TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 24, No. 6, December 2026, pp. 1976~1985
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v24i6.28101

Longitudinal k-means clustering for socio-economic profiling of Indonesian regions

Md. Wira Putra Dananjaya[1], I Gusti Ngurah Darma Paramartha[2], Made Ari Riska Dayanti[3], Caren Angellina Mimaki[4]

[1]Department of Business Digital, Faculty of Economics and Business, Universitas Pendidikan Nasional, Bali, Indonesia
[2]Department of Information Technology, Faculty of Engineering and Informatics, Universitas Pendidikan Nasional, Bali, Indonesia
[3]Department of Environmental Engineering, Faculty of Engineering and Informatics, Universitas Pendidikan Nasional, Bali, Indonesia
[4]Department of Management, Faculty of Economic and Business, Universitas Pendidikan Nasional, Bali, Indonesia

Article Info

Article history:

Received Jun 8, 2026
Revised Sep 17, 2026
Accepted Oct 11, 2026

Keywords:

Computational social science
Feature engineering
K-means clustering
Longitudinal data analysis
Regional resilience

ABSTRACT

Conventional descriptive statistical methods are limited in their ability to capture non-linear patterns in high-dimensional regional development data. This study proposes an Unsupervised Machine Learning framework to computationally map regional resilience typologies in Indonesia. Using a time-series dataset from Statistics Indonesia (*Badan Pusat Statistik*/BPS) spanning 2015–2024 across all 514 regencies and cities, we constructed an automated processing pipeline that reconciles divergent regional nomenclature and resolves records duplicated by provincial reorganisation, yielding a balanced panel of 5,137 region–year observations. Poverty and human development index (HDI) features were standardized using Z-score scaling to ensure equidistant contribution within the Euclidean vector space. K-means clustering identified three regional archetypes: advanced (HDI 75.94; poverty 6.82%), Transitioning (HDI 67.66; poverty 12.39%), and Lagging (HDI 57.48; poverty 30.58%). Fitting centroids once on the pooled panel and holding them fixed across all ten years reveals a decade that was simultaneously successful and static: a third of all regions crossed into a higher absolute tier and none regressed, yet 95.5% occupied the same position relative to their contemporaries in 2024 as in 2015. Convergence occurred in levels without convergence in ranks, a distinction that a single cross-sectional partition cannot recover and that bears directly on how rank-based eligibility criteria for regional programmes should be designed.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Md. Wira Putra Dananjaya
Department of Business Digital, Faculty of Economic and Business, Universitas Pendidikan Nasional
Bedugul St., No.39, Sidakarya, Denpasar Selatan, Kota Denpasar, Bali 80224, Indonesia
Email: putradananjaya@undiknas.ac.id

1. INTRODUCTION

Indonesia, as the world's largest archipelagic state, presents a complex spatial challenge in maintaining equitable socio-economic development [1], [2]. Historically, national development strategies have contended with a pronounced "Java-centric" disparity, creating a significant developmental gap between Western and Eastern Indonesia [3]-[5]. Within the context of the built environment, this disparity is not merely a statistical anomaly but a reflection of unequal access to physical infrastructure, digital connectivity, and essential services. However, policy interventions aimed at mitigating these inequalities often rely on conventional descriptive statistics and static snapshot analysis [6], [7]. While these traditional

methods provide a general overview, they frequently fail to capture the latent, non-linear patterns and temporal dynamics embedded within high-dimensional socio-economic data over extended periods [8], [9].

Computational social science offers a means of evaluating regional resilience at a granularity that administrative classification cannot reach [10]-[12]. Traditional administrative classification often suffers from subjective biases and lacks the granularity needed to identify "positive outliers" regions that defy geographical disadvantages through specific industrial or infrastructural variables. Consequently, there is an urgent need for an algorithmic framework capable of processing longitudinal time-series data to construct a more objective, data-driven taxonomy of regional performance [13], [14].

While advanced techniques such as density-based spatial clustering of applications with noise (DBSCAN), hierarchical clustering, or principal component analysis (PCA) offer robust analytical capabilities, they often lack the straightforward interpretability required by government stakeholders [15]. Therefore, a gap exists in deploying computationally rigorous yet highly interpretable algorithms. This study proposes a computational profiling approach using unsupervised machine learning, specifically the k-means clustering algorithm, to analyze the trajectory of 514 districts and cities in Indonesia from 2015 to 2024. Unlike supervised learning, which relies on pre-defined labels, unsupervised learning allows the data to self-organize based on feature similarity in a Euclidean vector space, thereby revealing natural groupings that may not be apparent through manual observation. By implementing a rigorous data processing pipeline explicit synonym mapping across divergent regional spellings, deduplication of records duplicated by provincial reorganisation, a missingness threshold applied in place of imputation, and Z-score feature scaling this research addresses the technical challenges of handling messy administrative datasets.

This study is guided by two questions. First, what socio-economic archetypes emerge when a decade of regency-level human development index (HDI) and poverty data is partitioned algorithmically, and which regions depart from the pattern their geographic position would predict? Second, do regions change archetype over the observation period, and does the answer depend on whether mobility is measured against fixed developmental thresholds or against the shifting national distribution? The second question motivates the principal contribution of this work. By fitting centroids once on the pooled panel and holding them fixed across all ten years, the analysis shows that a third of Indonesian regencies and cities crossed into a higher tier between 2015 and 2024 while none regressed, yet 95.5% occupied the same position relative to their contemporaries at the end of the decade as at the start. Convergence occurred in levels without convergence in ranks, a distinction that a single cross-sectional partition cannot recover and that carries direct implications for how regional programmes are evaluated. Alongside this finding, the study contributes a reproducible pipeline for constructing longitudinal panels from administrative exports, and a data-driven taxonomy that supports decision support systems (DSS) for precision-based regional intervention.

2. METHOD

This study employs a quantitative computational approach, utilizing unsupervised machine learning to uncover latent patterns in longitudinal socio-economic data [16]. The methodological framework is structured into a three-stage pipeline: (i) data acquisition and preprocessing, (ii) feature engineering and normalization, and (iii) algorithmic clustering.

2.1. Data acquisition and structure

The final additions to the regency roster took effect in 2014, so the 2015–2024 window covers a stable set of 514 units [17]. The 2022 Papua reorganisation altered provincial attribution but not the regency roster, and is handled by deduplication at the parsing stage (section 2.2). The selection of solely HDI and poverty is deliberate; these two metrics serve as the primary macroeconomic key performance indicators (KPIs) mandated by the Indonesian government for regional transfer funds (*Dana Alokasi Umum* or DAU / *Dana Alokasi Khusus* or DAK) allocation, ensuring the model's output remains directly actionable [18]. The dataset consists of two high-impact feature vectors:

- Poverty rate (X_1): the percentage of the population living below the national poverty line.
- HDI (X_2): a composite index measuring life expectancy, education, and per capita income.

2.2. Computational data pipeline (ETL)

Raw administrative data often contains noise, inconsistencies, and structural irregularities arising from changes in administrative geography. To address this, we developed a Python-based extract, transform, load (ETL) pipeline using the Pandas library. The complete workflow, from raw Statistics Indonesia (*Badan Pusat Statistik*/BPS) exports to the clustered analysis panel, is presented in Figure 1.

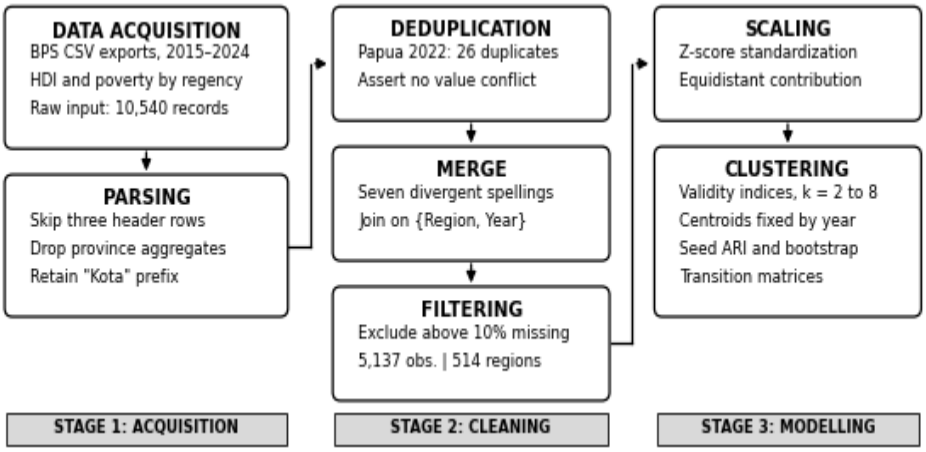


Figure 1. Three-stage computational pipeline for constructing and clustering the longitudinal panel

The preprocessing logic involved four critical operations:

- Parsing and structuring. Each BPS export interleaves province rows with regency and city rows within a single column, preceded by three header rows and closed by a national aggregate. Province rows were identified by capitalization, retained as an attribute, and then removed along with the national aggregate. The Kota prefix was deliberately preserved, since stripping it would conflate distinct entities sharing a name, such as Bogor Regency and Bogor City.
- Deduplication of administrative proliferation. Law No. 14 of 2022 partitioned Papua into four additional provinces. BPS lists the 26 affected regencies under both the parent and the successor province, producing duplicate region–year records. Inspection confirmed these pairs to be complementary rather than contradictory: exactly one side carries an observation, and in the 2022–2023 transition years both carry identical values. An assertion verifying the absence of value conflicts was embedded in the pipeline before collapsing each pair to a single record.
- Name harmonization and merge. Seven regions are rendered differently across the two series, reflecting abbreviation (*Kep. Seribu* versus *Kepulauan Seribu*), spacing (*Kotabaru* versus *Kota Baru*), an orthographic inconsistency in the HDI series (*Kota Makasar*), and three regions whose HDI entries carry both historic and current designations following renaming. An explicit synonym mapping was applied prior to an inner join on the composite key {Region, Year}, yielding a merged panel of 5,140 region–year observations.
- Missingness filter. Regions exceeding a 10% threshold of missing observations within the temporal window were excluded rather than imputed, since heavy imputation on sparse socio-economic series risks introducing synthetic bias that would alter the very resilience profiles the analysis seeks to detect. No region met the exclusion criterion. Three individual records were removed, corresponding to poverty observations unavailable for Buton Selatan, Buton Tengah, and Muna Barat in 2015, three regencies established in 2014 for which disaggregated figures were not yet published. The resulting analysis panel comprises 5,137 observations across 514 regencies and cities over ten years.

2.3. Feature engineering and normalization

The k-means algorithm relies on Euclidean distance to determine cluster membership. A significant challenge in this dataset is the disparity in feature magnitude: poverty rate is a percentage (0–100 scale, typically < 30), while HDI is an index (0–100 scale, typically > 60). Without normalization, the algorithm would be biased towards the feature with larger numerical values (HDI) [19].

To resolve this, we applied Z-score scaling (StandardScaler) to project both features onto a standard normal distribution with a mean (μ) of 0 and a standard deviation (σ) of 1 [20], [21]. The transformation for each data point x is defined in (1).

$$z = \frac{x-\mu}{\sigma} \quad (1)$$

This step ensures that both poverty and HDI contribute equidistantly to the calculation of the cluster centroids.

2.4. K-means clustering algorithm

We employed the k-means algorithm, a partition-based clustering method that iteratively minimizes the within-cluster sum of squares (WCSS), also known as inertia [22], [23]. The objective function J is defined in (2). The model was fitted with k-means++ initialization and a fixed random seed ($random_state = 42$) to ensure deterministic, reproducible results.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} ||x_i^{(j)} - \mu_j||^2 \quad (2)$$

The determination of the optimal number of clusters (k) is a critical step in unsupervised learning. To prevent subjective bias, four internal validity indices were computed across $k = 2$ to 8, as reported in Table 1 and visualized in [24].

Table 1. Cluster validity indices for k = 2 to 8

k	WCSS	Silhouette	Davies-Bouldin	Calinski-Harabasz
2	5273.3	0.5454	0.7487	4869.6
3	3231.8	0.4073	0.8055	5593.6
4	2260.8	0.3991	0.7896	6064.5
5	1780.3	0.3846	0.8111	6121.0
6	1476.7	0.3600	0.8294	6113.7
7	1260.9	0.3637	0.8018	6111.8
8	1115.6	0.3533	0.8203	6015.5

The Elbow criterion shows a pronounced inflection at $k = 3$, beyond which additional partitions produce diminishing reductions in within-cluster variance. The Calinski-Harabasz index rises substantially from $k = 2$ to $k = 3$ (4869.6 to 5593.6), indicating improved between-cluster separation relative to within-cluster dispersion. The Silhouette coefficient, however attains its maximum at $k = 2$ (0.5454) and declines to 0.4073 at $k = 3$. This pattern is characteristic of continuous socio-economic data, in which a single binary split between advantaged and disadvantaged regions constitutes the sharpest geometric partition while offering limited analytical granularity.

The configuration $k = 3$ was therefore retained as a reasoned trade-off rather than as an optimum across all indices. A two-cluster solution would collapse the substantial heterogeneity within the intermediate group, which accounts for the majority of Indonesian regencies and cities, into either extreme. The three-cluster solution preserves this distinction and corresponds to the advanced, transitioning, and lagging taxonomy already employed in Indonesian regional planning, thereby satisfying both algorithmic adequacy and policy interpretability.

3. RESULTS AND DISCUSSION

3.1. Cluster centroid analysis

The execution of the K-means algorithm converged successfully, partitioning the 5.137 data points into three distinct clusters. The analysis of the cluster centroids (the arithmetic mean position of all the points in each cluster) reveals three dominant typologies of regional resilience in Indonesia. The characteristics of these centroids are summarized in Table 2.

Table 2. Centroid characteristics of Indonesian regional clusters (2015-2024)

Label (typology)	Avg. HDI (index)	Avg. poverty (%)	n	Description
Advanced (*maju*)	75.94 ± 4.02	6.82 ± 2.83	1,732	High human capital, urban centers
Transitioning (*transisi*)	67.66 ± 2.78	12.39 ± 4.48	2,914	National average, resource/industrial
Lagging (*tertinggal*)	57.48 ± 7.67	30.58 ± 5.48	491	Infrastructure deficit, remote areas

Advanced regions ($n = 1{,}732$) comprise major metropolitan areas and established economic hubs, and occupy a dense region of the feature space. Transitioning regions ($n = 2{,}914$) form the largest cluster, positioned near the national average: most Indonesian regencies have moved past critical poverty without reaching advanced development markers. Lagging regions ($n = 491$) are the most vulnerable archetype, with wide internal dispersion, and are typically remote archipelagic or highland areas with limited infrastructure connectivity.

Because K-means converges to a local optimum that depends on centroid initialization, the stability of the reported partition was assessed before interpreting it substantively [25]. The algorithm was refitted across 20 random seeds and the resulting assignments compared pairwise using the adjusted rand index, yielding a mean agreement of 0.9621 with a minimum of 0.9043. Sensitivity to sample composition was examined through 100 bootstrap resamples drawn at 80% of the panel, in which regions retained their original cluster assignment in 97.43% of cases. Both diagnostics indicate that the three-cluster structure is a property of the data rather than an artefact of a particular initialization or subsample, which is a precondition for reading the centroids in Table 2 as substantive regional archetypes.

3.2. Algorithmic outlier detection: the "east-west" paradox

Conventionally, Indonesian development is viewed through a West–East dichotomy, in which Western regions (Java and Sumatra) are advanced and Eastern regions are lagging [26], [27]. The clustering challenges this spatial bias. Mimika Regency, located in the Papuan highlands where every neighbouring regency falls in the Lagging cluster, recorded an HDI of 76.85 in 2024 against a provincial mean of 61.98 and a national mean of 71.92, and is assigned to the advanced cluster under the absolute specification. Its feature vector lies closer to the centroid of java's industrial regencies than to that of its immediate geographic neighbours (Figure 2).

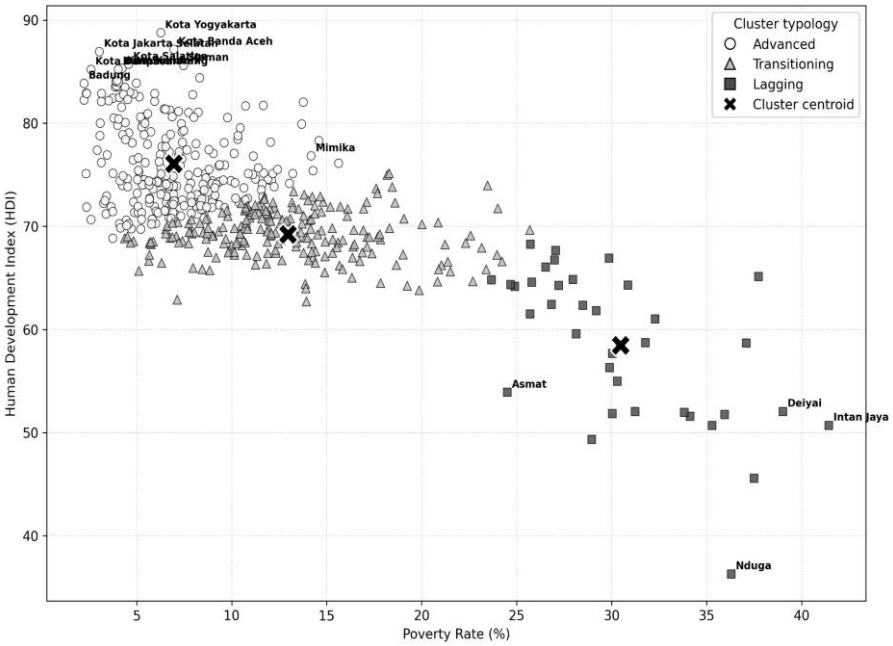


Figure 2. Cluster assignment of Indonesian regencies and cities

The dual specification introduced in section 3.3 qualifies this result in an instructive way. Mimika crossed into the absolute advanced tier in 2021 and remained there through 2024, yet under the relative specification it stayed in the transitioning tier throughout the decade. Extractive industry was therefore sufficient to detach the region from its geographic environment, but not to alter its standing within the national hierarchy a pattern consistent with the enclave economy literature, in which resource rents raise local aggregates without producing the broader structural transformation that would shift a region's rank [28]-[30]. The algorithm identifies development resilience as a function of infrastructural capacity rather than location alone, a nuance lost in aggregate regional statistics.

Cluster assignment of Indonesian regencies and cities in the poverty HDI feature space, 2024. Circles denote advanced, triangles transitioning, and squares lagging regions; crosses mark the cluster centroids. Note the positive outliers from Eastern Indonesia positioned above their geographic neighbours.

3.3. Longitudinal cluster transition analysis

The preceding sections characterize the cluster structure of the pooled decade. This section examines whether individual regions moved between clusters over the observation period, which is the

property that distinguishes a longitudinal application from a static snapshot. Because the centroids were fitted once on the pooled data and then held fixed across all ten years, any change in a region's cluster membership reflects a change in that region's own indicators rather than a redrawn partition boundary.

Transitions are reported under two complementary specifications. Absolute mobility applies the fixed centroids directly to the untransformed features, and therefore answers whether a region crossed a stable development threshold. Relative mobility standardizes both features within each year before clustering, and therefore answers whether a region improved its position with respect to its contemporaries. The distinction matters because national development levels rose throughout the period, so absolute improvement may occur without any change in relative standing.

3.3.1. Absolute mobility

Table 3 presents the transition matrix between the initial and terminal years of the observation window. Of the 511 regions observed in both 2015 and 2024, 334 (65.4%) remained in their original cluster and 177 (34.6%) advanced to a higher tier. No region moved downward.

Table 3. Absolute cluster transition matrix, 2015–2024 ($n = 511$)

2015	Lagging	Transitioning	Advanced	Total
Lagging	36	28	0	64
Transitioning	0	189	149	338
Advanced	0	0	109	109
Total	36	217	258	511

The absence of downward transitions is a substantive finding rather than an artefact. It indicates that no Indonesian regency or city experienced a deterioration in the joint HDI–poverty profile sufficient to fall below a fixed threshold during the decade, including through the COVID-19 contraction of 2020. Cluster composition shifted accordingly: the advanced cluster expanded from 109 to 258 regions, while the lagging cluster contracted from 64 to 36. Figure 3 traces this annual reconfiguration.

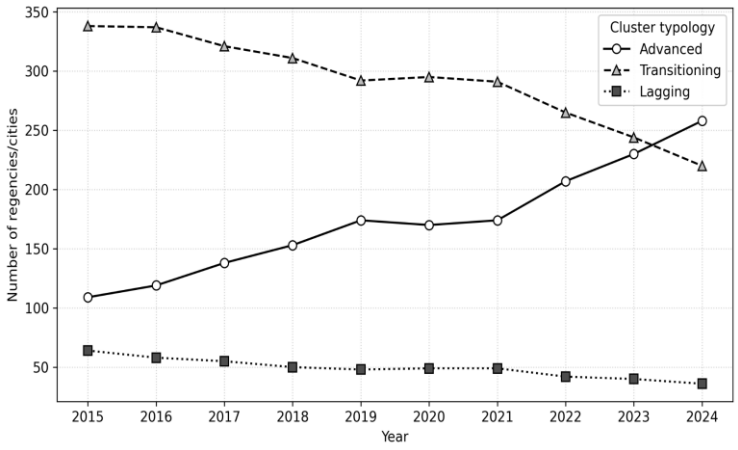


Figure 3. Annual reconfiguration

Among the 177 upwardly mobile regions, the largest HDI gains were recorded in Manggarai Barat (+6.61), Sampang (+6.47), Nias Barat (+6.42), and Manggarai (+6.41), all of which moved from Lagging to Transitioning. This concentration of gains in eastern Nusa Tenggara, Madura, and the Nias archipelago suggests that convergence was strongest precisely where initial conditions were weakest.

3.3.2. Relative mobility

Table 4 repeats the analysis with features standardized within each year. The picture inverts. Under this specification, 488 regions (95.5%) occupy the same relative tier in 2024 as in 2015. Only 6 regions (1.2%) improved their standing, while 17 (3.3%) declined. Of the 49 regions in the relative Lagging tier at the start of the period, 48 remained there a decade later; a single region exited.

Table 4. Relative cluster transition matrix, 2015–2024 ($n = 511$)

2015/2024	Lagging	Transitioning	Advanced	Total
Lagging	48	1	0	49
Transitioning	4	298	5	307
Advanced	0	13	142	155
Total	52	312	147	511

3.3.3. Interpretation

Read together, the two matrices describe a development trajectory that is simultaneously successful and static. In absolute terms, Indonesian regional development advanced broadly: a third of all regencies and cities crossed into a higher development tier and none regressed. In relative terms, the distribution barely moved: the ordering of regions with respect to one another was preserved almost intact. These findings are not contradictory but complementary. They indicate that the gains of the past decade were distributed broadly enough to lift regions across the national distribution, yet not differentially enough to alter the structure of inequality among them. Convergence in levels occurred without convergence in ranks [31]. For policy, this implies that programs succeeding at raising absolute welfare may nonetheless leave the spatial hierarchy of Indonesian development intact, and that interventions targeting relative disadvantage require a different design logic from those targeting absolute deprivation.

3.4. Discussion: implications for data-driven policy

The separation between the three archetypes indicates that a uniform national policy design is inefficient, and each archetype implies a different intervention logic. For lagging regions, the weight carried by the poverty feature suggests that digital infrastructure investment will yield low returns unless preceded by basic physical connectivity. Advanced regions, where HDI is approaching saturation, are positioned for industry 4.0 interventions. Transitioning regions the majority of the country represent the middle-income trap risk and require targeted economic diversification to close the remaining gap. The transition analysis adds a constraint that a static taxonomy would not reveal. Because the archetypes were broadly ascended in absolute terms but almost never reordered in relative terms, allocation rules keyed to fixed thresholds will mechanically shrink their own eligible population over time: the Lagging cluster contracted from 64 regions to 36 without any region falling into it. A programme evaluated against absolute criteria will therefore register success even where the spatial hierarchy it was designed to flatten remains intact. Rank-based eligibility criteria, recomputed annually, would target relative disadvantage more reliably. The Mimika case sharpens the same point: a region can cross an absolute threshold on the strength of a single sector while its position in the national distribution is unchanged, so cluster membership should inform prioritisation rather than substitute for diagnosis of local constraints.

3.5. Limitations

Four qualifications bound the interpretation of these results. First, k-means partitions the feature space using Euclidean distance, which implicitly assumes clusters of comparable and isotropic dispersion [24]. This assumption is not satisfied here: the standard deviation of HDI within the lagging cluster (±7.67) is nearly twice that of the advanced cluster (±4.02). Because the algorithm tends to impose boundaries of similar geometric extent, regions lying between the lagging and transitioning centroids may be drawn toward the tighter of the two, and the lagging cluster of 491 observations should accordingly be read as a conservative delineation. A finite mixture model with unconstrained covariance would relax this assumption, at the cost of the direct interpretability that motivated the present choice of algorithm. Second, observations are treated as independent points. The transition matrices therefore characterise net displacement between endpoints rather than the path taken between them, as discussed in section 3.3.3. Third, a three-cluster taxonomy compresses substantial within-group heterogeneity. The lagging cluster in particular contains regions separated by a wide margin of HDI, so a cluster label locates a region within the national distribution rather than diagnosing the specific constraints it faces; it is an instrument for prioritisation, not for programme design. Fourth, the feature space is restricted to HDI and poverty. These indicators were selected for their role in regional transfer fund allocation (section 2.1), but they omit dimensions directly relevant to the infrastructural interpretation advanced in section 3.2, including unemployment, physical and digital connectivity indices, and local fiscal capacity. Widening the feature vector to incorporate such indicators is a direct extension of this work.

4. CONCLUSION

Returning to the questions posed at the outset, the analysis identifies three stable regional archetypes advanced, transitioning, and lagging across 5,137 region–year observations covering all 514 Indonesian regencies and cities from 2015 to 2024, and shows that geographic position is an imperfect predictor of membership: resource-intensive regions such as Mimika sit well above their immediate neighbours. On the question of mobility, the decade was simultaneously successful and static. A third of all regions crossed into a higher absolute tier and none regressed, yet 95.5% occupied the same position relative to their contemporaries in 2024 as in 2015. Convergence occurred in levels without convergence in ranks a distinction invisible to a single cross-sectional partition, and one that bears directly on how regional programmes should be evaluated.

These findings are bounded by the limitations set out in section 3.5, of which the treatment of observations as independent points is the most consequential for a longitudinal application. Future work will extend the feature space beyond HDI and poverty to include connectivity and fiscal capacity indicators, and will explore sequential architectures such as long short-term memory (LSTM) or gated recurrent unit (GRU) models to capture year-to-year trajectories rather than endpoint displacement.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the Digital Business Study Program, Faculty of Economics and Business, Universitas Pendidikan Nasional (UNDIKNAS) for providing the academic environment and resources necessary to conduct this research. This research received no specific research grant or contract from any funding agency in the public, commercial, or not-for-profit sectors. We also acknowledge Statistics Indonesia (*Badan Pusat Statistik* / BPS) for making the socio-economic datasets publicly available, which served as the empirical foundation of this study.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Md. Wira Putra Dananjaya	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
I Gusti Ngurah Darma Paramartha		✓	✓	✓		✓		✓	✓	✓	✓	✓		
Made Ari Riska Dayanti	✓				✓		✓			✓				
Caren Angellina Mimaki					✓		✓	✓		✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

The datasets analyzed during the current study, along with the complete reproducible Python clustering pipeline, are publicly available in the GitHub repository at: https://github.com/Putradananjaya/indonesian-socio-

economic-clustering. The raw data were originally derived from resources available in the public domain Statistics Indonesia (*Badan Pusat Statistik* / BPS): https://www.bps.go.id.

REFERENCES

[1] Y. Rahmawati, I. R. Jamil, and I. Hidayah, "A Spatial Analysis on Heterogenous Determinant of Dengue Fever Cases in Indonesia," *Journal of Geovisualization and Spatial Analysis*, vol. 9, no. 1, 2025, doi: 10.1007/s41651-024-00212-1.

[2] R. K. K. Sitepu, Mhd. Asaad, M. Panjaitan, and J. R. Saragih, "Spatial Agglomeration and Regional Disparities: A Comparative Analysis Based on the Interregional Input-Output Model," *International Journal of Sustainable Development and Planning*, vol. 20, no. 11, pp. 4653–4668, 2025, doi: 10.18280/ijsdp.201110.

[3] M. I. Winata and M. F. Hasani, "Classifying Regional Poverty Levels in Indonesia: A Random Forest Analysis of Key Contributing Factors," *2025 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD)*, doi: 10.1109/ICoABCD67551.2025.11470881.

[4] R. C. Miranti and C. Mendez, "Social and Economic Convergence Across Districts in Indonesia: A Spatial Econometric Approach," *Bulletin of Indonesian Economic Studies*, vol. 59, no. 3, pp. 421–445, Sep. 2023, doi: 10.1080/00074918.2022.2071415.

[5] F. Santos-Marquez, A. B. Gunawan, and C. Mendez, "Regional income disparities, distributional convergence, and spatial effects: evidence from Indonesian regions 2010–2017," *GeoJournal*, vol. 87, no. 3, pp. 2373–2391, Jun. 2022, doi: 10.1007/s10708-021-10377-7.

[6] R. S. Aminda *et al.*, "Analysis of the influence of poverty, human development index, and unemployment on economic growth in the regency/city of Gorontalo Province: 2012–2021 study," *Journal of Infrastructure Policy and Development*, vol. 8, no. 4, 2024, doi: 10.24294/jipd.v8i4.3513.

[7] R. S. Dewi, "Urban vulnerability along the northern coast of Java: the intersection of development and resilience," *Geojournal*, vol. 90, no. 5, 2025, doi: 10.1007/s10708-025-11507-1.

[8] A. Setiawan, "Relationship model among Human Development Index, Gini Coefficient, Per capita Non-Food Expenditure, Per capita Gross Regional Domestic Product and Development Financing in Provinces in Indonesia for the 2015-2019 Period," *AIP Conference Proceedings*, vol. 2931, no. 1, 2023, doi: 10.1063/5.0177428.

[9] G. Z. Shafa and A. Fauzy, "K-means clustering analysis based on 2021 human development index components in regencies/cities Jawa Tengah," *AIP Conference Proceedings*, vol. 3024, no. 1, 2024, doi: 10.1063/5.0204364.

[10] T. Tumiran *et al.*, "Regional Clustering for Developing Electricity Systems in Archipelagic Area: A Case Study of Maluku and Papua Islands," *2021 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP)*, pp. 242–247, 2021, doi: 10.1109/ICT-PEP53949.2021.9601113.

[11] X. Du, "Impact of Climate Change and Regional Conflicts on Goal Interconnectivity under Network Analysis of the Sustainable Development Goals (SDGs)," *2024 3rd International Conference on Smart City Challenges & Outcomes for Urban Transformation (SCOUT)*, pp. 103–107, 2024, doi: 10.1109/SCOUT64349.2024.00030.

[12] S. Astari and Chotib, "Spatial Analysis of The Human Development Index in Indonesia Before and During The Covid-19 Pandemic," *IOP Conference Series: Earth and Environmental Science*, vol. 1291, no. 1, 2024, doi: 10.1088/1755-1315/1291/1/012002.

[13] R. Sijabat, "Democracy, Human Development, Income Distribution and Regional Economic Performance: A Panel Data Analysis of 34 Provinces in Indonesia," *Jurnal Ilmu Sosial dan Ilmu Politik*, vol. 28, no. 1, pp. 73–93, 2024, doi: 10.22146/jsp.76139.

[14] N. Aisyah, M. Y. Jinca, and L. Caroles, "Performance of Inter-City Transport within the Province in Improving Regional Accessibility," *International Journal of Civil Engineering*, vol. 12, no. 5, pp. 219–225, 2025, doi: 10.14445/23488352/IJCE-V12I5P117.

[15] R. N. M. Sanusi, D. R. S. Saputro, and R. Setiyowati, "Application of Robust Principal Component Analysis for Gross Regional Domestic Product of Provinces in Indonesia," *Journal of Physics: Conference Series*, vol. 1776, no. 1, 2021, doi: 10.1088/1742-6596/1776/1/012056.

[16] M. N. Alam, V. Laxmi, N. Kumar, R. Kumari, and S. Sharma, "Unsupervised Machine Learning for Anomaly Detection: A Systematic Review," *International Journal of Electrical and Electronics Engineering*, vol. 12, no. 7, pp. 34–72, 2025, doi: 10.14445/23488379/IJEEE-V12I7P104.

[17] BPS-Statistics Indonesia, "Human Development Index (HDI) and Poverty Profile in Indonesia 2015 – 2024," Statistics Indonesia, Jakarta, Indonesia, 2024. [Online]. Available: https://www.bps.go.id

[18] S. D. Putra, F. Utomo, F. Fitriani, I. Noer, and H. Heriansyah, "Predictive modeling of regional economic growth using agricultural and socioeconomic indicators," *International Journal of Advances in Applied Sciences (IJAAS)*, vol. 15, no. 2, pp. 872–882, 2026, doi: 10.11591/ijaas.v15.i2.pp872-882.

[19] R. Shetty, S. Gupta, V. Mediratta, S. Rai, and M. Geetha, "Optimizing Machine Learning-Based Ovarian Cancer Prediction Through Normalization Strategies," *IEEE Access*, vol. 13, pp. 128974–128995, 2025, doi: 10.1109/ACCESS.2025.3590871.

[20] S. Jena and S. Basel, "Classifying Global Economies Based on Sustainable Development Goals: A Data-Driven Clustering Approach," *Sustainable Development*, vol. 33, no. 3, pp. 4543–4556, 2025, doi: 10.1002/sd.3362.

[21] Q. Huang, "Multidimensional Data Mining Model Analysis of Financial Structure Optimization of Universities in Government Accounting Comprehensive Budget Performance Management," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 18, p. 15, 2026, doi: 10.70917/ijcisim-2026-0035.

[22] V. Singh, G. C. Hajare, S. A. Khan, and A. Kumari, "Decoding socio-economic disparities in Uttar Pradesh: a spatio-temporal analysis using Wroclaw Taxonomy and K-means unsupervised machine learning clustering," *Mathematical Population Studies*, 2026, doi: 10.1080/08898480.2025.2612641.

[23] M. Fatayat, S. H. Supangkat, and D. W. Anggara, "Clustering Socio-Economic Risk Zones for School Dropout: A Data-Driven Approach for Education Intervention in Indonesia," *2025 International Conference on ICT for Smart Society (ICISS)*, doi: 10.1109/ICISS66954.2025.11389433.

[24] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Information Sciences*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/j.ins.2022.11.139.

[25] C. Hennig, "Cluster-wise assessment of cluster stability," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 258–271, Sep. 2007, doi: 10.1016/j.csda.2006.11.025.

[26] R. J. Yuhan, P. M. Kutanegara, and S. R. Budiani, "Clustering of Stunting Prevalence with Gender Equality in Households Index and Socio-Demographic Variables in Regency/Cities of Bali, West Nusa Tenggara and East Nusa Tenggara Provinces," *IOP Conference Series: Earth and Environmental Science*, vol. 1443, no. 1, 2025, doi: 10.1088/1755-1315/1443/1/012005.

[27] Z. Azzahro, L. F. A. R. Chyntia, M. Nasar, and W. Widianto, "Implementation of the k-means clustering method for grouping poverty data: Case studies in Indonesia," *AIP Conference Proceedings*, vol. 2927, no. 1, 2024, doi: 10.1063/5.0193516.
[28] G. Hilson, Y. Hu, A. Hilson, J. R. Owen, É. Lèbre, and T. Sauerwein, "Rethinking resource enclavity in developing countries: Embedding Global Production Networks in gold mining regions," *Journal of Economic Geography*, vol. 24, no. 1, pp. 95–116, Feb. 2024, doi: 10.1093/jeg/lbad028.
[29] A. F. Anwar, D. W. Sari, H. A. T. Islamiya, R. A. R. Ahmad, and N. S. Alias, "Understanding the impact of mining activities and human capital improvement on achieving sustainable development goals; evidence from East Luwu, Indonesia," *Cogent Economics & Finance*, vol. 12, no. 1, p. 2386401, Dec. 2024, doi: 10.1080/23322039.2024.2386401.
[30] E. Warburton, "Nationalist enclaves: Industrialising the critical mineral boom in Indonesia," *The Extractive Industries and Society*, vol. 20, p. 101564, Dec. 2024, doi: 10.1016/j.exis.2024.101564.
[31] H. Aginta, A. B. Gunawan, and C. Mendez, "Regional income disparities and convergence clubs in Indonesia: new district-level evidence," *Journal of the Asia Pacific Economy*, vol. 28, no. 1, pp. 101–132, Jan. 2023, doi: 10.1080/13547860.2020.1868107.

BIOGRAPHIES OF AUTHORS



Md. Wira Putra Dananjaya received the Master's degree in Computer Science (M.Kom). He is currently a lecturer in the Department of Business Digital, Faculty of Economic and Business, Universitas Pendidikan Nasional, Bali, Indonesia. In addition, he serves as an active IT assessor specializing in the program analyst certification scheme. He has successfully led applied technological projects such as the Sabda Jari sign language application and a clean architecture-based tracer study platform. His primary academic and research interests are deeply rooted in information management, UI/UX design, and data science. In this study, he served as the principal investigator for formal data analysis and methodology. He can be contacted at email: putradananjaya@undiknas.ac.id.



I Gusti Ngurah Darma Paramartha received the D.Sc. degree (Doctor Habilitatus) in Computer Science from the Polish Academy of Sciences, Warsaw, Poland, with the dissertation "Soft Computing and Fractal Theory for Intelligent Manufacturing". He is a Professor of Computer Science in the Graduate Division, Tijuana Institute of Technology, Tijuana, Mexico. In addition, he is serving as research director of computer science and head of the research group on fuzzy logic and genetic algorithms. He is currently the vice-president of Hispanic American Fuzzy Systems Association (HAFSA) and president elect of International Fuzzy Systems Association (IFSA). He has published over 80 journal papers, 6 authored books, 20 edited books, and 200 papers in conference proceedings. His research interests are in type-2 fuzzy logic, fuzzy control, neuro-fuzzy and genetic-fuzzy hybrid approaches. He can be contacted at email: ocastillo@tectijuana.mx.



Made Ari Riska Dayanti received the Master's degree (M.T.) from Institut Teknologi Bandung (ITB), specializing in Environmental Engineering. She is an active lecturer in the Department of Environmental Engineering, Faculty of Engineering and Informatics, Universitas Pendidikan Nasional, Bali, Indonesia. In addition, she leads several community empowerment initiatives focusing on health resilience and environmental quality assessments. Her research interests encompass sustainable urban development, environmental spatial analysis, and community health. Her multidisciplinary expertise provided a crucial perspective for analyzing spatial disparities within this paper. She can be contacted at email: aririska@undiknas.ac.id.



Caren Angellina Mimaki received the Bachelor's degree in Accounting and the Master's degree in Management from Universitas Pendidikan Nasional, Bali, Indonesia. She is currently a lecturer in the Department of Management, Faculty of Economic and Business at the same university. In addition, she is actively involved in the university's international office. She is a prolific researcher with multiple high-impact publications focusing on wellness tourism, organizational behavior, and flexible work arrangements. Her research explores the intersection of digital marketing strategies, SME development, and inclusive entrepreneurship. She can be contacted at email: caren.mimaki@undiknas.ac.id.