



Analisis Disparitas Sosial-Ekonomi Antarprovinsi di Indonesia Menggunakan Pendekatan K-Means Clustering

Westri Andayanti

Pendidikan Fisika, FMIPA, Universitas Indraprasta PGRI, Jakarta, Indonesia

*Email Korespondensi: westri_andayanti@yahoo.com

Diterima: 14-08-2026 | Disetujui: 19-08-2026 | Diterbitkan: 21-08-2026

ABSTRACT

Socio-economic disparities across provinces remain an important issue in regional development due to differences in human development, labor market conditions, and poverty levels. This study aims to analyze socio-economic disparities across 38 provinces in Indonesia in 2025 using the K-Means clustering approach. Secondary data from Statistics Indonesia (BPS) were analyzed using three indicators: Human Development Index (HDI), Open Unemployment Rate (OUR), and percentage of poor population. The data were standardized using Z-score transformation, while the optimal number of clusters was determined using the Elbow Method and Silhouette Score. The results identified two clusters as the best configuration, with the highest Silhouette Score of 0.4961. Cluster 0 comprised 33 provinces with average HDI, OUR, and poverty rates of 75.70, 4.62%, and 8.40%, respectively. Cluster 1 comprised five provinces with averages of 64.69, 3.44%, and 22.60%, respectively. Principal Component Analysis showed that the first two components explained 94.40% of the total variance. These findings reveal distinct socio-economic typologies across Indonesian provinces and highlight the importance of multidimensional and differentiated regional development strategies.

Keywords: K-Means; socio-economic disparities; human development; unemployment; poverty..

ABSTRAK

Disparitas sosial-ekonomi antarprovinsi menjadi isu penting dalam pembangunan wilayah karena adanya perbedaan capaian pembangunan manusia, kondisi ketenagakerjaan, dan tingkat kemiskinan. Penelitian ini bertujuan menganalisis disparitas sosial-ekonomi pada 38 provinsi di Indonesia tahun 2025 menggunakan pendekatan *K-Means clustering*. Data sekunder yang bersumber dari Badan Pusat Statistik (BPS) dianalisis berdasarkan tiga indikator, yaitu Indeks Pembangunan Manusia (IPM), Tingkat Pengangguran Terbuka (TPT), dan persentase penduduk miskin. Data distandardisasi menggunakan Z-score, sedangkan jumlah kluster optimal ditentukan menggunakan *Elbow Method* dan *Silhouette Score*. Hasil penelitian menunjukkan dua kluster sebagai konfigurasi terbaik dengan *Silhouette Score* tertinggi sebesar 0,4961. Cluster 0 terdiri atas 33 provinsi dengan rata-rata IPM 75,70, TPT 4,62%, dan kemiskinan 8,40%, sedangkan Cluster 1 terdiri atas lima provinsi dengan rata-rata masing-masing sebesar 64,69, 3,44%, dan 22,60%. Hasil *Principal Component Analysis* menunjukkan bahwa dua komponen utama mampu menjelaskan 94,40% variasi data. Temuan menunjukkan adanya tipologi sosial-ekonomi yang berbeda antarprovinsi dan menegaskan pentingnya strategi pembangunan wilayah yang multidimensional dan terdiferensiasi.

Katakunci: *K-Means*; disparitas sosial-ekonomi; pembangunan manusia; pengangguran; kemiskinan.

Bagaimana Cara Sitasi Artikel ini:

Andayanti, W. (2026). Analisis Disparitas Sosial-Ekonomi Antarprovinsi di Indonesia Menggunakan Pendekatan K-Means Clustering. *Jurnal Ilmiah Ekonomi Dan Manajemen Indonesia*, 2(2), 4678-4691. <https://doi.org/10.63822/p3gk1679>

PENDAHULUAN

Pembangunan di Indonesia tidak hanya diukur melalui pencapaian pertumbuhan ekonomi, tetapi juga perlu mempertimbangkan peningkatan kesejahteraan masyarakat, kualitas sumber daya manusia, dan kondisi pasar tenaga kerja. Perbedaan kapasitas ekonomi, struktur ketenagakerjaan, ketersediaan infrastruktur, serta karakteristik geografis menyebabkan capaian pembangunan antarprovinsi menunjukkan tingkat yang beragam (Azim et al., 2022; Hafid et al., 2025). Wilayah dengan dukungan infrastruktur dan aktivitas investasi yang lebih baik cenderung memiliki peluang lebih besar dalam meningkatkan kualitas pembangunan manusia, sedangkan wilayah yang menghadapi berbagai kendala struktural masih mengalami tantangan dalam memperbaiki kondisi sosial dan ketenagakerjaan (Bina et al., 2025; Rahmah et al., 2025). Heterogenitas tersebut menunjukkan bahwa kebijakan pembangunan yang diterapkan secara seragam belum tentu menghasilkan dampak yang sama pada setiap wilayah. Oleh karena itu, identifikasi terhadap karakteristik dan disparitas sosial-ekonomi antarprovinsi menjadi penting untuk memperoleh gambaran yang lebih komprehensif mengenai perbedaan capaian pembangunan sekaligus menjadi dasar dalam merumuskan strategi pembangunan yang lebih tepat sasaran dan sesuai dengan karakteristik masing-masing wilayah.

Indeks Pembangunan Manusia (IPM), Tingkat Pengangguran Terbuka (TPT), dan persentase penduduk miskin merupakan indikator yang saling melengkapi dalam menggambarkan kondisi sosial-ekonomi suatu wilayah. IPM mencerminkan capaian pembangunan manusia melalui dimensi kesehatan, pendidikan, dan standar hidup layak yang berkaitan dengan peningkatan kualitas serta produktivitas sumber daya manusia (Suhendra et al., 2025; Yustie, 2017). Sementara itu, TPT menggambarkan proporsi angkatan kerja yang belum terserap dalam kegiatan ekonomi dan dapat digunakan untuk melihat dinamika pasar tenaga kerja suatu wilayah (Afifah & Imaningsih, 2025; Zahra et al., 2024). Persentase penduduk miskin merepresentasikan dimensi kesejahteraan masyarakat dengan menunjukkan proporsi penduduk yang masih menghadapi keterbatasan dalam memenuhi kebutuhan dasar (Dzahabiy et al., 2025). Ketiga indikator tersebut merepresentasikan dimensi pembangunan yang berbeda, yaitu kualitas manusia, kondisi ketenagakerjaan, dan kesejahteraan ekonomi. Oleh karena itu, analisis ketiganya secara simultan dapat memberikan gambaran yang lebih komprehensif mengenai heterogenitas kondisi sosial-ekonomi antarprovinsi dibandingkan dengan penggunaan satu indikator secara terpisah.

Kondisi empiris menunjukkan adanya variasi yang cukup lebar dalam karakteristik sosial-ekonomi antarprovinsi di Indonesia. Berdasarkan data tahun 2025 yang digunakan dalam penelitian ini, IPM pada 38 provinsi memiliki rata-rata sebesar 74,25, dengan nilai minimum 54,91 dan maksimum 85,05, yang memperlihatkan adanya perbedaan capaian pembangunan manusia antarwilayah. Kondisi ketenagakerjaan juga menunjukkan variasi, dengan rata-rata TPT sebesar 4,47% dan rentang antara 1,49% hingga 6,96%. Sementara itu, persentase penduduk miskin memiliki rata-rata sebesar 10,27%, dengan nilai minimum 3,42% dan maksimum 29,45%. Rentang yang cukup lebar pada ketiga indikator tersebut menunjukkan bahwa kondisi sosial-ekonomi provinsi di Indonesia bersifat heterogen dan tidak dapat dijelaskan secara memadai melalui satu indikator saja. Selain itu, perbedaan karakteristik pembangunan manusia, ketenagakerjaan, dan kemiskinan antarwilayah juga telah menjadi perhatian dalam berbagai penelitian terdahulu (Bina et al., 2025; Rachmawati et al., 2025; Wahyuni et al., 2025). Oleh karena itu, analisis yang mempertimbangkan ketiga indikator secara simultan diperlukan untuk mengidentifikasi pola disparitas sosial-ekonomi antarprovinsi dan menyediakan informasi yang lebih komprehensif sebagai dasar penyusunan kebijakan pembangunan yang terdiferensiasi (Harahap & Khaira, 2025; Yusman & Berliana,

2025).

Sejumlah penelitian telah menerapkan metode klasterisasi untuk mengidentifikasi heterogenitas dan mengelompokkan wilayah di Indonesia berdasarkan berbagai indikator sosial-ekonomi. Wahyuni et al. (2025), misalnya, menerapkan algoritma K-Means untuk mengelompokkan provinsi berdasarkan indikator kemiskinan dan menghasilkan tiga cluster dengan karakteristik yang berbeda. Penelitian lain menggunakan Fuzzy C-Means untuk mengelompokkan wilayah di Jawa Timur berdasarkan Indeks Pembangunan Manusia (IPM) dan menunjukkan adanya kelompok wilayah dengan capaian pembangunan manusia yang memerlukan perhatian lebih besar (Qori'atunnadyah, 2023). Sementara itu, Yusman dan Berliana (2025) menggunakan pendekatan klasterisasi berdasarkan tujuh indikator pembangunan untuk mengidentifikasi disparitas regional dan menunjukkan pentingnya intervensi kebijakan yang mempertimbangkan karakteristik masing-masing kelompok wilayah. Meskipun berbagai penelitian tersebut menunjukkan relevansi metode klasterisasi dalam menganalisis disparitas wilayah, masih terdapat ruang untuk mengembangkan analisis yang secara khusus mengintegrasikan dimensi pembangunan manusia, kondisi pasar tenaga kerja, dan kemiskinan dalam satu kerangka pengelompokan antarprovinsi. Integrasi ketiga dimensi tersebut penting karena masing-masing merepresentasikan aspek sosial-ekonomi yang berbeda sehingga pengelompokan secara simultan dapat menghasilkan tipologi wilayah yang lebih komprehensif.

Penggunaan algoritma K-Means dalam penelitian ini didasarkan pada karakteristik data sosial-ekonomi antarprovinsi yang bersifat multidimensional dan memiliki pola heterogen. K-Means memungkinkan provinsi dikelompokkan berdasarkan tingkat kemiripan karakteristik tanpa memerlukan kategori yang ditentukan sebelumnya, sehingga tipologi wilayah dapat terbentuk berdasarkan pola empiris yang terdapat dalam data (Vidyananta et al., 2025). Sebelum proses klasterisasi, standardisasi dilakukan untuk menyetarakan skala antarindikator sehingga perbedaan satuan pengukuran tidak memberikan pengaruh yang tidak proporsional terhadap pembentukan cluster (Nurba et al., 2025). Jumlah cluster optimal selanjutnya dievaluasi menggunakan *Elbow Method* dan *Silhouette Score* untuk memperoleh konfigurasi pengelompokan yang memiliki tingkat kohesi dan separasi yang memadai (Utami & Pramana, 2024). Selain itu, *Principal Component Analysis* (PCA) digunakan untuk mereduksi dimensi dan memvisualisasikan struktur pengelompokan dalam ruang dua dimensi sehingga pola kedekatan dan pemisahan antarprovinsi dapat diamati dengan lebih jelas (Ayberkin & Sebetci, 2026; Bhahari & Kusnawi, 2024). Dengan demikian, kombinasi standardisasi, K-Means, evaluasi jumlah cluster, dan PCA memungkinkan pembentukan tipologi sosial-ekonomi yang didasarkan pada karakteristik empiris wilayah, bukan pada kategori administratif yang telah ditentukan sebelumnya.

Berdasarkan uraian tersebut, penelitian ini bertujuan menganalisis disparitas sosial-ekonomi antarprovinsi di Indonesia melalui pengelompokan 38 provinsi berdasarkan tiga indikator, yaitu Indeks Pembangunan Manusia (IPM), Tingkat Pengangguran Terbuka (TPT), dan persentase penduduk miskin tahun 2025 menggunakan algoritma K-Means. Penelitian ini menawarkan pendekatan multidimensional dengan mengintegrasikan dimensi pembangunan manusia, kondisi ketenagakerjaan, dan kemiskinan dalam satu kerangka analisis untuk menghasilkan tipologi sosial-ekonomi antarprovinsi. Berbeda dengan pengelompokan yang hanya bertumpu pada satu indikator, penelitian ini menerapkan standardisasi data, mengevaluasi jumlah cluster menggunakan *Elbow Method* dan *Silhouette Score*, serta memanfaatkan *Principal Component Analysis* (PCA) untuk memvisualisasikan struktur pengelompokan. Pendekatan tersebut diharapkan mampu memberikan gambaran yang lebih komprehensif mengenai heterogenitas karakteristik sosial-ekonomi antarwilayah. Hasil penelitian diharapkan dapat memberikan bukti empiris

mengenai tipologi dan karakteristik kelompok provinsi serta menjadi bahan pertimbangan dalam merumuskan strategi pembangunan dan kebijakan sosial-ekonomi yang lebih terarah, terdiferensiasi, dan sesuai dengan karakteristik masing-masing wilayah.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan analisis kluster untuk mengidentifikasi pola disparitas sosial-ekonomi antarprovinsi di Indonesia. Unit analisis mencakup **38 provinsi**, dengan menggunakan data sekunder tahun **2025**. Analisis dilakukan menggunakan bahasa pemrograman **Python** melalui Google Colab. Tahapan analisis meliputi pengumpulan dan prapemrosesan data, analisis statistik deskriptif, standardisasi data, penentuan jumlah kluster optimal, pembentukan kluster menggunakan algoritma *K-Means*, analisis karakteristik kluster, serta visualisasi menggunakan *Principal Component Analysis* (PCA).

1. Data dan Variabel Penelitian

Data penelitian bersumber dari **Badan Pusat Statistik (BPS)** dan mencakup tiga indikator yang merepresentasikan dimensi pembangunan manusia, ketenagakerjaan, dan kesejahteraan ekonomi. Indikator yang digunakan adalah **Indeks Pembangunan Manusia (IPM)**, **Tingkat Pengangguran Terbuka (TPT)**, dan **persentase penduduk miskin**. Pemilihan ketiga indikator dimaksudkan untuk memperoleh gambaran multidimensional mengenai karakteristik sosial-ekonomi antarprovinsi. IPM merepresentasikan capaian pembangunan manusia, TPT menggambarkan kondisi pasar tenaga kerja, sedangkan persentase penduduk miskin menunjukkan kondisi kesejahteraan ekonomi masyarakat.

Data dari masing-masing indikator digabungkan berdasarkan nama provinsi sehingga menghasilkan satu set data yang terdiri atas 38 observasi. Agregat nasional Indonesia tidak dimasukkan dalam proses klusterisasi karena unit analisis penelitian adalah provinsi. Sebelum dianalisis, data diperiksa untuk memastikan konsistensi nama wilayah, kesesuaian tipe data, dan tidak adanya nilai yang hilang (*missing values*).

2. Statistik Deskriptif dan Standardisasi Data

Analisis statistik deskriptif dilakukan untuk memberikan gambaran awal mengenai karakteristik setiap indikator. Statistik yang dihitung meliputi jumlah observasi, nilai rata-rata, standar deviasi, nilai minimum, dan maksimum. Analisis ini digunakan untuk mengidentifikasi variasi awal kondisi sosial-ekonomi antarprovinsi sebelum dilakukan proses klusterisasi.

Karena IPM, TPT, dan persentase penduduk miskin memiliki rentang dan distribusi nilai yang berbeda, seluruh variabel kemudian distandardisasi menggunakan metode **Z-score**. Standardisasi diperlukan agar tidak terdapat variabel yang memberikan pengaruh lebih besar terhadap penghitungan jarak hanya karena perbedaan skala pengukuran. Standardisasi dirumuskan sebagai:

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}$$

dengan z_{ij} merupakan nilai terstandarisasi indikator j pada provinsi i , x_{ij} merupakan nilai asli,

μ_j merupakan rata-rata j indikator σ_j , dan merupakan standar deviasi indikator tersebut. Setelah standarisasi, setiap variabel memiliki rata-rata mendekati nol dan standar deviasi sebesar satu.

3. Penentuan Jumlah Kluster Optimal

Jumlah kluster optimal tidak ditetapkan secara subjektif, tetapi dievaluasi menggunakan Elbow Method dan Silhouette Score. Pengujian dilakukan untuk jumlah kluster $K = 2$ sampai $K = 6$. *Elbow Method* mengevaluasi perubahan *within-cluster sum of squares* atau *inertia* pada setiap jumlah kluster. Nilai *inertia* dapat dinyatakan sebagai:

$$WCSS = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2$$

Dengan K merupakan jumlah kluster, merupakan kluster ke- C_k , x_i adalah observasi dalam kluster, dan μ_k merupakan centroid kluster ke k . Selanjutnya, *Silhouette Score* digunakan untuk mengevaluasi tingkat kekompakan observasi dalam kluster sekaligus keterpisahannya dari kluster lain. Nilai *silhouette* untuk observasi dihitung dengan:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Dengan $a(i)$ merupakan rata-rata jarak observasi terhadap seluruh observasi lain dalam kluster yang sama, sedangkan $b(i)$ merupakan rata-rata jarak minimum observasi terhadap observasi pada kluster terdekat.

4. Klusterisasi Menggunakan K-Means

Klusterisasi dilakukan menggunakan algoritma K-Means terhadap data yang telah distandarisasi. Algoritma ini mengelompokkan provinsi berdasarkan kedekatan karakteristik ketiga indikator sosial-ekonomi. Secara matematis, *K-Means* meminimalkan fungsi objektif:

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2$$

Proses K-Means dilakukan secara iteratif melalui penentuan centroid awal, pengalokasian setiap provinsi ke centroid terdekat, penghitungan kembali centroid berdasarkan anggota kluster, dan pengulangan proses sampai keanggotaan kluster mencapai kondisi konvergen. Untuk meningkatkan reproduibilitas hasil, pemodelan dilakukan menggunakan `random_state=42` dan `n_init=10`.

Setelah kluster terbentuk, karakteristik masing-masing kluster dianalisis berdasarkan rata-rata nilai asli IPM, TPT, dan persentase penduduk miskin, bukan semata-mata berdasarkan nomor kluster. Pendekatan tersebut digunakan agar interpretasi tipologi wilayah didasarkan pada karakteristik empiris setiap kelompok.

5. Principal Component Analysis (PCA)

Principal Component Analysis digunakan sebagai teknik reduksi dimensi untuk memvisualisasikan struktur hasil klasterisasi. Tiga indikator yang telah distandardisasi ditransformasikan menjadi dua komponen utama, yaitu PC1 dan PC2. Secara umum, komponen utama dapat dinyatakan sebagai:

$$PC_m = a_{m1}Z_1 + a_{m2}Z_2 + \dots + a_{mp}Z_p$$

Dengan PC_m merupakan komponen utama ke-m, a_{mj} merupakan bobot atau loading variabel j , dan Z_j merupakan variabel yang telah distandardisasi. Proporsi varians yang dapat dijelaskan oleh setiap komponen dievaluasi menggunakan explained variance ratio. PCA pada penelitian ini digunakan terutama sebagai alat visualisasi dan interpretasi struktur klaster, sedangkan pembentukan klaster tetap dilakukan menggunakan tiga indikator terstandardisasi.

HASIL DAN PEMBAHASAN

1. Statistik Deskriptif Kondisi Sosial-Ekonomi Antarprovinsi

Analisis statistik deskriptif dilakukan untuk memberikan gambaran awal mengenai karakteristik sosial-ekonomi 38 provinsi di Indonesia tahun 2025. Indikator yang dianalisis meliputi Indeks Pembangunan Manusia (IPM), Tingkat Pengangguran Terbuka (TPT), dan persentase penduduk miskin. Hasil statistik deskriptif disajikan pada Tabel 1.

Tabel 1. Statistik Deskriptif Indikator Sosial-Ekonomi Antarprovinsi di Indonesia Tahun 2025

Indikator	N	Rata-rata	Std. Deviasi	Minimum	Maksimum
IPM	38	74,25	5,20	54,91	85,05
TPT (%)	38	4,47	1,38	1,49	6,96
Kemiskinan (%)	38	10,27	6,27	3,42	29,45

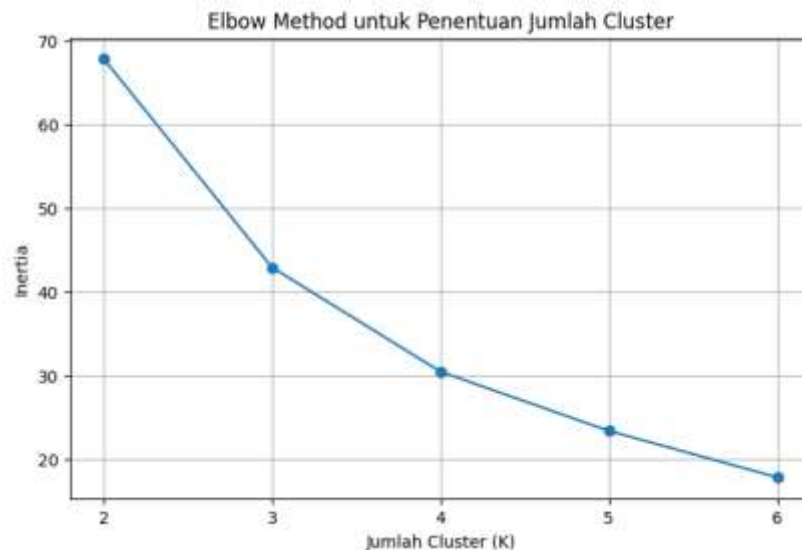
Sumber: Data BPS 2025, diolah menggunakan Python.

Hasil pada Tabel 1 menunjukkan adanya variasi kondisi sosial-ekonomi yang cukup besar antarprovinsi. IPM memiliki rata-rata sebesar 74,25 dengan nilai minimum 54,91 dan maksimum 85,05. Rentang yang mencapai 30,14 poin menunjukkan bahwa capaian pembangunan manusia belum merata di seluruh provinsi. Perbedaan tersebut menggambarkan heterogenitas kualitas pembangunan manusia yang menjadi salah satu dimensi penting dalam memahami disparitas sosial-ekonomi antarwilayah.

TPT memiliki rata-rata sebesar 4,47%, dengan nilai minimum 1,49% dan maksimum 6,96%. Sementara itu, persentase penduduk miskin memiliki rata-rata 10,27%, dengan rentang yang relatif lebar antara 3,42% hingga 29,45%. Standar deviasi kemiskinan sebesar 6,27 juga memperlihatkan variasi yang cukup tinggi antarprovinsi. Kondisi tersebut mengindikasikan bahwa disparitas wilayah tidak hanya terlihat dari pembangunan manusia, tetapi juga dari kondisi pasar tenaga kerja dan tingkat kesejahteraan masyarakat. Oleh karena itu, ketiga indikator dianalisis secara simultan menggunakan pendekatan klasterisasi.

2. Penentuan Jumlah Kluster Optimal

Sebelum penerapan algoritma *K-Means*, jumlah kluster optimal ditentukan menggunakan *Elbow Method* dan *Silhouette Score*. Pengujian dilakukan pada jumlah kluster hingga . *Elbow Method* menunjukkan bahwa nilai *inertia* mengalami penurunan seiring dengan meningkatnya jumlah kluster. Namun, penambahan jumlah kluster setelah titik tertentu menghasilkan penurunan *inertia* yang semakin kecil.



Gambar 1. Elbow Method untuk Penentuan Jumlah Kluster Optimal

Untuk memperkuat penentuan jumlah kluster, evaluasi dilanjutkan menggunakan *Silhouette Score*. Hasil pengujian disajikan pada Tabel 2.

Tabel 2. Hasil Evaluasi Silhouette Score

Jumlah Kluster (K)	Silhouette Score
2	0,4961
3	0,4013
4	0,4065
5	0,4206
6	0,3827

Sumber: Hasil pengolahan data menggunakan Python.

Nilai *Silhouette Score* tertinggi diperoleh pada K=2 sebesar 0,4961. Nilai tersebut lebih tinggi dibandingkan K=3 sebesar 0,4013, K=4 sebesar 0,4065, K=5 sebesar 0,4206, dan K=6 sebesar 0,3827. Berdasarkan hasil tersebut, dua kluster dipilih sebagai konfigurasi yang paling baik di antara alternatif yang diuji. Nilai 0,4961 menunjukkan bahwa struktur kluster memiliki pemisahan moderat; dengan demikian, hasil pengelompokan cukup informatif untuk mengidentifikasi tipologi utama, tetapi tidak menunjukkan pemisahan yang sepenuhnya tegas pada seluruh observasi.

3. Hasil Klasterisasi K-Means

Penerapan algoritma *K-Means* dengan menghasilkan dua kelompok provinsi dengan karakteristik sosial-ekonomi yang berbeda. Cluster 0 terdiri atas 33 provinsi, sedangkan Cluster 1 terdiri atas 5 provinsi. Untuk memahami karakteristik masing-masing kelompok, centroid dikembalikan dan diinterpretasikan dalam skala asli sebagaimana disajikan pada Tabel 3.

Tabel 3. Karakteristik Centroid Hasil Klasterisasi

Klaster	IPM	TPT (%)	Kemiskinan (%)	Jumlah Provinsi
Cluster 0	75,70	4,62	8,40	33
Cluster 1	64,69	3,44	22,60	5

Sumber: Hasil pengolahan data menggunakan Python.

Cluster 0 menunjukkan rata-rata IPM sebesar 75,70, TPT sebesar 4,62%, dan kemiskinan sebesar 8,40%. Sebaliknya, Cluster 1 memiliki rata-rata IPM sebesar 64,69, TPT 3,44%, dan kemiskinan 22,60%. Perbedaan yang paling menonjol terlihat pada IPM dan kemiskinan. Selisih rata-rata IPM antarklaster mencapai 11,01 poin, sedangkan tingkat kemiskinan Cluster 1 lebih tinggi 14,20 poin persentase dibandingkan Cluster 0.

Berdasarkan profil tersebut, Cluster 0 dapat diinterpretasikan sebagai kelompok dengan pembangunan manusia relatif lebih tinggi dan kemiskinan relatif lebih rendah, sedangkan Cluster 1 merupakan kelompok dengan pembangunan manusia relatif lebih rendah dan kemiskinan relatif lebih tinggi. Penamaan tersebut lebih tepat dibandingkan menggunakan kategori “maju” dan “tertinggal”, karena *K-Means* membentuk kelompok berdasarkan kemiripan statistik dan bukan berdasarkan klasifikasi pembangunan yang ditetapkan pemerintah.

4. Karakteristik Provinsi pada Cluster 1

Hasil klasterisasi menunjukkan bahwa Cluster 1 terdiri atas lima provinsi, yaitu Nusa Tenggara Timur, Papua Barat, Papua Selatan, Papua Tengah, dan Papua Pegunungan. Karakteristik masing-masing provinsi disajikan pada Tabel 4.

Tabel 4. Anggota dan Karakteristik Cluster 1

Provinsi	IPM	TPT (%)	Kemiskinan (%)
Nusa Tenggara Timur	69,89	3,31	17,50
Papua Barat	68,48	4,55	19,58
Papua Selatan	69,54	4,04	19,26
Papua Tengah	60,64	3,62	29,45
Papua Pegunungan	54,91	1,68	27,21

Sumber: Hasil pengolahan data menggunakan Python.

Kelima provinsi tersebut memiliki pola yang relatif serupa terutama dari sisi IPM dan kemiskinan. Papua Tengah memiliki tingkat kemiskinan tertinggi di antara lima anggota cluster, yaitu 29,45%, dengan IPM sebesar 60,64. Papua Pegunungan menunjukkan IPM terendah sebesar 54,91 dan tingkat kemiskinan sebesar 27,21%. Nusa Tenggara Timur, Papua Barat, dan Papua Selatan memiliki IPM yang lebih tinggi

dibandingkan Papua Tengah dan Papua Pegunungan, tetapi masih tergabung dalam kelompok yang sama berdasarkan kemiripan kombinasi ketiga indikator.

Temuan tersebut memperlihatkan bahwa karakteristik sosial-ekonomi wilayah tidak sebaiknya dibaca hanya berdasarkan satu indikator. K-Means mempertimbangkan posisi setiap provinsi secara simultan pada tiga dimensi, sehingga keanggotaan klaster merupakan hasil kombinasi IPM, TPT, dan kemiskinan.

5. Hubungan Pengangguran dengan Tipologi Sosial-Ekonomi

Salah satu temuan yang menarik adalah bahwa Cluster 1 justru mempunyai rata-rata TPT yang lebih rendah, yaitu 3,44%, dibandingkan Cluster 0 sebesar 4,62%. Dengan demikian, rendahnya tingkat pengangguran terbuka dalam data ini tidak secara otomatis berjalan bersama dengan IPM yang lebih tinggi atau kemiskinan yang lebih rendah.

Temuan ini perlu diinterpretasikan secara hati-hati. TPT mengukur proporsi pengangguran dalam angkatan kerja, tetapi tidak secara langsung menggambarkan kualitas pekerjaan, produktivitas tenaga kerja, tingkat pendapatan, proporsi pekerja informal, maupun setengah pengangguran. Oleh karena itu, suatu wilayah dapat memiliki TPT relatif rendah tetapi tetap menghadapi tingkat kemiskinan yang tinggi dan pembangunan manusia yang rendah. Hasil clustering memperlihatkan pentingnya penggunaan indikator multidimensional ketika mengevaluasi disparitas sosial-ekonomi wilayah.

Dari perspektif manajemen pembangunan, kondisi tersebut juga menunjukkan bahwa strategi kebijakan ketenagakerjaan tidak cukup hanya berorientasi pada penurunan tingkat pengangguran. Perhatian juga perlu diberikan pada kualitas dan produktivitas pekerjaan, peningkatan kompetensi sumber daya manusia, serta kemampuan aktivitas ekonomi menghasilkan pendapatan yang memadai. Dengan demikian, pengelolaan pembangunan wilayah membutuhkan strategi yang disesuaikan dengan profil sosial-ekonomi masing-masing kelompok provinsi.

6. Visualisasi Struktur Klaster Menggunakan PCA

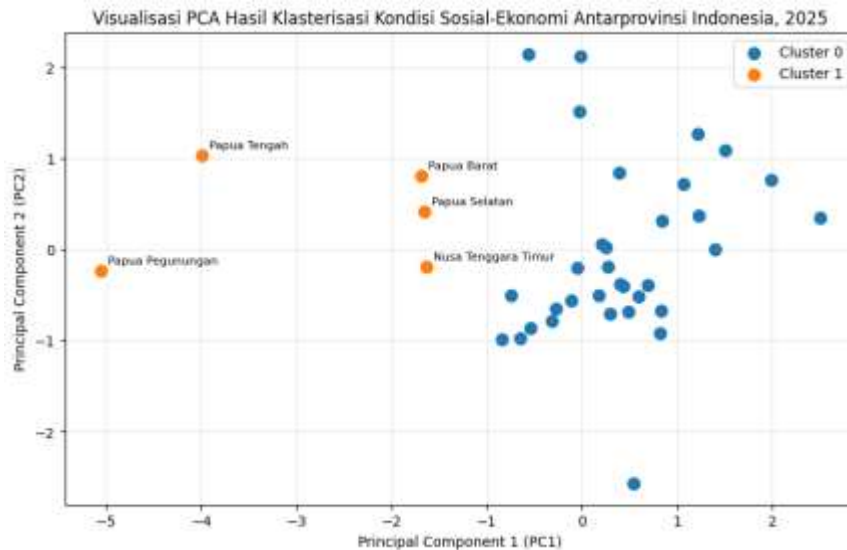
Principal Component Analysis digunakan untuk melihat struktur hasil klasterisasi dalam ruang dua dimensi. Hasil PCA menunjukkan bahwa PC1 mampu menjelaskan 65,81% variasi data, sedangkan PC2 menjelaskan 28,60%. Secara kumulatif, kedua komponen utama tersebut menjelaskan 94,40% dari total variasi data.

Tabel 5. Explained Variance Principal Component Analysis

Komponen	Explained Variance (%)
PC1	65,81
PC2	28,60
Kumulatif	94,40

Sumber: Hasil pengolahan data menggunakan Python.

Proporsi kumulatif sebesar 94,40% menunjukkan bahwa dua komponen utama mampu mempertahankan sebagian besar informasi yang terkandung dalam tiga indikator awal. Dengan demikian, visualisasi dua dimensi dapat digunakan untuk membantu memahami struktur pengelompokan tanpa kehilangan sebagian besar variasi data.



Gambar 2. Visualisasi PCA Hasil Klasterisasi Kondisi Sosial-Ekonomi Antarprovinsi di Indonesia Tahun 2025

Visualisasi PCA memperlihatkan adanya pemisahan antara sebagian besar anggota Cluster 0 dan lima provinsi pada Cluster 1. Papua Tengah dan Papua Pegunungan tampak berada relatif jauh dari konsentrasi utama provinsi, sedangkan Nusa Tenggara Timur, Papua Barat, dan Papua Selatan berada lebih dekat terhadap batas kelompok. Pola tersebut konsisten dengan variasi nilai indikator asli, terutama IPM dan kemiskinan.

7. Pembahasan Disparitas Sosial-Ekonomi Antarprovinsi

Secara keseluruhan, hasil penelitian menunjukkan bahwa kondisi sosial-ekonomi 38 provinsi di Indonesia dapat dibedakan menjadi dua tipologi utama berdasarkan kombinasi pembangunan manusia, pengangguran terbuka, dan kemiskinan. Sebagian besar provinsi, yaitu 33 provinsi atau sekitar 86,84%, berada pada kelompok dengan IPM relatif lebih tinggi dan kemiskinan lebih rendah. Sebanyak 5 provinsi atau sekitar 13,16% berada pada kelompok dengan IPM relatif lebih rendah dan kemiskinan lebih tinggi.

Perbedaan centroid memperlihatkan bahwa IPM dan kemiskinan menjadi karakteristik yang paling jelas membedakan kedua kelompok secara deskriptif. Namun, pernyataan tersebut tidak berarti bahwa kedua variabel secara formal terbukti sebagai variabel “paling berpengaruh”, karena K-Means bukan metode pengujian kausal. TPT memperlihatkan pola berbeda, dengan nilai rata-rata lebih rendah pada Cluster 1. Hal ini memperkuat argumentasi bahwa kondisi sosial-ekonomi wilayah merupakan fenomena multidimensional dan tidak dapat disimpulkan hanya berdasarkan rendah atau tingginya angka pengangguran.

Hasil tersebut memiliki implikasi terhadap perumusan strategi pembangunan wilayah. Provinsi dengan kombinasi IPM relatif rendah dan kemiskinan tinggi membutuhkan perhatian pada penguatan kualitas sumber daya manusia, peningkatan akses terhadap layanan dasar, serta pengembangan aktivitas ekonomi yang mampu meningkatkan produktivitas dan kesejahteraan. Sementara itu, provinsi pada Cluster 0 tidak dapat dianggap bebas dari permasalahan sosial-ekonomi, khususnya karena rata-rata TPT kelompok

tersebut justru lebih tinggi. Oleh sebab itu, hasil klasterisasi lebih tepat digunakan sebagai instrumen segmentasi awal wilayah untuk mendukung penyusunan kebijakan yang terdiferensiasi daripada sebagai peringkat keberhasilan pembangunan.

KESIMPULAN

Penelitian ini menunjukkan bahwa disparitas sosial-ekonomi antarprovinsi di Indonesia tahun 2025 dapat diidentifikasi melalui pengelompokan berdasarkan Indeks Pembangunan Manusia (IPM), Tingkat Pengangguran Terbuka (TPT), dan persentase penduduk miskin menggunakan algoritma *K-Means*. Hasil evaluasi menunjukkan bahwa dua klaster merupakan konfigurasi terbaik di antara $K=2$ hingga $K=6$, dengan *Silhouette Score* tertinggi sebesar 0,4961. Hasil klasterisasi membentuk Cluster 0 yang terdiri atas 33 provinsi dengan rata-rata IPM 75,70, TPT 4,62%, dan kemiskinan 8,40%, serta Cluster 1 yang terdiri atas lima provinsi, yaitu Nusa Tenggara Timur, Papua Barat, Papua Selatan, Papua Tengah, dan Papua Pegunungan, dengan rata-rata IPM 64,69, TPT 3,44%, dan kemiskinan 22,60%. Perbedaan karakteristik antarklaster terutama terlihat pada capaian pembangunan manusia dan tingkat kemiskinan, sedangkan TPT menunjukkan pola yang berbeda karena kelompok dengan IPM lebih rendah dan kemiskinan lebih tinggi justru memiliki rata-rata TPT yang lebih rendah. Temuan ini menegaskan pentingnya pendekatan multidimensional dalam memahami kondisi sosial-ekonomi wilayah. Hasil PCA turut mendukung representasi struktur klaster, dengan dua komponen utama mampu menjelaskan 94,40% variasi data. Hasil penelitian dapat digunakan sebagai dasar segmentasi awal wilayah untuk mendukung perumusan strategi pembangunan yang lebih sesuai dengan karakteristik sosial-ekonomi masing-masing kelompok provinsi. Penelitian selanjutnya dapat memperluas indikator dengan memasukkan PDRB per kapita, ketimpangan pendapatan, produktivitas tenaga kerja, TPAK, dan struktur pekerjaan serta menggunakan data multiyear untuk mengidentifikasi dinamika perubahan tipologi sosial-ekonomi antarprovinsi.

DAFTAR PUSTAKA

- Afifah, D. N., & Imaningsih, N. (2025). Analisis pengaruh laju pertumbuhan penduduk, indeks pembangunan manusia, pertumbuhan ekonomi, dan tingkat pengangguran terbuka terhadap tingkat kemiskinan di kabupaten pasuruan dan kabupaten gresik. *Deleted Journal*, 6(1), 109–122. <https://doi.org/10.57094/jpe.v8i2.3006>
- Ayberkin, D., & Sebetci, Ö. (2026). Analysis of Turkey's Demographic and Social Structure Using PCA And K-Means Clustering. *MANAS Sosyal Araştırmalar Dergisi*. <https://doi.org/10.33206/mjss.1481386>
- Azim, A. N., Sutjipto, H., & Ginanjar, R. A. F. (2022). *Determinan ketimpangan pembangunan ekonomi antarprovinsi di indonesia*. 2(1), 1–16. <https://doi.org/10.23969/jrie.v2i1.23>
- Bhahari, R. H., & Kusnawi, K. (2024). Clustering Analysis of Socio-Economic Districts/Cities In East Java Province Using PCA And Hierarchical Clustering Methods. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, 8(4), 2242–2251. <https://doi.org/10.33395/sinkron.v8i4.14078>
- Bina, A., Optari, L. K., & Widiya, H. (2025). Income inequality across provinces in Indonesia: An empirical analysis of human development, economic performance, and labor market conditions. *Priviet Social Sciences Journal*, 5(6), 119–126. <https://doi.org/10.55942/pssj.v5i6.617>

- Dzahabiy, A. S. S. F., Harini, S., & Juhari, J. (2025). Uji Kausalitas Granger dalam Menganalisis Hubungan Indeks Pembangunan Manusia, Tingkat Pengangguran Terbuka, dan Kemiskinan. *Jurnal Riset Mahasiswa Matematika*, 4(4), 149–160. <https://doi.org/10.18860/jrmm.v4i4.30406>
- Hafid, H., Meliyana, S. M., Muthahharah, I., & Mar'ah, Z. (2025). Implementation K-Medoids Algorithm for Clustering Indonesian Provinces by Poverty and Economic Indicators. *Quantitative Economics and Management Studies*, 6(2), 219–225. <https://doi.org/10.35877/454ri.qems3940>
- Harahap, E. S., & Khaira, I. (2025). Economic Growth and Income Inequality in Indonesia: A Systematic Review and Meta-Analysis of Two Decades of Research. *Outline Journal of Economic Studies*, 4(2), 98–107. <https://doi.org/10.61730/ax9kt271>
- Nurba, A. S. W. A., Fathahillah, D., Pratama, M. S., Bachtiar, M. R., Putra, M. C. A., & Sitio, S. L. M. (2025). Klasterisasi data : analisis kinerja k-means pada sektor pajak, ekspor, perikanan, modal, dan sumber daya. *JRIS: Jurnal Rekayasa Informasi Swadharma*, 5(2), 108–120. <https://doi.org/10.56486/jris.vol5no2.883>
- Qori'atunnadyah, M. (2023). Fuzzy C-Means for Regional Clustering in East Java Province Based on Human Development Index Indicators. *J Statistika*, 16(2), 524–534. <https://doi.org/10.36456/jstat.vol16.no2.a8240>
- Rachmawati, L. N., Lubis, F. R. A., Azzakiyyah, N. A., Rachmawati, L. N., Lubis, F. R. A., & Azzakiyyah, N. A. (2025). Regional income disparities in Indonesia: Insights from the Williamson index and panel data analysis. *Optimum: Jurnal Ekonomi & Pembangunan*, 15(2), 264–271. <https://doi.org/10.12928/optimum.v15i2.12953>
- Rahmah, H. A., Tamamudin, T., & Sadali, A. (2025). Analisis Proyeksi Tingkat Pengangguran Terbuka di 38 Provinsi Periode 2021-2025. *Journal Social Society*, 5(1), 241–259. <https://doi.org/10.54065/jss.5.1.2025.674>
- Suhendra, H., Sakti, A. I., Khaffi, M. A., Dalimunthe, D. Y., Zirtana, H., Juniar, R., Sari, D. N., Tama, M. R., & Verahditiya, K. (2025). Analysis of Population, Poverty, Unemployment Rate, and Gini Ratio on Human Development Index in Bangka Belitung. *Journal of Natural Sciences and Mathematics Research*. <https://doi.org/10.21580/jnsmr.v11i1.23671>
- Utami, N. R., & Pramana, S. (2024). Klasterisasi Bencana dan Dampaknya di Indonesia: Evaluasi Metode K-means dengan Integrasi PCA. *Prosiding Seminar Nasional Sains Data.*, 4(1), 367–377. <https://doi.org/10.33005/senada.v4i1.220>
- Vidyananta, I. P. A., Dermawan, K. T., & Vidyananta, I. P. A. (2025). Perbandingan algoritma k-means dan k-medoids dalam pengelompokan provinsi di indonesia berdasarkan indikator keadaan sekolah dasar. *JITET: Jurnal Informatika Dan Teknik Elektro Terapan*, 13(3S1). <https://doi.org/10.23960/jitet.v13i3s1.8145>
- Wahyuni, S., Hananto, A., Huda, B., Apriani, F. N., & Tukino, T. (2025). Identifying Regional Patterns of Poverty in Indonesia: a Clustering Approach Using K-Means. *International Journal of Computer and Information System*, 6(1), 45–52. <https://doi.org/10.29040/ijcis.v6i1.218>
- Yusman, A., & Berliana, S. M. (2025). Unveiling Regional Disparities in Indonesia: Clustering Provinces by Development Indicators. *Proceedings of The International Conference on Data Science and Official Statistics*. <https://doi.org/10.34123/icdsos.v2025i1.645>

- Yustie, R. (2017). *Analisis pengaruh indeks pembangunan manusia (ipm) dan tingkat pengangguran terbuka (tpt) terhadap kemiskinan kabupaten dan kota di provinsi jawa timur*. 49–57. <https://doi.org/10.30742/EQUILIBRIUM.V0I0.418>
- Zahra, B. S., Mevia, R. D. A., Sona, R. A., & Rohmi, M. L. (2024). *Pengaruh Indeks Pembangunan Manusia Dan Tingkat Pengangguran Terbuka Terhadap Kemiskinan Di Provinsi Lampung Pada Priode 2012-2023*. 1(3), 01–15. <https://doi.org/10.61132/jbep.v1i3.111>