

## PERBANDINGAN METODE K-MEANS DAN K-MEDOIDS DALAM CLUSTERING KETIMPANGAN PEMBANGUNAN PROVINSI DI INDONESIA TAHUN 2024 BERDASARKAN INDIKATOR SOSIAL EKONOMI

Fadhilah Putri Talsi<sup>1\*</sup>, Fonda Leviany<sup>2</sup>

<sup>1,2</sup>*Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas Terbuka*

\*Penulis Korepondensi : [fadhilahptalsi@gmail.com](mailto:fadhilahptalsi@gmail.com)

### ABSTRAK

Ketimpangan pembangunan antarprovinsi di Indonesia masih menjadi permasalahan yang ditandai oleh perbedaan kondisi ekonomi, pendidikan, dan kesejahteraan masyarakat. Penelitian ini mengelompokkan 38 provinsi Indonesia menurut tingkat ketimpangan pembangunan tahun 2024, memakai K-Means dan K-Medoids, lalu membandingkan performa keduanya. Data berasal dari Badan Pusat Statistik (BPS) tahun 2024 dan mencakup lima indikator: PDRB per kapita, Indeks Pembangunan Manusia (IPM), tingkat pengangguran, tingkat kemiskinan, dan rata-rata lama sekolah. Penelitian menstandarisasi data dengan *Z-score*, menentukan jumlah cluster optimal, menjalankan proses clustering, dan mengevaluasi hasil dengan *Silhouette Coefficient* dan *Davies-Bouldin Index* (DBI). Kedua metode menghasilkan empat cluster. K-Means menempatkan 5 provinsi pada kategori tinggi, 9 pada menengah tinggi, 22 pada menengah, dan 2 pada rendah. K-Medoids menempatkan 12 provinsi pada kategori tinggi, 17 pada menengah tinggi, 7 pada menengah, dan 2 pada rendah. K-Means mengungguli K-Medoids pada kedua metrik evaluasi, sehingga metode ini lebih cocok untuk mengelompokkan ketimpangan pembangunan antarprovinsi di Indonesia.

**Kata kunci:** *Clustering*, Indikator Sosial Ekonomi, *K-Means*, *K-Medoids*

### 1 PENDAHULUAN

Pembangunan antarprovinsi di Indonesia masih timpang. Sejumlah provinsi mencatat pertumbuhan pesat. Provinsi lain masih menghadapi kemiskinan tinggi, kualitas pendidikan rendah, dan kesempatan kerja yang terbatas. Dewanto (2021) menemukan ketimpangan paling dalam antara wilayah barat dan timur, akibat konsentrasi aktivitas ekonomi di barat, kualitas SDM yang tidak merata, infrastruktur yang belum menjangkau timur, serta akses pendidikan dan layanan publik yang terbatas. Todaro dan Smith (2021) menegaskan bahwa pembangunan bukan sekadar pertumbuhan ekonomi, melainkan juga pendidikan, kesehatan, dan kesejahteraan sosial.

BPS mengukur kondisi pembangunan daerah lewat lima indikator: PDRB per kapita, IPM, tingkat pengangguran terbuka, tingkat kemiskinan, dan rata-rata lama sekolah. Data BPS 2024 memperlihatkan ketimpangan tajam antarprovinsi. DKI Jakarta dan Kalimantan Timur mencatat IPM dan PDRB per kapita tertinggi. Provinsi di kawasan timur tertinggal pada kemiskinan dan capaian pendidikan.

Tan, et al. (2019) mendefinisikan clustering sebagai teknik pengelompokan data yang menempatkan objek dengan karakteristik serupa ke dalam satu kelompok dan memisahkannya dari kelompok lain. K-Means banyak dipakai karena sederhana dan efisien pada data besar, namun nilai ekstrem dapat menggeser posisi centroid-nya. K-Medoids memakai objek data nyata (*medoid*) sebagai pusat cluster, sehingga lebih tahan terhadap *outlier*.

. Saputri dan Arianto (2023) membuktikan ketahanan K-Medoids lebih baik dari K-Means, sementara Mohammadi et al. (2024) menunjukkan metode ini menghasilkan cluster lebih stabil pada data dengan struktur kompleks. Menurut Pratiwi et al. (2024) menguji kedua metode pada pengelompokan kemiskinan kabupaten/kota di Indonesia dan menemukan keduanya menghasilkan cluster yang baik.

Penelitian ini mengelompokkan provinsi di Indonesia tahun 2024 berdasarkan indikator sosial ekonomi menggunakan K-Means dan K-Medoids, serta membandingkan performa keduanya dalam memetakan ketimpangan pembangunan antarprovinsi. Hasilnya diharapkan menjadi dasar perumusan kebijakan pembangunan yang lebih tepat sasaran.

## 2 METODE

### 2.1 Cakupan Penelitian

Data yang digunakan adalah data sekunder indikator sosial ekonomi 38 provinsi di Indonesia tahun 2024, bersumber dari publikasi resmi BPS. Data ini menjadi dasar *clustering* ketimpangan pembangunan antarprovinsi menggunakan K-Means dan K-Medoids. Lima variabel bebas ( $X$ ) digunakan untuk menggambarkan kondisi sosial ekonomi tiap provinsi: PDRB per kapita atas dasar harga berlaku, IPM, tingkat pengangguran terbuka, tingkat kemiskinan, dan rata-rata lama sekolah penduduk usia 15 tahun ke atas.

Sebelum *clustering*, data diproses melalui *preprocessing* dan standardisasi *Z-score* agar seluruh variabel berskala sama. Jumlah cluster optimal ditentukan terlebih dahulu sebelum K-Means dan K-Medoids dijalankan. Definisi dan skala pengukuran tiap variabel disajikan pada Tabel 1.

**Tabel 1.** Variabel Indikator Sosial Ekonomi yang Digunakan dalam Penelitian

| Variabel | Nama Variabel                            | Skala Pengukuran | Keterangan                                                       |
|----------|------------------------------------------|------------------|------------------------------------------------------------------|
| $X_1$    | PDRB per Kapita Atas Dasar Harga Berlaku | Rasio            | Nilai PDRB per kapita provinsi tahun 2024                        |
| $X_2$    | Indeks Pembangunan Manusia (IPM)         | Rasio            | Nilai IPM provinsi tahun 2024                                    |
| $X_3$    | Tingkat Pengangguran Terbuka             | Rasio            | Persentase pengangguran terbuka provinsi tahun 2024              |
| $X_4$    | Tingkat Kemiskinan                       | Rasio            | Persentase penduduk miskin provinsi tahun 2024                   |
| $X_5$    | Rata-Rata Lama Sekolah                   | Rasio            | Rata-rata lama sekolah penduduk umur 15 tahun ke atas tahun 2024 |

### 2.2 Metode Analisis

#### 2.2.1 Standardisasi Data Menggunakan Z-Score

Setiap variabel memiliki satuan yang berbeda, sehingga data distandarisasi menggunakan *Z-score* sebelum *clustering*. Standardisasi menyamakan skala antarvariabel agar tidak ada yang mendominasi proses pengelompokan (Han et al., 2022):

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

Keterangan:

- $Z$  = nilai hasil standardisasi
- $X$  = nilai data asli
- $\mu$  = rata-rata data
- $\sigma$  = standar deviasi

### 2.2.2 Penentuan Jumlah Cluster

Jumlah cluster optimal ditentukan dengan mengombinasikan dua pendekatan: *Elbow Method* dan *Silhouette Coefficient*. Kombinasi keduanya membantu memilih jumlah kelompok yang menghasilkan karakteristik wilayah yang mudah diinterpretasikan.

Menurut James et al., (2021) *Elbow Method* mengevaluasi grafik penurunan nilai *Within Cluster Sum of Squares* (WCSS) seiring bertambahnya jumlah kelompok. WCSS mengukur total variansi internal cluster dengan rumus:

$$WCSS = \sum_{i=1}^k \sum_{x_j \in C_i} (x_j - c_i)^2 \quad (2)$$

Keterangan:

$WCSS$  = total variasi dalam cluster

$x_j$  = data ke-  $j$

$c_i$  = *centroid cluster* ke-  $i$

$k$  = jumlah cluster

### 2.2.3 Analisis Data dengan Metode Clustering K-Means

Metode K-Means dipilih untuk memetakan provinsi-provinsi berdasarkan kemiripan profil sosial ekonominya. Secara teknis, algoritma ini mempartisi data secara berulang dengan berpatokan pada jarak terpendek objek menuju pusat massa klaster (*centroid*).

Langkah-langkah metode K-Means adalah sebagai berikut (Han et al., 2022):

1. Menentukan jumlah ( $k$ )
2. Menentukan centroid awal secara acak.
3. Menghitung jarak setiap objek terhadap centroid menggunakan Euclidean Distance. Rumus *Euclidean Distance* adalah sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - y_{jk})^2} \quad (3)$$

Keterangan:

$d_{ij}$  = jarak objek ke-  $i$  dengan centroid ke- $j$

$x_{ik}$  = nilai objek ke-  $i$  pada variabel ke- $k$

$y_{jk}$  = centroid ke  $j$  pada variabel ke- $k$

$p$  = jumlah variabel

4. Mengelompokkan data ke centroid terdekat.
5. Menghitung centroid baru menggunakan rata-rata data pada setiap cluster. Rumus *centroid* adalah sebagai berikut:

$$C_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i \quad (4)$$

Keterangan:

$C_j$  = *centroid cluster* ke-  $j$

$n_j$  = jumlah data pada cluster

$x_i$  = data pada cluster

6. Mengulangi proses hingga cluster stabil.

### 2.2.4 Analisis Data dengan Metode Clustering K-Medoids

Sebagai pembanding, penelitian ini juga menerapkan algoritma K-Medoids atau (*Partitioning Around Medoids* atau PAM). Berbeda dengan K-Means yang berbasis nilai rata-rata fiktif, K-Medoids menggunakan objek riil dari dataset sebagai pusat kelompok (*medoid*).

Langkah-langkah metode K-Medoids adalah sebagai berikut (Han et al., 2022):

1. Menentukan jumlah *cluster* ( $k$ )
2. Menentukan medoid awal secara acak.
3. Menghitung jarak antar objek menggunakan *Euclidean Distance*.

Rumus *Euclidean Distance* adalah sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - y_{jk})^2} \quad (5)$$

4. Mengelompokkan data ke *medoid* terdekat.
5. Menghitung total *cost* cluster.

Menurut Geron (2022) Rumus total *cost* adalah sebagai berikut:

$$TC = \sum_{i=1}^n d_i \quad (6)$$

Keterangan:

$TC$  = total *cost*

$d_i$  = jarak data ke *medoid*

$n$  = jumlah data

6. Melakukan pertukaran medoid apabila menghasilkan total *cost* lebih kecil.
7. Mengulangi proses hingga medoid stabil.

### 2.2.5 Evaluasi Cluster

K-Means dan K-Medoids dievaluasi menggunakan dua ukuran: *Silhouette Coefficient* dan *Davies-Bouldin Index* (DBI) (James et al., 2021).

Rumus *Silhouette Coefficient* adalah sebagai berikut:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (7)$$

Keterangan:

$s(i)$  = nilai *silhouette*

$a(i)$  = rata-rata jarak dalam cluster

$b(i)$  = rata-rata jarak dengan cluster terdekat

Rumus *Davies-Bouldin Index* adalah sebagai berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left( \frac{S_i + S_j}{M_{ij}} \right) \quad (8)$$

Keterangan:

$DBI$  = *Davies-Bouldin Index*

$S_i$  = rata-rata jarak cluster ke-iii

$M_{ij}$  = jarak antarcluster

$k$  = jumlah cluster

Metode terbaik ditentukan berdasarkan nilai *Silhouette Coefficient* tertinggi dan nilai *Davies-Bouldin Index* terendah.

### 3 HASIL DAN PEMBAHASAN

#### 3.1 Pengumpulan Data

Tabel 2 menyajikan data indikator sosial ekonomi 38 provinsi di Indonesia tahun 2024, meliputi PDRB per kapita atas dasar harga berlaku, IPM, tingkat pengangguran, tingkat kemiskinan, dan rata-rata lama sekolah penduduk usia 15 tahun ke atas.

**Tabel 2.** Ketimpangan Pembangunan Provinsi Di Indonesia Tahun 2024 Berdasarkan Indikator Sosial Ekonomi

| Provinsi                  | Produk Domestik Regional Bruto per Kapita Atas Dasar Harga Berlaku | IPM (%) | Pengangguran (%) | Kemiskinan (%) | Rata-Rata Lama Sekolah Penduduk Umur 15 Tahun ke Atas Menurut Provinsi |
|---------------------------|--------------------------------------------------------------------|---------|------------------|----------------|------------------------------------------------------------------------|
| Aceh                      | 45.770.000                                                         | 74,03   | 5,75             | 12,64          | 9.95                                                                   |
| Sumatera Utara            | 78.310.000                                                         | 74,02   | 5,6              | 7,19           | 10.18                                                                  |
| Sumatera Barat            | 59.549.000                                                         | 74,49   | 5,75             | 5,42           | 9.72                                                                   |
| Riau                      | 1763.85.000                                                        | 74,79   | 3,7              | 6,36           | 9.69                                                                   |
| Jambi                     | 92.786.000                                                         | 73,43   | 4,48             | 7,26           | 9.26                                                                   |
| Sumatera Selatan          | 80.663.000                                                         | 72,3    | 3,86             | 10,51          | 8.98                                                                   |
| Bengkulu                  | 52.305.000                                                         | 73,39   | 3,11             | 12,52          | 9.4                                                                    |
| Lampung                   | 55.009.000                                                         | 71,81   | 4,19             | 10,62          | 8.8                                                                    |
| Kepulauan Bangka Belitung | 75.323.000                                                         | 73,33   | 4,63             | 5,08           | 8.78                                                                   |
| Kepulauan Riau            | 172.459.000                                                        | 77,97   | 6,39             | 4,78           | 10.65                                                                  |
| DKI Jakarta               | 367.687.000                                                        | 83,08   | 6,21             | 4,14           | 11.49                                                                  |
| Jawa Barat                | 59.865.000                                                         | 74,43   | 6,75             | 7,08           | 9.24                                                                   |
| Jawa Tengah               | 50.824.000                                                         | 73,88   | 4,78             | 9,58           | 8.47                                                                   |
| DI Yogyakarta             | 55.039.000                                                         | 81,55   | 3,48             | 10,4           | 10.23                                                                  |
| Jawa Timur                | 80.856.000                                                         | 74,09   | 4,19             | 9,56           | 8.69                                                                   |
| Banten                    | 74.673.000                                                         | 74,48   | 6,68             | 5,7            | 9.55                                                                   |
| Bali                      | 72.658.000                                                         | 77,76   | 1,79             | 3,8            | 9.87                                                                   |
| Nusa Tenggara Barat       | 33.667.000                                                         | 70,93   | 2,73             | 11,91          | 8.5                                                                    |
| Nusa Tenggara Timur       | 25.837.000                                                         | 67,39   | 3,02             | 19,02          | 8.48                                                                   |
| Kalimantan Barat          | 56.984.000                                                         | 70,13   | 4,86             | 6,25           | 8.28                                                                   |
| Kalimantan Tengah         | 84.730.000                                                         | 72,73   | 4,01             | 5,26           | 9.17                                                                   |

| Provinsi           | Produk Domestik Regional Bruto per Kapita Atas Dasar Harga Berlaku | IPM (%) | Pengangguran (%) | Kemiskinan (%) | Rata-Rata Lama Sekolah Penduduk Umur 15 Tahun ke Atas Menurut Provinsi |
|--------------------|--------------------------------------------------------------------|---------|------------------|----------------|------------------------------------------------------------------------|
| Kalimantan Selatan | 70.577.000                                                         | 73,03   | 4,2              | 4,02           | 9.02                                                                   |
| Kalimantan Timur   | 208.331.000                                                        | 78,83   | 5,14             | 5,51           | 10.22                                                                  |
| Kalimantan Utara   | 208.210.000                                                        | 73,02   | 3,9              | 5,38           | 9.58                                                                   |
| Sulawesi Utara     | 752.35.000                                                         | 75,03   | 5,85             | 6,7            | 10.02                                                                  |
| Sulawesi Tengah    | 131.643.000                                                        | 71,56   | 2,94             | 11,04          | 9.28                                                                   |
| Sulawesi Selatan   | 78.748.000                                                         | 74,05   | 4,19             | 7,77           | 9.22                                                                   |
| Sulawesi Tenggara  | 72.031.000                                                         | 73,48   | 3,09             | 10,63          | 9.74                                                                   |
| Gorontalo          | 47.477.000                                                         | 71,23   | 3,13             | 13,87          | 8.64                                                                   |
| Sulawesi Barat     | 46.654.000                                                         | 68,2    | 2,68             | 10,71          | 8.56                                                                   |
| Maluku             | 33.652.000                                                         | 71,57   | 6,11             | 15,78          | 10.44                                                                  |
| Maluku Utara       | 97.262.000                                                         | 71,03   | 4,03             | 6,03           | 9.7                                                                    |
| Papua Barat        | 139.213.000                                                        | 67,02   | 4,13             | 21,09          | 9.87                                                                   |
| Papua Barat Daya   | 61.624.000                                                         | 68,63   | 6,48             | 16,95          | 10.48                                                                  |
| Papua              | 84.305.000                                                         | 73      | 6,48             | 18,09          | 10.7                                                                   |
| Papua Selatan      | 64.928.000                                                         | 67,9    | 4,05             | 19,35          | 8.99                                                                   |
| Papua Tengah       | 107.566.000                                                        | 59,75   | 2,75             | 27,6           | 6.17                                                                   |
| Papua Pegunungan   | 19.111.000                                                         | 53,42   | 1,32             | 29,66          | 5.1                                                                    |

Sumber: Data BPS Tahun 2024

### 3.2 Standardisasi Data Menggunakan Z-Score

Standardisasi *Z-score* menghasilkan nilai rata-rata mendekati 0 dan standar deviasi mendekati 1 untuk seluruh variabel. Dengan skala yang seragam, tidak ada variabel yang mendominasi proses clustering (Han et al., 2021). Perbandingan statistik deskriptif sebelum dan sesudah standardisasi disajikan pada Tabel 3.

**Tabel 3.** Perbandingan Statistik Deskriptif Sebelum dan Sesudah Standardisasi Data

| Variabel                                                           | Sebelum    |             |              |              | Sesudah |    |
|--------------------------------------------------------------------|------------|-------------|--------------|--------------|---------|----|
|                                                                    | Minimum    | Maksimum    | Mean         | SD           | Mean    | SD |
| Produk Domestik Regional Bruto per Kapita Atas Dasar Harga Berlaku | 19.111.000 | 367.687.000 | 89.419.631,6 | 651.234.02,8 | 0       | 1  |
| IPM (%)                                                            | 53.42      | 83.08       | 72.3884211   | 5.15022672   | 0       | 1  |
| Pengangguran(%)                                                    | 1.32       | 6.75        | 4.37973684   | 1.41422214   | 0       | 1  |
| Kemiskinan (%)                                                     | 3.8        | 29.66       | 10.6647368   | 6.36675433   | 0       | 1  |

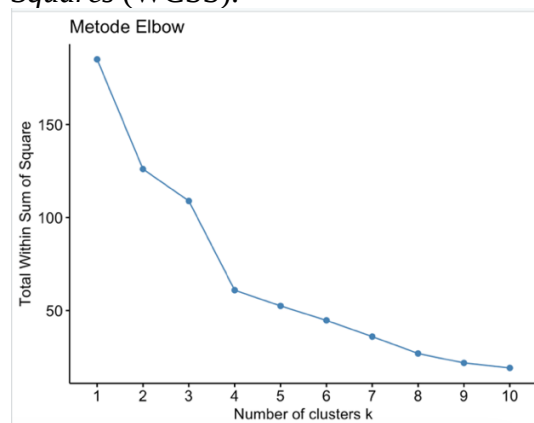
|                                                                              |     |       |            |            |   |   |
|------------------------------------------------------------------------------|-----|-------|------------|------------|---|---|
| Rata-Rata Lama Sekolah<br>Penduduk Umur 15 Tahun ke<br>Atas Menurut Provinsi | 5.1 | 11.49 | 9.29236842 | 1.14888767 | 0 | 1 |
|------------------------------------------------------------------------------|-----|-------|------------|------------|---|---|

Sumber: Hasil Pengolahan Data Eksperimen (2026)

Kelima variabel memiliki skala yang jauh berbeda sebelum standardisasi. PDRB per kapita bernilai puluhan hingga ratusan juta rupiah, sedangkan IPM, tingkat pengangguran, tingkat kemiskinan, dan rata-rata lama sekolah bernilai jauh lebih kecil. Tanpa penyeragaman skala, PDRB per kapita akan mendominasi proses *clustering* dan menghasilkan pengelompokan yang bias. Skala yang seragam membuat setiap variabel berkontribusi seimbang dalam pembentukan cluster, sehingga hasil pengelompokan mencerminkan kemiripan karakteristik antarprovinsi secara objektif.

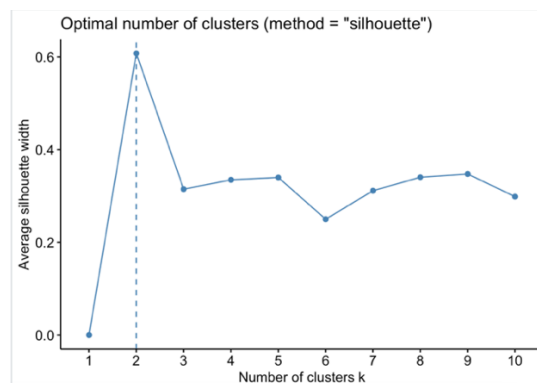
### 3.3 Penentuan Jumlah Cluster Optimal

Jumlah cluster optimal ditentukan sebelum K-Means dan K-Medoids dijalankan. Menurut Thorndike (2019), menentukan jumlah cluster dengan mencari titik siku pada grafik penurunan nilai *Within Cluster Sum of Squares* (WCSS).



**Gambar 1.** Jumlah Cluster Menggunakan *Elbow Method*

Gambar 1 menunjukkan bahwa nilai WSS mengalami penurunan tajam dari cluster 1 hingga 4, lalu melambat dan membentuk pola landai. Penambahan cluster setelah  $k = 4$  tidak meningkatkan homogenitas secara berarti. *Elbow Method* merekomendasikan 4 cluster. Namun demikian, jumlah cluster yang digunakan dalam penelitian tidak hanya mempertimbangkan hasil *Elbow Method*, tetapi juga mempertimbangkan hasil metode *Silhouette Method* untuk mengevaluasi kualitas cluster yang terbentuk.



**Gambar 2.** Jumlah Cluster Menggunakan *Silhouette Method*

Gambar 2 menunjukkan nilai rata-rata silhouette tertinggi pada  $k = 2$ . Setelah titik itu, nilai turun dan berfluktuasi. Pemisahan antarcluster melemah seiring bertambahnya  $k$ . Pada  $k = 2$ , cohesi dan separation cluster sama-sama kuat. Kedua metode menghasilkan rekomendasi berbeda karena setiap indeks validitas mengukur kualitas cluster dari sudut pandang berbeda (Akhanli & Hennig, 2020). Nilai indeks saja tidak cukup untuk memilih jumlah cluster; karakteristik data dan kemudahan interpretasi ikut menentukan (Hassan et al., 2024).

Penelitian ini menetapkan empat cluster. Empat kelompok menggambarkan keragaman kondisi sosial ekonomi antarprovinsi dengan lebih spesifik dan memungkinkan analisis perbedaan antarprovinsi yang lebih tajam dibanding dua cluster.

### 3.4 Metode Clustering K-Means

K-Means mengelompokkan data berdasarkan jarak terdekat setiap objek ke *centroid* clusternya (MacQueen, 2020). Penelitian ini membagi 34 provinsi ke dalam 4 cluster, memakai lima variabel: PDRB per kapita, IPM, tingkat pengangguran, tingkat kemiskinan, dan rata-rata lama sekolah. Tabel 4 menyajikan hasil *clustering* K-Means.

**Tabel 4.** Pembagian Provinsi Berdasarkan Tingkat Pembangunan Hasil *Clustering* K-Means

| Kelompok | Nama Cluster | Karakteristik Tingkat Pembangunan | Jumlah Provinsi | Provinsi                                                                                                                                                                                                                                                                                                                                          |
|----------|--------------|-----------------------------------|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| I        | Cluster 1    | Pembangunan Tinggi                | 5               | Riau, Kepulauan Riau, DKI Jakarta, Kalimantan Timur, Kalimantan Utara                                                                                                                                                                                                                                                                             |
| II       | Cluster 2    | Pembangunan Menengah Tinggi       | 9               | Aceh, Sumatera Utara, Sumatera Barat, Jawa Barat, Banten, Sulawesi Utara, Maluku, Papua Barat Daya, Papua                                                                                                                                                                                                                                         |
| III      | Cluster 3    | Pembangunan Menengah              | 22              | Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara, Papua Barat, Papua Selatan |
| IV       | Cluster 4    | Pembangunan Rendah                | 2               | Papua Tengah, Papua Pegunungan                                                                                                                                                                                                                                                                                                                    |

Sumber: Hasil Pengolahan Data Eksperimen (2026)

Hasil *centroid cluster* metode K-Means disajikan pada tabel 5 berikut.

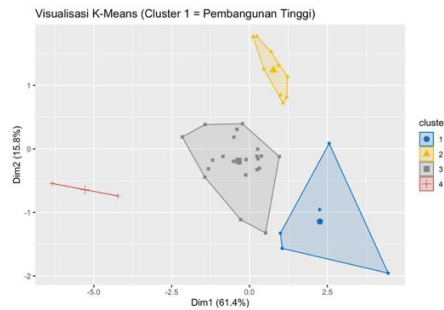
**Tabel 5.** Hasil *Centroid Cluster* Metode K-Means

| Kelompok | Nama Cluster | PDRB per Kapita | IPM   | Pengangguran (%) | Kemiskinan (%) | Rata-rata Lama Sekolah (Tahun) |
|----------|--------------|-----------------|-------|------------------|----------------|--------------------------------|
| I        | Cluster 1    | 226.614.400     | 77,54 | 5,07             | 5,23           | 10,33                          |
| II       | Cluster 2    | 63.664.778      | 73,30 | 6,16             | 10,62          | 10,03                          |
| III      | Cluster 3    | 71.146.091      | 72,28 | 3,71             | 10,29          | 9,09                           |
| IV       | Cluster 4    | 63.338.500      | 56,59 | 2,04             | 28,63          | 5,64                           |

Tabel 5 menunjukkan empat kelompok provinsi dengan karakteristik pembangunan yang berbeda. Cluster 1 mencakup provinsi dengan PDRB per kapita tertinggi, IPM tertinggi kedua, dan

angka kemiskinan terendah. Cluster 2 unggul di IPM, tapi menanggung pengangguran tertinggi di antara keempat kelompok. Cluster 3 punya pengangguran rendah meski PDRB per kapitanya berada di urutan ketiga. Cluster 4 tertinggal paling jauh: rata-rata lama sekolah terendah dan angka kemiskinan tertinggi.

Perbedaan antarcluster yang tajam mengonfirmasi bahwa K-Means berhasil memisahkan provinsi berdasarkan kemiripan karakteristik sosial ekonomi (Hair et al., 2022). Gambar 3 menyajikan sebaran spasial keempat cluster.



**Gambar 3.** Visualisasi Cluster K-Means

Gambar 3 menunjukkan 38 provinsi terbagi ke dalam empat cluster tanpa tumpang tindih. Cluster 1 (pembangunan tinggi) dan Cluster 4 (pembangunan rendah) berjauhan, mencerminkan ketimpangan sosial ekonomi antara dua wilayah ekstrem. Cluster 2 (pembangunan menengah tinggi) dan Cluster 3 (pembangunan menengah) berdekatan di area tengah karena karakteristiknya mirip. Pemisahan yang tegas antarcluster menunjukkan bahwa K-Means mengenali homogenitas internal setiap kelompok sekaligus membedakan kelompok yang berbeda.

### 3.5 Metode Clustering K-Medoids

K-Medoids adalah metode clustering non-hierarki yang mirip K-Means, tetapi pusat clusternya menggunakan objek data asli (medoid). Kaufman dan Rousseeuw (2021) menyebut medoid sebagai sumber daya tahan metode ini terhadap *outlier*. Penelitian ini menerapkan K-Medoids dengan 4 cluster dan lima variabel yang sama dengan K-Means: PDRB per kapita, IPM, tingkat pengangguran, tingkat kemiskinan, dan rata-rata lama sekolah. Tabel 6 menyajikan pembagian provinsi dan karakteristik tiap cluster.

**Tabel 6.** Pembagian Provinsi Berdasarkan Tingkat Pembangunan Hasil Clustering K-Medoids

| Kelompok | Nama Cluster | Karakteristik Tingkat Pembangunan | Jumlah Provinsi | Provinsi                                                                                                                                                                                             |
|----------|--------------|-----------------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| I        | Cluster 1    | Pembangunan Tinggi                | 12              | Aceh, Sumatera Utara, Sumatera Barat, Kepulauan Riau, DKI Jakarta, Jawa Barat, Banten, Kalimantan Timur, Sulawesi Utara, Maluku, Papua Barat Daya, Papua                                             |
| II       | Cluster 2    | Pembangunan Menengah Tinggi       | 17              | Riau, Jambi, Sumatera Selatan, Lampung, Kepulauan Bangka Belitung, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi |

|     |           |                      |   |                                                                                                           |
|-----|-----------|----------------------|---|-----------------------------------------------------------------------------------------------------------|
|     |           |                      |   | Tengah, Sulawesi Selatan, Sulawesi Tenggara, Maluku Utara                                                 |
| III | Cluster 3 | Pembangunan Menengah | 7 | Bengkulu, Nusa Tenggara Barat, Nusa Tenggara Timur, Gorontalo, Sulawesi Barat, Papua Barat, Papua Selatan |
| IV  | Cluster 4 | Pembangunan Rendah   | 2 | Papua Tengah, Papua Pegunungan                                                                            |

Sumber: Hasil Pengolahan Data Eksperimen (2026)

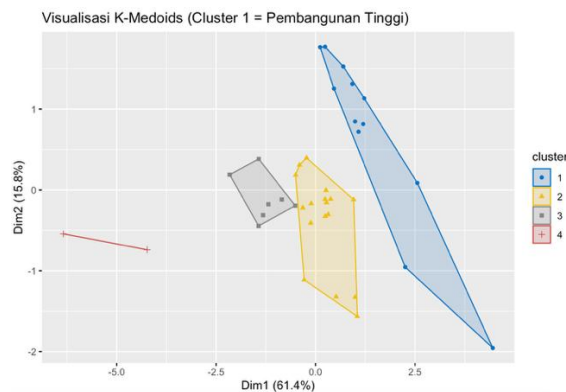
Hasil centroid cluster metode K-Medoids ditunjukkan pada tabel 7 berikut.

**Tabel 7.** Hasil *Centroid Cluster* Metode K-Medoids

| Kelompok | Nama Cluster | PDRB per Kapita | IPM   | Pengangguran (%) | Kemiskinan (%) | Rata-rata Lama Sekolah (Tahun) |
|----------|--------------|-----------------|-------|------------------|----------------|--------------------------------|
| I        | Cluster 1    | 110.121.667     | 74,96 | 6,10             | 9,17           | 10,22                          |
| II       | Cluster 2    | 90.572.235      | 73,65 | 3,90             | 7,62           | 9,22                           |
| III      | Cluster 3    | 58.583.000      | 69,44 | 3,26             | 15,50          | 8,92                           |
| IV       | Cluster 4    | 63.338.500      | 56,59 | 2,04             | 28,63          | 5,64                           |

Tabel 7 menunjukkan empat kelompok provinsi dengan ketimpangan yang masih jelas antarwilayah. Cluster 1 (pembangunan tinggi) mencatat nilai tertinggi pada seluruh indikator sosial ekonomi kecuali pengangguran. Cluster 2 (menengah tinggi) memiliki kondisi ekonomi baik dan kemiskinan rendah; kelompok ini juga beranggotakan provinsi paling banyak, mencerminkan kondisi mayoritas wilayah di Indonesia. Cluster 3 (menengah) berada pada kondisi ekonomi sedang dengan tantangan pada indikator kesejahteraan dan kualitas SDM. Cluster 4 (pembangunan rendah) tertinggal pada seluruh indikator dan membutuhkan intervensi kebijakan yang lebih terarah untuk meningkatkan kualitas SDM dan kesejahteraan masyarakat.

Sebagian besar provinsi masuk ke Cluster 2 dan Cluster 1, menandakan pembangunan di banyak wilayah sudah berjalan. Keberadaan Cluster 4 menunjukkan ketimpangan antarprovinsi masih nyata dan perlu masuk pertimbangan dalam perumusan kebijakan pembangunan nasional. Gambar 4 menyajikan distribusi keempat cluster.



**Gambar 4.** Visualisasi Cluster K-Medoids

Berdasarkan Gambar 4 menunjukkan 38 provinsi terbagi ke dalam empat cluster. Cluster 1 dan Cluster 4 berada jauh dari cluster lain, mencerminkan perbedaan karakteristik yang tajam di antara keduanya. Cluster 2 dan Cluster 3 berdekatan namun tetap membentuk kelompok yang berbeda. Pemisahan yang jelas antarcluster mengonfirmasi bahwa K-Medoids berhasil mengelompokkan provinsi berdasarkan kemiripan karakteristik sosial ekonomi dan memperkuat adanya ketimpangan pembangunan antarprovinsi.

### 3.6 Evaluasi Hasil Clustering

K-Means dan K-Medoids dievaluasi menggunakan dua ukuran yaitu *Silhouette Coefficient* dan *Davies-Bouldin Index* (DBI). Rousseeuw (2022) menjelaskan bahwa *Silhouette Coefficient* mengukur kedekatan objek dalam satu cluster dibanding cluster lain, dengan rentang nilai -1 hingga 1. Nilai mendekati 1 menandakan cluster berkualitas baik. Davies dan Bouldin (2021) menjelaskan bahwa DBI mengukur pemisahan antarcluster. Nilai DBI yang lebih kecil menandakan perbedaan karakteristik antarcluster yang lebih jelas.

Hasil evaluasi disajikan pada Tabel 8.

**Tabel 8.** Hasil Evaluasi Kualitas Cluster Menggunakan *Silhouette Coefficient* dan *Davies-Bouldin Index*

| Metode    | <i>Silhouette Coefficient</i> | <i>Davies-Bouldin Index</i> |
|-----------|-------------------------------|-----------------------------|
| K-Means   | 0,3348601                     | 0,9422484                   |
| K-Medoids | 0,2436963                     | 1,2749150                   |

Sumber: Hasil Pengolahan Data Eksperimen (2026)

Davies dan Bouldin (2021) menjelaskan bahwa nilai DBI yang lebih kecil menandakan pemisahan antarcluster yang lebih jelas. K-Means mencatat DBI lebih rendah daripada K-Medoids, yang berarti pemisahan clusternya lebih baik. Han et al. (2021) menegaskan bahwa metode dengan *silhouette* lebih tinggi dan DBI lebih rendah menghasilkan pengelompokan yang lebih optimal. Berdasarkan kedua ukuran itu, K-Means tampil sebagai metode terbaik dalam penelitian ini: clusternya lebih kompak dan pemisahan antarkelompoknya lebih tajam.

### 3.7 Interpretasi hasil cluster

Berdasarkan *Silhouette Coefficient* dan *Davies-Bouldin Index*, penelitian ini memilih K-Means sebagai metode terbaik dan mendasarkan interpretasi cluster pada hasil pengelompokannya. K-Means mengelompokkan 38 provinsi ke dalam empat kategori: pembangunan tinggi, menengah tinggi, menengah, dan rendah. Perbedaan antarkelompok tercermin dari lima indikator: PDRB per kapita, IPM, tingkat pengangguran, tingkat kemiskinan, dan rata-rata lama sekolah.

Pembangunan antarprovinsi di Indonesia belum merata. Ketimpangan antarkategori mengindikasikan bahwa kebijakan perlu mengarah pada peningkatan kualitas SDM, pemerataan akses pendidikan, pengurangan kemiskinan, dan penguatan ekonomi daerah. Pemerintah dapat memakai pengelompokan ini sebagai dasar untuk merumuskan kebijakan yang sesuai dengan karakteristik dan kebutuhan tiap wilayah.

## 4 KESIMPULAN

Berdasarkan hasil penelitian mengenai perbandingan metode K-Means dan K-Medoids dalam clustering ketimpangan pembangunan provinsi di Indonesia tahun 2024 berdasarkan indikator sosial ekonomi, diperoleh beberapa kesimpulan sebagai berikut.

1. Kedua metode menghasilkan empat tingkat pembangunan: tinggi, menengah tinggi, menengah, dan rendah. Kelompok pembangunan tinggi mencatat PDRB per kapita, IPM, dan rata-rata lama sekolah tertinggi dengan kemiskinan terendah; kelompok pembangunan rendah mencatat pola sebaliknya.
2. K-Means membagi 38 provinsi menjadi: pembangunan tinggi (5 provinsi), menengah tinggi (9 provinsi), menengah (22 provinsi), dan rendah (2 provinsi). Sebagian besar provinsi berada di kelompok menengah.
3. K-Medoids menghasilkan pembagian berbeda: pembangunan tinggi (12 provinsi), menengah tinggi (17 provinsi), menengah (7 provinsi), dan rendah (2 provinsi).
4. K-Means menghasilkan Silhouette Coefficient 0,3348601 dan DBI 0,9422484, sementara K-Medoids menghasilkan Silhouette Coefficient 0,2436963 dan DBI 1,2749150. K-Means unggul pada kedua metrik.
5. K-Means lebih sesuai untuk mengelompokkan ketimpangan pembangunan provinsi di Indonesia berdasarkan indikator sosial ekonomi.

### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Terbuka atas fasilitas yang mendukung penyelesaian jurnal ini, serta kepada Ibu Fonda Leviany, S.Stat., M.Stat yang telah meluangkan waktu memberikan koreksi dan saran penyusunan artikel ini. Kepada Badan Pusat Statistik, penulis berterima kasih atas ketersediaan data yang menjadi fondasi penelitian ini. Terima kasih kepada keluarga dan kerabat atas doa dan dukungan. Sebagai bentuk transparansi akademik, penulis menyatakan menggunakan teknologi *Artificial Intelligence* (AI) secara terbatas hanya untuk membantu penyuntingan tata bahasa dan penyempurnaan struktur kalimat.

### DAFTAR PUSTAKA

- Akhanli, S. E., & Hennig, C. (2020). *Comparing clusterings and numbers of clusters by aggregation of calibrated clustering validity indexes*.
- Akhanli, S., & Hennig, C. (2020). *Comparing clusterings and numbers of clusters by aggregation of calibrated clustering validity indexes*. *Statistics and Computing*, 30(5), 1523–1544
- Badan Pusat Statistik. (2024). *Indikator Sosial Ekonomi Indonesia Tahun 2024*. Jakarta: Badan Pusat Statistik.
- Davies, D. L., & Bouldin, D. W. (2021). *A Cluster Separation Measure*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Dewanto, F. W. (2021). Analisis Ketimpangan Pembangunan Ekonomi Antarwilayah (Studi di Kawasan Barat Indonesia dan Kawasan Timur Indonesia) Tahun 2014–2018. *JIMEK*, 4(1), 1–14.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2022). *Multivariate Data Analysis* (8th ed.). Pearson Education.
- Han, J., Kamber, M., & Pei, J. (2021). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Hassan, B. A., Tayfor, N. B., Hassan, A. A., Ahmed, A. M., Rashid, T. A., & Abdalla, N. N. (2024). *From A-to-Z Review of Clustering Validation Indices*
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: with applications in R* (2nd ed.). Springer.
- Kaufman, L., & Rousseeuw, P. J. (2021). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.
- Lantz, B. (2023). *Machine Learning with R* (4th ed.). Packt Publishing.

- MacQueen, J. (2020). *Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.*
- Mohammadi, M., et al., (2024). *A Novel K-Means and K-Medoids Algorithms for Clustering Non-Spherical-Shape Clusters Non-Sensitive to Outliers. Pattern Recognition*, 156, 110834.
- Pratiwi, G. R., Wahiddin, D., Awal, E. E., & Fauzi, A. (2024). Klasterisasi Tingkat Kemiskinan Kabupaten/Kota di Indonesia Menggunakan Algoritma K-Means dan K-Medoids. *Jurnal Algoritma*, 21(2), 197–208.
- Rousseeuw, P. J. (2022). *Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. Journal of Computational and Applied Mathematics.*
- Saputri, F. W., & Arianto, D. B. (2023). Perbandingan Performa Algoritma K-Means, K-Medoids, dan DBSCAN dalam Penggerombolan Provinsi di Indonesia Berdasarkan Indikator Kesejahteraan Masyarakat. *Jurnal Teknologi Informasi*, 17(2), 85–94.
- Tan, P. N., Steinbach, M., & Kumar, V. (2019). *Introduction to Data Mining* (2nd ed.). *Pearson Education.*
- Todaro, M. P., & Smith, S. C. (2021). *Economic Development* (13th ed.). *Pearson Education.*