

Implementasi K-Means Mengelompokkan Kabupaten/Kota Berdasarkan Faktor Sosial-Ekonomi untuk Prioritas Alokasi Bantuan Sumatera Selatan 2023

Asa Do'a Uyi¹, Siti Nur Aarifah¹, Dwi Ratna Anggraeni¹, M. Syamsuddin Wisnubroto^{*1}, Fajri Farid¹

¹Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera, Jl. Terusan Ryacudu, Way Huwi, Kec. Jati Agung, Kab. Lampung Selatan, Lampung, Indonesia

ARTICLE INFO

Keywords:

K-Means Clustering, Human Development Index, Socio-Economic, South Sumatra, Aid Allocation

Received: November 25, 2025

Revised: December 17, 2025

Accepted: April 19, 2026

*Corresponding author:

E-mail: syamsuddin.wisnubroto@sd.itera.ac.id

DOI: [10.37253/telcomatics.v1i1.11546](https://doi.org/10.37253/telcomatics.v1i1.11546)

ABSTRAK

Pembangunan manusia merupakan salah satu indikator utama keberhasilan pembangunan suatu daerah yang menjadi dasar perencanaan pembangunan berkelanjutan. Penelitian ini bertujuan mengelompokkan kabupaten/kota di Provinsi Sumatera Selatan berdasarkan indikator sosial-ekonomi yang membentuk Indeks Pembangunan Manusia (IPM) dengan menerapkan algoritma *K-Means Clustering*. Data yang digunakan berasal dari Badan Pusat Statistik (BPS) tahun 2023. Kualitas hasil klusterisasi dievaluasi melalui *Davies-Bouldin Index* dan *Silhouette Score* untuk menilai tingkat pemisahan antar klaster dan konsistensi data. Hasil analisis menunjukkan terbentuknya tiga kelompok wilayah rendah, sedang, dan tinggi dengan distribusi masing-masing 3, 13, dan 1 daerah. Temuan ini mengindikasikan masih adanya ketimpangan kualitas pembangunan manusia di Sumatera Selatan, terutama pada aspek pendidikan dan pengeluaran riil per kapita. Nilai *Davies-Bouldin Index* sebesar 0,9698 dan *Silhouette Score* sebesar 0,3313 menunjukkan bahwa hasil pengelompokan cukup baik dan dapat digunakan sebagai acuan dalam penentuan prioritas alokasi bantuan. Dengan demikian, penerapan algoritma *K-Means* dapat membantu pemerintah daerah dalam memetakan kondisi pembangunan manusia secara objektif dan mendukung pengambilan kebijakan yang lebih tepat sasaran.

I. PENDAHULUAN

Indeks Pembangunan Manusia (IPM) merupakan indikator komposit yang digunakan untuk menilai sejauh mana kemajuan pembangunan manusia di suatu wilayah, mencakup dimensi kesehatan, pendidikan, dan ekonomi. Secara operasional, IPM dihitung melalui empat indikator utama, yaitu angka harapan hidup (AHH), harapan lama sekolah (HLS), rata-rata lama sekolah (RLS), serta pengeluaran per kapita yang telah disesuaikan [1][2]. Indeks ini berfungsi tidak hanya sebagai alat ukur keberhasilan pembangunan, tetapi juga sebagai dasar bagi pemerintah dalam menyusun perencanaan strategis, evaluasi kebijakan, hingga alokasi anggaran termasuk penentuan Dana Alokasi Umum (DAU) di Indonesia [3].

Berbagai penelitian sebelumnya menegaskan bahwa analisis IPM diperlukan untuk memahami ketimpangan pembangunan antar daerah. Studi-studi tersebut menunjukkan bahwa perbedaan capaian IPM sering kali mencerminkan variasi kondisi sosial-ekonomi, ketersediaan layanan publik, serta efektivitas intervensi pembangunan [4]. Pembangunan nasional pada dasarnya memiliki tujuan untuk meningkatkan kesejahteraan masyarakat secara adil dan merata [5]. Namun demikian, sebagian besar kajian masih terbatas pada analisis deskriptif tanpa mempertimbangkan kemungkinan pengelompokan wilayah berdasarkan karakteristik pembangunan yang serupa. Padahal, pendekatan pengelompokan (*clustering*) dapat memberikan gambaran

yang lebih komprehensif mengenai pola pembangunan regional sebagai dasar perumusan kebijakan yang lebih tepat sasaran.

Sumatera Selatan merupakan contoh wilayah yang menunjukkan variasi capaian IPM antar Kabupaten/Kota. Pada tahun 2023, sejumlah daerah telah mencapai indikator kesehatan dan pendidikan yang relatif baik, sementara daerah lainnya masih berjuang meningkatkan kualitas layanan dasar dan ketahanan ekonomi masyarakat [6]. Ketimpangan ini menegaskan perlunya analisis berbasis data untuk mengelompokkan daerah sesuai karakteristik pembangunannya, sehingga pemerintah dapat menentukan prioritas intervensi secara lebih efektif.

Untuk menjawab kebutuhan tersebut, penelitian ini menerapkan algoritma K-Means dalam pengelompokan kabupaten/kota di Provinsi Sumatera Selatan berdasarkan indikator-indikator penentu IPM, yaitu AHH, HLS, RLS, dan pengeluaran per kapita [7]. Pendekatan clustering ini diharapkan dapat memberikan pemahaman baru tentang struktur pembangunan manusia di tingkat regional serta membantu pemerintah dalam merancang kebijakan yang lebih terarah. Penelitian ini bertujuan:

- (1) mengidentifikasi kelompok wilayah berdasarkan tingkat pembangunan manusianya,
- (2) menilai variasi capaian IPM antar kelompok, dan
- (3) menegaskan kembali pentingnya IPM sebagai data strategis dalam perumusan kebijakan publik dan alokasi anggaran, termasuk DAU [8].

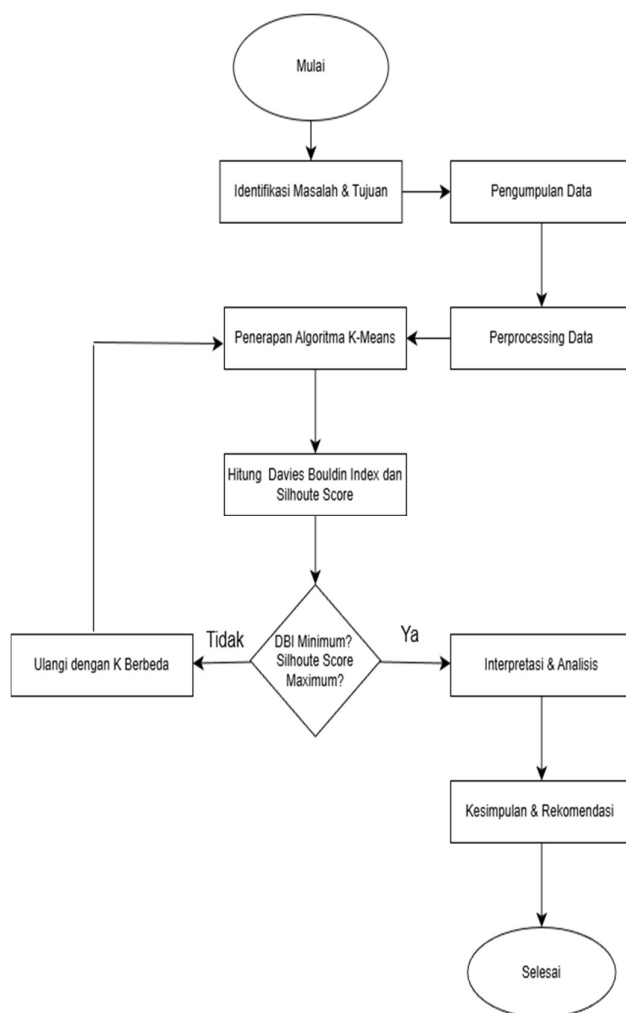
Secara praktis, penelitian ini memberikan kontribusi bagi pemerintah daerah dalam pengambilan keputusan berbasis bukti (*evidence-based policy*), sementara secara akademik, penelitian ini memperluas pemanfaatan metode data mining dalam analisis pembangunan regional. Artikel ini disusun melalui paparan metode penelitian, hasil analisis dan pembahasan, serta ditutup dengan kesimpulan dan implikasi kebijakan.

II. METODE PENELITIAN

Penelitian ini menggunakan metode *K-Means clustering* untuk mengelompokkan kabupaten/kota di Provinsi Sumatera Selatan berdasarkan indikator sosial-ekonomi. Metode ini dipilih karena dapat mengelompokkan data dengan karakteristik yang serupa secara otomatis.

A. Tahapan Penelitian

Penelitian ini dilakukan berdasarkan tahapan penelitian yang terdiri dari langkah-langkah yang terstruktur yang digambarkan melalui *flowchart* pada Gambar 1.



Gambar 1. Tahapan Penelitian

B. Data

Data yang digunakan pada penelitian ini diperoleh dari laman resmi Badan Pusat Statistik (BPS) Indonesia. Penelitian ini menggunakan data yang dikumpulkan dari seluruh kabupaten di Provinsi Sumatera Selatan sejumlah 10 kabupaten/kota dengan lima variabel atau lima indikator. Indikator yang digunakan yaitu Angka Harapan Lama Sekolah, Rata-Rata Lama Sekolah, Rata-Rata Pengeluaran per Kapita, Rata-rata Upah/Gaji Bersih Pekerja Informal, Umur Harapan Hidup.

C. Algoritma K-Means

K-Means Clustering merupakan salah satu algoritma dalam unsupervised learning yang digunakan untuk mengelompokkan sekumpulan data berdasarkan persamaan karakteristiknya. Proses klusterisasi pada algoritma ini dijelaskan dalam langkah-langkah berikut [8]:

- Menentukan jumlah kluster, langkah pertama yaitu menentukan berapa banyak kluster.
- Menghitung jarak setiap data ke pusat kluster.

$$d(x_i, c_j) = \sqrt{\sum_{m=1}^n (x_{im} - c_{jm})^2} \quad (1)$$

Pada persamaan (1) untuk menghitung jarak kluster dengan pusat kluster dengan $d(x_i, c_j)$ adalah jarak titik data ke *centroid*, x_{im} adalah nilai fitur m dari titik data x_i , dan c_{jm} adalah nilai fitur m dari titik data c_j .

- Mengelompokkan data ke kluster terdekat.

$$Cluster(x_i) = \operatorname{argmin}_{\{j\}} d(x_i, c_j) \quad (2)$$

Pada persamaan (2), $Cluster(x_i) = \operatorname{argmin}_{\{j\}} d(x_i, c_j)$ menjelaskan tahapan untuk mengelompokkan data ke kluster terdekat dalam algoritma K-Means. Rumus ini berfungsi untuk menetapkan setiap titik data (x_i) ke kluster yang memiliki jarak paling minimum ke pusat kluster (*centroid*). Di sini, x_i adalah titik data yang akan dikelompokkan, dan c_j adalah *centroid* dari kluster ke- j yang menjadi pembanding. Fungsi $\operatorname{argmin}_{\{j\}}$ secara spesifik bertugas untuk menentukan kluster (j) mana yang memberikan jarak terpendek (d), sehingga data dikelompokkan secara optimal.

- Memperbarui posisi pusat kluster (*centroid*).

$$c_j = \left(\frac{1}{n_j} \right) \sum_{x \in Cluster\ j} x \quad (3)$$

Dari persamaan (3), dengan c_j adalah *centroid* baru untuk kluster j , n_j adalah jumlah titik data pada kluster j , $x \in cluster\ j$ adalah titik data termasuk ke dalam kluster j .

- Mengulangi proses hingga hasil clustering sama dengan kluster sebelumnya.

D. Davies-Bouldin Index

David L dan Donald Bouldin memperkenalkan metode *Davies-Bouldin Index* (DBI) pada tahun 1979 sebagai rasio

rata-rata jarak dalam dan antar kluster untuk setiap kluster dengan kluster tetangga terdekatnya. DBI digunakan dalam mengukur mengevaluasi kinerja pengelompokkan dengan cara pengelompokkan validasi yang berisi dua kategori baik eksternal dan internal [9]. DBI dihitung dengan persamaan sebagai berikut.

$$DB = \left(\frac{l}{2}\right) \sum_{i=1}^N \max_{j \neq i} (R_{ij}) \quad (4)$$

Dengan,

$$R_{ij} = \frac{s_i + s_j}{d_{ij}} \quad (5)$$

Davies-Bouldin Index (DB), yang didefinisikan secara keseluruhan dalam persamaan (4), adalah metrik evaluasi klusterisasi yang mengukur kualitas pemisahan kluster, di mana N merupakan jumlah kluster. Nilai DB ini merupakan rata-rata dari $\max_j$ dari rasio R_{ij} . Rasio R_{ij} dalam persamaan (5) itu sendiri dihitung dari perbandingan antara penjumlahan jarak rata-rata titik data ke *centroid* di kluster i dan j (s_i dan s_j). Nilai DB yang semakin rendah menunjukkan hasil klusterisasi yang semakin optimal karena kluster yang terbentuk padat dan terpisah dengan baik.

E. Silhouette Score

Kualitas dan kekuatan suatu kluster dapat ditentukan menggunakan metode *Silhouette Coefficient*. Peran *Euclidean distance* Sangat berpengaruh dalam mengukur baik tidaknya suatu objek dalam sebuah kluster. Kualitas dan Kekuatan objek dilihat dari seberapa dekat jarak antara objek dalam sebuah kluster [10].

Tahapan untuk menghitung nilai *Silhouette Coefficient* adalah sebagai berikut.

- a) Menghitung jarak rata-rata data pada kluster yang sama $a(i)$.

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (6)$$

Persamaan (6) digunakan untuk menghitung $a(i)$, yaitu nilai yang mengukur seberapa baik suatu objek (titik data i) dikelompokkan ke dalam klasternya sendiri (A). Nilai $a(i)$ didefinisikan sebagai jarak rata-rata antara titik data i dengan semua titik data lain (j) yang berada dalam kluster yang sama. Dalam rumus ini, j anggota A dan j tidak sama dengan i , berarti perhitungan hanya mencakup data lain di kluster A , dan $|A|$ adalah jumlah total data di kluster A . Semakin kecil nilai $a(i)$, semakin tinggi tingkat kekompakan (kohesivitas) titik data i di dalam klasternya.

- b) Menghitung rata-rata jarak dari data i dengan semua data kluster lain. menghasilkan nilai terkecil.

$$d(i, C) = \frac{1}{|A|-1} \sum_{j \in C, j \neq i} d(i, j) \quad (7)$$

Dari persamaan (7) dengan C adalah total data pada kluster C , dengan $d(i, c)$ adalah arak rata-rata data I dengan semua. Terdapat syarat yaitu $b(i) \min C \neq Ad(i, C)$ yaitu objek pada kluster lain C dimana.

- c) Proses *Silhouette Coefficient*.

$$s(i) = \frac{(b(i) - a(i))}{\max(a(i), b(i))} \quad (8)$$

Berdasarkan Persamaan (8), $s(i)$ merupakan nilai *silhouette coefficient* untuk data ke- i , yang dihitung berdasarkan perbandingan antara $a(i)$ (rata-rata jarak data ke- i terhadap data lain dalam kluster yang sama) dan $b(i)$ (rata-rata jarak data ke- i terhadap data pada kluster terdekat lainnya).

III. HASIL DAN PEMBAHASAN

A. Data

Data yang digunakan dalam penelitian ini disajikan pada Tabel 1 yang mencakup berbagai indikator sosial-ekonomi untuk kabupaten/kota di Provinsi Sumatera Selatan. Tiap Variabel Kabupaten/Kota dipakai untuk membedakan wilayah administratif sebagai objek pengamatan, dikarenakan tiap wilayah memiliki kondisi sosial yang berbeda. Angka Harapan Lama Sekolah (HLS) menunjukkan potensi lama pendidikan yang ditempuh seseorang sejak usia sekolah, yang dinyatakan dalam tahun. Rata-Rata Lama Sekolah (RLS) menunjukkan rata-rata lama pendidikan formal yang telah dijalani masyarakat pertahunnya [11]. Menurut perspektif ekonomi, Rata-Rata Pengeluaran Kapita menunjukkan banyaknya pengeluaran rata-rata tiap individu dalam satu tahunan dapat menjadi indikator kesejahteraan masyarakat melalui pola konsumsi, sedangkan Rata-Rata Upah/Gaji Bersih Pekerja Informal menunjukkan pendapatan rata-rata pertahun yang diterima oleh pekerja informal yang dapat menjadi cerminan kesejahteraan kelompok pekerja di sektor informasi [12].

Tabel 1. Data Indeks Pembangunan Manusia

Kabupaten/kota	HLS	RLS	Pengeluaran per Kapita	Rata-rata Upah Bersih	UHH
OKU	0,469	0,370	1154192	1389032	68,85
OKI	0,098	0,460	1160304	2105053	69,31
Muara Enim	0,112	0,000	1072846	1487470	69,72
Lahat	0,269	0,276	1374192	1597063	66,87
Musi Rawas	0,189	0,385	993094	1694338	68,95
Musi Banyuasin	0,221	0,125	1249650	2415678	69,51
Banyuasin	0,141	0,156	1244567	2397537	69,76
OKU Selatan	0,032	0,098	984573	1167018	67,88
OKU Timur	0,280	0,252	960288	1322883	69,78
Ogan Ilir	0,258	0,257	897070	1639946	66,33
Empat Lawang	0,145	0,260	1000059	1364698	65,73
Pali	0,287	0,151	904973	1455086	68,96
Musi Rawas Utara	0,000	0,000	1071605	1968165	66,37
Palembang	1,000	0,109	1607054	2321726	71,99
Prabumulih	0,585	1,000	1201533	1552232	71,27
Pagar Alam	0,650	0,851	978948	1865778	67,70
Lubuk Linggau	0,672	0,612	1218088	2373799	70,25

Sedangkan dari perspektif kesehatan yaitu Umur Harapan Hidup menunjukkan perkiraan atau prediksi rata-rata usia yang dicapai oleh masyarakat sejak lahir yang dinyatakan dalam tahun. Variabel ini memberikan gambaran mengenai tingkat kesehatan, kualitas hidup, serta kesejahteraan masyarakat di suatu wilayah [13].

B. Data Preprocessing

Berdasarkan data yang diperoleh, terdapat lima variabel utama yang dipakai untuk melihat kondisi sosial dan ekonomi masyarakat, yaitu Angka Harapan Lama Sekolah (HLS), Rata-Rata Lama Sekolah (RLS), Rata-Rata Pengeluaran per Kapita, Rata-Rata Upah/Gaji Bersih Pekerja Informal dan Umur Harapan Hidup (UHH). Variabel-variabel tersebut kemudian diubah ke dalam skala 0-1 melalui proses normalisasi, sehingga nilai dari setiap indikator mempunyai rentang yang sama dan dapat dibandingkan secara proporsional.

Selain itu, diterapkan pembobotan pada setiap variabel untuk mencerminkan tingkat pengaruhnya terhadap kualitas pembangunan manusia. Bobot yang digunakan adalah sebagai berikut: Harapan Lama Sekolah (HLS) (0,20), Rata-Rata Lama Sekolah (RLS) (0,15), Rata-Rata Pengeluaran per Kapita (0,25), Rata-Rata Upah/Gaji Bersih Pekerja Informal (0,20) dan Umur Harapan Hidup (UHH) (0,20). Selanjutnya, dilakukan perhitungan skor total untuk setiap kabupaten/kota untuk mendapatkan gambaran secara menyeluruh mengenai tingkat pembangunan manusia di masing-masing daerah. Perhitungan dilakukan dengan mengalikan setiap variabel dengan bobot yang telah ditetapkan, kemudian hasilnya dijumlahkan. Berikut merupakan hasil data yang telah dilakukan pembobotan dan normalisasi yang dapat dilihat pada Tabel 2.

Tabel 2. Hasil Normalisasi

Kabupaten/kota	HLS	RLS	Pengeluaran per Kapita	Rata-rata Upah Bersih	UHH
OKU	0,469	0,370	0,362	0,177	0,498
OKI	0,098	0,460	0,370	0,751	0,571
Muara Enim	0,112	0,000	0,247	0,256	0,637
Lahat	0,269	0,276	0,672	0,344	0,182
Musi Rawas	0,189	0,385	0,135	0,422	0,514
Musi Banyuasin	0,221	0,125	0,496	1,000	0,603
Banyuasin	0,141	0,156	0,489	0,985	0,643
OKU Selatan	0,032	0,098	0,123	0,000	0,343
OKU Timur	0,280	0,252	0,089	0,124	0,646
Ogan Ilir	0,258	0,257	0,000	0,378	0,095
Empat Lawang	0,145	0,260	0,145	0,158	0,000
Pali	0,287	0,151	0,011	0,230	0,515
Musi Rawas Utara	0,000	0,000	0,245	0,641	0,102
Palembang	1,000	0,109	1,000	0,924	1,000
Prabumulih	0,585	1,000	0,428	0,308	0,884
Pagar Alam	0,650	0,851	0,115	0,559	0,314
Lubuk Linggau	0,672	0,612	0,452	0,966	0,722

C. Menerapkan Algoritma K-Means Clustering

Pada penelitian ini jumlah yang digunakan sebanyak tiga kluster yaitu rendah, sedang, dan tinggi. Setelah pemilihan jumlah kluster sebanyak tiga kluster sehingga terdapat tiga *centroid* awal sebagai titik pusat tiap kluster. Adapun kabupaten yang digunakan sebagai *centroid* adalah Kota Palembang sebagai *centroid* kluster tinggi, Kabupaten Muara Enim sebagai *centroid* pada kluster sedang dan Kota Pagar Alam sebagai *centroid* pada kluster rendah.

Tabel 3. Jarak Terhadap Centroid Tinggi Palembang

Kabupaten/kota	HLS	RLS	Pengeluaran per Kapita	Rata-rata Upah Bersih	UHH
OKU	0,530	0,260	0,637	0,746	0,501
OKI	0,901	0,351	0,629	0,173	0,428
Muara Enim	0,887	0,109	0,752	0,668	0,362
Lahat	0,730	0,166	0,327	0,580	0,817
Musi Rawas	0,810	0,276	0,864	0,502	0,485
Musi Banyuasin	0,77	0,015	0,503	0,075	0,396
Banyuasin	0,858	0,046	0,510	0,060	0,356
OKU Selatan	0,96	0,010	0,876	0,924	0,656
OKU Timur	0,720	0,143	0,910	0,799	0,353
Ogan Ilir	0,741	0,148	1,000	0,546	0,904
Empat Lawang	0,854	0,151	0,854	0,766	1,000
Pali	0,712	0,041	0,988	0,694	0,484
Musi Rawas Utara	1,00	0,109	0,754	0,283	0,897
Palembang	0,000	0,000	0,000	0,000	0,000
Prabumulih	0,414	0,890	0,571	0,616	0,115
Pagar Alam	0,349	0,742	0,884	0,365	0,685
Lubuk Linggau	0,327	0,502	0,547	0,041	0,277

Tabel 4. Jarak Terhadap Centroid Sedang Muara Enim

Kabupaten/kota	HLS	RLS	Pengeluaran per Kapita	Rata-rata Upah Bersih	UHH
OKU	0,356	0,370	0,114	0,078	0,138
OKI	0,014	0,460	0,123	0,494	0,065
Muara Enim	0,000	0,000	0,000	0,000	0,000
Lahat	0,156	0,276	0,424	0,087	0,455
Musi Rawas	0,076	0,385	0,112	0,165	0,123
Musi Banyuasin	0,109	0,125	0,249	0,743	0,033
Banyuasin	0,029	0,156	0,241	0,728	0,006
OKU Selatan	0,080	0,098	0,124	0,256	0,293
OKU Timur	0,167	0,252	0,158	0,131	0,009
Ogan Ilir	0,145	0,257	0,247	0,122	0,541
Empat Lawang	0,032	0,260	0,102	0,098	0,637
Pali	0,174	0,151	0,236	0,025	0,121
Musi Rawas Utara	0,112	0,000	0,001	0,385	0,535
Palembang	0,887	0,109	0,752	0,668	0,362
Prabumulih	0,472	1,000	0,181	0,051	0,247
Pagar Alam	0,538	0,851	0,132	0,303	0,322
Lubuk Linggau	0,560	0,612	0,204	0,710	0,084

Langkah selanjutnya adalah menghitung jarak euclidean antara setiap titik data dengan semua *centroid* yang ada. Seperti langkah berikut ini dilakukan perhitungan jarak antara semua titik data dengan *centroid* Palembang. Proses ini bertujuan untuk menentukan jarak antara semua titik data dengan *centroid* Palembang yang nantinya akan menjadi dasar pengelompokan. Hasil perhitungan jarak dari setiap data ke *centroid* Palembang disajikan pada Tabel 3. *Centroid* Muara Enim sebagai perwakilan klaster sedang ditunjukkan pada Tabel 4.

Centroid Pagar Alam sebagai perwakilan klaster rendah ditunjukkan pada Tabel 5.

Tabel 5. Jarak Terhadap *Centroid* Rendah Pagar Alam

Kabupaten/kota	HLS	RLS	Pengeluaran per Kapita	Rata-rata Upah Bersih	UHH
OKU	0,181	0,481	0,246	0,381	0,183
OKI	0,552	0,390	0,255	0,191	0,257
Muara Enim	0,538	0,851	0,132	0,302	0,322
Lahat	0,381	0,575	0,556	0,215	0,132
Musi Rawas	0,461	0,466	0,019	0,137	0,199
Musi Banyuasin	0,429	0,726	0,381	0,440	0,289
Banyuasin	0,509	0,695	0,374	0,425	0,329
OKU Selatan	0,618	0,752	0,007	0,559	0,028
OKU Timur	0,370	0,598	0,026	0,434	0,332
Ogan Ilir	0,392	0,593	0,115	0,180	0,218
Empat Lawang	0,505	0,591	0,029	0,401	0,314
Pali	0,363	0,700	0,104	0,328	0,201
Musi Rawas Utara	0,650	0,851	0,130	0,081	0,212
Palembang	0,349	0,742	0,884	0,365	0,685
Prabumulih	0,065	0,148	0,313	0,251	0,570
Pagar Alam	0,000	0,000	0,000	0,000	0,000
Lubuk Linggau	0,021	0,239	0,336	0,406	0,407

Pengelompokan hasil berdasarkan jarak yang diperoleh dari perhitungan di atas akan dimasukkan ke dalam setiap klaster ditunjukkan pada Tabel 6.

Tabel 6. Hasil klaster Terhadap *Centroid*

No	Kabupaten/Kota	Klaster
1	OKU	1
2	OKI	1
3	Muara Enim	1
4	Lahat	1
5	Musi Rawas	1
6	Musi Banyuasin	1
7	Banyuasin	1
8	OKU Selatan	1
9	OKU Timur	1
10	Ogan Ilir	1
11	Empat Lawang	1
12	Pali	1
13	Musi Rawas Utara	1
14	Palembang	2
15	Prabumulih	0
16	Pagar Alam	0
17	Lubuk Linggau	0

Klaster yang bernilai 0 menyatakan klaster rendah, klaster yang bernilai 1 menyatakan klaster sedang dan klaster yang bernilai 2 menyatakan klaster tinggi. *Centroid* yang telah dikelompokkan dilakukan pembaruan *centroid* kembali yang ditunjukkan pada Tabel 7.

Setelah mendapatkan *centroid* baru langkah berikutnya adalah menghitung ulang jarak tiap titik ke *centroid* baru. Iterasi akan diberhentikan saat kondisi *centroid* tidak mengalami perubahan pada *centroid* sebelumnya dengan iterasi berikutnya. Pada langkah ini iterasi berhenti pada iterasi ke 2. Berikut adalah jarak dari tiap data ke setiap *centroid* didapatkan hasil seperti Tabel 8.

Tabel 7. *Centroid* yang digunakan

Klaster	HLS	RLS	Pengeluaran per Kapita	Rata-rata Upah Bersih	UHH
0	0,63	0,82	0,33	0,61	0,64
1	0,192	0,21	0,26	0,42	0,41
2	1	0,10	1	0,92	1

Tabel 8. Jarak Data terhadap Tiap *Centroid*

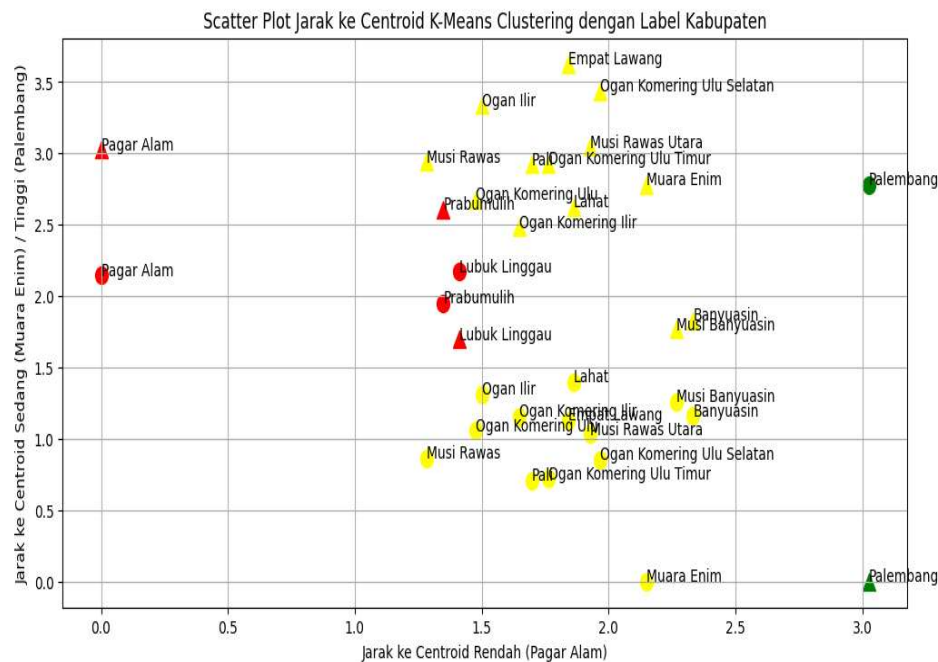
Kabupaten/kota	Centroid 1	Centroid 2	Centroid 3
OKU	0,664	0,420	1,251
OKI	0,667	0,465	1,243
Muara Enim	1,040	0,361	1,393
Lahat	0,910	0,487	1,294
Musi Rawas	0,693	0,235	1,403
Musi Banyuasin	0,914	0,660	1,010
Banyuasin	0,922	0,656	1,062
OKU Selatan	1,179	0,489	1,729
OKU Timur	0,863	0,425	1,460
Ogan Ilir	0,960	0,419	1,639
Empat Lawang	1,098	0,506	1,752
Pali	0,913	0,349	1,484
Musi Rawas Utara	1,173	0,477	1,570
Palembang	1,145	1,345	0,000
Prabumulih	0,442	1,017	1,297
Pagar Alam	0,396	0,815	1,434
Lubuk Linggau	0,438	0,904	0,859

Tabel 9. Jumlah Kabupaten/Kota Tiap Klaster

Provinsi	Jumlah Klaster Rendah	Jumlah Klaster Sedang	Jumlah Klaster Tinggi	Jumlah Kabupaten/Kota
Sumatera Selatan	3	13	1	17

D. Hasil K-Means Clustering

Berdasarkan skema *clustering* yang telah diterapkan, diperoleh hasil sebagai berikut. Jumlah kabupaten/kota provinsi Sumatera Selatan yang tergabung dalam klaster dinyatakan dalam Tabel 9.



Gambar 2. Scatter Plot Jarak ke Centroid K-Means Clustering dengan Label Kabupaten

Berdasarkan pengelompokkan yang disajikan pada Tabel 9, terdapat 3 daerah yang termasuk dalam kluster rendah, 13 daerah yang termasuk dalam kluster sedang dan 1 daerah yang termasuk dalam kluster tinggi dari total 17 kabupaten/kota di Provinsi Sumatera Selatan. Hal ini dapat menyatakan bahwa sebagian besar wilayah di Sumatera Selatan mempunyai tingkat pembangunan manusia yang cukup baik tetapi belum baik secara menyeluruh, dengan sebagian besar wilayah tergolong pada kluster sedang. Sebaliknya, masih terdapat beberapa daerah yang perlu mendapatkan penanganan khusus dalam peningkatan kualitas pendidikan, ekonomi, maupun kesehatan supaya dapat seimbang dengan daerah pada kluster tinggi.

Berdasarkan *scatter plot* pada Gambar 2 titik berwarna merah mewakili kabupaten/kota yang tergolong dalam kluster rendah, kuning untuk kluster sedang, dan hijau untuk kluster tinggi. Terlihat bahwa sebagian besar kabupaten/kota masih berada pada kluster sedang. Beberapa titik kabupaten/kota pada kluster sedang juga tampak menyebar ke arah nilai yang lebih tinggi, menandakan potensi untuk naik ke kluster tinggi. Namun demikian, kabupaten/kota yang termasuk dalam kluster tinggi (berwarna hijau) jumlahnya masih sangat sedikit dan umumnya berada di area dengan nilai jarak ke centroid yang lebih besar pada kedua sumbu, menunjukkan performa atau karakteristik yang lebih unggul dibandingkan kluster lainnya.

E. Pengujian Hasil K-Means Clustering

Berdasarkan hasil penerapan algoritma K-Means, dilakukan proses evaluasi untuk menilai kualitas klusterisasi yang dihasilkan. Nilai masing-masing metrik evaluasi disajikan pada Tabel 10.

Tabel 10. Perbandingan Hasil Pengujian K-Means Clustering

Metriks Evaluasi	Nilai
Davies-Bouldin Index (DBI)	0,9698
Silhouette Score	0,3313

Evaluasi kualitas kluster dilakukan menggunakan Davies-Bouldin Index (DBI) dan Silhouette Score. DBI mengukur rasio antara sebaran data dalam kluster dan jarak antar kluster, di mana nilai yang lebih rendah menunjukkan klusterisasi yang semakin optimal karena kluster bersifat padat dan terpisah jelas [14]. Sementara itu, Silhouette Score mengevaluasi seberapa baik objek sesuai dengan klusternya sendiri dibandingkan dengan kluster tetangganya. Nilai positif pada hasil pengujian menunjukkan bahwa sebagian besar objek lebih dekat ke kluster mereka sendiri, meskipun skor yang masih relatif rendah menandakan bahwa pemisahan antar kluster belum sepenuhnya tegas dan kemungkinan masih terdapat tumpang tindih [15].

Berdasarkan kedua metrik tersebut, hasil klusterisasi dapat dikategorikan cukup baik namun belum sepenuhnya optimal. Nilai DBI yang mendekati 1 menunjukkan bahwa jarak antar kluster sudah cukup terpisah, meskipun masih terdapat beberapa kemiripan antar kelompok data. Sementara itu, nilai Silhouette Score di kisaran 0,3 mengindikasikan bahwa struktur kluster sudah mulai terbentuk dengan pola yang dapat dibedakan, tetapi belum menunjukkan pemisahan yang sangat kuat. Dengan demikian, model K-Means ini sudah mampu memberikan pemetaan awal yang representatif terhadap data, namun masih memiliki ruang untuk peningkatan kualitas klusterisasi, misalnya melalui optimasi jumlah kluster atau normalisasi fitur yang lebih tepat.

IV. KESIMPULAN

Berdasarkan hasil analisis menggunakan metode *K-Means Clustering* terhadap 17 kabupaten/kota di Provinsi Sumatera Selatan, diperoleh tiga kluster utama yaitu rendah, sedang, dan tinggi yang mencerminkan tingkat perkembangan masing-masing daerah. Sebagian besar wilayah, yaitu 13 kabupaten/kota, termasuk dalam kluster sedang, sementara 3 daerah berada pada kluster rendah dan hanya 1 daerah pada kluster tinggi. Hasil ini menunjukkan bahwa secara umum tingkat pembangunan manusia di Provinsi Sumatera Selatan masih berada pada kategori menengah. Sebagai saran, pemerintah daerah diharapkan dapat lebih fokus dalam meningkatkan kualitas pendidikan, pemerataan ekonomi, serta pelayanan kesehatan di wilayah yang termasuk dalam kluster rendah. Upaya peningkatan tersebut diharapkan dapat memperkecil kesenjangan antar daerah dan mendorong pemerataan pembangunan manusia di seluruh Provinsi Sumatera Selatan.

Sebagai pengembangan dari penelitian ini, disarankan agar penelitian selanjutnya menggunakan rentang waktu data yang lebih luas sehingga dapat mengamati perkembangan dan tren pembangunan manusia dari tahun ke tahun. Peneliti juga dapat menambahkan variabel sosial-ekonomi lain, seperti tingkat kemiskinan, partisipasi kerja, serta akses terhadap fasilitas publik untuk memperoleh hasil klusterisasi yang lebih mendalam. Peneliti juga dapat mempertimbangkan penggunaan pendekatan analisis yang berbeda untuk memberikan wawasan yang lebih mendalam terhadap perbedaan antar metode. Selain itu, penerapan metode klusterisasi lain seperti Hierarchical Clustering atau DBSCAN dapat menjadi perbandingan untuk meningkatkan akurasi dan validitas hasil analisis mengenai ketimpangan pembangunan manusia antar wilayah.

V. DAFTAR PUSTAKA

- [1] M. Fahrurrozi, U. Hamzanwadi, and U. Hamzanwadi, "Peningkatan Indeks Pembangunan Manusia Regional Dan Implikasinya Terhadap Ketahanan Ekonomi Wilayah," *Jurnal Ketahanan Nasional*, vol. 29, no. 1, pp. 70–89, 2023. DOI: <https://doi.org/10.22146/jkn.83425>
- [2] D. I. Ginting and I. Lubis, "Pengaruh Angka Harapan Hidup dan Harapan Lama Sekolah," no. 2, pp. 519–528, 2023. DOI: <https://doi.org/10.46576/bn.v6i2.3884>
- [3] Y. Yulianti and S. Qomariah, "Indeks Pembangunan Manusia Ilmu Pengetahuan," *Jurnal Ilmu Sosial, Bahasa dan Pendidikan*, vol. 5, no. 1, 2025. DOI: <https://doi.org/10.55606/cendekia.v5i1.3512>
- [4] I. Aprilliansyah, P. I. Pemerintahan, J. A. Yani, I. P. Manusia, and K. B. Barat, "Strategi Peningkatan Indeks Pembangunan," *Jurnal Praxis Idealis*, vol. 02, no. 01, 2025. DOI: <https://doi.org/10.36859/jp.v1i2.2662>
- [5] A. Rizkina, S. Siregar, N. H. Lubis, and F. Isfa, "Pentingnya Pemerataan Pembangunan Ekonomi dalam Meningkatkan Kesejahteraan Masyarakat Indonesia," *CEMERLANG : Jurnal Manajemen dan Ekonomi Bisnis* vol. 5, 2025. DOI: <https://doi.org/10.55606/cemerlang.v5i1.3424>
- [6] A. G. Pandiangan, D. S. Priyarsono, and M. Probokawuryan, "Pengaruh Dana Desa terhadap Kesenjangan Indeks Pembangunan Manusia Wilayah Desa Kota di Indonesia," *Jurnal Ekonomi dan Kebijakan Pembangunan*, vol. 10, no. 2, pp. 134–153, 2021. DOI: <https://doi.org/10.29244/jekp.10.2.2021.114-133>
- [7] N. R. Saputra, G. Z. Muflih, and T. Informatika, "Pengelompokan Wilayah Indonesia Berdasarkan Komponen Indeks Pembangunan Manusia dengan Pendekatan Algoritma K-Means Clustering," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 8, pp. 156–167, 2025. DOI: <https://doi.org/10.36080/skanika.v8i1.3318>
- [8] P. Sektor, P. Dan, and K. Terhadap, "Analisis Pengaruh Pertumbuhan Ekonomi, Pengeluaran, Irman F. Ismail," *Jurnal Berkala Ilmiah Efisiensi*, vol. 21, no. 03, pp. 103–114, 2021.
- [9] Y. Sopyan, A. D. Lesmana, and C. Juliane, "Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan," *BITS*, vol. 4, no. 3, pp. 1464–1470, 2022, doi: [10.47065/bits.v4i3.2697](https://doi.org/10.47065/bits.v4i3.2697).
- [10] A. Akram et al., "Implementasi K-Means Clustering Untuk Rekomendasi Kelas Unggulan di SMK 1 Teknologi dan Rekayasa Mimika," vol. 5, no. 3, pp. 255–261, 2024. DOI: <https://doi.org/10.59562/jessi.v5i3.5446>.
- [11] B. Priyono, "Ekonomi desa," in *Pembangunan Desa*, S. Herman, Ed. Jakarta: Rajagrafindo, pp. 45–62, 2021. Tersedia online di <https://repository.iainpare.ac.id/id/eprint/1912/1/13.2200.160.pdf#:~:text=Potensi%20ekonomi%20desa,Desa>.
- [12] N. Aisyah, Q. Saiful, and N. April, "Analisis Faktor-Faktor yang Mempengaruhi Tingkat Kesejahteraan Masyarakat Kabupaten Jeneponto," 2023. Inisiatif: *Jurnal Ekonomi, Akuntansi dan Manajemen*, vol. 2, no.2, 2023. DOI: <https://doi.org/10.30640/inisiatif.v2i2.1581>
- [13] L. Belakang, "Determinan Umur Harapan Hidup di Indonesia," *Jurnal Ekonomi Trisakti*, vol. 4, no. 1, pp. 217–224, 2024. DOI: <https://doi.org/10.25105/jet.v4i1.18676>.
- [14] S. Harini et al., "Evaluasi Clustering K-Means dan K-Medoid pada Persebaran," *Jurnal Mnemonic*, vol. 6, no. 2, pp. 117–128, 2023. DOI: <https://doi.org/10.36040/mnemonic.v6i2.6642>.
- [15] S. Mutiah, Y. Hasnataeni, A. Fitrianto, and L. M. R. D. Jumansyah, "Perbandingan Metode Klastering K-Means dan DBSCAN dalam Identifikasi Kelompok Rumah", *Teorema: Teori dan Riset Matematika* vol.9, no. 2, 2024. DOI: <http://dx.doi.org/10.25157/teorema.v9i2.16290>.