Eduvest – Journal of Universal Studies
Volume 5 Number 8, August, 2025
p- ISSN 2775-3735- e-ISSN 2775-3727

Socio-Economic Welfare Clustering: a Sub-National Governments Analysis in Indonesia

Ayu Aryawati1*, Mulya Amri1, Raden Aswin Rahadi1
Institute Technology Bandung, Indonesia
*Email: ayu_aryawati@sbm-itb.ac.id

ABSTRACT

Indonesia, as the world's largest archipelagic nation, encounters challenges in achieving equitable development and socio-economic well-being across its diverse regions. To address developmental disparities, the central government allocates specific purpose grants, the amount of which is tailored to the development needs of each region. Additionally, Indonesia has established a development financial institution aimed at providing financing and capacity-building support to SNGs in planning and implementing regional development initiatives. In determining the priority of grant allocation, financing, and capacity building, a tool is needed to identify disparities in socio-economic welfare between one SNG and other SNGs in Indonesia. This study examines the use of the clustering method with a K-means approach to group SNGs based on similarity in socio-economic welfare. This research introduces a novel approach by combining Human Development Index (HDI), Gross Regional Development Product (GRDP) per capita, poverty rate, and unemployment rate in a comprehensive K-means clustering model specifically designed for sub-national government analysis. Unlike previous studies that focus on single indicators or provincial-level analysis, this research provides a multi-dimensional assessment across all administrative levels in Indonesia. This study employs Poverty rate, Human Development Index and Gross Regional Development Product per capita from 2021 to 2023 as variables input. The clustering robustness is validated through Davies-Bouldin Index (1.00) and Silhouette scores (0.16), demonstrating stable groupings across the three-year period. At the end of the study, four clusters of SNGs in Indonesia were obtained based on their level of socio-economic welfare, providing a comprehensive framework for targeted policy interventions and resource allocation.

KEYWORDS *Socio-economic welfare, Clustering analysis, Sub-national government, Regional development, Disparity*

INTRODUCTION

In 2024, Indonesia had implemented fiscal decentralization for more than two decades. At present, the sources of regional revenue consist of: (1) local own-source revenue, (2) intergovernmental transfers, and (3) other legitimate sources of regional revenue. Transfer mechanisms aim to improve the welfare of the entire Indonesian population (USAID EGSA, 2021). Akai et al. (2002) proved that fiscal decentralization contributes to economic growth. Furthermore, Suwanan et al. (2009) found that fiscal decentralization reduces socio-economic welfare disparities between regions, based on dynamic panel data from 33 provinces for the period of 2001–2008.

Central government transfers, such as physical *DAK*, fund approximately 90% of subnational public investment (World Bank, 2023). However, the unpredictable allocation and lengthy approval process for these transfers discourage subnational governments from planning complex, multi-year infrastructure projects, despite their potential to improve welfare (Klytchnikova & Lokshin, 2009b). In this case, financing from development financial institutions can be an alternative to accelerate infrastructure development. This approach would help achieve better socio-economic welfare and minimize regional disparities in Indonesia (Alisjahbana & Akita, 2020; Massenga, 2025).

Furthermore, socio-economic welfare, which is the primary goal of fiscal decentralization, still shows significant disparities across regions in Indonesia (Brodjonegoro & Martinez-Vazquez, 2016; Lewis, 2017). Differences in gross regional domestic product (GRDP) per capita remain pronounced among provinces, reflecting structural inequalities in economic development (Sodik et al., 2019; Nugraha & Lewis, 2022). The Human Development Index (HDI) displays similar patterns, where regions with high GRDP per capita do not necessarily achieve high HDI, and vice versa, although there is generally a positive correlation between the two indicators (BPS, 2023; Hartono et al., 2020). These disparities highlight the need for more targeted fiscal policies that account for regional heterogeneity. Consequently, grouping subnational governments (SNGs) based on their socio-economic profiles can facilitate decision-making in prioritizing the distribution of physical DAK transfer funds, accessing financing from development finance institutions (DFIs), and selecting infrastructure projects suited to local conditions (Nasution et al., 2020; Suryahadi et al., 2021). This approach also offers insights for SNGs in formulating long-term infrastructure development plans that align with their respective socio-economic capacities.

This study contributes to the existing literature by uniquely combining *GRDP*, HDI, poverty rate, and unemployment rate in a single clustering framework designed specifically for Indonesian subnational governments. While previous clustering studies, such as Morales et al. (2020), focused on municipal management in Peru using different variables, and Gružauskas et al. (2021) examined regional socio-economic indicators in European municipalities using time-series analysis, this research fills a critical gap by providing a comprehensive, multidimensional assessment across all administrative levels in Indonesia. The novelty lies in integrating economic (*GRDP* per capita), social (HDI), and welfare indicators (poverty and unemployment rates) to create distinct clusters that better reflect the complex socio-economic landscape of Indonesia's diverse regions.

Unlike traditional provincial categorizations that rely on administrative boundaries, this clustering approach groups regions based on actual socio-economic performance similarities, enabling more targeted and effective policy interventions. This methodology provides decision-makers with a data-driven tool that transcends conventional geographic divisions and focuses on development needs and characteristics.

Given the diverse levels of socio-economic welfare across Indonesia, a grouping of *SNGs* was conducted. This is expected to facilitate decision-making in prioritizing the distribution of physical *DAK* transfer funds, financing from DFIs, and selecting infrastructure projects relevant to the respective *SNGs*. This study can also provide insights for *SNGs* when planning long-term infrastructure development in their regions.

METHOD

The research was designed through 6 steps as shown in the following figure.

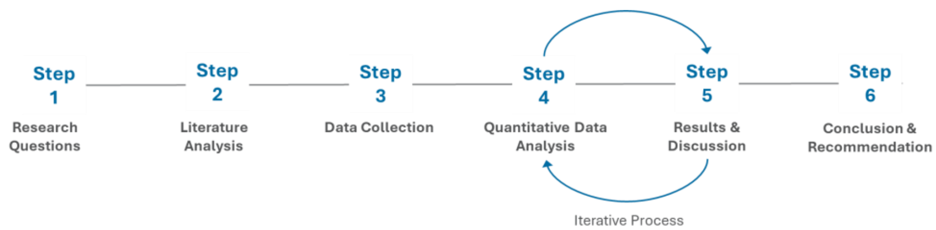

Figure 1. Research Design for Clustering SNGs

This research investigates which regions in Indonesia are less fortunate than others in terms of socio-economic welfare. By understanding this, development acceleration can be carried out more purposefully to reduce disparities.

Research questions were compiled to guide the study:

Q1: What variables can be used to cluster regions in Indonesia based on socio-economic welfare?

Q2: What is the clustering assignment for each region in Indonesia?

The analysis was performed on all regions, from the provincial level to the city and district levels.

K-means clustering was selected over alternative clustering methods due to its scalability and computational efficiency when handling large datasets with 548 sub-national governments. Unlike hierarchical clustering, which becomes computationally intensive with large datasets, or *DBSCAN*, which requires density parameters that may not be suitable for the varied geographic and demographic characteristics of Indonesian regions, *K-means* provides clear, interpretable cluster assignments while maintaining stability across multiple years of data.

Literature research was conducted to identify common measurements for socio-economic welfare, including metrics used by international organizations. Data from 2021 to 2023 was collected from government publications accessible through the official website of *BPS*. A quantitative clustering analysis was conducted and refined through expert consultations for sample checks.

The expert validation process involved three development economists and two regional planning specialists who independently reviewed a random sample of 50 cluster assignments to verify the logical consistency of the groupings, based on their knowledge of regional characteristics. The validation criteria included assessing whether regions with similar development challenges were grouped together and whether outliers were appropriately classified.

Statistical tests were performed to assess the separation of clustering results and their year-to-year stability. The study concludes by presenting answers to each research question and recommending cluster-specific corporate strategies for *PT SMI* to achieve its mandate of enhancing national welfare while maintaining asset quality and profitability.

RESULT AND DISCUSSION

To assess socio-economic welfare in Indonesia, this research will exercise four primary variables. Gross Regional Domestic Product (GRDP) per capita will measure the economic output of a region. However, recognizing that economic growth does not necessarily translate directly into improved well-being, the Human Development Index (HDI) will be incorporated to capture the social dimension of welfare. Specifically, the HDI considers factors such as life expectancy at birth, educational attainment (literacy rates and years of schooling), and standard of living (income per capita), offering a multidimensional perspective on societal progress. Poverty rate (POV) and unemployment rate (UNEM) as part of people welfare index also exercised in this research.

To ensure the robustness of the clustering results, a preliminary correlation analysis was performed to identify and remove high correlated variables which convey similar information. Variables with absolute Pearson correlation >0.7 will not be used together in the clustering model.

Table 1: Correlation test of socio-economic welfare metrics

	POV	*UNEM*	*HDI*	*GRDP*
POV	1			
UNEM	-0,37481919	1		
HDI	-0,66629241	0,545491	1	
GRDP	-0,21688937	0,126665	0,29396	1

As presented in table 1, all variables exhibited absolute correlation values below 0.7. However, the unemployment rate (UNEM) demonstrated a negative correlation with HDI and GRDP per capita, which contradicts the expectation of regions with higher socio-economic welfare will have lower unemployment rates.

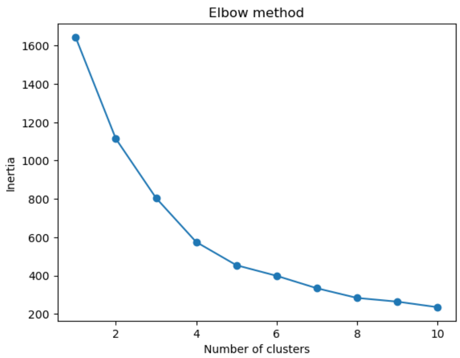

Figure 2. Inertia by Number of Clusters

Clustering began by normalizing the data so that each variable had the same scale. Subsequently, the inertia for each number of clusters was calculated as shown in figure 4 to determine the optimal number of clusters. Based on the chart, clustering will be optimal if 4 distinct clusters are formed. Adding more than 4 clusters does not provide significant additional information. Based on the study conducted, the cluster centroids obtained are shown in table 2. Based on 2023 data, clustering has DBI 1.00 and Silhouette score 0.16 (close to zero), meaning that similar SNGs has grouped within a cluster. These statistics and its centroid stable during 2021 to 2023. The clustering result is stable during 2021 to 2023, with only 19 SNGs that experience transition between clusters. It will benefit financing policy can be consistent from year to year.

Table 2: Cluster's centroid based on 2023 data

Clusters	POV	HDI	GRDP per Capita (IDR million)	#SNG
0	6.72	77.17	88.82	187
1	12.01	70.29	44.01	300
2	8.42	76.54	423.21	17
3	29.40	59.83	24.43	44

Based on the centroids in Table 2, the relationships between variables are then shown in the pair plot in Figure 5. In the pair plot, the level of socio-economic welfare can be ordered from highest to lowest, with the order of cluster 2, cluster 0, cluster 1, then cluster 3.

Cluster Analysis and Characteristics

1. Cluster 2 or ‘High Socio-Economic Development’ cluster is region with very high GRDP per capita, HDI tend to be high or very high, accompanied by low poverty rate. Only 1 province assigned to this cluster, DKI Jakarta. Further, 6 cities and 10 districts clustered in this group either because the region is metropolitan city (South Jakarta, Central Jakarta, North Jakarta), industrial city (Cilegon and Kediri), or having rich natural resources like most of cities and districts that fall in this cluster (Mimika, Anambas islands, Teluk Bintuni, etc).
2. Cluster 0 or ‘Moderate Economic Base’ cluster, similar to the ‘High Socio-Economic Development’ cluster but has lower GRDP per capita. In this cluster, GRDP per capita are varied from Rp25 million to Rp245 million, HDI in high or very high category, and poverty rate is below 15%. Most of the cities are classified in this cluster (82.83%) consist of provincial capital. While almost half of provinces (47.06%) located in Sumatera, Java, Bali, Kalimantan and Sulawesi assigned to this cluster. In districts level, only 21.45% districts that spread across 24 provinces belong to this group. Bali and most of its cities/districts are classified in this cluster.
3. Cluster 1 or ‘Transitional Regions’ cluster has lower welfare compared to previous cluster with lower GRDP per capita and HDI, accompanied by a higher poverty rate. Most of districts and provinces have belonged to this group, spread from West to East Indonesia. In this cluster, can be seen that lower HDI has correlation with higher poverty rate. Regions included in this cluster can emulate development in other regions categorized as cluster ‘Moderate Economic Base’, in accordance with the characteristics of their respective regions. However, basic infrastructure development might still be required.
4. Cluster 3 or ‘Vulnerable Regions’ cluster has lowest GRDP per capita, lowest HDI, and highest poverty rate. Strong negative correlation between the Human Development Index (HDI) and the poverty limit indicate that basic infrastructure development such as healthcare, education, roads and water sanitation is needed. Neither province nor cities assigned to this cluster. Meanwhile, most of districts belong to this clusters are located in Maluku, Nusa Tenggara and Papua. In addition, SNGs in areas categorized as vulnerable regions require greater transfer support to catch up. Meanwhile, DFIs in Indonesia can implement a number of exceptions in their financing policies in these areas, provide technical assistance, or even allocate grants to improve the fulfilment of the community's basic needs, such as water and sanitation.

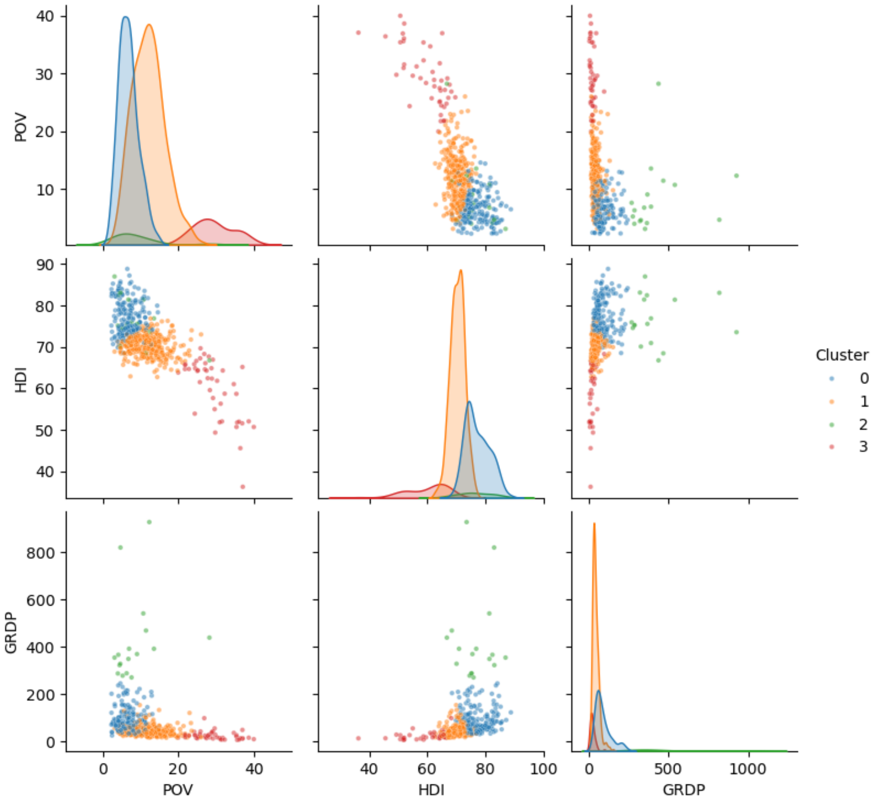

Figure 3. Pair Plot socio-economic welfare clustering of SNGs in Indonesia 2023

The distribution of socio-economic clustering is presented in table 3 below. No GRDP per capita available for new established Provinces e.g. Papua Barat Daya, Papua Pegunungan, Papua Tengah, and Papua Selatan. So, socio-economic welfare clustering can follow the majority cluster of its districts and cities.

Table 3. Number of SNGs clustered by Socio-Economic Welfare

Cluster	Districts (*Kabupaten*)		Cities (*Kota*)		Provinces	
High Socio-Economic Development (2)	10	2,41%	6	6,06%	1	2,94%
Moderate Economic Base (0)	89	21,45%	82	82,83%	16	47,06%
Transitional Regions (1)	272	65,54%	11	11,11%	17	50,00%
Vulnerable Regions (3)	44	10,60%				
Total	**415**	**100.00%**	**99**	**100.00%**	**34**	**100.00%**

CONCLUSION

Poverty level, Human Development Index (*HDI*), and *GRDP* per capita variables can be used to group regions in Indonesia based on socio-economic welfare. Optimal clustering was achieved by forming four clusters. The first cluster is *High Socio-Economic Development* regions, characterized by very high *GRDP* per capita, *HDI* that tends to be high or very high, and a low poverty rate. The second cluster is *Moderate Economic Base*, which is similar to the first cluster but has moderate *GRDP* per capita. The third cluster is *Transitional Regions*, which have moderate to high *GRDP* per capita and *HDI*, accompanied by a higher poverty rate. The last cluster is *Vulnerable Regions*, which have low *GRDP* per capita, an *HDI* that tends to be low, and a very high poverty rate. Each of these clusters has distinct characteristics and, therefore, requires different infrastructure development plans.

Further research can be conducted by incorporating other welfare indicators, such as the percentage of households with electricity as a source of lighting, income inequality measured by the *Gini index*, and others. Additional variables may enrich the clustering results and the strategies that can be developed from them. Another important consideration when creating an infrastructure development plan is aligning the development with local potential. Each region has its own characteristics and direction, such as being a cultural city evolving into a tourism destination, an industrial city, a livable city, or an agricultural city. Therefore, studies aimed at diagnosing a region's potential development will be useful in guiding infrastructure planning.

REFERENCES

Alisjahbana, A., & Akita, T. (2020). Economic tertiarization and regional income inequality in a decentralized Indonesia: A bi-dimensional inequality decomposition analysis. *Social Indicators Research*, *151*(1), 51–80.

Akai, N., & Sakata, M. (2002). Fiscal decentralization contributes to economic growth: Evidence from state-level cross-section data for the United States. *Journal of Economics, 52*(1), 93–108.

Badan Pusat Statistik. (2023). *Indeks pembangunan manusia 2023*. https://www.bps.go.id

Brodjonegoro, B., & Martinez-Vazquez, J. (2016). An analysis of Indonesia's transfer system: Recent performance and future prospects. *Bulletin of Indonesian Economic Studies, 52*(3), 329–353. https://doi.org/10.1080/00074918.2016.1249262

Gružauskas, V., Čalnerytė, D., Fyleris, T., & Kriščiūnas, A. (2021). Application of multivariate time series cluster analysis to regional socio-economic indicators of municipalities. *Real Estate Management and Valuation, 29*(3), 39–51. https://doi.org/10.2478/remav-2021-0020

Hartono, D., Resosudarmo, B. P., & Siregar, H. (2020). Inequality and human development in Indonesia. *Asian Development Review, 37*(1), 1–28. https://doi.org/10.1162/adev_a_00142

José Morales, N. V. (2020). Classification model of municipal management in local governments of Peru based on K-means clustering algorithm. *International Journal of Advanced Computer Science and Applications, 11*(7).

Klytchnikova, I., & Lokshin, M. (2009). Measuring welfare gains from better quality infrastructure. *Journal of Infrastructure Development, 1*(2), 87–109. https://doi.org/10.1177/097493060900100201

Lewis, B. D. (2017). Local government spending and service delivery in Indonesia: The perverse effects of substantial fiscal resources. *Regional Studies, 51*(11), 1695–1707. https://doi.org/10.1080/00343404.2016.1216957

Nasution, A., Harahap, N., & Digdowiseiso, K. (2020). Fiscal decentralization, inequality, and poverty in Indonesia. *Economic Change and Restructuring, 53*(2), 235–252. https://doi.org/10.1007/s10644-019-09251-5

Nugraha, A. T., & Lewis, B. D. (2022). Fiscal capacity, expenditure needs, and equalization transfers in Indonesia. *International Journal of Public Administration, 45*(4), 311–322. https://doi.org/10.1080/01900692.2021.1884725

Massenga, T. W. (2025). Village policy study in overcoming social and economic inequality in local communities. *Jurnal Konseling dan Pendidikan*, *13*(2), 521–530.

Sodik, J., Syahputra, R., & Haris, A. (2019). Analysis of GRDP and its impact on regional inequality in Indonesia. *Journal of Applied Economic Sciences, 14*(3), 857–865. https://doi.org/10.14505/jaes.v14.3(65).20

Suryahadi, A., Akhmadi, & Yumna, A. (2021). Social protection and the welfare of the poor in Indonesia: Lessons from the COVID-19 pandemic. *Bulletin of Indonesian Economic Studies, 57*(3), 319–349. https://doi.org/10.1080/00074918.2021.2005513

USAID EGSA. (2021). *Two decades of fiscal decentralization implementation in Indonesia*. DevTech Systems, Inc.

World Bank. (2023). *Subnational debt financing in Indonesia*. World Bank Publications.

Xiao, J., Lu, J., & Li, X. (2017). Davies–Bouldin index based hierarchical initialization K-means. *Intelligent Data Analysis, 21*(6), 1301–1524.