

## Analisis Cluster K-Means dan K-Medoids Untuk Pengelompokan Provinsi di Indonesia Berdasarkan Pembangunan Tik Tahun 2024

Dian Natasia Dewanti Sukarman<sup>1,\*</sup>, Radian Januari Situmeang<sup>2</sup>

<sup>1,2</sup>Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Cenderawasih Jayapura, Indonesia

\*Correspondance author: [diannatasia0502@gmail.com](mailto:diannatasia0502@gmail.com)

### ABSTRACT

Inequality in access and digital infrastructure among provinces in Indonesia has become a significant challenge in the digital economy era. This study aims to cluster 38 provinces in Indonesia based on Information and Communication Technology (ICT) development indicators and human resource quality in 2024, as well as to conduct a performance comparison test between the K-Means and K-Medoids algorithms. The variables used include the percentage of cellular phone ownership, computer ownership, internet access, Mean Years of Schooling (MYS), and the Human Development Index (HDI). The determination of the optimal number of clusters was evaluated using the Elbow Method. The Elbow Method results established 4 optimal clusters for K-Means and 5 optimal clusters for K-Medoids. In the K-Means modeling, the regions were distinctly partitioned by isolating Central Papua and Highland Papua Provinces into the Very Disadvantaged cluster due to an ICT infrastructure crisis. The validity of the clustering structure was tested using five matrix instruments, namely the Silhouette Score, Davies-Bouldin Index (DBI), Dunn Index, SW/SB Ratio, and Separation Index. The performance testing results showed that the K-Means algorithm absolutely outperformed K-Medoids across all evaluation metrics. K-Means recorded a higher Silhouette Score 0.3473, a smaller DBI 0.7853, and a very minimal SW/SB ratio 0.1647 compared to K-Medoids, which experienced bias due to the SW/SB ratio value exploding to 1.9135. Thus, the K-Means algorithm is proven to have a more solid and accurate goodness of fit in mapping the portrait of digital inequality in Indonesia, so its clustering results can be used as a strategic reference for the government in formulating well-targeted technology equalization policies.

**Keywords:** *K-Means, K-Medoids, ICT Development, Clustering, Performance Comparison.*

### ABSTRAK

Ketimpangan akses dan infrastruktur digital antarprovinsi di Indonesia menjadi tantangan signifikan dalam era ekonomi digital. Penelitian ini bertujuan untuk mengelompokkan 38 provinsi di Indonesia berdasarkan indikator pembangunan Teknologi Informasi dan Komunikasi (TIK) serta kualitas sumber daya manusia tahun 2024, sekaligus melakukan uji komparasi kinerja antara algoritma *K-Means* dan *K-Medoids*. Variabel yang digunakan meliputi persentase kepemilikan telepon seluler (HP), kepemilikan komputer, akses internet, Rata-rata Lama Sekolah (RLS), dan Indeks Pembangunan Manusia (IPM). Penentuan jumlah kelompok optimal dievaluasi menggunakan *Elbow Method*. Hasil *Elbow Method* menetapkan 4 klaster optimal untuk *K-Means* dan 5 klaster optimal untuk *K-Medoids*. Pada pemodelan *K-Means*, wilayah terpartisi secara tegas dengan mengisolasi Provinsi Papua Tengah dan Papua Pegunungan ke dalam klaster Sangat Tertinggal akibat krisis infrastruktur TIK. Validitas struktur pengelompokan diuji menggunakan lima instrumen matriks, yaitu *Silhouette Score*, *Davies-Bouldin Index* (DBI), *Dunn Index*, Rasio SW/SB, dan *Separation Index*. Hasil pengujian performa menunjukkan bahwa algoritma *K-Means* secara mutlak mengungguli *K-Medoids* pada seluruh metrik evaluasi. *K-Means* mencatatkan *Silhouette Score* lebih tinggi (0,3473), DBI lebih kecil (0,7853), dan rasio SW/SB yang sangat minimalis (0,1647) dibandingkan *K-Medoids* yang mengalami pembiasan akibat nilai rasio SW/SB yang meledak hingga 1,9135. Dengan demikian, algoritma *K-Means* terbukti memiliki kecocokan (*goodness of fit*) yang lebih solid dan akurat dalam memetakan potret ketimpangan digital di Indonesia, sehingga hasil klasternya dapat dijadikan acuan strategis bagi pemerintah dalam merumuskan kebijakan pemerataan teknologi yang tepat sasaran.

**Keywords:** *K-Means, K-Medoids, Pembangunan TIK, Clustering, Komparasi Kinerja.*

---

#### ARTICLE INFO

---

Submission received: 1 June 2025

Accepted: 15 August 2025

Revised: 2 July 2025

Published: 30 August 2025

Available on: <https://doi.org/10.32493/sm.v7i2.xxxxx>

---

**StatMat: Jurnal Statistika dan Matematika** is licenced under a Creative Commons Attribution-ShareAlike 4.0 International License.

---

## PENDAHULUAN

Perkembangan ilmu pengetahuan dan teknologi berkembang dengan sangat cepat. Pada saat era globalisasi turut andil dalam mempercepat perkembangan tersebut. Era globalisasi diawali dengan adanya kemajuan pada bidang komunikasi dan informasi. Masalah yang cukup pokok adalah dampak dari adanya perkembangan ilmu pengetahuan dan teknologi tersebut. Ilmu pengetahuan dan teknologi saat ini dapat menawarkan berbagai macam kemudahan bagi kehidupan manusia, tetapi pada waktu bersamaan terjadi ketidakjelasan kecendrungan kemanusiaan. Kemajuan yang seharusnya perlu dimaknai secara global. Seharusnya manusia yang dapat mengendalikan kemajuan ilmu pengetahuan dan teknologi bukan teknologi yang bisa mengendalikan manusia (Sudarsih et al., 2024)

Kualitas hidup masyarakat pada suatu negara atau suatu wilayah dapat diukur secara komprehensif melalui Indeks Pembangunan Manusia (IPM). Konsep IPM didesain untuk dapat melampaui pendekatan yang hanya memperhitungkan pertumbuhan ekonomi atau pendapat per kapita yang sebagai satu-satunya indikator pembangunan utama. Terdapat tiga aspek kunci dalam Indeks Pembangunan Manusia (IPM), yaitu taraf hidup, kesehatan, dan pendidikan (Fadilla et al., 2024).

Menurut definisi Badan Pusat Statistik (BPS), Rata-rata Lama Sekolah (RLS) merupakan indikator jumlah tahun yang dihabiskan penduduk dalam sistem pendidikan formal untuk menggambarkan kualitas pendidikan di suatu daerah. Dalam perhitungannya, durasi tinggal kelas tidak dimasukkan sebagai komponen penambah tahun sekolah. Patokan yang digunakan adalah akumulasi waktu berdasarkan jenjang kelulusan, yakni 6 tahun bagi tamatan SD, 9 tahun bagi tamatan SMP, dan 12 tahun bagi masyarakat yang menyelesaikan pendidikan SMA (Dewi et al., 2022).

Kesenjangan digital menggambarkan ketimpangan akses dan pemanfaatan teknologi informasi antar kelompok masyarakat. Di Indonesia, fenomena ini terlihat jelas melalui perbedaan infrastruktur digital antara kawasan perkotaan dan perdesaan. Berdasarkan Indeks Masyarakat Digital Indonesia 2023 dari Kemenkominfo, masyarakat desa masih mengalami hambatan besar dalam penggunaan teknologi. Meskipun akses di kota terus berkembang, penduduk di daerah perdesaan masih tertinggal jauh dalam mengadopsi teknologi digital untuk mendukung sektor ekonomi mereka (Koswara et al., 2024).

Dalam upaya merumuskan kebijakan yang efisien untuk mengatasi ketimpangan tersebut, pemerintah memerlukan pemetaan wilayah yang akurat dan objektif. Mengingat banyaknya jumlah provinsi di Indonesia dengan karakteristik data multivariat yang kompleks, pendekatan statistik modern diperlukan untuk mengekstraksi informasi dari data tersebut (Nabila & Mardhotillah, 2025). Pemetaan ini memungkinkan pemerintah untuk beralih dari kebijakan yang bersifat seragam menjadi strategi yang lebih spesifik berdasarkan karakteristik klaster wilayah (Widiasanti et al., 2025).



Pembangunan Teknologi Informasi dan Komunikasi (TIK) merupakan indikator fundamental dalam mengukur daya saing serta tingkat kesejahteraan suatu wilayah. Berbagai kajian terdahulu telah berupaya memetakan tingkat pembangunan TIK di Indonesia menggunakan pendekatan *data mining*. Sebagai contoh, penelitian Suharni (2020) mengelompokkan Indeks Pembangunan TIK pada 34 provinsi di Indonesia menggunakan algoritma *K-Means*. Pendekatan pengelompokan sejenis juga dilakukan oleh Novianto et al., (2023) dengan menerapkan algoritma *K-Medoids* untuk tingkat nasional, serta Hidayati et al., (2024) yang menggunakan *K-Means* untuk melihat ketimpangan infrastruktur TIK pada level daerah di Jawa Timur. Meskipun literatur-literatur tersebut memberikan landasan awal yang sangat berharga mengenai pemetaan wilayah, terdapat beberapa keterbatasan mendasar yang masih memerlukan pembahasan lebih mendalam. Mayoritas studi terdahulu cenderung berfokus eksklusif pada penggunaan satu jenis algoritma saja, tanpa menguji performanya secara berpasangan atau melakukan perbandingan langsung pada lingkup indeks pembangunan wilayah yang sama. Di samping itu, analisis yang dilakukan masih bergantung pada basis data lama dari rentang tahun 2019 hingga 2022. Rentang waktu tersebut dinilai belum sepenuhnya mampu memotret realitas kondisi infrastruktur saat ini, mengingat adanya akselerasi transformasi digital yang sangat masif di berbagai daerah pasca-pandemi. Kondisi ini diperumit oleh adanya disparitas pembangunan yang tajam antarpulau—terutama antara wilayah Jawa dan kawasan Indonesia Timur seperti Papua—yang sering kali menghasilkan variasi data ekstrem atau pencilan (*outlier*). Keberadaan data pencilan ini berisiko membiaskan pusat pengelompokan apabila hanya bersandar pada satu metode konvensional seperti *K-Means*.

Berangkat dari keterbatasan tersebut, penelitian ini diarahkan untuk mengisi ruang kosong yang belum terselesaikan dengan melakukan pengujian komparatif secara langsung antara algoritma *K-Means* dan *K-Medoids*. Melalui integrasi kedua metode ini, ketahanan algoritma *K-Medoids* dalam menangani variansi data ekstrem dapat diuji secara objektif guna meminimalkan distorsi akibat kesenjangan infrastruktur antardaerah. Sisi pembeda utama dari analisis ini juga terletak pada pemanfaatan data Indeks Pembangunan TIK terbaru tahun 2024. Langkah ini diambil agar pengelompokan yang dihasilkan benar-benar mencerminkan potret ketimpangan digital yang aktual dan kontekstual dengan kondisi riil di lapangan saat ini. Melalui pengukuran validitas pengelompokan yang terukur, studi ini diharapkan dapat memberikan alternatif metodologi pengelompokan yang lebih akurat sekaligus menyajikan rekomendasi yang lebih tajam bagi pemerintah dalam merumuskan kebijakan pemerataan akses teknologi secara nasional.

Pemetaan melalui *clustering* ini sangat esensial agar intervensi kebijakan pemerintah dapat difokuskan secara spesifik sesuai dengan karakteristik dan tingkat kebutuhan masing-masing kelompok provinsi. Berdasarkan uraian latar belakang tersebut, dipandang perlu untuk melakukan kajian pengelompokan wilayah secara statistik. Oleh karena itu, penelitian ini dilakukan dengan mengambil judul "Analisis *Cluster K-Means* dan *K-Medoids* Untuk Pengelompokan Provinsi di Indonesia Berdasarkan Pembangunan TIK Tahun 2024".

## METODOLOGI

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis data sekunder yang dikumpulkan untuk periode tahun 2024. Subjek penelitian mencakup seluruh wilayah administratif tingkat satu di Indonesia, yaitu 38 provinsi. Teknik pengambilan sampel yang diterapkan adalah *sampling* jenuh atau sensus, di mana seluruh anggota populasi dijadikan sampel guna meminimalisir *sampling error* dan menjamin representasi data secara nasional.

### Data dan Sumber Data

Penelitian ini menggunakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS) Republik Indonesia untuk tahun 2024. Populasi yang digunakan mencakup seluruh wilayah administratif tingkat satu di Indonesia, yaitu 38 provinsi. Penggunaan seluruh provinsi sebagai populasi bertujuan untuk memberikan gambaran yang menyeluruh mengenai kondisi pembangunan TIK secara nasional tanpa adanya bias pemilihan wilayah.

### Variabel Penelitian

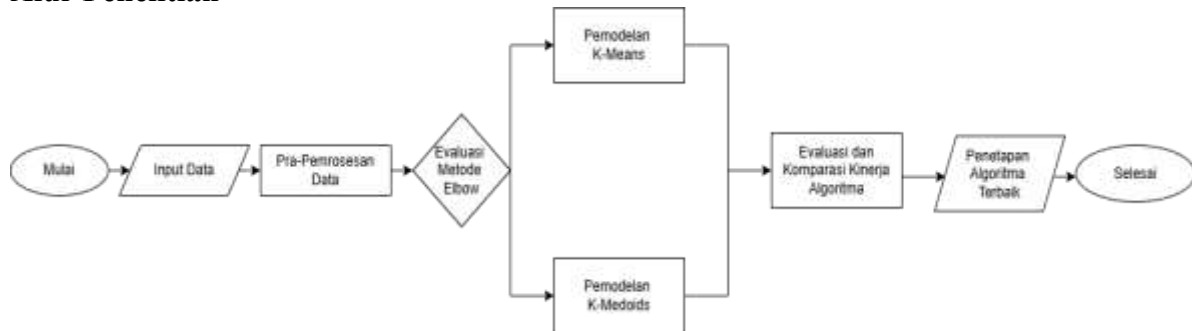
Variabel yang dianalisis dalam penelitian ini meliputi lima indikator utama yang merepresentasikan akses teknologi, kualitas pendidikan, dan kualitas hidup masyarakat.

**Tabel 1.** Kelima Variabel

| Variabel | Definisi                                                                                                              | Satuan           | Sumber                      |
|----------|-----------------------------------------------------------------------------------------------------------------------|------------------|-----------------------------|
| HP       | Menunjukkan tingkat kepemilikan dan kemudahan akses alat komunikasi dasar pada tingkat masyarakat maupun rumah tangga | Persen (%)       | Badan Pusat Statistik (BPS) |
| Komputer | Menunjukkan kepandaian digital dan kemampuan teknologi tingkat lanjut                                                 | Persen (%)       | Badan Pusat Statistik (BPS) |
| Internet | Menunjukkan seberapa jauh akses internet telah menyebar dan menjangkau masyarakat                                     | Persen (%)       | Badan Pusat Statistik (BPS) |
| RLS      | Menunjukkan rata-rata masa menempuh pendidikan penduduk 25 tahun ke atas                                              | Tahun            | Badan Pusat Statistik (BPS) |
| IPM      | Menunjukkan ukuran gabungan pada pencapaian kualitas hidup, pendidikan, kesehatan dan kesejahteraan masyarakat        | Indeks (0 – 100) | Badan Pusat Statistik (BPS) |



## Alur Penelitian



Gambar 1. Alur Penelitian

## Tahapan Analisis Data

Proses analisis data dalam penelitian ini dilaksanakan melalui serangkaian tahapan komputasi matematis yang terstruktur untuk mempartisi data wilayah berdasarkan indikator pembangunan TIK dan kualitas manusia. Adapun penjabaran dari masing-masing tahapan analisis tersebut adalah sebagai berikut:

### Tahap Pra-Pemrosesan Data (Standarisasi Z-Score)

Tahap paling awal sebelum algoritma diterapkan adalah penyeragaman rentang nilai antarvariabel. Hal ini mutlak diperlukan agar komputasi jarak spasial tidak didominasi oleh indikator tertentu yang memiliki skala ukur absolut lebih besar. Proses normalisasi dalam penelitian ini menggunakan teknik *Z-Score Standardization* (Maulana et al., 2024).

Formula matematis *Z-Score* dirumuskan sebagai berikut:

$$Z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (2.1)$$

Keterangan:

$Z_{ij}$  = Nilai data baru hasil normalisasi untuk wilayah ke- $i$  pada variabel ke- $j$ .

$x_{ij}$  = Nilai observasi asli wilayah ke- $i$  pada variabel ke- $j$ .

$\mu_j$  = Nilai rata-rata (*mean*) dari variabel ke- $j$  secara keseluruhan.

$\sigma_j$  = Standar deviasi (simpangan baku) dari variabel ke- $j$ .

### Penentuan Jumlah *Cluster* Optimal

Langkah krusial sebelum melakukan partisi data adalah menetapkan jumlah *cluster* ( $K$ ) yang paling optimal. Penelitian ini menerapkan *Elbow Method* dengan mengevaluasi tingkat kepadatan *cluster* melalui perhitungan *Sum of Squared Errors* (SSE) (Nabila & Mardhotillah, 2025). Titik optimal ditentukan pada area di mana kurva grafik mengalami penurunan tajam yang kemudian melandai hingga membentuk patahan menyerupai siku. Formula perhitungan SSE dirumuskan sebagai:

$$SSE = \sum_{k=1}^K \sum_{x_i \in C_k} d(x_i, c_k)^2 \quad (2.2)$$

Keterangan:

$SSE$  = Nilai total kuadrat galat (*Sum of Squared Errors*).

$K$  = Jumlah *cluster* yang sedang diuji.



$x_i$  = Titik data observasi wilayah yang tergabung dalam *cluster*  $C_k$ .

$c_k$  = Titik pusat rata-rata (*centroid*) dari *cluster*  $C_k$ .

$d(x_i, c_k)$  = Jarak spasial antara data observasi wilayah dengan titik pusat *cluster*.

### Penerapan Algoritma K-Means

Pendekatan K-Means Clustering yang merupakan bagian dari metode non-hierarki dipilih karena keunggulannya dalam mengelompokkan data berskala besar secara efisien dan objektif (Saputra et al., 2025). Algoritma ini mempartisi wilayah dengan meminimalkan jarak objek terhadap pusatnya menggunakan formula jarak metrik Euclidean:

$$d(x_i, c_k) = \sqrt{\sum_{j=1}^n (x_{ij} - c_{kj})^2} \quad (2.3)$$

Keterangan:

$d(x_i, c_k)$  = Jarak Euclidean antara wilayah ke- $i$  dengan pusat cluster ke- $k$ .

$n$  = Jumlah variabel (dimensi) indikator yang dianalisis.

$x_{ij}$  = Nilai dari variabel ke- $j$  pada wilayah ke- $i$ .

$c_{kj}$  = Nilai dari variabel ke- $j$  pada pusat cluster ke- $k$ .

Pada akhir setiap iterasi komputasi, posisi titik pusat cluster (*centroid*) akan selalu diperbarui menggunakan perhitungan nilai rata-rata (*mean*):

$$c_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i \quad (2.4)$$

Keterangan:

$c_k$  = Titik pusat (*centroid*) yang baru diperbarui.

$|C_k|$  = Jumlah anggota provinsi yang berada di dalam cluster  $C_k$ .

$x_i$  = Vektor data anggota yang berada di dalam cluster  $C_k$ .

### Penerapan Algoritma K-Medoids

Guna menjamin validitas dan memberikan instrumen komparasi, pemodelan data juga dieksekusi menggunakan algoritma K-Medoids. Berbeda dengan K-Means yang memunculkan titik rata-rata maya sebagai pusat, K-Medoids secara spesifik menggunakan objek observasi riil yang ada di dalam basis data sebagai pusat cluster (*medoid*). Objektif utama algoritma ini adalah meminimalkan total fungsi biaya atau simpangan jarak absolut (*Cost Function*) (Erlangga & Octavia, 2022). Rumus fungsi objektif tersebut adalah:

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} d(x_i, m_k) \quad (2.4)$$

Keterangan:

$J$  = Total fungsi biaya dari keseluruhan model *clustering*.

$K$  = Jumlah *cluster* yang terbentuk.

$x_i$  = Titik data observasi di dalam *cluster*  $SC\_k$ .

$m_k$  = Titik data riil yang ditunjuk sebagai *medoid* (pusat) pada *cluster*  $SC\_k$ .

$d(x_i, m_k)$  = Jarak fisik antara titik data wilayah dengan *medoid*.

### Evaluasi dan Komparasi Validitas *Cluster*

Untuk menentukan algoritma dengan tingkat akurasi struktur terbaik, dilakukan pengujian kinerja komparatif menggunakan lima instrumen matriks validasi *cluster*:

#### Silhouette Score

Metrik ini digunakan untuk mengukur seberapa baik dan presisi sebuah objek ditempatkan di dalam *cluster*-nya sendiri (kerapatan internal) dibandingkan dengan *cluster* tetangga di sekitarnya (pemisahan eksternal) (Erlangga & Octavia, 2022). Pemodelan dinyatakan sangat akurat apabila nilai *Silhouette Score* semakin mendekati angka 1.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2.5)$$

Keterangan:

$s(i)$  = Nilai evaluasi *Silhouette* untuk wilayah ke- $i$ .

$a(i)$  = Rata-rata jarak wilayah  $i$  ke seluruh anggota lain di dalam *cluster* yang sama.

$b(i)$  = Rata-rata jarak wilayah  $i$  ke seluruh anggota di *cluster* tetangga yang paling dekat letaknya.

#### Davies Bouldin Index (DBI)

DBI berfokus pada evaluasi rasio antara tingkat penyebaran (*scatter*) anggota di dalam satu *cluster* dengan jarak pisah menuju *cluster* lain (Fitriana et al., 2022). Syarat parameter keberhasilan pada metrik ini adalah semakin kecil nilainya (semakin mendekati 0), maka pembentukan struktur *cluster* tersebut dinilai semakin ideal.

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{j \neq i} \left( \frac{S_i + S_j}{d(c_i, c_j)} \right) \quad (2.6)$$

Keterangan:

$K$  = Jumlah total *cluster*.

$S_i, S_j$  = Nilai sebaran jarak spasial anggota di dalam *cluster*  $i$  dan *cluster*  $j$

$d(c_i, c_j)$  = Jarak fisik spasial antara titik pusat *cluster*  $i$  dan pusat *cluster*  $j$ .

#### Dunn Index

*Dunn Index* secara khusus membandingkan jarak spasial terpendek antar-*cluster* dengan jarak sebaran maksimum intra-*cluster* (Erlangga & Octavia, 2022). Algoritma dinyatakan memiliki kinerja yang lebih tangguh apabila menghasilkan nilai *Dunn Index* yang lebih besar.

$$DI = \frac{\min_{1 \leq i < j \leq K} d(c_i, c_j)}{\max_{1 \leq k \leq K} \Delta_k} \quad (2.7)$$

Keterangan:

$DI$  = Nilai evaluasi *Dunn Index*.

$d(c_i, c_j)$  = Jarak terpendek yang memisahkan antara dua pusat *cluster* yang berbeda.

$\Delta_k$  = Jarak spasial terjauh antara dua wilayah yang berada di dalam satu *cluster*  $k$  yang sama.

### Rasio SW/SB

Pengujian ini secara khusus difungsikan untuk melihat sejauh mana kualitas pemisahan variansi data secara keseluruhan (Nabila & Mardhotillah, 2025). Terdapat dua elemen perhitungan, yakni Sum of Squares Within (SW) yang mengukur tingkat variansi di dalam internal cluster (merepresentasikan kerapatan internal), dan Sum of Squares Between (SB) yang mengukur tingkat variansi antar-cluster terhadap titik pusat global (merepresentasikan pemisahan eksternal).

Rumus untuk menghitung nilai Sum of Squares Within (SW) adalah:

$$SW = \sum_{k=1}^K \sum_{x_i \in C_k} d(x_i, c_k)^2 \quad (2.8)$$

Keterangan Rumus SW:

$SW$  = Total variansi kuadrat galat internal di dalam seluruh *cluster*.

$x_i$  = Titik data objek provinsi ke- $i$  yang terdaftar sebagai anggota di dalam *cluster*  $C_k$ .

$c_k$  = Titik pusat kelompok (*centroid*) dari *cluster*  $C_k$ .

$d(x_i, c_k)$  = Jarak metrik antara objek provinsi  $x_i$  dengan titik pusat kelompok  $c_k$ .

Rumus untuk menghitung nilai Sum of Squares Between (SB) adalah:

$$SB = \sum_{k=1}^K n_k \cdot d(c_k, \bar{x})^2 \quad (2.9)$$

Keterangan Rumus SB:

$SB$  = Total variansi kuadrat galat eksternal antar-*cluster*.

$n_k$  = Jumlah total anggota provinsi yang berada di dalam *cluster* ke- $k$ .

$c_k$  = Titik pusat kelompok dari *cluster*  $C_k$ .

$\bar{x}$  = Titik pusat rata-rata global (*grand centroid*) dari keseluruhan data observasi.

$d(c_k, \bar{x})$  = Jarak metrik antara titik pusat kelompok  $c_k$  dengan titik pusat acuan global  $\bar{x}$

$$Rasio = \frac{SW}{SB} \quad (2.10)$$

Keterangan Rumus Rasio:

$Rasio$  = Nilai rasio evaluasi kualitas kepadatan.

$SW$  = Nilai total variansi internal kelompok.

$SB$  = Nilai total variansi eksternal antar-kelompok.

Syarat utama agar hasil partisi *cluster* dinilai valid dan solid secara statistik adalah nilai rasio harus jatuh sejauh mungkin di bawah angka 1. Kondisi tersebut membuktikan bahwa sebaran data objek di dalam kelompok cenderung sangat rapat, sementara jarak pemisah antar-*cluster* terpaut sangat jauh secara proporsional.

### Separation Index

Indeks pemisahan digunakan secara mendetail untuk mengukur batas aman jarak terpendek di antara dua titik pusat (*centroid* atau *medoid*) dari seluruh *cluster* yang terbentuk (Fitriana et al., 2022). Evaluasi ini bertujuan untuk mendeteksi dan mencegah zona wilayah yang saling bertabrakan (*overlapping*). Nilai Separation Index yang lebih besar mengindikasikan bahwa setiap *cluster* berlokasi pada posisi koordinat spasial yang sangat tegas dan mandiri.

$$Sep = \min_{1 \leq i < j \leq K} d(c_i, c_j) \quad (2.9)$$

Keterangan:

$Sep$  = Nilai evaluasi Indeks Pemisahan.



$d(c_i, c_j)$  = Jarak absolut terdekat antara titik pusat *cluster* ke- $i$  dan titik pusat *cluster* ke- $j$ .

## HASIL DAN PEMBAHASAN

### Deskripsi Data Penelitian

Penelitian ini menggunakan data capaian indikator Teknologi Informasi dan Komunikasi (TIK) serta Indeks Pembangunan Manusia (IPM) yang mencakup 38 provinsi di Indonesia. Variabel yang digunakan dalam proses analisis terdiri atas lima indikator utama, yaitu tingkat kepemilikan telepon seluler (HP), kepemilikan komputer, akses internet, Rata-rata Lama Sekolah (RLS), dan skor IPM.

Tahap awal sebelum melakukan pemodelan data adalah melakukan analisis statistik deskriptif untuk mengetahui gambaran umum dan ukuran pemusatan data secara nasional. Hasil perhitungan statistik deskriptif disajikan pada Tabel 1 berikut:

**Tabel 2.** Statistik Deskriptif Global (38 Provinsi)

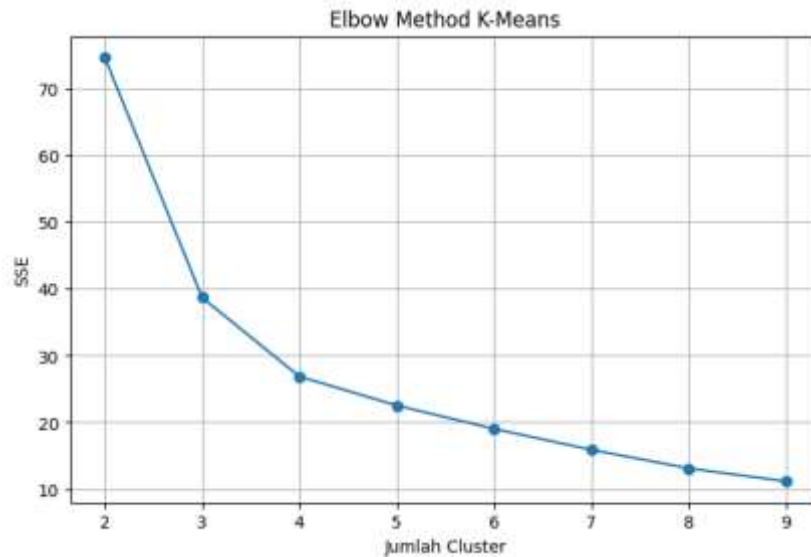
| Statistik                   | HP    | Komputer | Internet | RLS   | IPM   |
|-----------------------------|-------|----------|----------|-------|-------|
| <i>Count</i>                | 38    | 38       | 38       | 38    | 38    |
| Rata-rata<br>(Mean)         | 91,29 | 18,89    | 86,27    | 8,84  | 72,39 |
| Std<br>(Standar<br>Deviasi) | 13,08 | 6,67     | 15,99    | 1,23  | 5,15  |
| Minimum                     | 31,66 | 2,31     | 12,15    | 4,21  | 53,42 |
| Maksimum                    | 99,15 | 39,33    | 98,18    | 11,49 | 83,08 |
| Koefisien<br>Korelasi       | 0,79  | 0,82     | 0,83     | 0,89  | 1,00  |

Tabel 2 memaparkan ringkasan numerik dari kondisi 38 provinsi di Indonesia menunjukkan bahwa rata-rata Indeks Pembangunan Manusia (IPM) berada pada angka 72,39, dengan rentang nilai antara 53,42 hingga 83,08. Pada indikator TIK, penetrasi *handphone* dan akses internet sudah tergolong tinggi dengan rata-rata masing-masing mencapai angka 91,29 dan 86,27. Meskipun demikian, akses internet memiliki tingkat kesenjangan yang cukup ekstrem antarwilayah, ditunjukkan oleh nilai standar deviasi tertinggi sebesar 15,99 dan nilai minimum yang anjlok di angka 12,15. Sebaliknya, kepemilikan komputer masih menjadi indikator dengan capaian terendah, yakni dengan rata-rata hanya sebesar 18,89. Sementara itu, indikator pendidikan dasar yang direpresentasikan oleh Rata-rata Lama Sekolah (RLS) mencatatkan angka nasional sebesar 8,84 tahun. Lebih lanjut, hasil analisis korelasi mengonfirmasi bahwa seluruh variabel independen memiliki hubungan positif yang kuat hingga sangat kuat terhadap IPM. Variabel RLS memiliki korelasi linier terkuat di angka 0,89, yang sejalan dengan perannya sebagai salah satu dimensi fundamental pembentuk IPM. Indikator infrastruktur digital lanjutan, yakni akses internet (0,83) dan kepemilikan komputer (0,82), juga menunjukkan hubungan yang sangat erat dengan tingkat kesejahteraan. Kepemilikan *handphone* secara statistik turut memiliki korelasi yang kuat di angka 0,79 terhadap pencapaian IPM, menegaskan bahwa seluruh variabel tersebut sangat relevan untuk menjelaskan ketimpangan pembangunan antarprovinsi.

## Hasil *Clustering* Algoritma K-Means

### Penentuan Jumlah *Cluster* Optimal (K-Means)

Langkah pertama dalam menerapkan algoritma K-Means adalah menentukan jumlah *cluster* yang paling optimal. Penelitian ini menggunakan pendekatan *Elbow Method* dengan mengevaluasi pergerakan nilai kesalahan kuadrat atau *Sum of Squared Errors* (SSE).



**Gambar 2.** Grafik *Elbow Method* K-Means

Visualisasi Gambar 2 menampilkan kurva penurunan nilai SSE seiring dengan penambahan jumlah *cluster* pada sumbu horizontal. Kurva tersebut memperlihatkan penurunan yang sangat curam dari titik 2 ke titik 3, lalu mulai bergerak melandai secara konsisten dan membentuk sebuah patahan sudut (siku) yang paling stabil tepat pada absis angka 4. Mengacu pada prinsip heuristik *Elbow Method*, parameter jumlah *cluster* optimal yang ditetapkan untuk pemodelan K-Means dalam penelitian ini adalah 4 *cluster*.

### Profil dan Distribusi Anggota *K-Means*

Berdasarkan parameter berjumlah 4 *cluster*, algoritma mengeksekusi proses partisi terhadap 38 provinsi. Hasil pemetaan anggota untuk setiap *cluster* dirangkum dalam Tabel 3.

**Tabel 3.** Distribusi Anggota *Cluster K-Means*

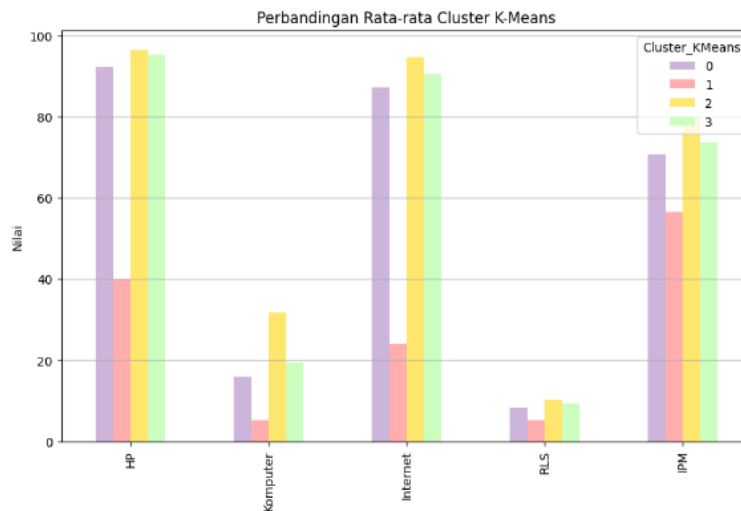
| <i>Cluster</i>   | Jumlah Anggota | Daftar Provinsi                                                                                                                                                                                                                                    |
|------------------|----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Cluster 0</i> | 15 Provinsi    | Sumatera Selatan, Lampung, Kep. Bangka Belitung, Jawa Tengah, Jawa Timur, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Selatan, Sulawesi Tengah, Gorontalo, Sulawesi Barat, Papua Barat, Papua Barat Daya, Papua Selatan |
| <i>Cluster 1</i> | 2 Provinsi     | Papua Tengah, Papua Pegunungan .                                                                                                                                                                                                                   |
| <i>Cluster 2</i> | 5 Provinsi     | Kep. Riau, DKI Jakarta, DI Yogyakarta, Bali, Kalimantan Timur                                                                                                                                                                                      |
| <i>Cluster 3</i> | 16 Provinsi    | Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Bengkulu, Jawa Barat, Banten, Kalimantan Tengah, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Maluku, Maluku Utara, Papua .                                           |

Tabel 2 menunjukkan sebaran wilayah ke dalam empat klasifikasi yang terbentuk secara otomatis oleh algoritma. *Cluster 3* menjadi klasifikasi dengan anggota terbanyak, mencakup 16 provinsi dari wilayah barat hingga timur Indonesia. Di sisi lain, algoritma secara spesifik mengisolasi Provinsi Papua Tengah dan Papua Pegunungan secara eksklusif ke dalam *Cluster 1*, menjadikannya *cluster* dengan keanggotaan terkecil, yakni hanya 2 provinsi. Guna mengidentifikasi karakteristik dari keempat *cluster* tersebut, dihitung nilai rata-rata (*centroid*) dari masing-masing variabel yang disajikan pada Tabel 3.

**Tabel 3.** Rata-rata Variabel per *Cluster K-Means*

| <b>Cluster</b>   | <b>HP</b> | <b>Komputer</b> | <b>Internet</b> | <b>RLS</b> | <b>IPM</b> | <b>Kategori</b>   |
|------------------|-----------|-----------------|-----------------|------------|------------|-------------------|
| <i>Cluster 0</i> | 92,15     | 15,97           | 87,25           | 8,26       | 70,76      | Menengah          |
| <i>Cluster 1</i> | 39,92     | 5,09            | 24,11           | 5,16       | 56,58      | Sangat Tertinggal |
| <i>Cluster 2</i> | 96,51     | 31,67           | 94,57           | 10,29      | 79,83      | Sangat Maju       |
| <i>Cluster 3</i> | 95,26     | 19,35           | 90,53           | 9,38       | 73,56      | Maju              |

Data pada Tabel 3 memperlihatkan nilai absolut yang mendasari penjenjangan kategori wilayah. *Cluster 2* mencatatkan rata-rata tertinggi secara konsisten di seluruh kolom indikator, sehingga diklasifikasikan sebagai kategori Sangat Maju. Pada kutub yang berlawanan, *Cluster 1* mencatatkan capaian terendah dengan angka kepemilikan komputer yang hanya menyentuh titik 5,09 dan akses internet tertahan di angka 24,11. Angka krisis ini menempatkan *Cluster 1* ke dalam kategori Sangat Tertinggal.



**Gambar 2. Grafik Perbandingan Rata-rata *Cluster K-Means***

Gambar 2 merupakan proyeksi visual dari nilai sentroid yang terdapat pada Tabel 3 Melalui diagram batang ber-*cluster* ini, keunggulan *Cluster 2* (direpresentasikan oleh batang kuning) tampak sangat mendominasi tinggi grafik pada setiap variabel. Sebaliknya, ketertinggalan infrastruktur pada *Cluster 1* (direpresentasikan oleh batang merah muda) tervisualisasi jelas dengan tinggi batang yang selalu berada di urutan paling bawah, jauh tertinggal dari ketiga *cluster* lainnya.

### Peta Persebaran Cluster Provinsi di Indonesia Berdasarkan Pembangunan TIK Tahun 2024 dengan Algoritma *K-Means*



**Gambar 3. Peta Persebaran *Cluster K-Means***

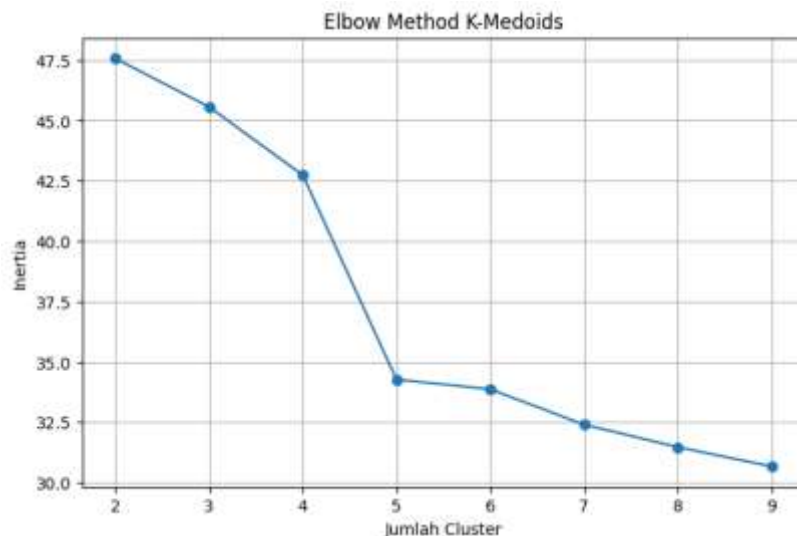
Berdasarkan Peta persebaran *Cluster K-Means*, ketimpangan infrastruktur TIK dan kesejahteraan di Indonesia dapat diklasifikasikan ke dalam empat kelompok wilayah. Wilayah kategori sangat maju (*Cluster 2*) terlihat memusat secara eksklusif di pusat pemerintahan dan ekonomi utama, meliputi DKI Jakarta, DI Yogyakarta, Bali, Kepulauan Riau, dan Kalimantan Timur. Selanjutnya, wilayah kategori maju (*Cluster 3*) memiliki sebaran yang paling luas, mendominasi sebagian besar Pulau Sumatera, sebagian wilayah Jawa, hingga Semenanjung Sulawesi. Di sisi lain, wilayah kategori menengah (*Cluster 0*) tersebar memanjang melintasi kawasan selatan Sumatera, Jawa bagian tengah dan timur, Kepulauan Nusa Tenggara, hingga pesisir Pulau Papua. Ketimpangan yang paling ekstrem ditunjukkan oleh wilayah kategori sangat tertinggal (*Cluster 1*) yang terisolasi secara

geografis hanya di pedalaman timur Indonesia, yakni di Provinsi Papua Tengah dan Papua Pegunungan. Secara keseluruhan, pemetaan ini menegaskan bahwa sentra kemajuan masih sangat terkonsentrasi di kawasan barat dan aglomerasi perkotaan, berbanding terbalik dengan kawasan pedalaman di timur yang masih sangat membutuhkan pemerataan pembangunan.

### Hasil Clustering Algoritma K-Medoids

#### Penentuan Jumlah Cluster Optimal K-Medoids

Pada algoritma K-Medoids, penentuan jumlah parameter *cluster* dievaluasi melalui pergerakan penurunan nilai variansi internal atau *Inertia*.



**Gambar 4 Grafik Elbow Method K-Medoids**

Grafik *Elbow* pada Gambar 4 menampilkan bentuk kurva yang memiliki karakteristik berbeda dari model sebelumnya. Penurunan nilai *Inertia* yang tajam terlihat terus berlanjut hingga akhirnya menciptakan sudut landai (siku) yang paling kentara tepat pada absis angka 5. Berdasarkan titik kestabilan tersebut, parameter *clustering* yang diaplikasikan pada pemodelan K-Medoids adalah sebanyak 5 *cluster*.

### Profil dan Distribusi Anggota K-Means

Pemodelan K-Medoids mempartisi ke-38 provinsi di Indonesia menjadi 5 entitas terpisah. Struktur distribusi anggota dijabarkan pada Tabel 4.

**Tabel 4. Distribusi Anggota Cluster K-Medoids**

| Cluster   | Jumlah Anggota | Daftar Provinsi                                                                       |
|-----------|----------------|---------------------------------------------------------------------------------------|
| Cluster 0 | 5              | Kep. Riau, DKI Jakarta, DI Yogyakarta, Bali, Kalimantan Timur                         |
| Cluster 1 | 5              | Aceh, Sumatera Utara, Sulawesi Utara, Maluku, Maluku Utara                            |
| Cluster 2 | 7              | Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Barat, Papua Barat, Papua Selatan, |



|                  |    |                                                                                                                                                                                                                                                                                                                                              |
|------------------|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Cluster 3</i> | 12 | Papua Tengah, Papua Pegunungan<br>Sumatera Barat, Riau,<br>Jambi, Bengkulu, Jawa Barat, Banten, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Selatan, Sulawesi Tenggara, Papua Sumatera Selatan, Lampung, Kep. Bangka Belitung, Jawa Tengah, Jawa Timur, Kalimantan Barat, Sulawesi Tengah, Gorontalo, Papua Barat Daya |
| <i>Cluster 4</i> | 9  |                                                                                                                                                                                                                                                                                                                                              |

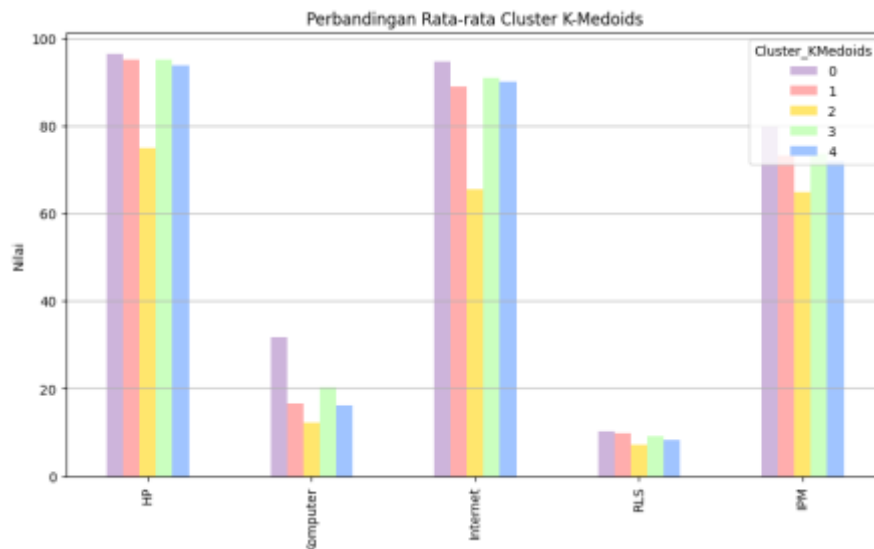
Berdasarkan Tabel 4, hasil pemetaan K-Medoids menghasilkan sebaran anggota yang cukup merata. *Cluster 3* menjadi himpunan dengan anggota paling banyak, yakni 12 provinsi. Disusul oleh *Cluster 4* sebanyak 9 provinsi, dan *Cluster 2* sebanyak 7 provinsi. Sementara itu, *Cluster 0* dan *Cluster 1* masing-masing menampung porsi terkecil yaitu 5 provinsi.

Untuk melihat jenjang performa tiap klasifikasi, dihitung nilai titik pusat riil (*medoid*) pada masing-masing variabel yang disajikan dalam Tabel 5.

**Tabel 5. Rata-rata Variabel per Cluster K-Medoids**

| <i>Cluster</i>   | HP    | Komputer | Internet | RLS   | IPM   | Kategori          |
|------------------|-------|----------|----------|-------|-------|-------------------|
| <i>Cluster 0</i> | 96,51 | 31,67    | 94,57    | 10,29 | 79,83 | Sangat Maju       |
| <i>Cluster 1</i> | 95,05 | 16,67    | 89,02    | 9,80  | 73,13 | Menengah Atas     |
| <i>Cluster 2</i> | 74,91 | 12,13    | 65,44    | 7,23  | 64,94 | Sangat Tertinggal |
| <i>Cluster 3</i> | 95,14 | 20,40    | 91,05    | 9,14  | 73,69 | Maju              |
| <i>Cluster 4</i> | 93,89 | 16,25    | 89,96    | 8,34  | 71,88 | Menengah Bawah    |

Tabel 5 menyajikan hierarki pencapaian kelima *cluster* secara berurutan. *Cluster 0* menempati peringkat tertinggi dengan nilai kepemilikan HP mencapai 96,51 dan IPM 79,83, menjadikannya kategori Sangat Maju. Penjenjangan ini terus menurun hingga bermuara pada *Cluster 2*, di mana angka capaian akses internet anjlok ke titik 65,44 dan nilai IPM berada di angka 64,94, sehingga *Cluster 2* dilabeli sebagai kategori Sangat Tertinggal.



**Gambar 5. Grafik Perbandingan Rata-rata *Cluster* K-Medoids**

Visualisasi grafik batang pada Gambar 5 mendukung interpretasi numerik dari tabel sebelumnya. Batang berwarna ungu (*Cluster* 0) secara konsisten berdiri paling tinggi pada setiap indikator bangunan TIK maupun manusia. Sebaliknya, batang berwarna kuning (*Cluster* 2) selalu menampilkan tinggi yang paling kerdil dibandingkan keempat *cluster* lainnya, merepresentasikan kondisi wilayah dengan performa terendah pada model K-Medoids.

### **Peta Persebaran Cluster Provinsi di Indonesia Berdasarkan Pembangunan TIK Tahun 2024 dengan Algoritma *K-Medoids***



**Gambar 6. Peta Persebaran *Cluster* K-Medoids**

Berdasarkan peta persebaran K-Medoids, 38 provinsi di Indonesia terbagi ke dalam lima tingkatan pembangunan. Cluster 0 (Kategori Sangat Maju) memiliki rata-rata tertinggi pada seluruh indikator (seperti IPM 79,83 dan akses internet 94,57), yang terpusat secara eksklusif di wilayah aglomerasi utama meliputi DKI Jakarta, DI Yogyakarta, Bali, dan Kalimantan Timur. Menyusul di bawahnya, Cluster 3 (Kategori Maju) dengan rata-rata IPM 73,69 mendominasi sebagian besar daratan Sumatra, Kalimantan Tengah, hingga Sulawesi Selatan. Pada kelompok menengah, Cluster 1 (Menengah Atas) dengan IPM 73,13 tersebar di utara Sumatra, utara Sulawesi, hingga Maluku. Sementara itu, Cluster 4 (Menengah

Bawah) yang memiliki IPM 71,88 lebih banyak terkonsentrasi di kawasan padat penduduk seperti Pulau Jawa, selatan Sumatera, dan pesisir barat Kalimantan. Sebaliknya, ketimpangan paling tajam terlihat pada Cluster 2 (Kategori Sangat Tertinggal). Kelompok ini mencatatkan angka terendah pada seluruh variabel, khususnya akses internet (65,44) dan IPM (64,94), yang secara visual terisolasi masif di kawasan timur Indonesia, membentang dari Kepulauan Nusa Tenggara hingga hampir seluruh daratan Papua.

### Evaluasi dan Komparasi Kinerja Algoritma

Guna membuktikan validitas matematis serta menentukan algoritma manakah yang memiliki kecocokan (*goodness of fit*) terbaik terhadap pola sebaran data, dilakukan pengujian perbandingan kinerja menggunakan lima instrumen matriks evaluasi *cluster*. Hasil kalkulasi disajikan dalam Tabel 6.

**Tabel 6. Komparasi Evaluasi Kinerja Algoritma**

| Metrik Validasi             | Nilai K-Means | Nilai K-Medoids | Algoritma Terbaik |
|-----------------------------|---------------|-----------------|-------------------|
| <i>Silhouette Score</i>     | 0,347335      | 0,163293        | K-Means           |
| <i>Davies Bouldin Index</i> | 0,785325      | 1,065251        | K-Means           |
| <i>Dunn Index</i>           | 0,112289      | 0,039472        | K-Means           |
| <i>SW/SB Ratio</i>          | 0,164702      | 1,913505        | K-Means           |
| <i>Separation Index</i>     | 1,236856      | 0,837684        | K-Means           |

Tabel 6 memberikan gambaran yang jelas mengenai hasil pengujian performa kedua algoritma. Berdasarkan perhitungan tersebut, K-Means secara mutlak mengungguli K-Medoids pada seluruh metrik evaluasi. Hal ini dibuktikan dengan perolehan *Silhouette Score* K-Means 0,3473 yang jauh lebih tinggi dari K-Medoids 0,1632, yang menandakan tingkat kedekatan antaranggota di dalam klaster yang sangat erat. Keunggulan ini didukung oleh *Davies Bouldin Index* yang lebih kecil 0,7853 dan *Separation Index* yang lebih besar 1,2368. Angka-angka tersebut menegaskan bahwa K-Means mampu membentuk batas pemisah klaster yang sangat tegas sehingga terhindar dari tumpang tindih wilayah. Secara praktis, K-Means terbukti lebih sensitif memotret realitas ketimpangan dengan mengisolasi Provinsi Papua Tengah dan Papua Pegunungan secara spesifik ke dalam kategori sangat tertinggal akibat krisis infrastruktur TIK. Tingginya akurasi K-Means ini sangat dipengaruhi oleh kemampuannya dalam merespons data yang memiliki rentang kesenjangan ekstrem. Pendekatan nilai rata-rata maya (*centroid*) pada K-Means terbukti lebih adaptif menyeimbangkan sebaran data dibandingkan K-Medoids yang menggunakan data observasi riil (*medoid*) yang cenderung kaku. Bukti nyata kelemahan K-Medoids terlihat pada rasio SW/SB yang meledak hingga 1,9135. Angka yang melebihi batas validitas di atas 1 ini menunjukkan bahwa variansi data di dalam klaster justru lebih berantakan daripada jarak antar-klasternya, sehingga memicu pembiasan. Sebaliknya, K-Means tampil sangat solid dengan menekan rasio SW/SB hingga ke angka minimalis sebesar 0,1647. Keseluruhan hasil ini menyimpulkan bahwa K-Means merupakan algoritma yang paling ideal dan akurat untuk memetakan potret ketimpangan digital di Indonesia.

### SIMPULAN

Berdasarkan analisis klaster terhadap data pembangunan TIK dan kualitas manusia di 38 provinsi Indonesia tahun 2024, pendekatan *Elbow Method* menghasilkan jumlah kelompok optimal yang berbeda, yaitu 4 klaster untuk *K-Means* dan 5 klaster untuk *K-*



*Medoids*. Pemetaan menggunakan *K-Means* terbukti lebih kontekstual dalam memotret ketimpangan digital, di mana Klaster 2 (5 provinsi) masuk kategori Sangat Maju, sementara Klaster 1 secara spesifik mengisolasi Provinsi Papua Tengah dan Papua Pegunungan sebagai kategori sangat tertinggal akibat krisis teknologi ekstrem dengan tingkat akses internet yang tertahan di angka 24,11. Atas dasar ketimpangan riil ini, pemerintah dan pemangku kebijakan disarankan untuk menghentikan kebijakan penyamarataan infrastruktur (*one-size-fits-all*) dan segera beralih ke strategi intervensi asimetris yang berfokus penuh pada wilayah pedalaman Papua tersebut melalui percepatan pembangunan BTS siber, subsidi perangkat komputer sekolah, dan perluasan jaringan *broadband*. Dari sisi metodologi, hasil pengujian validitas menggunakan lima instrumen matriks menunjukkan bahwa algoritma *K-Means* secara mutlak mengungguli *K-Medoids* pada seluruh indikator performa. Keunggulan *K-Means* ditunjukkan oleh *Silhouette Score* yang lebih tinggi 0,3473 dan 0,1632 dan nilai *Davies-Bouldin Index* yang lebih kecil 0,7853 dan 1,0652. Pembuktian paling krusial terletak pada rasio SW/SB, di mana *K-Means* tampil sangat solid di bawah angka satu 0,1647, sedangkan *K-Medoids* melemah dan tidak valid secara statistik karena rasionya meledak hingga 1,9135 akibat terdistorsi oleh variansi data ekstrem. Oleh karena itu, peneliti selanjutnya disarankan untuk mengembangkan efisiensi *K-Means* dengan mengintegrasikan algoritma optimasi heuristik seperti *Genetic Algorithm* (GA) atau *Particle Swarm Optimization* (PSO) untuk mengatasi kelemahan inisialisasi pusat klaster awal yang acak. Selain itu, Badan Pusat Statistik (BPS) direkomendasikan untuk memperluas indikator pendataan daerah dengan memasukkan variabel literasi siber dan pemanfaatan teknologi pada UMKM guna mendukung analisis multivariat yang lebih mikro di masa mendatang.

## 5. DAFTAR PUSTAKA

- Dewi, D. M., Saingan, A. F., & Fahmi, Y. (2022). Kontribusi teknologi informasi dan komunikasi terhadap rata-rata lama sekolah di Pulau Jawa. *Jurnal Pendidikan*, 20, 24–36.
- Erlangga, W., & Octavia, S. F. (2022). Perbandingan K-Means dan K-Medoids pada pengelompokan data miskin di Indonesia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(2), 35–41.
- Fitriana, A., Yanti, N., & Zulkarnaen, I. (2022). Komparasi K-Means Clustering dan K-Medoids dalam mengelompokkan produksi susu segar di Indonesia. *BITE: Jurnal Bumigora Information Text and Engineering*, 4(2), 154-162.
- Koswara, A., et al. (2024). Digitalisasi ekonomi di pedesaan: Mengkaji kesenjangan infrastruktur digital di Indonesia. *Jurnal Manajemen*, 5(5).
- Maulana, M. S., Huda, M., & Fauzi, A. (2024). Simulation of the K-Means clustering algorithm with the Elbow method in making clusters of provincial poverty levels in Indonesia. *Journal of Mathematics and Informatics (Mantik)*, 8(4), 1580-1587.
- Nabila, A. P., & Mardhotillah, B. (2025). Segmentasi provinsi di Indonesia menggunakan metode K-Means berdasarkan tujuan mengakses internet tahun 2024. *Jurnal Komputasi*, 5(November), 123–141.
- Saputra, N. R., Muflih, G. Z., & Informatika, T. (2025). Pengelompokan wilayah Indonesia berdasarkan komponen Indeks Pembangunan Manusia dengan pendekatan algoritma K-Means clustering. *Jurnal Informatika*, 8, 156–167.
- Widiasanti, I., Rahmadani, S., & Nur, D. A. (2025). Kesetaraan akses internet dan tantangan literasi digital di Indonesia. *Jurnal Sains Data*, 9, 19631–19637.