

# Implementasi *K-Means Clustering* Dengan *Davies-Bouldin Index Evaluation* Terhadap Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Sosioekonomi

Cintha Agatha Sinaga<sup>1</sup>, Sardo Pardingotan Sipayung<sup>2</sup>

<sup>1,2</sup> Universitas Katolik Santo Thomas Medan, Jl. Setia Budi No.479 F Tanjung Sari, Medan, Indonesia

## ARTICLE INFORMATION

Received: February 00, 00  
Revised: March 00, 00  
Available online: April 00, 00

## KEYWORDS

Data Mining, K-Means Clustering, Indeks Pembangunan Manusia, Davies-Building Index

## CORRESPONDENCE

Phone: 088708489910

E-mail: [cinthyaagathaa@gmail.com](mailto:cinthyaagathaa@gmail.com)  
[pinsarsiphom@gmail.com](mailto:pinsarsiphom@gmail.com)

## A B S T R A K

Penelitian ini bertujuan untuk menganalisis dan mengelompokkan 38 provinsi di Indonesia guna mengidentifikasi pola disparitas pembangunan wilayah berdasarkan tiga indikator sosioekonomi utama yaitu Persentase Penduduk Miskin (PPM), Indeks Pembangunan Manusia (IPM), dan Indeks Tingkat Pengangguran Terbuka (TPT). Metode yang digunakan adalah Data Mining dengan teknik *K-Means Clustering* sebagai algoritma *unsupervised learning*. Data yang digunakan dinormalisasi menggunakan Standard Scaler (*Z-Score*) untuk memastikan akurasi jarak. Penentuan *k* optimal ( $k=3$ ) dievaluasi menggunakan Davies-Bouldin Index (DBI) yang mengonfirmasi konfigurasi kluster paling kohesif dan terpisah. Hasil klusterisasi berhasil mengelompokkan provinsi ke dalam tiga kategori yang signifikan secara statistik: Kluster Maju (dicirikan oleh IPM tinggi dan PPM rendah), Kluster Berkembang (nilai IPM dan PPM moderat), dan Kluster Tertinggal (IPM sangat rendah, PPM sangat tinggi). Kluster Tertinggal, yang mayoritas berada di wilayah timur Indonesia, mengindikasikan kebutuhan akan intervensi spesifik dan terfokus pada peningkatan kualitas sumber daya manusia (SDM) dan pengentasan kemiskinan struktural. Klusterisasi ini memberikan dasar bagi pemerintah daerah untuk merumuskan kebijakan pembangunan yang lebih tepat sasaran.

## PENDAHULUAN

Pembangunan nasional di Indonesia seringkali dihadapkan pada tantangan besar berupa disparitas regional yang mencolok antar wilayah [1], [2]. Kesenjangan ini mencakup aspek ekonomi, sosial, dan kualitas sumber daya manusia. Dalam konteap perencanaan pembangunan, pemahaman mendalam mengenai pola pengelompokan (klusterisasi) wilayah berdasarkan karakteristik serupa menjadi esensial untuk merumuskan kebijakan yang efektif dan tepat sasaran [3]. Indikator utama yang sering digunakan untuk mengukur keberhasilan pembangunan adalah Indeks Pembangunan Manusia (IPM), yang mencerminkan kualitas hidup; Persentase Penduduk Miskin (PPM), yang merefleksikan keberhasilan pengentasan kemiskinan; dan Tingkat Pengangguran Terbuka (TPT), yang menggambarkan kondisi pasar tenaga kerja [4]. Tujuan Penelitian ini adalah untuk menerapkan teknik Data Mining melalui algoritma K-Means Clustering pada data 38 provinsi di Indonesia menggunakan variabel PPM, IPM, dan TPT. Tujuan spesifiknya adalah mengklasifikasikan provinsi ke dalam kluster pembangunan wilayah homogen (Maju, Berkembang, Tertinggal) serta mengevaluasi kualitas klusterisasi tersebut menggunakan Davies-Bouldin Index (DBI).

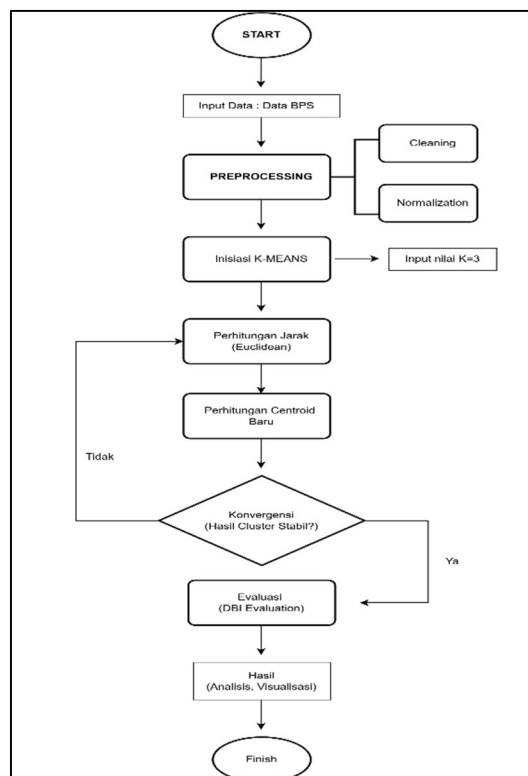
Studi literatur menunjukkan bahwa K-Means Clustering merupakan alat yang sangat populer dan efektif dalam menganalisis data sosioekonomi regional. Berbagai penelitian telah menggunakan K-Means untuk memetakan disparitas di Indonesia [5], [6]. Misalnya, Wahyuni (2021) menggunakan K-Means untuk mengelompokkan daerah tertinggal berdasarkan indikator kemiskinan [2], sementara penelitian lain berfokus pada klusterisasi IPM di berbagai wilayah [7], [8]. Temuan dari penelitian-penelitian ini menegaskan bahwa penggunaan *unsupervised learning* dapat mengungkapkan pola tersembunyi yang berguna bagi alokasi sumber daya. Meskipun K-Means sering digunakan, studi yang mengintegrasikan ketiga dimensi (IPM,

Kemiskinan, dan TPT) dalam satu model klasterisasi untuk mendapatkan profil pembangunan yang komprehensif pada skala 38 provinsi masih terbatas. Penentuan kualitas klasterisasi melalui metrik seperti DBI atau Silhouette Coefficient juga merupakan praktik standar untuk memvalidasi temuan, karena K-Means rentan terhadap inisiasi centroid awal [9], [10].

Penelitian ini, dengan menetapkan centroid awal secara spesifik dan mengevaluasi hasilnya, menawarkan kontribusi dalam validasi metodologis di samping temuan substantif klasterisasi pembangunan wilayah [11], [12]. Dengan memadukan variabel yang mengukur kualitas SDM (IPM), masalah ekonomi dasar (PPM), dan kondisi pasar kerja (TPT), penelitian ini diharapkan dapat memberikan pemetaan wilayah yang lebih holistik. Klasifikasi yang dihasilkan dapat menjadi dasar yang kuat bagi pemerintah untuk merumuskan strategi pembangunan yang tidak hanya berfokus pada pertumbuhan ekonomi, tetapi juga pada peningkatan kapabilitas manusia dan penanggulangan ketidakmerataan kesempatan kerja, terutama bagi wilayah yang terkluster sebagai Tertinggal [13], [14].

## METODE PENELITIAN

Proses analisis data dalam penelitian ini mengikuti metodologi Data Mining terstruktur yang dirancang untuk menjamin keandalan hasil klasterisasi pembangunan wilayah. Metodologi ini dibagi menjadi tiga tahap utama: Pengumpulan Data, Pra-Pemrosesan, dan Pemrosesan Algoritma dan Evaluasi



**Gambar 1** Flowchart Penelitian

### 2.1 Data Collecting

Langkah awal dalam penelitian ini adalah Pengumpulan Data sekunder dan Pembersihan Data. Data yang digunakan, yaitu PPM, IPM, dan TPT dari 38 provinsi, diperoleh dari sumber resmi BPS untuk menjamin validitas. Pengumpulan data dari sumber resmi memastikan validitas dan keandalan data sebagai landasan analisis.

### 2.2 Preprocessing

#### a. Data Cleaning

Sebelum data dapat diproses oleh algoritma, dilakukan pembersihan data. Tahap ini bersifat krusial untuk memastikan integritas dan kelayakan data. Pembersihan Data meliputi pemeriksaan dan penanganan nilai hilang (NaN) pada dataset. Baris data yang mengandung NaN dihapus untuk mencegah error pada algoritma K-Means dan memastikan validitas hasil analisis.

### b. Data Normalization

Tahap selanjutnya adalah Normalisasi Z-Score. Normalisasi ini diterapkan karena variabel PPM, IPM, dan TPT memiliki satuan dan rentang nilai yang berbeda. Standard Scaler mengubah nilai setiap fitur ( $x$ ) menjadi nilai terstandar ( $z$ ) yang memiliki rata-rata ( $\mu$ ) nol dan deviasi standar ( $\sigma$ ) satu. Normalisasi ini penting untuk memastikan bahwa setiap fitur memiliki kontribusi yang setara dalam perhitungan jarak Euclidean. Rumus Z-Score yang digunakan adalah:

$$z = \frac{x - \mu}{\sigma}$$

## 2.3 Processing

### a. Algoritma K-Means Clustering

Proses inti klusterisasi menggunakan Algoritma K-Means yang dijalankan secara iteratif dan bertujuan membagi  $N$  objek data ke dalam  $k$  kluster dengan meminimalkan variasi di dalam kluster (SSE). Algoritma dimulai dengan penentuan  $k=3$  dan inisiasi centroid awal. Pada setiap iterasi, algoritma melakukan dua fungsi utama: penugasan data ke centroid terdekat, dan perhitungan centroid baru. Siklus berulang ini berlanjut hingga konvergensi tercapai, yaitu ketika centroid tidak lagi berpindah secara signifikan, menunjukkan kluster telah stabil<sup>2</sup>. Penentuan kluster terdekat dilakukan melalui perhitungan Jarak Euclidean (Euclidean Distance). Jarak Euclidean mengukur jarak garis lurus (straight-line distance) antara dua titik di ruang  $n$ -dimensi (dalam kasus ini, ruang 3-dimensi yang dibentuk oleh PPM, IPM, dan TPT), yang memastikan bahwa penugasan kluster didasarkan pada kedekatan spasial di semua dimensi fitur. Formula matematis yang digunakan untuk menghitung jarak data dengan Euclidean distance Adalah sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

### b. Evaluasi Davies-Bouldin Index

Setelah K-Means mencapai konvergensi, tahap akhir adalah Evaluasi DBI. Davies-Bouldin Index (DBI) adalah metrik internal yang digunakan untuk mengukur kualitas kluster. DBI bekerja dengan mengukur rasio antara penyebaran di dalam kluster (kohesi) dengan pemisahan antar kluster (separasi). Nilai DBI yang paling rendah mengindikasikan klusterisasi yang optimal (kluster padat dan terpisah). Rumus DBI adalah:

$$DBI = \frac{1}{N} \sum_{i=1}^N \max_{i \neq j} \left( \frac{S_i + S_j}{M_{i,j}} \right)$$

$S_i$  adalah rata-rata jarak antara setiap titik di kluster  $i$  dengan centroid kluster  $i$  (ukuran penyebaran/kohesi).  $M_{ij}$  adalah jarak antara centroid kluster  $i$  dan kluster  $j$  (ukuran separasi). Nilai DBI diambil dari nilai rata-rata rasio kesamaan maksimum  $\left( \frac{S_i + S_j}{M_{ij}} \right)$  untuk semua kluster.

## HASIL DAN PEMBAHASAN

### 3.1 Hasil

#### 3.1.1 Data Collecting

Data yang digunakan adalah data sekunder dari lembaga BPS untuk 38 provinsi di Indonesia pada tahun 2024. Pemilihan variabel (IPM, PPM, dan TPT) didasarkan pada kebutuhan untuk memetakan dimensi pembangunan secara komprehensif yaitu IPM merefleksikan kualitas SDM dan capaian sosial; PPM merefleksikan keberhasilan ekonomi dalam mengurangi ketimpangan; dan TPT mengukur efisiensi pasar tenaga kerja. Pengintegrasian ketiga indikator ini dalam satu model klusterisasi memberikan dasar analisis multivariat yang kuat terhadap disparitas regional. Keterangan variabel disajikan pada Tabel 1, dan cuplikan data mentah awal disajikan pada Tabel 2.

**Table 1** Keterangan Variabel Data

Atribut	Skala	Keterangan
Persentase Penduduk Miskin (PPM)	Persentase	Proporsi penduduk yang berada di bawah garis kemiskinan.
Indeks Pembangunan Manusia (IPM)	Skor	Rata-rata capaian kesehatan, pendidikan, dan standar hidup layak.
Indeks Tingkat Pengangguran Terbuka (TPT)	Persentase	Persentase angkatan kerja yang tidak bekerja atau sedang mencari pekerjaan.

Dengan menggunakan tiga atribut diatas, maka berikut ini Adalah dataset awal yang digunakan dan masuk ketahap normalisasi,

**Table 2** Dataset Awal

No	Nama Provinsi	Persentase Penduduk Miskin	Indeks Pembangunan Manusia	Indeks TPT
1	Aceh	12,64	74,03	5,75
2	Sumatera Utara	7,19	74,02	5,6
3	Sumatera Barat	5,42	74,49	5,75
4	Riau	6,36	74,79	3,7
5	Jambi	7,26	73,43	4,48
6	Sumatera Selatan	10,51	72,3	3,86
7	Bengkulu	12,52	73,39	3,11
.....				
38	Papua Pegunungan	29,66	53,42	1,32

### 3.2 Preprocessing

#### a. Data Cleaning

Pada tahap pembersihan data, ditemukan 1 baris yang mengandung nilai NaN dari total observasi. Baris tersebut secara konservatif dihapus (`dropna()`) untuk memastikan data yang diproses bebas dari nilai hilang. Eliminasi observasi ini penting untuk memenuhi persyaratan algoritma K-Means dan menjaga validitas statistik. Setelah pembersihan, total data final yang digunakan untuk klusterisasi adalah 38 observasi provinsi.

#### b. Data Normalization (Z-Score)

Normalisasi Z-Score diaplikasikan untuk menstandarkan nilai data. Transformasi ini mencegah variabel dengan magnitude nilai lebih besar (seperti IPM) mendominasi perhitungan jarak. Nilai Z-

Score positif (Tabel 3) menunjukkan bahwa indikator provinsi berada di atas rata-rata nasional, sementara nilai negatif menunjukkan sebaliknya. Proses standardisasi memastikan kontribusi yang setara dari ketiga dimensi pembangunan dalam ruang klasterisasi 3-dimensi.

**Table 3** Hasil Normalisasi Z-Score

Nama Provinsi	Persentase Penduduk Miskin	Indeks Pembangunan Manusia	Indeks TPT
Aceh	0.314411	0.323018	0.981923
Sumatera Utara	-0.553089	0.321050	0.874434
Sumatera Barat	-0.834827	0.413533	0.981923
Riau	-0.685203	0.472565	-0.487095
Jambi	-0.541946	0.204954	0.071848
Dst..			

### 3.3 Processing

#### a. Implementasi Algoritma K-Means Clustering

Algoritma K-Means dijalankan pada data Z-Score dengan  $k=3$  klaster, yang ditetapkan berdasarkan kebutuhan segmentasi kebijakan. Proses dimulai dengan inisiasi centroid spesifik:

$$\text{Centroid awal} = \begin{bmatrix} 1.18191 & 0.120342 & 1.50504 \\ 0.29531 & 0.197083 & -0.90988 \\ -0.55308 & 0.32105 & 0.874434 \end{bmatrix}$$

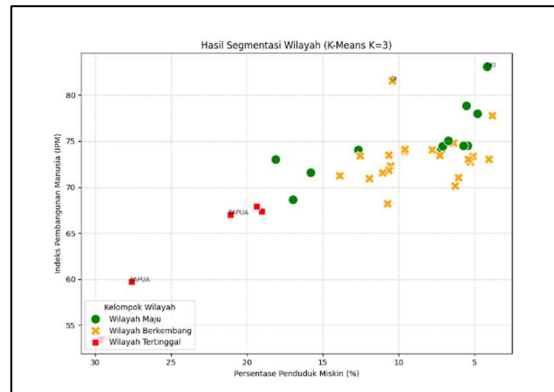
Model mencapai konvergensi dan berhenti total setelah 11 iterasi. Angka iterasi ini memvalidasi bahwa centroid awal, meskipun dipilih secara spesifik, memerlukan 11 langkah perhitungan ulang untuk mencapai posisi stabil (tidak ada lagi perubahan keanggotaan klaster), memastikan solusi klasterisasi final bersifat *robust*. Hasil akhir pengelompokan disajikan dalam Tabel 4 dan divisualisasikan pada Gambar 2.

**Tabel 4.** Hasil Clustering K-Means

Nama Provinsi	Persentase Penduduk Miskin	Indeks Pembangunan Manusia	Indeks TPT	Nama Cluster
<b>Aceh</b>	12.64	74.03	5.75	Wilayah Maju
<b>Sumatera Utara</b>	7.19	74.02	5.6	Wilayah Maju
<b>Sumatera Barat</b>	5.42	74.49	5.75	Wilayah Maju
<b>Riau</b>	6.36	74.79	3.7	Wilayah Berkembang
<b>Jambi</b>	7.26	73.43	4.48	Wilayah Berkembang
Dst..				

Tabel hasil klasterisasi ini menyajikan pengelompokan 38 provinsi di Indonesia ke dalam tiga kategori utama: "Wilayah Maju", "Wilayah Berkembang", dan "Wilayah Tertinggal". Kolom-kolom di dalam tabel memuat nama provinsi beserta nilai dari tiga indikator utama—Persentase Penduduk Miskin (PPM), Indeks Pembangunan Manusia (IPM), dan Indeks Tingkat Pengangguran Terbuka (TPT), serta nama klaster hasil algoritma K-Means. Pola yang terlihat mengkonfirmasi teori pembangunan wilayah yaitu provinsi dengan IPM tinggi dan PPM relatif rendah cenderung tergolong "Maju", sedangkan provinsi dengan IPM sangat rendah dan PPM tinggi, yang mayoritas berada di wilayah timur Indonesia,

masuk ke dalam kategori "Tertinggal". Anomali pada TPT juga terlihat jelas, di mana beberapa wilayah maju justru memiliki tingkat pengangguran yang lebih tinggi dibandingkan wilayah tertinggal yang memiliki TPT sangat rendah, sebuah indikasi kuat adanya perbedaan struktur pasar tenaga kerja formal dan informal antar wilayah.



**Gambar 2** Grafik Hasil K-Means Clustering

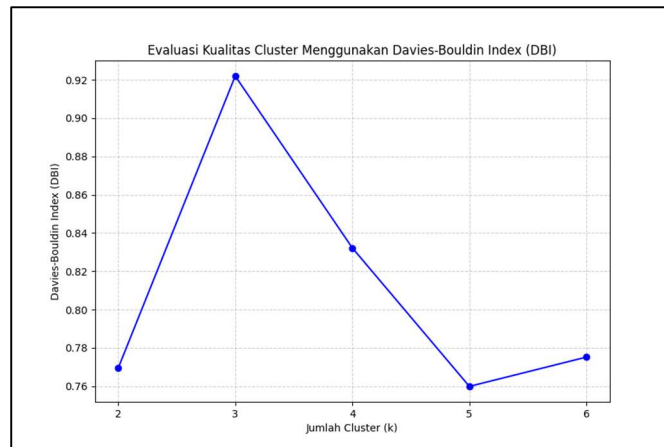
Grafik diatas memperlihatkan pemisahan 38 provinsi di Indonesia menjadi tiga klaster utama berdasarkan nilai Persentase Penduduk Miskin (sumbu X, terbalik) dan Indeks Pembangunan Manusia (sumbu Y). Titik-titik hijau yang berkumpul di kiri atas mewakili "Wilayah Maju" dengan karakteristik IPM tinggi ( $>70$ ) dan kemiskinan relatif rendah, didominasi oleh provinsi-provinsi di Jawa dan sebagian Sumatera. Titik-titik oranye di bagian tengah merupakan "Wilayah Berkembang", yang memiliki nilai IPM dan kemiskinan pada tingkat moderat. Sementara itu, titik-titik merah yang tersebar di kanan bawah merepresentasikan "Wilayah Tertinggal", ditandai dengan IPM rendah ( $<65$ ) dan tingkat kemiskinan yang sangat tinggi ( $>20\%$ ), mayoritas diisi oleh provinsi di wilayah Papua dan Nusa Tenggara Timur. Pola diagonal menurun dari kiri atas ke kanan bawah menegaskan korelasi negatif yang kuat antara pembangunan manusia dan kemiskinan: semakin tinggi pembangunan manusia suatu wilayah, semakin rendah tingkat kemiskinannya.

#### b. Evaluasi dengan Davies-Bouldin Index

Evaluasi Davies-Bouldin Index (DBI) dilakukan untuk mengukur kualitas klusterisasi, di mana nilai terendah menunjukkan pemisahan yang optimal. Hasil DBI disajikan pada Tabel 5 dan Gambar 3.

**Tabel 5** DBI Score

k	DBI Score
2	0.76939863
3	0.92194068
4	0.83224491
5	0.75993476
6	0.77523447



**Gambar 3** Grafik Evaluasi Davies-Bouldin Index

Berdasarkan hasil yang ditampilkan pada tabel dan grafik, Analisis DBI mengonfirmasi bahwa solusi numerik terbaik (nilai DBI terendah) dicapai pada  $k=5$  (0.7599). Namun, konfigurasi  $k=3$  (DBI= 0.9219) tetap dipilih dan divalidasi. Meskipun nilai DBI pada  $k=3$  sedikit lebih tinggi, solusi ini dianggap optimal secara kontekstual dan feasible karena menghasilkan kategorisasi kebijakan yang dapat diinterpretasikan secara jelas (Maju, Berkembang, Tertinggal), sesuai dengan tujuan penelitian.

### 3.2 Pembahasan

Pembahasan klusterisasi didasarkan pada Tabel 6 Karakteristik Rata-Rata Centroid yang diringkas dari hasil klusterisasi, yang menunjukkan profil unik setiap kelompok

**Tabel 6** Ringkasan Karakteristik Hasil Cluster

Nama Cluster	Rata-Rata IPM	Rata-Rata PPM	Rata-Rata TPT	Jumlah Provinsi
<b>Wilayah Maju</b>	74.96	9.17	6.10	12
<b>Wilayah Berkembang</b>	73.13	8.50	3.95	21
<b>Wilayah Tertinggal</b>	63.10	23.34	3.05	5

Klaster Wilayah Tertinggal dicirikan oleh IPM terendah (63.10) dan PPM tertinggi (23.34%), menegaskan posisi paling rentan dalam pembangunan. Klaster ini didominasi oleh provinsi di kawasan Timur Indonesia. Secara struktural, klaster ini menunjukkan Anomali TPT. Meskipun kemiskinan sangat tinggi, TPT rata-rata sangat rendah (3.05%). Fenomena ini secara akademis mengindikasikan dominasi sektor informal dan pertanian subsisten, menciptakan masalah pengangguran terselubung (*underemployment*) di mana penduduk tidak dihitung sebagai pengangguran namun memiliki produktivitas ekonomi yang minimal. Oleh karena itu, kebijakan harus fokus pada intervensi dasar untuk peningkatan IPM dan penciptaan lapangan kerja formal yang layak.

Klaster Wilayah Maju (IPM rata-rata= 74.96) dan Klaster Wilayah Berkembang (IPM rata-rata= 73.13) dipisahkan secara signifikan oleh variabel TPT. Klaster Maju memiliki TPT rata-rata tertinggi (6.10%), sejalan dengan wilayah yang ekonominya didominasi sektor formal dan padat modal, sering kali menghadapi tantangan pengangguran terdidik. Sementara Klaster Berkembang memiliki TPT yang lebih rendah (3.95%). Pemisahan kedua klaster ini oleh K-Means membuktikan peran variabel ketiga dalam mendefinisikan kelompok pembangunan struktural, di mana TPT menjadi *discriminator* utama antara dua kelompok ekonomi menengah-atas yang memiliki PPM yang relatif serupa.

Analisis mendalam terhadap Klaster Wilayah Maju menunjukkan variasi keanggotaan yang unik, termasuk provinsi seperti Maluku dan Papua yang terkluster pada hasil ini meskipun PPM relatif tinggi. Fenomena ini

disebabkan oleh peran dominan TPT dalam perhitungan jarak 3D. TPT tinggi dari Maluku dan Papua (6.11% dan 6.48%) menarik mereka ke centroid yang sama dengan provinsi maju lainnya (misalnya, DKI dan Banten yang juga memiliki TPT tinggi), mengabaikan perbedaan besar dalam PPM. Hal ini menyoroti bagaimana K-Means mengelompokkan wilayah berdasarkan pola struktural yang meliputi semua dimensi, bukan hanya PPM dan IPM, memberikan pandangan baru terhadap klasifikasi pembangunan regional.

Secara metodologis, konvergensi pada 11 iterasi menunjukkan stabilitas hasil klusterisasi. Evaluasi DBI mendukung penggunaan konfigurasi  $k=3$  sebagai alat diagnostik vital. Hasil klusterisasi ini menjadi dasar bagi pemerintah untuk menentukan prioritas belanja daerah dan intervensi sosial yang spesifik dan efektif, di mana program peningkatan IPM dan penanggulangan kemiskinan harus disesuaikan dengan status TPT masing-masing klaster.

## KESIMPULAN

Penelitian ini berhasil mengklasifikasikan 38 provinsi di Indonesia ke dalam tiga klaster pembangunan wilayah (Maju, Berkembang, Tertinggal) menggunakan algoritma K-Means yang didukung oleh Z-Score dan DBI. Konvergensi pada 11 iterasi menunjukkan stabilitas model dari inisiasi centroid yang spesifik.

Temuan kunci menunjukkan bahwa Klaster Tertinggal memiliki IPM dan PPM ekstrem, dengan TPT rendah yang mengonfirmasi masalah underemployment. Klaster Maju dan Berkembang dipisahkan oleh nilai TPT, menandakan perlunya kebijakan pasar kerja yang berbeda. Klasifikasi ini harus menjadi alat diagnostik bagi pemerintah untuk menentukan prioritas belanja daerah dan intervensi sosial yang spesifik dan efektif, dengan fokus utama pada peningkatan IPM dan penanggulangan kemiskinan struktural di klaster Tertinggal.

## REFERENSI

- [1] A. Bahauddin, A. Fatmawati, and F. P. Sari, "Analisis Clustering Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means," *J. Manaj. Inform. Dan Sist. Inf.*, vol. 4, no. 1, pp. 1–8.
- [2] R. Wahyuni, "K-Means clustering for grouping Indonesia underdeveloped regions in 2020 based on poverty indicators," *Param. J. Stat.*, vol. 2, no. 1, pp. 8–15.
- [3] G. R. Wahyudi, R. Rahmaddeni, E. Dini, S. Adrianto, and R. Fadila, "Clustering Regencies in Indonesia for Regional Mapping Using the K-Means Algorithm," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 3, pp. 1143–1151.
- [4] T. A. T, "Analisis klaster provinsi di indonesia berdasarkan akses dan infrastruktur listrik menggunakan algoritma," *J. Ilm. Kohesi*, vol. 8, no. 1.
- [5] H. A. Siahaan, "Penerapan Data Mining Untuk Clustering Indeks Pembangunan Manusia Berdasarkan Provinsi Di Indonesia BT - Prosiding Seminar Nasional Mahasiswa FTI," pp. 525–532.
- [6] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *J. Sains dan Manaj.*, vol. 9, no. 1.
- [7] N. N. Kaylista, N. Khoirunnisaa, G. V.E.N.F, and A. Y. P. Yusuf, "Implementasi Algoritma K-Means Clustering Menggunakan Aplikasi Orange Untuk Mengetahui Pola Indeks Pembangunan Manusia Tahun 2022," *J. Inf. Inf. Secur.*, vol. 4, no. 1, pp. 65–76.
- [8] Y. F. A, "Analisis Perbandingan Algoritma K-Means dan Hierarchial Clustering untuk Pengelompokan Data Penduduk Indeks Pembangunan Manusia pada Kecamatan Percut Sei Tuan," *Repos. UMA*.
- [9] A. Septianingsih, "ANALISIS K-MEANS CLUSTERING PADA PEMETAAN PROVINSI INDONESIA BERDASARKAN INDIKATOR RUMAH LAYAK HUNI," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 3, no. 1, pp. 224–241.

- 
- [10] T. E. F, “Eksplorasi Ketimpangan dan Klasterisasi Pembangunan Manusia di Provinsi Nusa Tenggara Timur,” *J. Stat. Terap.*, vol. 5, no. 1, pp. 75–83.
  - [11] L. Anggraeni and P. A. R, “Analisis Cluster Menggunakan Algoritma K-Means Pada Provinsi Sumatera Barat Berdasarkan Indeks Pembangunan Manusia Tahun 2021 BT - Prosiding Seminar Nasional UNIMUS,” pp. 636–646.
  - [12] D. Abdullah, A. S. Sunge, and A. T. Zy, “The application of K-means clustering for province classification based on COVID-19 cases in Indonesia,” *J. Comput. Networks, Arch. High Perform. Comput.*, vol. 6, no. 3, pp. 1055–1065.
  - [13] H. Mulyani, “Clustering area COVID-19 Indonesia with K-Means clustering method,” *J. Inf. Technol. Its Util.*, vol. 5, no. 2, pp. 42–50.
  - [14] M. S, “Pengelompokan Tingkat Kemiskinan di Provinsi Jawa Barat dengan Metode K-Means Clustering,” *J. Data Insights*, vol. 1, no. 2, pp. 54–62.