

SEGMENTASI DAERAH DI JAWA BARAT MENGGUNAKAN *K-MEANS* DAN *HIERARCHICAL CLUSTERING* BERDASARKAN FAKTOR SOSIAL EKONOMI TAHUN 2024

Martha Erlita Sabadtini^{1*}, Dewi Juliah Ratnaningsih²

¹Mahasiswa Program Studi Statistika, Universitas Terbuka, Tangerang Selatan, Indonesia

²Dosen Program Studi Statistika, Universitas Terbuka, Tangerang Selatan, Indonesia

*Penulis korespondensi: marthaerls11@gmail.com

ABSTRAK

Permasalahan ekonomi dan ketenagakerjaan masih menjadi isu krusial di Provinsi Jawa Barat. Hal ini ditunjukkan oleh Tingkat Pengangguran Terbuka (TPT) Jawa Barat sebesar 6,74% pada Februari 2025 yang menempatkannya sebagai provinsi dengan TPT tertinggi ke-3 di Indonesia. Kondisi ini menandakan adanya ketimpangan sosial-ekonomi antardaerah yang perlu dipetakan secara komprehensif untuk mendukung kebijakan pembangunan daerah yang tepat sasaran. Penelitian ini bertujuan untuk mengelompokkan daerah di Jawa Barat berdasarkan kesamaan karakteristik sosial-ekonomi menggunakan metode *K-Means* dan *Hierarchical Clustering*. Data yang digunakan berasal dari publikasi Badan Pusat Statistik (BPS) tahun 2024 yang mencakup indikator sosial-ekonomi, seperti Indeks Pembangunan Manusia (IPM), PDRB per kapita, dan indikator lainnya. Hasil klasterisasi menunjukkan bahwa daerah di Jawa Barat terbagi menjadi tiga klaster, yaitu Daerah Maju (15%), Daerah Berkembang (44%), dan Daerah Tertinggal (41%). Penelitian ini diharapkan dapat menjadi dasar bagi perumusan kebijakan pembangunan ekonomi dan ketenagakerjaan yang lebih terarah, efisien, serta sesuai dengan karakteristik dan kebutuhan masing-masing daerah.

Kata kunci: *k-means*, *hierarchical*, segmentasi, sosial, ekonomi

1 PENDAHULUAN

Pembangunan ekonomi daerah merupakan fokus utama dalam upaya meningkatkan kesejahteraan masyarakat yang dipengaruhi oleh berbagai faktor sosial-ekonomi, seperti tingkat pendidikan, tingkat kemiskinan, maupun tingkat pengangguran. Data Badan Pusat Statistik (BPS) per Februari 2025 menunjukkan bahwa Jawa Barat merupakan provinsi dengan Tingkat Pengangguran Terbuka (TPT) sebesar 6,74% atau sebagai provinsi dengan TPT tertinggi ke-3 di Indonesia. Kondisi ini menegaskan permasalahan ketenagakerjaan masih menjadi tantangan serius di Jawa Barat, terutama akibat ketidaksesuaian antara tingginya jumlah usia produktif (siap bekerja) dengan penyediaan lapangan pekerjaan yang memadai. Ketimpangan ini juga dipengaruhi oleh perbedaan struktur ekonomi dan kualitas sumber daya manusia antardaerah yang menyebabkan pendistribusian kesempatan kerja belum merata di seluruh daerah Jawa Barat. Oleh karena itu, perlu pendekatan analitis untuk memahami kondisi sosial-ekonomi secara mendalam guna menetapkan intervensi kebijakan yang tepat (Rachma et al., 2025).

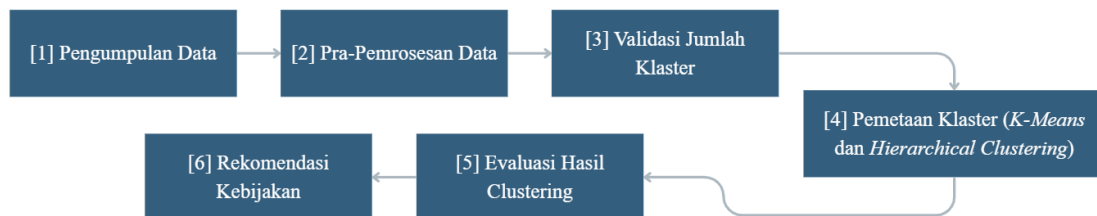
Segmentasi daerah merupakan langkah strategis untuk mengidentifikasi perbedaan karakteristik sosial-ekonomi antardaerah di Jawa Barat. Salah satu pendekatan analitis yang digunakan untuk melakukan segmentasi daerah dalam penelitian ini adalah algoritma *K-Means* dan *Hierarchical Clustering* yang dianggap memiliki keunggulan dalam efisiensi pengelompokan data dan kemudahan dalam interpretasi hasilnya. Algoritma ini bekerja dengan cara mengelompokkan data menjadi beberapa klaster berdasarkan kesamaan karakteristik antarvariabel (Putra et al., 2023).

Penelitian sebelumnya oleh Maulana et al. (2025) menunjukkan bahwa segmentasi wilayah berdasarkan indikator sosial-ekonomi di Indonesia dapat mendukung pengambilan kebijakan dan perencanaan pembangunan wilayah yang tepat sasaran. Selain itu, penerapan metode *K-Means Clustering* untuk segmentasi wilayah di Indonesia juga telah dilakukan oleh Rachma et al. (2025), Wahyudi et al. (2025), dan Valentina et al. (2025) yang membuktikan efektivitas metode ini dalam menemukan pola ekonomi regional. Sementara itu, segmentasi dengan metode *Hierarchical Clustering* telah dilakukan oleh Prihastari et al. (2024), Rohmatulillah et al. (2024), dan Wijaya et al. (2025) untuk menganalisis perbedaan karakteristik sosial-ekonomi antardaerah dengan memberikan visualisasi hubungan antardaerah melalui dendrogram.

Penelitian ini bertujuan untuk menganalisis kondisi tingkat pengangguran dan faktor-faktor sosial-ekonomi di Jawa Barat, serta memberikan pemahaman yang komprehensif mengenai variasi tingkat pengangguran antardaerah dengan mempertimbangkan aspek seperti pendidikan, kemiskinan, dan struktur ekonomi daerah. Penelitian ini semestinya dapat mengidentifikasi pola segmentasi atau pengelompokan daerah berdasarkan kesamaan karakteristik sosial-ekonomi dan tingkat penganggurannya, serta menjelaskan ciri khas dari setiap kluster yang terbentuk. Hasil analisis tersebut diharapkan mampu memberikan implikasi strategis bagi perumusan arah kebijakan pembangunan ekonomi dan prioritas ketenagakerjaan di Jawa Barat agar lebih tepat sasaran sesuai dengan karakteristik dan kebutuhan masing-masing daerah.

2 METODE

Pendekatan utama yang digunakan adalah pendekatan kuantitatif dengan metode *unsupervised learning* (lihat Gambar 1), yaitu *K-Means* dan *Hierarchical Clustering*, untuk mengelompokkan data berdasarkan kesamaan karakteristik tertentu. Implementasinya dalam penelitian ini adalah melakukan segmentasi daerah di Jawa Barat berdasarkan indikator sosial-ekonomi tahun 2024.



Gambar 1. Diagram Alir Metodologi *Unsupervised Learning* (Valentina et al., 2025)

2.1 Pengumpulan Data

Penelitian ini mencakup sebanyak 27 kabupaten/kota yang terdiri dari 18 kabupaten dan 9 kota di Jawa Barat dengan tahun 2024 sebagai tahun referensi. Data yang digunakan adalah data sekunder berupa dataset yang diambil dari Badan Pusat Statistik (BPS) Provinsi Jawa Barat (jabar.bps.go.id) dengan sembilan indikator sosial-ekonomi yang mencakup berbagai dimensi pembangunan daerah (lihat Tabel 1). Pemilihan indikator-indikator tersebut didasarkan pada kesesuaian kerangka konseptual Indeks Pembangunan Manusia (UNPD, 2020) yang dianggap dapat memberikan gambaran komprehensif terkait kondisi ketenagakerjaan dan pembangunan daerah di Jawa Barat. Seluruh data yang dikumpulkan telah melalui proses validasi oleh BPS, sehingga dapat dipastikan keakuratan dan reliabilitasnya untuk analisis lebih lanjut.

Tabel 1. Indikator Penelitian

Indikator	Satuan	Dimensi
Umur Harapan Hidup (UHH)	Tahun	Kesehatan
Harapan Lama Sekolah (HLS)	Tahun	Pendidikan

Indikator	Satuan	Dimensi
Rata-rata Lama Sekolah (RLS)	Tahun	Pendidikan
PDRB per Kapita	Juta Rupiah	Ekonomi
Pengeluaran per Kapita	Juta Rupiah	Ekonomi
Indeks Pembangunan Manusia (IPM)	Persen	Kesejahteraan
Tingkat Pengangguran Terbuka (TPT)	Persen	Ketenagakerjaan
Persentase Penduduk Miskin	Persen	Kemiskinan
Laju Pertumbuhan Penduduk	Persen	Demografi

Sumber: (Maulana et al., 2025)

2.2 Pra-Pemrosesan Data

Tahap pra-pemrosesan data mencakup *data cleaning* dan standarisasi data, sehingga dapat dipastikan data yang digunakan untuk penelitian telah akurat guna mendapatkan hasil klasterisasi yang tepat (Gori et al., 2024). Perlu atau tidaknya tahap ini pada suatu penelitian biasanya ditentukan oleh asal sumber data dan variabel yang digunakan. Data yang digunakan dalam penelitian ini diambil dari BPS, sehingga telah dipastikan tidak ada *missing values* ataupun data duplikat. Namun, tetap dilakukan *data cleaning* untuk menghapus kolom-kolom yang tidak relevan dan normalisasi data untuk menyamaratakan nama kabupaten/kota.

2.3 Validasi Jumlah Klaster

Untuk menentukan jumlah klaster optimal, digunakan *Elbow Method* dan *Silhouette Score* dalam penelitian ini. *Elbow Method* merupakan metode yang menggunakan *Within-Cluster Sum of Squares* (WCSS) dan tidak memiliki rumus tunggal. Sedangkan penggunaan *Silhouette Score* untuk menunjukkan pemisahan klaster yang baik berada dalam rentang $[-1, 1]$ dengan nilai mendekati 1. Namun, hasil yang direpresentasikan oleh kedua metode ini tetap dapat disesuaikan dengan data dan tujuan penelitian. Berikut rumus perhitungan *Elbow Method* dan *Silhouette Score*:

$$\text{Elbow Method} = \sum_{k=1}^K \sum_{x_i \in S_k} \|x_i - C_k\|_2^2 \quad (1)$$

Keterangan:

- K = jumlah klaster yang digunakan
- k = indeks klaster ke- k
- x_i = data ke- i
- S_k = himpunan data yang tergabung dalam klaster ke- k
- C_k = *centroid* klaster ke- k
- $\|x_i - C_k\|_2^2$ = jarak Euclidean kuadrat antara data ke- i dan *centroid* klaster ke- k

$$\text{Silhouette Score} = \frac{b(i) - a(i)}{\max[a(i), b(i)]} \quad (2)$$

Keterangan:

- i = indeks data ke- i
- $a(i)$ = rata-rata jarak antara data ke- i dengan data lain dalam satu klaster
- $b(i)$ = jarak rata-rata minimum antara data ke- i dengan data klaster terdekat
- $\max[a(i), b(i)]$ = nilai maksimum antara $a(i)$ dan $b(i)$

2.4 Pemetaan Klaster

2.4.1 K-Means Clustering

Algoritma *K-Means* merupakan bagian dari *unsupervised learning* yang cocok untuk mengelompokkan data berdasarkan karakteristik yang sama. Metode *clustering* ini termasuk dalam teknik klasterisasi non-hierarki. Tujuannya adalah memetakan data ke dalam jumlah

klaster yang diharapkan dengan jarak dekat *centroid* yang diukur secara iteratif (Rachma & Budy Santoso, 2025). Secara matematis, fungsi objektif *K-Means* dituliskan sebagai berikut:

$$J = \sum_{i=1}^k \sum_{x \in C_i} ||x_i - \mu_i||^2 \quad (3)$$

Keterangan:

- J = nilai fungsi objektif *K-Means*
- k = jumlah klaster
- i = indeks klaster ke- i
- x_i = data ke- i
- C_i = himpunan data yang tergabung dalam klaster ke- i
- μ_i = *centroid* klaster ke- i
- $||x_i - \mu_i||^2$ = jarak Euclidean kuadrat antara data ke- i dan *centroid* klaster ke- i

Langkah-langkah kerja *K-Means Clustering*:

- a. Inisialisasi *centroid*, yakni menentukan jumlah klaster optimal yang sesuai;
- b. Menghitung jarak setiap *data point* ke semua *centroid* dan mengalokasikan setiap *data point* ke klaster dengan *centroid* terdekat;
- c. Menghitung ulang posisi *centroid* untuk masing-masing klaster;
- d. Mengulangi langkah b dan c hingga mencapai jumlah iterasi maksimum yang telah ditentukan (konvergensi); dan
- e. Setelah konvergensi tercapai, posisi *centroid* yang akan merepresentasikan karakteristik rata-rata setiap klaster.

2.4.2 Hierarchical Clustering

Algoritma *Hierarchical Clustering* merupakan teknik klasterisasi hirarki yang menggabungkan data berdasarkan tingkat kemiripannya, sehingga membentuk sebuah struktur hirarki atau yang biasa disebut dendrogram. Algoritma ini memiliki banyak cara, salah satunya adalah *Average Linkage Method*, dengan rumus sebagai berikut:

$$d(C_i, C_j) = \frac{1}{|C_i| \times |C_j|} \sum_{x \in C_i} \sum_{y \in C_j} d(x, y) \quad (4)$$

Keterangan:

- $d(C_i, C_j)$ = jarak antara klaster C_i dan klaster C_j
- C_i, C_j = klaster ke- i dan klaster ke- j
- $|C_i|$ = jumlah data dalam klaster C_i
- $|C_j|$ = jumlah data dalam klaster C_j
- x = data dalam klaster C_i
- y = data dalam klaster C_j
- $d(x, y)$ = jarak Euclidean kuadrat antara data x dan y

Metode ini menggunakan pendekatan *bottom-up*, yakni menggabungkan setiap data secara bertahap menjadi satu klaster hingga seluruh data berhasil dikelompokkan (Justitia et al., 2021).

Langkah-langkah kerja *Hierarchical Clustering*:

- a. Menghitung matrik jarak antardata;
- b. Menggabungkan dua klaster terdekat berdasarkan parameter yang ditentukan;
- c. Memperbarui jarak antardata yang merepresentasikan kedekatan di antara klaster; dan
- d. Mengulangi langkah b dan c hingga tersisa satu kelompok terpisah.

2.5 Interpretasi Hasil

Interpretasi hasil dilakukan untuk memberitahukan arti dan pemahaman mendalam terhadap karakteristik setiap klaster yang terbentuk dengan menganalisis nilai rata-rata (*centroid*) untuk mengidentifikasi pola dan perbedaan antarklaster. Pada penelitian ini, didapatkan interpretasi dengan membandingkan indikator sosial-ekonomi daerah di Jawa Barat

yang menghasilkan tiga klaster (daerah maju, berkembang, dan tertinggal). Penelitian ini menggunakan dendrogram dan berbagai plot untuk memudahkan interpretasi.

2.6 Rekomendasi Kebijakan

Kebijakan disusun berdasarkan interpretasi klaster dengan pendekatan diferensiasi sesuai karakteristik masing-masing klaster melalui identifikasi permasalahan utama yang dihadapi. Penyusunan kebijakan disesuaikan dengan kondisi dan kebutuhan masing-masing klaster yang dapat berpengaruh secara positif terhadap pembangunan ekonomi dan ketenagakerjaan, serta perumusan kebijakan lintas klaster yang bertujuan untuk mempercepat pemerataan pembangunan melalui kolaborasi antardaerah. Rekomendasi yang dihasilkan dalam penelitian ini berupa implikasi strategis bagi pemerintah provinsi Jawa Barat dalam merumuskan arah kebijakan pembangunan ekonomi dan prioritas ketenagakerjaan yang lebih tepat sasaran.

3 HASIL DAN PEMBAHASAN

3.1 Gambaran Umum Data

Pada tahap sebelumnya telah dijelaskan bahwa penelitian ini menggunakan data sekunder dari BPS yang mencakup sembilan indikator sosial-ekonomi tahun 2024 dari 27 kabupaten/kota di Jawa Barat. Indikator yang digunakan dalam penelitian ini awalnya tersedia dalam dataset terpisah. Maka, diperlukan proses integrasi data (*merge data*) untuk menggabungkan seluruh indikator dalam satu dataset terpadu, sehingga memungkinkan analisis multidimensional yang komprehensif. Dataset terpadu akan memperlihatkan adanya variasi data antardaerah, sehingga perlu adanya identifikasi pola kesenjangan pembangunan melalui metode *clustering*. Tabel 2 menyajikan hasil integrasi data dengan menampilkan sampel data representatif (awal dan akhir), tetapi seluruh data lengkap tetap digunakan dalam proses pemetaan klaster.

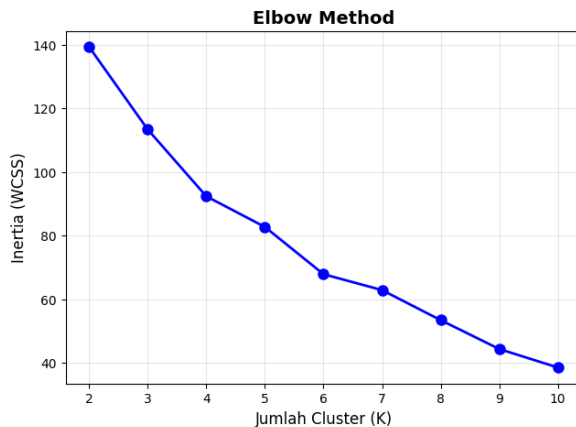
Tabel 2. Sampel Data Penelitian

Kode	Kabupaten/Kota	UHH	...	Laju Pertumbuhan Penduduk
3201	Bogor	74,9	...	1,23
3202	Sukabumi	74,85	...	0,99
3203	Cianjur	74,8	...	1,14
3204	Bandung	75,23	...	0,94
3205	Garut	74,87	...	1,33
⋮	⋮	⋮	⋮	⋮
3275	Kota Bekasi	76,14	...	1,35
3276	Kota Depok	75,82	...	0,95
3277	Kota Cimahi	75,55	...	0,79
3278	Kota Tasikmalaya	75,31	...	1,14
3279	Kota Banjar	74,99	...	1,12

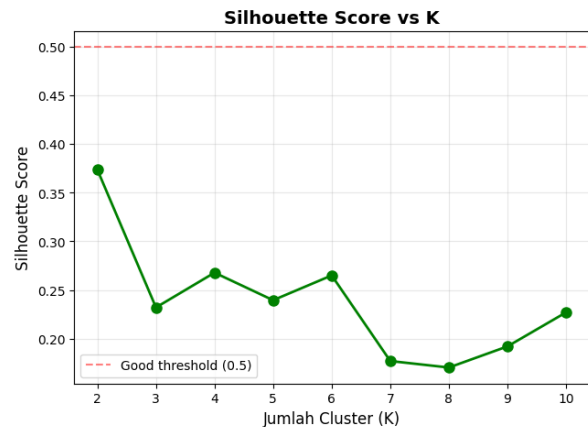
Sumber: (Badan Pusat Statistik, 2024)

3.2 Validasi Jumlah Klaster

Jumlah klaster optimal untuk analisis *K-Means* dan *Hierarchical Clustering* ditentukan berdasarkan metrik *inertia* (*Elbow Method*) dan metrik evaluasi *silhouette score*. Evaluasi ini dilakukan dengan menguji variasi jumlah klaster mulai dari K=2 hingga K=10 karena *clustering* dapat dilakukan dengan minimal dua klaster. Penentuan klaster optimal dilihat dari persentase grafik yang mulai membentuk titik siku pada *elbow method* (lihat Gambar 2) dan persentase nilai relatif tinggi pada *silhouette score* (lihat Gambar 3).



Gambar 2. Elbow Method
(Data diolah – Python, 2025)



Gambar 3. Silhouette Score
(Data diolah – Python, 2025)

Evaluasi *elbow method* menunjukkan titik siku mulai dari K=3, sementara *silhouette score* menunjukkan K=2 memiliki nilai tertinggi (0,3737), tetapi semua nilai berada di bawah *threshold*. Berdasarkan pertimbangan antara detail dan kemudahan interpretasi, dipilih K=3 sebagai jumlah kluster optimal. Keputusan ini didukung oleh K=3 yang masih berada di kategori *weak clustering* dengan *silhouette score* 0,232 dan dapat memberikan klasifikasi yang lebih informatif untuk pembangunan daerah, sehingga ketiga kluster ini dirasa cukup merepresentasikan variasi kondisi pembangunan di Jawa Barat tanpa *overfitting*.

3.3 Hasil dan Evaluasi Klusterisasi

Setelah ditentukan K=3 sebagai jumlah kluster optimal, dilakukan proses klusterisasi menggunakan metode *K-Means* dan *Hierarchical Clustering*. Pada penelitian ini, perhitungan dan klusterisasi dilakukan melalui Python di Google Colab. Pendistribusian daerah per kluster dapat dilihat pada Tabel 3. Sesuai dengan tujuan *clustering*, yaitu untuk mengelompokkan data berdasarkan kesamaan karakteristik tertentu, masing-masing kluster di atas mempunyai rata-rata karakteristik berdasarkan indikator sosial-ekonomi (lihat Tabel 4).

Tabel 3. Distribusi Daerah per Kluster

Kluster	Jumlah Anggota	Kabupaten/Kota
0	12	Kabupaten Bogor, Kabupaten Bandung, Kabupaten Ciamis, Kabupaten Sumedang, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kota Bogor, Kota Sukabumi, Kota Cirebon, Kota Tasikmalaya, Kota Banjar
1	4	Kota Bandung, Kota Bekasi, Kota Depok, Kota Cimahi
2	11	Kab Sukabumi, Kabupaten Cianjur, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Bandung Barat, Kabupaten Pangandaran.

Sumber: (Data diolah – Python, 2025)

Tabel 4. Rata-rata Karakteristik Indikator per Kluster

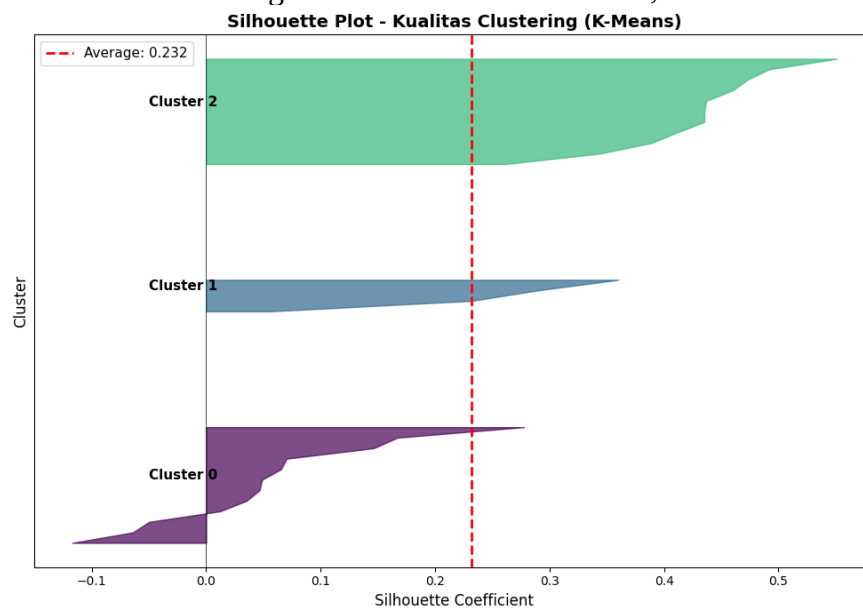
Indikator	Kluster 0	Kluster 1	Kluster 2
Umur Harapan Hidup (UHH) (Tahun)	75,23	75,84	74,89
Harapan Lama Sekolah (HLS) (Tahun)	13,13	14,09	12,27

Rata-rata Lama Sekolah (RLS) (Tahun)	9,20	11,49	7,67
PDRB per Kapita (Juta Rupiah)	63,57	78,01	32,63
Pengeluaran per Kapita (Juta Rupiah)	12,05	16,37	10,17
Indeks Pembangunan Manusia (IPM) (%)	75,57	82,66	70,80
Tingkat Pengangguran Terbuka (TPT) (%)	6,74	5,56	6,64
Persentase Penduduk Miskin (%)	7,54	3,65	10,12
Laju Pertumbuhan Penduduk (%)	1,11	1,18	1,08

Sumber: (Data diolah – Python, 2025)

3.3.1 K-Means Clustering

Untuk memperkuat interpretasi hasil klasterisasi dan mengetahui kualitas *K-Means*, diperlukan evaluasi melalui *silhouette plot* (lihat Gambar 4). Pada $K=3$, *silhouette plot* menunjukkan kualitas klaster dengan *silhouette score* sebesar 0,232.

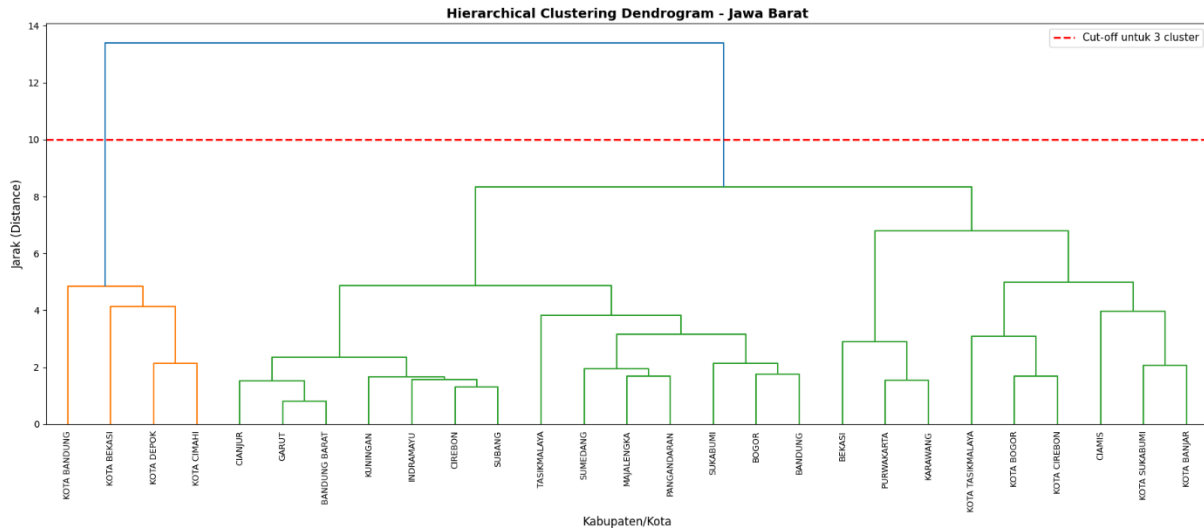


Gambar 4. *Silhouette Plot* (Data diolah – Python, 2025)

Dari Gambar 4, terlihat klaster 2 (hijau) memiliki luas area terbesar dan sebagian besar bernilai positif di atas rata-rata yang mengindikasikan adanya kohesi internal yang baik. Klaster 1 (biru) mengindikasikan adanya daerah yang berada di perbatasan klaster, ditandai dengan distribusi yang relatif merata dan beberapa nilai mendekati nol. Klaster 0 (ungu) memiliki luas area terkecil dengan sebagian kecil bernilai negatif yang mengindikasikan adanya daerah yang mungkin lebih cocok untuk masuk ke klaster lain. Meskipun nilai *Silhouette Score* termasuk moderat, keseimbangan distribusi antarklaster menunjukkan bahwa pembagian menjadi tiga klaster mampu merepresentasikan variasi karakteristik pembangunan daerah di Jawa Barat.

3.3.2 Hierarchical Clustering

Pemetaan klaster melalui *Hierarchical Clustering* divisualisasikan dalam bentuk dendrogram hierarkis guna menunjukkan segmentasi daerah di Jawa Barat berdasarkan kemiripan indikator sosial-ekonomi dengan *cut-off* di jarak 10 dengan tiga klaster utama. Dendrogram hierarkis juga memberikan gambaran terkait adanya kesenjangan yang signifikan antara daerah perkotaan dengan daerah kabupaten, terlihat dari heterogenitas yang tinggi dalam daerah kabupaten itu sendiri (lihat Gambar 5).

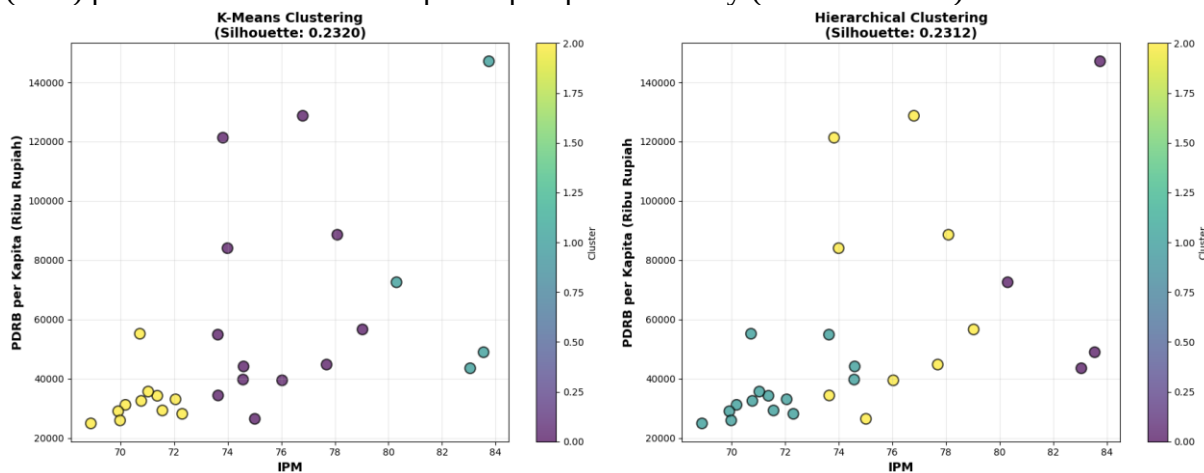


Gambar 5. Dendrogram Hierarkis (Data diolah – Python, 2025)

Dilihat dari Gambar 5, kluster 1 (*orange*) yang terdiri dari 4 kota terbentuk sebagai kluster yang terpisah dari kluster lainnya dengan jarak >13 melebihi *cut-off*. Kluster ini mengindikasikan karakteristik perkotaan yang sangat maju. Kluster 0 dan kluster 2 (hijau) berisi mayoritas daerah kabupaten dengan sub-kluster yang cukup beragam. Hal ini mengindikasikan adanya campuran karakteristik pembangunan dari daerah yang berkembang hingga daerah yang tertinggal.

3.3.3 Perbandingan *K-Means* dan *Hierarchical Clustering*

Kedua metode *clustering* menghasilkan *silhouette score* yang hampir identik, yaitu 0,232 untuk *K-Means* dan 0,2312 untuk *Hierarchical Clustering*. Hal ini mengindikasikan konsistensi dari hasil klusterisasi. Namun, diperlukan adanya *scatter plot* untuk merepresentasikan pola persebaran daerah masing-masing kluster. Visualisasi *scatter plot* untuk kedua metode *clustering* tersebut menggunakan distribusi daerah berdasarkan Indeks Pembangunan Manusia (IPM) pada sumbu x dan PDRB per Kapita pada sumbu y (lihat Gambar 6).



Gambar 6. Perbandingan Klusterisasi (Data diolah – Python, 2025)

Pada sisi kiri, yaitu *scatter plot K-Means Clustering*, kluster 0 (ungu) mencakup daerah dengan IPM dan PDRB yang relatif stabil (IPM 73-80% dan PDRB 25-130 juta). Kluster 1 (hijau) mencakup daerah dengan IPM relatif menengah-tinggi dan variasi PDRB yang lebih luas (IPM 80-85% dan PDRB 40-150 juta). Sedangkan kluster 2 (kuning) mencakup daerah dengan IPM dan PDRB yang relatif rendah (IPM <73% dan PDRB 20-60 juta). Lalu pada sisi

kanan, yaitu *scatter plot Hierarchical Clustering*, terlihat pola klusterisasi yang menunjukkan kemiripan dengan beberapa perbedaan alokasi daerah. Pemisahan klaster lebih berfokus pada keragaman IPM dan PDRB secara bertahap.

Perbedaan sangat terlihat pada distribusi daerah dengan karakteristik perbatasan bahwa *K-Means* lebih kuat dalam pemisahan klaster, sehingga merepresentasikan batas klaster yang lebih jelas. Sedangkan *Hierarchical Clustering* mempertimbangkan pendekatan bertahap antardaerah sehingga perubahan antarklaster lebih natural. Perbedaan alokasi daerah berkembang antara *K-Means* dengan *Hierarchical Clustering* mengindikasikan bahwa daerah-daerah ini memiliki karakteristik yang berada di zona netral antara dua kategori.

3.4 Interpretasi Hasil

Berdasarkan hasil klusterisasi, daerah di Jawa Barat terbagi menjadi tiga klaster, yaitu klaster 1 sebagai daerah maju, klaster 0 sebagai daerah berkembang, dan klaster 2 sebagai daerah tertinggal. Kategori klaster ini disesuaikan dengan indikator sosial-ekonomi dan kualitas sumber ketenagakerjaan periode terkini, yaitu status Indeks Pembangunan Manusia (IPM), PDRB per kapita, dan Tingkat Pengangguran Terbuka (TPT).

Klaster 1 (**Daerah Maju**) didominasi oleh daerah perkotaan dengan IPM tertinggi (82,66%) dan PDRB per kapita tertinggi (78,01 juta rupiah), tetapi dengan TPT terendah (5,56%). Daerah-daerah di klaster ini merupakan pusat-pusat ekonomi yang memiliki infrastruktur, fasilitas pendidikan, dan akses layanan kesehatan yang memadai bahkan terbilang sangat baik. Tingginya pengeluaran per kapita (16,37 juta rupiah) mengindikasikan daya beli masyarakat yang kuat dan ekonomi dinamis. Oleh karena itu, klaster ini juga menjadi penggerak utama ekonomi Jawa Barat dengan maraknya sektor industri, jasa, dan perdagangan yang berkembang pesat, serta mampu menyerap tenaga kerja dengan kualifikasi tinggi.

Klaster 0 (**Daerah Berkembang**) mencakup daerah dengan IPM menengah (75,75%) dan PDRB per kapita menengah (63,57 juta rupiah), tetapi dengan TPT relatif tinggi (6,74%). Daerah-daerah di klaster ini berada dalam fase perubahan menuju kategori maju yang mengindikasikan pembangunan daerah sudah mulai berkembang, tetapi belum merata. Pengeluaran per kapita sebesar 12,05 juta rupiah menunjukkan kestabilan ekonomi masyarakat di tengah keterbatasan akses terhadap fasilitas pendidikan dan kesehatan dibandingkan daerah maju. Oleh karena itu, klaster ini mulai berkembang dalam sektor ekonomi tertentu, seperti industri manufaktur, agribisnis, atau pariwisata. Meskipun demikian, masih diperlukan peningkatan kualitas SDM dan daya saing ekonomi sebagai bentuk akselerasi pembangunan daerah agar dapat naik ke kategori daerah maju.

Klaster 2 (**Daerah Tertinggal**) didominasi oleh daerah kabupaten yang masih mengandalkan sektor pertanian tradisional atau pekerjaan informal dengan IPM terendah (70,80%) dan PDRB per kapita terendah (32,63 juta rupiah), tetapi dengan TPT yang relatif tinggi (6,64%). Daerah-daerah ini menghadapi tantangan pembangunan daerah yang kompleks dengan infrastruktur, fasilitas pendidikan, pusat ekonomi, dan akses layanan kesehatan yang sangat terbatas. Rendahnya pengeluaran per kapita sebesar 10,17 juta rupiah mengindikasikan daya beli masyarakat yang lemah dan aktivitas ekonomi terbatas. Masyarakat kesulitan untuk meningkatkan taraf hidup karena rendahnya kualitas pekerjaan dan tingkat upah.

3.5 Rekomendasi Kebijakan

Strategi kebijakan disesuaikan dengan kondisi karakteristik masing-masing klaster untuk mewujudkan pemerataan pembangunan ekonomi dan ketenagakerjaan yang berkualitas. Pada klaster 1 (daerah maju), perlu adanya peningkatan kualitas dan produktivitas tenaga kerja, seperti melalui pelatihan, sertifikasi kompetensi, serta kebijakan upah layak yang disesuaikan dengan inflasi dan standar hidup di perkotaan. Kebijakan ini dilakukan untuk meningkatkan daya saing ekonomi dan mempertahankan stabilitas ketenagakerjaan. Perusahaan yang berhasil

menciptakan lapangan kerja berkualitas tinggi dengan upah layak dan mampu mengembangkan pusat pelatihan bagi klaster lain, sebaiknya diberikan insentif fiskal oleh pemerintah. Dengan demikian, daerah maju juga berperan sebagai penggerak utama pembangunan ekonomi dan ketenagakerjaan melalui mobilitas tenaga kerja produktif.

Didukung oleh pembangunan infrastruktur dan kemudahan investasi, pada klaster 0 (daerah berkembang), diperlukan diversifikasi ekonomi melalui pengembangan sektor unggulan, yaitu sektor agribisnis dan pariwisata. Pemerintah juga perlu mendorong pertumbuhan UMKM, mengembangkan pelatihan kerja, dan memperluas lingkup jaminan sosial ketenagakerjaan, terutama bagi pekerja informal. Kebijakan ini dilakukan sebagai langkah awal penerapan perlindungan upah minimum layak yang dapat meningkatkan kesejahteraan pekerja, serta memperkuat inkubator bisnis dan kewirausahaan, sehingga kelak daerah berkembang dapat bertransformasi menuju kategori daerah maju.

Sementara itu, klaster 2 (daerah tertinggal) memerlukan transformasi ekonomi dari pertanian tradisional menuju agribisnis modern dengan nilai tambah tinggi melalui penyediaan pendampingan teknis, modal mikro berbunga rendah, bantuan teknologi pertanian, dan pengembangan pusat ekonomi di perdesaan. Perlu adanya implementasi untuk pembangunan infrastruktur untuk mengoptimalkan Badan Usaha Milik Desa (BUMDes) sebagai penggerak ekonomi desa, serta menyediakan bantuan berupa modal usaha dan pelatihan agar masyarakat dapat meningkatkan pendapatan secara berkelanjutan.

Selain rekomendasi kebijakan per klaster, perlu juga adanya kebijakan lintas klaster. Kebijakan yang dapat dilakukan antara lain ialah membangun sistem informasi pasar kerja yang menghubungkan pencari kerja dengan peluang pekerjaan di seluruh daerah, alokasi anggaran melalui Dana Alokasi Khusus (diutamakan untuk klaster 0 dan 2), memfasilitasi kemitraan industri dengan insentif pajak, menetapkan upah minimum yang sesuai dengan biaya hidup masing-masing klaster, serta evaluasi berbasis data secara berkala agar kebijakan berjalan efektif dan ada perkembangan merata di setiap daerah.

4 KESIMPULAN

Berdasarkan hasil dan pembahasan, dapat disimpulkan bahwa metode *K-Means* dan *Hierarchical Clustering* dengan $K=3$ menunjukkan segmentasi daerah di Jawa Barat dengan karakteristik yang berbeda, yakni daerah maju, berkembang, dan tertinggal. Kesenjangan yang signifikan dalam pembangunan ekonomi dan ketenagakerjaan antarklaster mengindikasikan perlunya pendekatan kebijakan yang sesuai dengan kondisi spesifik masing-masing klaster guna mewujudkan pemerataan pembangunan daerah di Jawa Barat. Untuk penelitian selanjutnya, disarankan untuk menambah variabel infrastruktur dan teknologi digital guna memperlihatkan profil klaster yang lebih komprehensif, serta menggunakan analisis *time-series* dan analisis spasial untuk melihat dinamika perubahan karakteristik masing-masing daerah dan mengidentifikasi pola geografis antardaerah.

DAFTAR PUSTAKA

- Badan Pusat Statistik Kabupaten Majalengka. (2024). *Indeks Pembangunan Manusia (IPM) Provinsi Jawa Barat Menurut Kabupaten/Kota*. BPS.
- Badan Pusat Statistik Provinsi Jawa Barat. (2024). *Penduduk, Laju Pertumbuhan Penduduk, Distribusi Persentase Penduduk Kepadatan Penduduk, Rasio Jenis Kelamin Penduduk Menurut Kabupaten/Kota di Provinsi Jawa Barat*. BPS.
- Badan Pusat Statistik Provinsi Jawa Barat. (2024). *Produk Domestik Regional Bruto per Kapita Atas Dasar Harga Berlaku Menurut Kabupaten/Kota di Provinsi Jawa Barat (ribu rupiah)*. BPS.
- Gatto, P. A., & Awangga, R. M. (2023). *PENGELOMPOKAN KEDISIPLINAN PEGAWAI BERDASARKAN ABSENSI MENGGUNAKAN ALGORITMA K-MEANS*.

- Gori, T., Sunyoto, A., & Al Fatta, H. (2024). Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(1), 215–224. <https://doi.org/10.25126/jtiik.20241118074>
- Justitia, R. P., Hidayat, N., & Santoso, E. (2021). Implementasi Metode Agglomerative Hierarchical Clustering Pada Segmentasi Pelanggan Barbershop (Studi Kasus: RichDjoe Barbershop Malang). *JIPTIK: Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(3).
- Lashiyanti, A. R., Munthe, I. R., & Nasution, F. A. (2023). Optimisasi Klasterisasi Nilai Ujian Nasional dengan Pendekatan Algoritma K-Means, Elbow, dan Silhouette. *Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI)*, 6(1), 14–20.
- Malik, A. A. (2022). *ANALISIS SUMBER DAYA MANUSIA, PENGANGGURAN DAN KEMISKINAN DI INDONESIA*. Jakarta: Universitas Bakrie Press.
- Maliki, R., Falgenti, K., Priani, S., Fithri, F., Suherman, M., & Nugraha, D. S. (2022). Perbandingan Tingkat Pengangguran Terbuka Provinsi di Indonesia Berbasis Metode K-Means Clustering. *Computer Science (CO-SCIENCE)*, 2(2), 109–116.
- Maulana, A., Albani, R., Susanti, E., Rahmaya, E. S., & Haris, I. A. (2025). Penerapan Algoritma K-Means Clustering Untuk Segmentasi Wilayah Berdasarkan Indikator Ekonomi di Indonesia. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(2), 4055–4062. <https://doi.org/10.31004/riggs.v4i2.1169>
- Muflihan, Y., Retnawati, H., & Kistian, A. (2022). Analisis cluster dengan metode hierarki untuk pengelompokan sekolah menengah atas berdasarkan raport mutu sekolah di Kabupaten Nagan Raya. *Measurement In Educational Research*, 2(1), 22–33. <https://doi.org/10.33292/meter.v2i1.155>
- Open Data Jabar. (2024). *Persentase Penduduk Miskin Berdasarkan Kabupaten/Kota di Jawa Barat*. BPS.
- Open Data Jabar. (2024). *Tingkat Pengangguran Terbuka Berdasarkan Kabupaten/Kota di Jawa Barat*. BPS.
- Prihastari, R. R., Azizah, S. A., & Wulandari, S. P. (2024). Analisis Cluster Kabupaten/Kota di Jawa Timur Berdasarkan Faktor Kesejahteraan Masyarakat Tahun 2023. *Mutiara Jurnal Ilmiah Multidisiplin Indonesia*, 2(4), 220–235. <https://doi.org/10.61404/jimi.v2i4.329>
- Putra, R. F., Zebua, R. S. Y., Budiman, & Rahayu, P. W. (2023). *DATA MINING: ALGORITMA DAN PENERAPANNYA*. Jambi: PT Sonpedia Publishing.
- Rachma, M., & Budy Santoso, C. (2025). KLASTERISASI DATA PENGANGGURAN DI PULAU JAWA MENGGUNAKAN ALGORITMA K-MEANS DALAM PENANGGULANGAN PENGANGGURAN TAHUN 2020-2023. *Idealis: Indonesia Journal Information System*, 8(2), 202–209.
- Radha, B. V., & Sugiarti, H. (2025). ANALISIS REGRESI LOGISTIK ORDINAL UNTUK MENENTUKAN FAKTOR-FAKTOR YANG MEMENGARUHI PERFORMA BELAJAR. *Prosiding Seminar Nasional Sains Dan Teknologi Seri III Fakultas Sains Dan Teknologi*, 2(1).
- Rahayu, P. W., Sudipa, I. G. I., Suryani, & Surachman, A. (2024). *BUKU AJAR DATA MINING*. Jambi: PT Sonpedia Publishing.
- Ramdan, A. M., Siwiyanti, L., Nurmilah, R., Komariah, K., Septian, M., & Euis, S. (2024). *KLASTERISASI UMKM DAN PRODUK UNGGULAN KOTA SUKABUMI*. Bandung: Widina Media Utama.
- Rohmatulillah, O. N., Nirmala, K. B., & Wulandari, S. P. (2024). Analisis Kluster pada Faktor-Faktor yang Mempengaruhi Indikator Kesejahteraan Sosial dan Ekonomi di Provinsi Jawa Timur Tahun 2023. *Jurnal Ekonomi, Bisnis Dan Manajemen*, 3(4), 307–325. <https://doi.org/10.58192/ebismen.v3i4.2784>

- Septiani, N. I., & Fitriana, I. N. L. (2025). PENGGUNAAN K-NEAREST NEIGHBORS ALGORITHM DALAM MENGIDENTIFIKASI KELAYAKAN PINJAMAN NASABAH. *Prosiding Seminar Nasional Sains Dan Teknologi Seri III Fakultas Sains Dan Teknologi*, 2(1).
- Valentina, T., Naurah Sabela, S., Shella, S., War, M., Syaifullah, W., & Saputra, J. (2025a). SEGMENTASI PROVINSI JAWA BARAT BERDASARKAN INDIKATOR SOSIAL-EKONOMI MENGGUNAKAN METODE K-MEANS DAN AGGLOMERATIVE CLUSTERING. *RAGAM: Journal of Statistics and Its Application*, 4(1).
- Wahyudi, G. R., Rahmadden, R., Dini, E., Adrianto, S., & Fadila, R. (2025). Pengelompokan Kabupaten di Indonesia untuk Pemetaan Pendapatan Daerah Menggunakan Algoritma K-Means. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(3), 1143–1151. <https://doi.org/10.57152/malcom.v5i3.2206>
- Wazari, E. Z., & Permatasari, S. M. (2025). PENDEKATAN SPASIAL TERHADAP FAKTOR PENENTU PENYERAPAN TENAGA KERJA SEKTOR PERTANIAN DI KABUPATEN SUMBAWA BARAT 2023. *Prosiding Seminar Nasional Sains Dan Teknologi Seri III Fakultas Sains Dan Teknologi*, 2(1).
- Wijaya, A. P., Kinanthi, A. S., Yanawati, D., Syamsidar, S., & Widodo, E. (2025). Pengelompokan Kabupaten/Kota di Pulau Jawa Berdasarkan Faktor Kemiskinan Menggunakan Metode Hierarchical Clustering. *Jurnal Sains Dan Manajemen*, 13(1).