

---

## OPTIMALISASI VALIDITAS KLASTERISASI IPM MELALUI PENERAPAN VARIASI DISTANCE MEASURE PADA ALGORITMA K-MEANS++

Sardo Pardingotan Sipayung<sup>1</sup>, Syahril Efendi<sup>2</sup>

<sup>1,2</sup>Universitas Sumatera Utara, Medan

<sup>1</sup>Universitas Katolik Santo Thomas, Medan

e-mail: <sup>1</sup>pinsarsiphom@gmail.com <sup>2</sup>syahril1@usu.ac.id

**Abstract:** *The Human Development Index (HDI) is an important indicator for measuring the quality of regional development through the dimensions of health, education, and decent living standards. In North Sumatra Province, HDI achievements between districts/cities still show significant disparities, requiring a data-based analytical approach to map development patterns objectively. This study aims to optimize the validity of regional HDI clustering through the application of the K-Means++ algorithm with distance measure variations. This study uses a quantitative approach with an unsupervised learning method. The data analyzed includes HDI, Average Length of Schooling (ALS), and Adjusted Per Capita Expenditure sourced from the Central Statistics Agency. The research stages include data preprocessing and standardization, determining the optimal number of clusters using the Elbow method, applying the K-Means++ algorithm, and evaluating cluster quality using the Davies–Bouldin Index (DBI) and Purity Index. In addition, a comparison of clustering performance based on Euclidean, Manhattan, and Cosine distances was conducted. The results of the study show that the optimal number of clusters is three clusters representing high, medium, and low levels of human development. A DBI value of 0.60 and a Purity Index of 0.61 indicate good clustering quality. Euclidean and Manhattan distances produced the best performance compared to Cosine distance.*

**Keyword:** *Human Development Index; K-Means++; Clustering; Distance Measure; Davies–Bouldin Index; Purity Index.*

**Abstrak:** Indeks Pembangunan Manusia (IPM) merupakan indikator penting untuk mengukur kualitas pembangunan wilayah melalui dimensi kesehatan, pendidikan, dan standar hidup layak. Di Provinsi Sumatera Utara, capaian IPM antar kabupaten/kota masih menunjukkan ketimpangan yang cukup signifikan, sehingga diperlukan pendekatan analitis berbasis data untuk memetakan pola pembangunan secara objektif. Penelitian ini bertujuan untuk mengoptimalkan validitas klasterisasi IPM wilayah melalui penerapan algoritma K-Means++ dengan variasi distance measure. Penelitian ini menggunakan pendekatan kuantitatif dengan metode unsupervised learning. Data yang dianalisis meliputi IPM, Rata Lama Sekolah (RLS), dan Pengeluaran per Kapita Disesuaikan yang bersumber dari Badan Pusat Statistik. Tahapan penelitian mencakup praproses dan standarisasi data, penentuan jumlah klaster optimal menggunakan metode Elbow, penerapan algoritma K-Means++, serta evaluasi kualitas klaster menggunakan Davies–Bouldin Index (DBI) dan Purity Index. Selain itu, dilakukan perbandingan kinerja klasterisasi berdasarkan Euclidean, Manhattan, dan Cosine distance. Hasil penelitian menunjukkan bahwa jumlah klaster optimal adalah tiga klaster yang merepresentasikan tingkat pembangunan manusia tinggi, menengah, dan rendah. Nilai DBI sebesar 0,60 dan Purity Index sebesar 0,61 menunjukkan kualitas klasterisasi yang baik. Euclidean dan Manhattan distance menghasilkan performa terbaik dibandingkan Cosine distance.

**Kata kunci:** Indeks Pembangunan Manusia; K-Means++; Klasterisasi; Distance Measure; Davies–Bouldin Index; Purity Index.

## PENDAHULUAN

Seluruh Indeks Pembangunan Manusia (IPM) merupakan indikator komposit yang digunakan secara global untuk mengukur kualitas hidup masyarakat melalui dimensi kesehatan, pendidikan, dan standar hidup layak (Badan Pusat Statistik 2024). IPM berperan strategis dalam perencanaan dan evaluasi pembangunan nasional maupun daerah karena mencerminkan kemampuan suatu wilayah dalam meningkatkan kesejahteraan sumber daya manusianya. Namun demikian, perkembangan IPM tidak berlangsung secara merata di setiap wilayah Indonesia. Pada konteks Provinsi Sumatera Utara, kesenjangan pembangunan masih tampak jelas antara wilayah perkotaan seperti Kota Medan dan Deli Serdang dengan wilayah kepulauan dan perdesaan seperti Nias, Nias Utara, dan Mandailing Natal (Badan Pusat Statistik 2024). Kondisi ini menunjukkan perlunya pemetaan pembangunan berbasis data untuk mendukung perumusan kebijakan yang lebih inklusif dan berkeadilan.

Dalam bidang analisis data, klasterisasi merupakan salah satu pendekatan yang efektif untuk mengidentifikasi kesamaan karakteristik antarwilayah melalui metode unsupervised learning. Pendekatan ini banyak digunakan dalam kajian pembangunan karena mampu menyederhanakan kompleksitas data multidimensi menjadi kelompok yang mudah diinterpretasikan (Aggarwal 2005; To and Mining n.d.). Algoritma K-Means masih menjadi metode klasterisasi yang paling banyak digunakan karena efisiensi komputasi dan kemudahan implementasinya. Namun, berbagai studi terkini menunjukkan bahwa K-Means memiliki keterbatasan utama berupa sensitivitas terhadap pemilihan centroid awal dan ukuran jarak, sehingga berpotensi menghasilkan solusi lokal yang kurang optimal (Hasugian et al. 2021; Jain 2010). Untuk mengatasi keterbatasan tersebut, algoritma K-

Means++ dikembangkan dengan mekanisme inisialisasi centroid yang lebih terstruktur sehingga mampu meningkatkan stabilitas dan kualitas klaster yang dihasilkan. Sejumlah penelitian mutakhir membuktikan bahwa K-Means++ menghasilkan pemisahan klaster yang lebih baik dibandingkan K-Means konvensional, khususnya pada data berdimensi ganda (Huang, Dai, and Li 2023; Maharani, Juni Yanris, and Aini Nasution 2025). Selain itu, pemilihan distance measure seperti Euclidean, Manhattan, Minkowski, dan Cosine juga terbukti berpengaruh signifikan terhadap struktur dan validitas klaster yang terbentuk (Aggarwal 2005; Idrus et al. 2022). Oleh karena itu, pengujian kombinasi K-Means++ dengan variasi ukuran jarak menjadi penting untuk memperoleh hasil klasterisasi yang lebih representatif.

Sejumlah penelitian telah menerapkan teknik klasterisasi pada analisis pembangunan manusia. (Saraiva and Caiado 2025) melakukan klasterisasi pembangunan manusia secara global dan berhasil mengidentifikasi ketimpangan antarnegara, namun metode yang digunakan masih berbasis K-Means standar. Pada tingkat nasional, (Fahmiyah, Ningrum, and Airlangga 2023) melakukan klasterisasi IPM provinsi di Indonesia, tetapi belum melibatkan evaluasi indeks validitas klaster secara komprehensif. Di tingkat regional, (Satyahadewi, Sinaga, and Perdana 2023) menunjukkan adanya disparitas IPM yang signifikan antarwilayah di Provinsi Sumatera Utara, namun pendekatan klasterisasi yang digunakan belum mengoptimalkan inisialisasi centroid maupun variasi ukuran jarak.

Penelitian oleh (Idrus et al. 2022) menegaskan bahwa variasi distance measure dapat meningkatkan pemisahan klaster, tetapi belum mengombinasikannya dengan algoritma K-Means++. Sementara itu, (Arthur and Vassilvitskii 2007; Huang et al. 2023) membuktikan bahwa K-Means++ mampu

meningkatkan kualitas klaster pada data sektor publik, namun evaluasi yang dilakukan masih terbatas pada satu indikator validitas. Sintesis penelitian terdahulu tersebut menunjukkan bahwa belum terdapat kajian yang secara spesifik mengombinasikan K-Means++ dan variasi distance measure dengan evaluasi komprehensif menggunakan Purity Index dan Davies–Bouldin Index (DBI) pada data IPM Provinsi Sumatera Utara. Oleh karena itu, penelitian ini memiliki kontribusi ganda. Secara praktis, penelitian ini menyajikan hasil segmentasi pembangunan manusia pada tingkat kabupaten/kota di Provinsi Sumatera Utara yang dapat dijadikan dasar perumusan kebijakan pembangunan daerah yang lebih tepat sasaran. Secara metodologis, penelitian ini memperkuat kajian optimasi klasterisasi dengan menganalisis pengaruh variasi distance measure terhadap validitas hasil klasterisasi IPM menggunakan algoritma K-Means++ melalui pendekatan evaluasi yang lebih komprehensif.

Dengan demikian, penelitian ini berfokus pada optimalisasi validitas klasterisasi IPM di Provinsi Sumatera Utara melalui penerapan variasi distance measure pada algoritma K-Means++ serta evaluasi menggunakan indikator Purity dan Davies–Bouldin Index (DBI) guna menghasilkan pendekatan analitis berbasis data yang lebih akurat, relevan, dan bermanfaat bagi pembangunan daerah.

## METODE

Penelitian ini menggunakan pendekatan kuantitatif eksploratif dengan metode klasterisasi tidak terawasi (unsupervised learning) menggunakan algoritma K-Means++. Pendekatan ini dipilih karena penelitian bertujuan untuk mengidentifikasi pola alami dan kesamaan karakteristik pembangunan manusia antarwilayah tanpa memerlukan label kelas sebelumnya. Klasterisasi memungkinkan pengelompokan

kabupaten/kota berdasarkan kemiripan indikator pembangunan manusia secara objektif dan berbasis data.

Desain penelitian bersifat komputasional, di mana analisis dilakukan melalui tahapan pengolahan data numerik dan penerapan algoritma klasterisasi. Hasil klasterisasi kemudian dianalisis secara deskriptif untuk menginterpretasikan karakteristik masing-masing klaster dalam konteks pembangunan manusia. Dengan demikian, penelitian ini tidak hanya menghasilkan pembagian klaster secara matematis, tetapi juga memberikan pemahaman mengenai perbedaan tingkat pembangunan manusia antarwilayah di Provinsi Sumatera Utara.

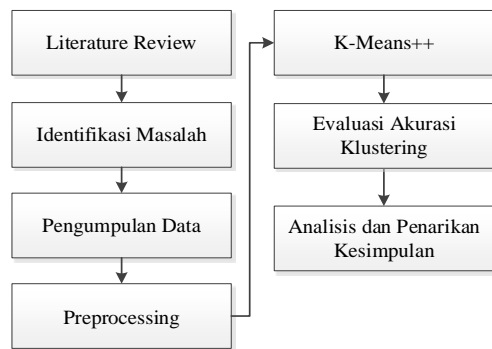
Data penelitian bersumber dari publikasi resmi Badan Pusat Statistik (BPS) yang mencakup seluruh kabupaten/kota di Provinsi Sumatera Utara. Data yang digunakan merupakan data sekunder terbaru yang mencerminkan kondisi pembangunan manusia wilayah secara aktual. Variabel penelitian dipilih berdasarkan relevansinya terhadap komponen utama Indeks Pembangunan Manusia, yaitu pendidikan dan standar hidup layak, serta telah dinormalisasi untuk memastikan keseragaman skala sebelum proses klasterisasi.

Variabel yang digunakan dalam penelitian ini disajikan pada Tabel 1.

**Tabel 1 Variabel Penelitian**

No	Atribut	Keterangan
1	Indeks Pembangunan Manusia (IPM)	Indikator komposit pembangunan manusia
2	Rata Lama Sekolah (RLS)	Rata-rata lama pendidikan penduduk (tahun)
3	Pengeluaran per Kapita	Indikator standar hidup layak (rupiah)

Kerangka kerja penelitian ditunjukkan pada Gambar 1.

**Gambar 1 Alur Penelitian**

Tahapan analisis dimulai dari pengumpulan data, praproses dan standarisasi data, penentuan jumlah kluster optimal menggunakan metode Elbow, penerapan algoritma K-Means++ dengan beberapa variasi distance measure, serta evaluasi kualitas kluster menggunakan Davies–Bouldin Index (DBI) dan Purity Index. Seluruh tahapan tersebut dilakukan untuk memastikan bahwa kluster yang dihasilkan memiliki kualitas yang baik dan dapat diinterpretasikan secara bermakna dalam konteks pembangunan manusia.

## HASIL DAN PEMBAHASAN

Bagian ini menjelaskan hasil penerapan algoritma K-Means++ pada data Indeks Pembangunan Manusia (IPM) kabupaten/kota di Provinsi Sumatera Utara. Pembahasan disusun secara bertahap sesuai dengan alur metode penelitian, mulai dari pengolahan data hingga penafsiran kluster yang terbentuk.

### Data Selection

Data yang digunakan dalam penelitian ini diperoleh dari Badan Pusat Statistik (BPS) dan mencakup 33 kabupaten/kota di Provinsi Sumatera Utara. Tiga variabel utama yang dianalisis adalah IPM, Rata Lama Sekolah (RLS), dan Pengeluaran per Kapita Disesuaikan. Ketiga variabel ini dipilih karena mewakili aspek pendidikan dan ekonomi yang menjadi komponen penting dalam pembangunan manusia. Dari data awal terlihat adanya perbedaan yang cukup jelas antarwilayah. Kota-kota besar

seperti Medan, Pematang Siantar, dan Binjai memiliki nilai IPM, RLS, dan pengeluaran yang tinggi. Sebaliknya, wilayah kepulauan seperti Nias Barat, Nias, dan Nias Utara menunjukkan nilai yang jauh lebih rendah. Kondisi ini menunjukkan bahwa pengelompokan wilayah berdasarkan kemiripan karakteristik pembangunan manusia perlu dilakukan. Untuk memberikan gambaran awal terhadap dataset yang digunakan, pada tabel 2 ditampilkan lima data pertama dari keseluruhan data penelitian

**Tabel 2 Data IPM Provinsi Sumatera Utara**

Wilayah	IPM	Rata Lama Sekolah (RLS)	Pengeluaran per Kapita Disesuaikan (Ribuan Rupiah)
Nias	0.0486	0.0000	0.0989
Mandailing Natal	0.2576	0.5000	0.4164
Tapanuli Selatan	0.4129	0.5830	0.5862
Tapanuli Tengah	0.3787	0.4649	0.4636
Tapanuli Utara	0.5837	0.7214	0.6170
Toba Samosir	0.6766	0.8063	0.6774
Labuhan Batu	0.5257	0.5720	0.5691
Asahan	0.4102	0.4520	0.5825
Simalungun	0.5694	0.6144	0.5773
Dairi	0.5113	0.6439	0.4937
...	...	...	...
Kota Gunungsitoli	0.3726	0.4354	0.2425

### Pre-processing

Sebelum dilakukan analisis, data terlebih dahulu diperiksa untuk memastikan kualitasnya. Pemeriksaan dilakukan terhadap kemungkinan adanya data kosong, data ganda, atau nilai yang tidak wajar. Hasil pemeriksaan menunjukkan bahwa data sudah lengkap dan konsisten. Seluruh variabel juga sudah berada pada skala yang sama (0–1), sehingga tidak diperlukan proses standarisasi tambahan. Dengan demikian,

data siap digunakan pada tahap klusterisasi.

### Penentuan Jumlah Klaster

Penentuan jumlah klaster yang optimal merupakan tahapan penting dalam proses klusterisasi karena akan memengaruhi kualitas pengelompokan wilayah. Pada penelitian ini, jumlah klaster ditentukan menggunakan metode Elbow dengan mengevaluasi nilai Within-Cluster Sum of Squares (WCSS) pada beberapa kemungkinan jumlah klaster. Hasil perhitungan nilai WCSS untuk jumlah klaster  $K = 1$  hingga  $K = 6$  disajikan pada Tabel 1. Berdasarkan tabel tersebut, terlihat bahwa nilai WCSS mengalami penurunan yang sangat tajam dari  $K = 1$  (5,2849) ke  $K = 2$  (2,4674) dan kembali menurun secara signifikan pada  $K = 3$  (1,0289). Penurunan ini menunjukkan bahwa pembentukan klaster hingga  $K = 3$  mampu mengurangi variasi dalam klaster secara substansial. Namun, setelah  $K > 3$ , penurunan nilai WCSS menjadi semakin kecil dan cenderung melandai, sebagaimana ditunjukkan pada nilai  $K = 4$  hingga  $K = 6$ . Kondisi ini mengindikasikan bahwa penambahan jumlah klaster di atas tiga tidak lagi memberikan peningkatan kualitas pengelompokan yang berarti. Berdasarkan pola penurunan nilai WCSS yang ditampilkan pada Tabel 3, titik tekuk (elbow point) secara jelas berada pada  $K = 3$ . Oleh karena itu, jumlah klaster optimal dalam penelitian ini ditetapkan sebanyak tiga klaster, yang selanjutnya digunakan pada proses klusterisasi menggunakan algoritma K-Means++.

**Tabel 3 Hasil Penentuan Jumlah Klaster Menggunakan Metode Elbow**

Jumlah Klaster (K)	Nilai WCSS	Interpretasi
1	5,2849	Variasi sangat tinggi, seluruh data berada dalam satu klaster
2	2,4674	Penurunan WCSS sangat signifikan,

		struktur awal klaster mulai terbentuk
3	1,0289	Titik tekuk (elbow point), keseimbangan terbaik
4	0,7137	Penurunan mulai melandai, peningkatan kualitas kecil
5	0,4372	Perbaikan klaster tidak signifikan
6	0,3523	Model mulai stabil (over-segmentation)

### Hasil Klusterisasi Menggunakan K-Means++

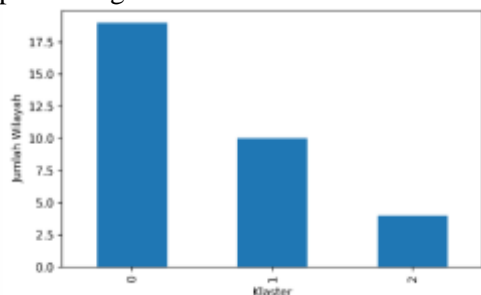
Setelah jumlah klaster optimal ditetapkan sebanyak tiga klaster ( $K = 3$ ), tahap selanjutnya adalah penerapan algoritma K-Means++ pada data IPM kabupaten/kota di Provinsi Sumatera Utara. Algoritma ini digunakan untuk mengelompokkan wilayah berdasarkan kemiripan nilai IPM, Rata Lama Sekolah (RLS), dan Pengeluaran per Kapita Disesuaikan. Hasil proses klusterisasi menunjukkan bahwa 33 wilayah berhasil dikelompokkan ke dalam tiga klaster dengan jumlah anggota yang berbeda pada setiap klaster. Distribusi jumlah wilayah pada masing-masing klaster disajikan pada Tabel 2.

**Tabel 4 Distribusi Jumlah Wilayah Hasil Klusterisasi**

Klaster	Jumlah Wilayah	Persentase (%)	Karakteristik Umum
K 0	19	57,6 %	IPM Menengah
K 1	10	30,3 %	IPM Tinggi
K 2	4	12,1 %	IPM Rendah
Total	33	100 %	—

Untuk memperjelas hasil pengelompokan tersebut, grafik distribusi jumlah wilayah per klaster ditampilkan pada Gambar 2. Grafik ini menunjukkan bahwa Klaster 0 merupakan kelompok dengan jumlah wilayah terbanyak, diikuti

oleh Klaster 1, sedangkan Klaster 2 memiliki jumlah wilayah paling sedikit. Distribusi ini mengindikasikan bahwa sebagian besar wilayah di Sumatera Utara berada pada kategori pembangunan manusia menengah, sementara hanya sebagian kecil wilayah yang tergolong pada kategori rendah.



**Gambar 2 Distribusi Jumlah Wilayah**

Selanjutnya, hasil klasterisasi divisualisasikan dalam grafik sebar (scatter plot) antara nilai IPM dan Pengeluaran per Kapita, sebagaimana ditampilkan pada Gambar 3. Pada grafik tersebut terlihat pemisahan klaster yang cukup jelas. Klaster 1 terkonsentrasi pada bagian kanan atas grafik, yang menunjukkan wilayah dengan IPM dan pengeluaran per kapita tinggi. Klaster 0 berada di area tengah grafik, mencerminkan wilayah dengan tingkat pembangunan manusia menengah. Sementara itu, Klaster 2 berada pada bagian kiri bawah grafik, yang menandakan wilayah dengan IPM dan pengeluaran per kapita rendah.

Pola pengelompokan yang tampak pada grafik ini memperlihatkan bahwa algoritma K-Means++ mampu mengidentifikasi struktur alami data dengan baik. Wilayah-wilayah yang memiliki karakteristik pembangunan manusia serupa cenderung berkumpul dalam klaster yang sama, sementara perbedaan antar klaster terlihat cukup kontras. Hal ini menunjukkan bahwa hasil klasterisasi tidak hanya valid secara numerik, tetapi juga mudah diinterpretasikan secara visual.

#### Evaluasi Kualitas Klaster

Evaluasi kualitas klaster

dilakukan untuk menilai sejauh mana hasil klasterisasi yang dihasilkan oleh algoritma K-Means++ memiliki tingkat kekompakan internal dan pemisahan antar klaster yang baik. Evaluasi ini penting untuk memastikan bahwa klaster yang terbentuk tidak hanya valid secara numerik, tetapi juga layak diinterpretasikan sebagai representasi kondisi pembangunan manusia antarwilayah. Pada penelitian ini, evaluasi kualitas klaster dilakukan menggunakan dua metrik utama, yaitu Davies–Bouldin Index (DBI) dan Purity Index. DBI digunakan sebagai ukuran validitas internal klaster yang mengukur rasio antara jarak dalam klaster dan jarak antar klaster, di mana nilai yang lebih kecil menunjukkan kualitas klaster yang lebih baik. Sementara itu, Purity Index digunakan untuk mengukur tingkat homogenitas anggota klaster berdasarkan kategori IPM. Hasil evaluasi kualitas klaster secara keseluruhan disajikan pada Tabel 5.

**Tabel 5 Hasil Evaluasi Kualitas Klaster**

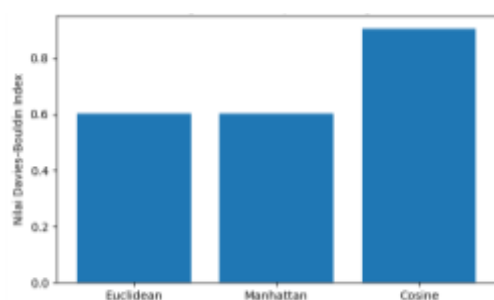
Indikator Evaluasi	Nilai	Interpretasi
Davies–Bouldin Index (DBI)	0,60	Klaster kompak dan terpisah dengan baik
Purity Index	0,61	Konsistensi klaster tergolong sedang–baik
Jumlah Klaster	3	Optimal (berdasarkan Elbow Method)

Berdasarkan Tabel 5, nilai Davies–Bouldin Index (DBI) sebesar 0,60 menunjukkan bahwa klaster yang terbentuk memiliki kekompakan internal yang baik dan pemisahan antar klaster yang jelas. Nilai DBI di bawah 1 mengindikasikan bahwa algoritma K-Means++ mampu menghasilkan struktur klaster yang baik pada data sosial-ekonomi seperti Indeks Pembangunan Manusia. Sementara itu, nilai Purity Index sebesar 0,61 menunjukkan bahwa klaster cukup konsisten dalam merepresentasikan

kelompok wilayah dengan tingkat pembangunan manusia yang serupa, meskipun masih terdapat wilayah transisi. Selain itu, dilakukan perbandingan nilai DBI pada berbagai distance measure, yaitu Euclidean, Manhattan, dan Cosine, dengan jumlah kluster tetap  $K = 3$ . Hasil perbandingan tersebut disajikan pada Tabel 6 dan divisualisasikan pada Gambar 3.

**Tabel 6. Perbandingan Nilai Davies–Bouldin Index (DBI) pada Berbagai Distance Measure.**

Distance Measure	Nilai DBI	Kualitas Kluster
Euclidean	0,60	Baik
Manhattan	0,60	Baik
Cosine	0,91	Cukup



**Gambar 3 Perbandingan Nilai Davies–Bouldin Index pada Berbagai Distance Measure**

#### Analisis Karakteristik Kluster

Analisis karakteristik kluster dilakukan untuk memahami ciri utama yang membedakan setiap kelompok wilayah hasil klusterisasi menggunakan algoritma K-Means++. Analisis ini berfokus pada nilai rata-rata indikator Indeks Pembangunan Manusia (IPM), Rata Lama Sekolah (RLS), dan Pengeluaran per Kapita Disesuaikan, sehingga dapat menggambarkan profil pembangunan manusia yang khas pada masing-masing kluster.

#### Ringkasan Karakteristik Setiap Kluster

Ringkasan nilai rata-rata indikator pada masing-masing kluster disajikan pada Tabel 7.

**Tabel 7 Rata-rata Indikator IPM per**

#### Kluster

K	Rata-rata IPM	Rata-rata RLS	Rata-rata Pengeluaran	Kategori Umum Pembangunan
K0	Sedang	Sedang	Sedang	Menengah
K1	Tinggi	Tinggi	Tinggi	Tinggi
K2	Rendah	Rendah	Rendah	Rendah

Tabel 7 menunjukkan bahwa setiap kluster memiliki perbedaan karakteristik yang cukup jelas, baik dari sisi pendidikan, ekonomi, maupun capaian pembangunan manusia secara keseluruhan. Perbedaan ini menjadi dasar dalam interpretasi lebih lanjut terhadap masing-masing kluster.

#### Karakteristik Kluster 0

##### (Pembangunan Manusia Menengah)

Kluster 0 merupakan kluster dengan jumlah wilayah terbanyak. Wilayah-wilayah dalam kluster ini memiliki nilai IPM, RLS, dan pengeluaran per kapita yang berada pada kategori menengah. Kondisi ini menunjukkan bahwa pembangunan manusia di wilayah tersebut telah berjalan cukup baik, namun belum mencapai tingkat optimal. Dari sisi pendidikan, nilai RLS pada kluster ini mencerminkan bahwa rata-rata penduduk telah menempuh pendidikan hingga jenjang menengah. Sementara itu, tingkat pengeluaran per kapita yang moderat mengindikasikan kemampuan ekonomi masyarakat yang relatif stabil, meskipun masih terbatas untuk mendorong percepatan peningkatan kualitas hidup secara signifikan. Kluster ini merepresentasikan wilayah-wilayah yang berpotensi untuk ditingkatkan melalui kebijakan penguatan kualitas pendidikan dan peningkatan daya beli masyarakat.

#### Karakteristik Kluster 1

##### (Pembangunan Manusia Tinggi)

Kluster 1 terdiri dari wilayah-wilayah dengan capaian pembangunan manusia tertinggi. Kluster ini ditandai oleh nilai IPM yang tinggi, didukung oleh rata-rata RLS yang lebih panjang serta

tingkat pengeluaran per kapita yang relatif besar. Tingginya nilai RLS menunjukkan akses pendidikan yang lebih baik dan berkelanjutan, sementara pengeluaran per kapita yang tinggi mencerminkan kondisi ekonomi masyarakat yang lebih mapan. Kombinasi kedua faktor ini berkontribusi secara langsung terhadap capaian IPM yang tinggi. Wilayah-wilayah dalam klaster ini dapat dipandang sebagai pusat pertumbuhan pembangunan manusia dan berpotensi menjadi model atau rujukan bagi wilayah lain dalam perumusan kebijakan pembangunan daerah.

### **Karakteristik Klaster 2 (Pembangunan Manusia Rendah)**

Klaster 2 merupakan klaster dengan jumlah wilayah paling sedikit, namun memiliki tingkat pembangunan manusia paling rendah. Wilayah dalam klaster ini dicirikan oleh nilai IPM, RLS, dan pengeluaran per kapita yang relatif rendah dibandingkan klaster lainnya. Rendahnya nilai RLS menunjukkan keterbatasan akses dan keberlanjutan pendidikan, sementara rendahnya pengeluaran per kapita mengindikasikan keterbatasan kemampuan ekonomi masyarakat. Kondisi ini berdampak langsung pada capaian IPM yang masih tertinggal. Klaster ini mencerminkan wilayah-wilayah yang memerlukan perhatian dan intervensi pembangunan yang lebih intensif, terutama pada sektor pendidikan dan peningkatan kesejahteraan ekonomi masyarakat.

### **Implikasi Analisis Klaster**

Secara keseluruhan, hasil analisis karakteristik klaster menunjukkan adanya ketimpangan tingkat pembangunan manusia antarwilayah di Provinsi Sumatera Utara. Klasterisasi berhasil mengelompokkan wilayah berdasarkan kesamaan karakteristik pembangunan, sehingga memberikan gambaran yang lebih terstruktur mengenai kondisi IPM regional. Hasil ini dapat dimanfaatkan sebagai dasar dalam penyusunan kebijakan pembangunan yang lebih tepat sasaran, di mana strategi intervensi dapat

disesuaikan dengan karakteristik masing-masing klaster, baik untuk mempertahankan capaian pembangunan yang sudah tinggi maupun untuk mempercepat peningkatan pembangunan pada wilayah yang masih tertinggal.

### **SIMPULAN**

Penelitian ini bertujuan untuk mengoptimalkan proses klasterisasi wilayah berdasarkan indikator Indeks Pembangunan Manusia (IPM) di Provinsi Sumatera Utara dengan menerapkan algoritma K-Means++. Klasterisasi dilakukan menggunakan variabel IPM, Rata Lama Sekolah (RLS), dan Pengeluaran per Kapita Disesuaikan, yang telah melalui tahap praproses dan standarisasi data untuk memastikan keseragaman skala dan kualitas analisis. Berdasarkan hasil penentuan jumlah klaster menggunakan metode Elbow, diperoleh jumlah klaster optimal sebanyak tiga klaster. Penerapan algoritma K-Means++ dengan jumlah klaster tersebut mampu mengelompokkan 33 kabupaten/kota ke dalam kelompok wilayah dengan karakteristik pembangunan manusia yang berbeda, yaitu klaster pembangunan manusia tinggi, menengah, dan rendah. Visualisasi hasil klasterisasi menunjukkan pemisahan klaster yang cukup jelas, baik secara kuantitatif maupun visual.

Evaluasi kualitas klaster menunjukkan bahwa hasil klasterisasi memiliki tingkat validitas yang baik. Nilai Davies–Bouldin Index (DBI) sebesar 0,60 mengindikasikan bahwa klaster yang terbentuk memiliki kekompakan internal yang baik serta pemisahan antar klaster yang jelas. Sementara itu, nilai Purity Index sebesar 0,61 menunjukkan bahwa klaster yang dihasilkan cukup konsisten dalam merepresentasikan kelompok wilayah dengan kategori IPM yang serupa, meskipun masih terdapat beberapa wilayah transisi. Analisis karakteristik klaster mengungkap adanya ketimpangan pembangunan manusia antarwilayah di



Provinsi Sumatera Utara. Klaster pembangunan manusia tinggi didominasi oleh wilayah perkotaan dengan capaian pendidikan dan tingkat pengeluaran yang relatif lebih baik. Klaster menengah mencerminkan wilayah dengan potensi pengembangan yang masih terbuka, sedangkan klaster rendah menunjukkan wilayah-wilayah yang memerlukan perhatian dan intervensi pembangunan yang lebih intensif, khususnya pada sektor pendidikan dan peningkatan kesejahteraan ekonomi. Secara keseluruhan, penelitian ini menunjukkan bahwa algoritma K-Means++ efektif digunakan untuk memetakan pola pembangunan manusia antarwilayah. Hasil klasterisasi yang diperoleh dapat dijadikan sebagai dasar pendukung pengambilan kebijakan pembangunan daerah yang lebih tepat sasaran, terutama dalam penentuan prioritas pembangunan berdasarkan karakteristik masing-masing klaster.

#### DAFTAR PUSTAKA

- Aggarwal, Charu C. 2005. "Data Mining Data Mining." *Mining of Massive Datasets* 2(January 2013):5–20. [https://www.cambridge.org/core/product/identifier/CBO9781139058452/A007/type/book\\_part](https://www.cambridge.org/core/product/identifier/CBO9781139058452/A007/type/book_part).
- Arthur, David, and Sergei Vassilvitskii. 2007. *K-Means++: The Advantages of Careful Seeding*. Vol. 8.
- Badan Pusat Statistik. 2024. "Berita Resmi Statistik Indeks Pembangunan Manusia 2023." 18:1–282.
- Fahmiyah, Indah, Ratih Ardiati Ningrum, and Universitas Airlangga. 2023. "Journal+5 (1)." 02(01):27–33.
- Hasugian, Paska Marto, Bosker Sinaga, Jonson Manurung, and Safa Ayoub Al Hashim. 2021. "Best Cluster Optimization with Combination of K-Means Algorithm And Elbow Method Towards Rice Production Status Determination." *International Journal of Artificial Intelligence Research* 5(1):102–10. doi:10.29099/ijair.v6i1.232.
- Huang, Jiale, Jingtong Dai, and Yanjin Li. 2023. "Research on PCA-Kmeans++ Clustering Algorithm Considering Spatiotemporal Dimension." Pp. 195–201 in *2023 2nd International Conference on 3D Immersion, Interaction and Multi-sensory Experiences (ICDIIME)*.
- Idrus, Ali, Nafan Tarihoran, Ucup Supriatna, Ahmad Tohir, Suwarni Suwarni, and Robbi Rahim. 2022. "Distance Analysis Measuring for Clustering Using K-Means and Davies Bouldin Index Algorithm." *TEM Journal* 11(4):1871–76. doi:10.18421/TEM114-55.
- Jain, Anil K. 2010. "Data Clustering: 50 Years beyond K-Means." *Pattern Recognition Letters* 31(8):651–66. doi:<https://doi.org/10.1016/j.patrec.2009.09.011>.
- Maharani, Aulia, Gomal Juni Yanris, and Fitri Aini Nasution. 2025. "Implementation of the K-Means Clustering Method in Clustering Poor Population in Bandar Kumbul Village, Labuhanbatu Regency." *International Journal of Science, Technology & Management* 6(1):248–56. doi:10.46729/ijstm.v6i1.1209.
- Saraiva, Carolina, and Jorge Caiado. 2025. "Global Development Patterns: A Clustering Analysis of Economic, Social and Environmental Indicators." *Sustainable Futures* 10(June):100907. doi:10.1016/j.sftr.2025.100907.
- Satyahadewi, Neva, Steven Jansen Sinaga, and Hendra Perdana. 2023. "Hierarchical Cluster Analysis of Districts/Cities in North Sumatra Province Based on Human Development Index Indicators Using Pseudo-F." *Barekeng* 17(3):1429–38. doi:10.30598/barekengvol17iss3pp1429-1438.
- To, Introduction, and Data Mining. n.d. *Data Mining Tan*.