahauddin, A. Fatmawati, and F. Permata Sari, "ANALISIS CLUSTERING PROVINSI DI INDONESIA BERDASARKAN TINGKAT KEMISKINAN MENGGUNAKAN ALGORITMA K-MEANS," *MISI (Jurnal Manaj. Inform. Sist. Informasi)*, vol. 4, no. 1, pp. 1–8, 2021, doi: https://doi.org/10.36595/misi.v4i1.216.

[2] R. B. Deviana and S. S. Putro, "PENGELOMPOKAN WILAYAH DI INDONESIA BERDASARKAN INDIKATOR KERAWANAN EKONOMI PASCA PANDEMI COVID-19 (K-MEANS CLUSTER ALGORITHM)," vol. 19, pp. 544–555, 2006.

[3] E. T. Maliku, S. Statistika, and J. Matematika, "PENGELOMPOKAN KABUPATEN / KOTA

DI PROVINSI SULAWESI TENGAH BERDASARKAN INDIKATOR PEMBANGUNAN EKONOMI MENGGUNAKAN," vol. 19, pp. 130–143, 2022, doi: doi.org/10.22487/2540766X.2022.v19.i1.15868 ISSN.

[4] S. N. Priansyah and D. U. Wutsqa, "PENERAPAN METODE ENSEMBLE CLUSTERING PADA PENGELOMPOKAN KABUPATEN/KOTA DI PROVINSI JAWA TIMUR BERDASARKAN INDIKATOR PEMBANGUNAN EKONOMI," *J. Stat. DAN SAINS DATA*, vol. 2, no. 1, pp. 1–14, 2024, doi: http://dx.doi.org/10.21831/jssd.v2i1.19804.

[5] Sugiarto and S. Wibowo, "KLASTERISASI KABUPATEN/KOTA DI JAWA TENGAH BERDASARKAN INDIKATOR KINERJA PEMBANGUNAN," *J. Litbang Sukowati*, vol. 3, no. 2, pp. 136–149, 2020, doi: https://doi.org/10.32630/sukowati.v3i2.161.

[6] G. E. Dirgantara, I. Novianti, and R. D. Bekti, "SPATIAL CLUSTER UNTUK PENGELOMPOKKAN WILAYAH SETIAP PROVINSI DI INDONESIA BERDASARKAN KARAKTERISTIK KESENJANGAN EKONOMI (SPATIAL CLUSTER FOR GROUPING REGIONS IN EVERY PROVINCE IN INDONESIA BASED ON THE CHARACTERISTICS OF ECONOMIC DISPARITY)," pp. 430–437, 2019, doi: https://doi.org/10.34123/semnasoffstat.v2019i1.63.

[7] T. A. Munandar, "PENERAPAN ALGORITMA CLUSTERING," vol. 9, no. 2, pp. 109–114, 2022, doi: 10.30656/jsii.v9i2.5099.

[8] Mahindah, F. Rossi, and M. Yasin, "Analisis Struktur Kinerja dan Kluster Industri Unggul," *J. Manaj. DAN BISNIS Ekon.*, vol. 2, no. 3, pp. 92–101, 2024, doi: https://doi.org/10.54066/jmbe-itb.v2i3.1950.

[9] M. Wardani, P. Andiny, and R. Syaputra, "ANALISIS PERKEMBANGAN EKONOMI WILAYAH," vol. 4307, no. June, pp. 3285–3290, 2026, doi: https://doi.org/10.54314/jssr.v9i3.6146.

[10] S. W. Hati and D. Y. P. Sari, "Factors Affecting Workers in Choosing a Boarding Services at Batam City - Indonesia," no. Icib 2019, pp. 683–692, 2020, doi: 10.5220/0008435306830692.

[11] D. S. Zaenab, Y. Farida, D. C. R. Novitasari, A. Lubab, and D. Yuliati, "The Implementation of Multi Criteria Decision Making (MCDM) for the Evaluation of Sustainable Regional Development in East Java by Using the Fuzzy C-Means Method and Technique for Order Preference By Similarity To Ideal Solution (TOPSIS) Evaluation of," *Proc. ofthe Built Environ. Sci. andTechnology Int. Conf. (BESTICON 2018)*, pp. 281–288, 2020, doi: 10.5220/0008906402810288.

[12] L. Maribel *et al.*, "Journal of Open Innovation : Technology , Market , and Complexity Digital knowledge transfer and regional development asymmetries : Empirical Evidence from European patent citations and geolocated economic activity," vol. 11, no. August, pp. 0–3, 2025, doi: 10.1016/j.joitmc.2025.100637.

[13] M. White, E. Heros, E. Graterol, N. Chirinda, and M. White, "Balancing Economic and Environmental Performance for Small-Scale Rice Farmers in Peru," vol. 4, no. October, pp. 1–14, 2020, doi: 10.3389/fsufs.2020.564418.