

Pengelompokkan Kabupaten/Kota di Jawa Barat Tahun 2024 Berdasarkan Anak Putus Sekolah dan Indeks Pembangunan Manusia Menggunakan Analisis *Hierarchical Agglomerative Clustering*

Dea Rohmayanti *, Nusar Hajarisman

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

dearohmayantii@gmail.com, nusarjarisman@yahoo.com

Abstract. The high number of school dropouts in several regions of West Java reflects ongoing disparities in access to education and human development. This condition affects the unequal achievement of the Human Development Index (HDI) across regions. This study aims to cluster districts and cities in West Java Province based on two variables: school dropout rates and the 2024 HDI. The method used is Agglomerative Hierarchical Clustering (AHC) with single linkage, average linkage, and complete linkage approaches. The data were obtained from the Center for Data and Technology of the Ministry of Education and Culture (Pusdatin Kemendikbud) and the West Java Statistics Agency (BPS Jawa Barat). Prior to analysis, the data underwent preprocessing stages, including completeness testing, outlier detection, standardization, and multicollinearity testing. The validity of the clustering results was evaluated using the cophenetic correlation coefficient and the silhouette coefficient. The average linkage method yielded the best result with the highest cophenetic coefficient of 0.8086. The clustering results show that areas with higher dropout rates tend to have lower HDI values. These findings provide an important basis for designing policies focused on equalizing access to education and improving the quality of human development at the regional level.

Keywords: *Hierarchical Clustering, Human Development Indeks, School Dropouts.*

Abstrak. Tingginya angka anak putus sekolah di beberapa wilayah Jawa Barat mencerminkan masih adanya ketimpangan dalam akses pendidikan dan pembangunan manusia. Kondisi ini berdampak pada pencapaian Indeks Pembangunan Manusia (IPM) yang tidak merata antar daerah. Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Jawa Barat berdasarkan dua variabel, yaitu angka putus sekolah dan IPM tahun 2024. Metode yang digunakan adalah *Agglomerative Hierarchical Clustering* (AHC) dengan pendekatan *single linkage*, *average linkage*, dan *complete linkage*. Data diperoleh dari Pusat Data dan Teknologi Informasi (Pusdatin) Kemendikbud dan Badan Pusat Statistik (BPS) Jawa Barat. Sebelum dianalisis, data melalui tahapan praproses, termasuk uji kelengkapan, deteksi *outlier*, standarisasi, dan uji multikolinearitas. Validitas hasil klastering diuji menggunakan koefisien korelasi *cophenetic* dan *silhouette coefficient*. Metode *average linkage* menghasilkan hasil terbaik dengan nilai koefisien *cophenetic* tertinggi sebesar 0,8086. Hasil klasterisasi menunjukkan bahwa wilayah dengan angka putus sekolah yang tinggi cenderung memiliki IPM yang lebih rendah. Temuan ini menjadi dasar penting bagi perumusan kebijakan yang berfokus pada pemerataan akses pendidikan dan peningkatan kualitas pembangunan manusia di tingkat daerah.

Kata Kunci: *Hierarchical Clustering, IPM, Putus Sekolah.*

A. Pendahuluan

Provinsi Jawa Barat merupakan salah satu wilayah dengan populasi terbesar di Indonesia. Kepadatan penduduk yang tinggi memberikan potensi besar dalam pembangunan sumber daya manusia, namun juga memunculkan tantangan yang kompleks. Salah satu dampak dari kondisi ini adalah munculnya berbagai dinamika sosial, terutama dalam aspek pendidikan dan kesejahteraan masyarakat. Isu utama yang menjadi perhatian saat ini adalah tingginya angka putus sekolah, serta kesenjangan dalam pencapaian Indeks Pembangunan Manusia (IPM) antar wilayah kabupaten/kota di provinsi ini (Pratikto & Damastuti, 2021). Kedua indikator tersebut mencerminkan kualitas pendidikan, akses terhadap layanan dasar, serta standar hidup masyarakat di suatu wilayah.

Pendidikan di Indonesia telah diatur dalam Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, yang menyatakan bahwa pendidikan merupakan proses yang dirancang secara sadar dan sistematis untuk mendukung peserta didik dalam mengembangkan potensi secara optimal. Tujuan pendidikan tidak hanya mencakup aspek intelektual, tetapi juga mencakup dimensi spiritual, moral, keterampilan, dan pengembangan kepribadian (Khoirunnisa' et al., 2023). Namun, dalam implementasinya, berbagai kendala masih membayangi sistem pendidikan nasional, salah satunya adalah tingginya angka putus sekolah.

Berdasarkan data dari Kementerian Pendidikan Dasar dan Menengah, pada tahun 2024 terdapat sekitar 66,8 ribu anak yang putus sekolah di Indonesia. Angka ini mencakup 38,5 ribu anak pada jenjang SD, 12,2 ribu anak pada jenjang SMP, 6,7 ribu anak pada jenjang SMA, dan 9,3 ribu anak pada jenjang SMK. Secara lebih spesifik, Provinsi Jawa Barat mencatat jumlah anak putus sekolah tertinggi di Indonesia, yaitu sebanyak 658.831 anak yang tidak bersekolah (BPS Jabar, 2025). Rinciannya mencakup 164.631 anak yang putus sekolah, 198.570 anak yang lulus tetapi tidak melanjutkan pendidikan, serta 295.530 anak yang belum pernah bersekolah (Adi Hidayat, 2025). Fakta ini mencerminkan permasalahan serius dalam aspek aksesibilitas dan keberlanjutan pendidikan, serta menunjukkan adanya ketimpangan sosial-ekonomi yang signifikan di tingkat daerah.

Dalam kerangka pembangunan manusia, pendekatan pembangunan tidak hanya berfokus pada pertumbuhan ekonomi, tetapi juga diarahkan pada peningkatan kualitas hidup manusia secara menyeluruh. Pembangunan manusia menempatkan individu sebagai subjek utama yang aktif, mandiri, dan memiliki kapasitas untuk berkontribusi secara produktif dalam masyarakat. Indeks Pembangunan Manusia (IPM), sebagai indikator utama dalam mengukur capaian pembangunan manusia, terdiri atas tiga dimensi penting, yaitu: umur panjang dan hidup sehat, akses terhadap pendidikan, serta standar hidup yang layak (Todaro & Smith, 2020).

Untuk memahami lebih dalam permasalahan ketimpangan pembangunan manusia dan pendidikan di Provinsi Jawa Barat, diperlukan pendekatan berbasis data yang mampu memetakan kondisi antarwilayah. Salah satu metode statistik yang dapat digunakan adalah analisis kluster, yang memungkinkan pengelompokan kabupaten/kota berdasarkan kemiripan karakteristik. Metode yang digunakan dalam penelitian ini adalah *Agglomerative Hierarchical Clustering* (AHC), yaitu salah satu teknik dari Hierarchical Clustering yang membentuk struktur hierarki dan biasanya divisualisasikan dalam bentuk dendrogram. (Suliadi & Marliana, 2024). Tujuan utama dari metode ini adalah untuk memperoleh gambaran pengelompokan objek yang memiliki kesamaan karakteristik sehingga dapat diidentifikasi kelompok-kelompok daerah dengan profil yang serupa. Hasil dari pengelompokan ini dapat memberikan pemahaman yang lebih mendalam mengenai distribusi wilayah dengan karakteristik sosial-ekonomi yang sebanding, dan pada akhirnya menjadi dasar bagi perumusan kebijakan yang lebih tepat sasaran, efisien, dan adil.

B. Metode

Dalam analisis ini, menggunakan pendekatan kuantitatif dengan data bersumber dari Badan Pusat Statistik Provinsi Jawa Barat dan Pusdatin Kemendikbud. Dengan variabel yang digunakan:

Tabel 1. Variabel Penelitian

Variabel	Nama Variabel	Satuan
X_1	Anak Putus Sekolah jenjang SD dan SMP	Rasio
X_2	Indeks Pembangunan Manusia	Rasio

Putus sekolah merupakan suatu proses terhentinya siswa dari lembaga pendidikan yang terjadi karena berbagai faktor, sehingga anak kehilangan kesempatan untuk memperoleh haknya atas pendidikan yang layak. Angka putus sekolah merupakan suatu proporsi dari penduduk yang berusia antara 7-15 tahun yang tidak menyelesaikan Sekolah Dasar atau Sekolah Menengah Pertama dan tidak terdaftar dalam berbagai tingkatan pendidikan. Badan Pusat Statistik menjelaskan bahwa Indeks Pembangunan Manusia (IPM) adalah indikator yang digunakan untuk menghitung keberhasilan pembangunan kualitas hidup manusia yang dengan setiap indikator yang termasuk kedalamnya. Metode yang digunakan yaitu Analisis Hierarchical Agglomerative Clustering.

Analisis Clustering

Analisis kluster adalah metode dalam statistik multivariat yang bertujuan mengelompokkan objek ke dalam kelompok (kluster) homogen berdasarkan kemiripan karakteristik antar objek (Hair et al., 2010). Menurut (Triayudi et al., 2024), metode ini efektif dalam mengidentifikasi pola pengelompokan dan dapat diterapkan di berbagai bidang seperti ekonomi wilayah, citra digital, dan segmentasi pasar.

Salah satu pendekatan yang umum digunakan dalam clustering adalah pendekatan berbasis jarak, terutama jarak Euclidean dan jarak Manhattan, yang membantu mengukur tingkat kemiripan antar objek secara matematis. Formulasi jarak *Euclidean* dapat dilakukan dengan rumus:.

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} ; i, j = 1, 2, 3, \dots, n; k = 1, 2, \dots, p \quad (1)$$

Hierarchical Clustering

Metode *Hierarchical Clustering* merupakan teknik pengelompokan data yang membentuk struktur hierarki, yang biasanya divisualisasikan dalam bentuk dendrogram. Tujuan utama dari metode ini adalah menyusun data dalam struktur seperti pohon yang merepresentasikan tingkat kemiripan antar kelompok data pada berbagai level hierarki (Suliadi & Marlina, 2024). Pendekatan ini terbagi menjadi dua jenis utama, yaitu *agglomerative*, yang memulai proses dari unit-unit data individual dan menggabungkannya secara bertahap (*bottom-up*), serta *divisif*, yang melakukan pemisahan dari satu kelompok besar menjadi kelompok-kelompok kecil (*top-down*).

Agglomerative Clustering

Metode *Agglomerative* adalah pendekatan klasterisasi *bottom-up*, dimulai dari setiap objek sebagai kluster tunggal yang secara bertahap digabungkan berdasarkan tingkat kemiripan hingga menjadi satu kluster besar (Triayudi et al., 2024). Dalam metode *agglomerative* dikenal beberapa metode untuk membentuk cluster, masing-masing memiliki pendekatan berbeda dalam menentukan jarak antar kluster yakni sebagai berikut.

a. Single linkage

Metode ini Menggabungkan kluster berdasarkan jarak terdekat antar anggotanya.

$$d_{(AB)W} = \min \{ d_{AW}, d_{BW} \} \quad (2)$$

b. Average linkage

Metode ini Menggabungkan kluster berdasarkan rata-rata jarak antar anggota.

$$d_{(AB)W} = \text{average} \{ d_{AW}, d_{BW} \} \quad (3)$$

c. Complete linkage

Metode ini Menggabungkan kluster berdasarkan jarak terjauh antar anggota.

$$d_{(AB)W} = \max \{ d_{AW}, d_{BW} \} \quad (4)$$

Uji Validitas Cluster

Uji validitas kluster bertujuan untuk menilai kualitas hasil analisis kluster. Salah satu indikator yang digunakan dalam metode kluster hierarki adalah koefisien korelasi *cophenetic*, yaitu korelasi antara matriks jarak asli (misalnya *Euclidean*) dengan matriks jarak yang dihasilkan dari dendrogram. Nilai koefisien ini berada antara -1 hingga 1, dan semakin mendekati 1 menunjukkan hasil kluster yang semakin baik (Fadliana, 2015). Rumus yang digunakan sebagai berikut

$$r_{Coph} = \frac{\sum_{i < k} (d_{ik} - \bar{d})(d_{cik} - \bar{d}_c)}{\sqrt{[\sum_{i < k} (d_{ik} - \bar{d})]^2 [\sum_{i < k} (d_{cik} - \bar{d}_c)]^2}} \quad (5)$$

Uji *Silhouette*

Silhouette Coefficient adalah salah satu metode evaluasi dalam analisis kluster yang digunakan untuk menilai sejauh mana suatu data cocok berada dalam kluster tertentu, dengan menghitung nilai masing-masing objek secara individu (Rahmawati et al., 2024). Tujuan dari pengukuran ini adalah untuk menentukan apakah kluster yang terbentuk memiliki kualitas yang baik atau tidak (Indra et al., 2023). Nilai *Silhouette Coefficient* berkisar antara -1 hingga +1, di mana nilai yang ideal berada antara 0 dan 1. Semakin mendekati angka 1, maka semakin baik kualitas pengelompokan data yang dihasilkan. Persamaan *Silhouette Coefficient* ditunjukkan sebagai berikut:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (6)$$

Tahapan Analisis

Berikut merupakan tahapan analisis yang akan dilakukan pada penelitian kali ini yaitu sebagaimana di bawah ini:

1. Pengumpulan Data
Mengambil data jumlah anak putus sekolah (SD dan SMP) dan Indeks Pembangunan manusia (IPM) dari sumber resmi.
2. Statistik Deskriptif
Menyajikan gambaran umum mengenai dua variabel utama, dengan analisis statistik dasar seperti *mean*, *max*, *min*, *standar deviasi* dan kuartil.
3. Preprocessing Data
Meliputi pemeriksaan data duplikat dan data hilang, deteksi *outlier* melalui boxplot, serta standardisasi menggunakan *Z-score* karena variabel memiliki satuan berbeda.
4. Uji Normalitas dan korelasi
Sebelumnya, pada data kondisi sosial-ekonomi dilakukan pengecekan distribusi data. Jika data tidak normal maka digunakan korelasi *Spearman* sebagai berikut:

$$r_s = \frac{\sum x^2 + \sum y^2 - \sum d_i^2}{2\sqrt{\sum x^2 \cdot \sum y^2}} \quad (7)$$

Sedangkan, jika data normal maka digunakan korelasi *Pearson* dengan rumus sebagaimana Persamaan berikut:

$$r_{xy} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{\{n\sum x^2 - (\sum x)^2\}\{n\sum y^2 - (\sum y)^2\}}} \quad (8)$$

5. Melakukan *Clustering*
Melakukan *agglomerative hierarchical clustering*. Dengan menggunakan tiga metode pendekatan *linkage*, yaitu *single linkage*, *average linkage* dan *complete linkage*.
6. Validasi *Cluster*
 - a. Menghitung nilai *Cophenetic correlation coefficient* untuk mengevaluasi struktur dendrogram. Metode *linkage* dengan nilai korelasi tertinggi dipilih sebagai metode terbaik
 - b. Setelah metode *linkage* terbaik ditentukan, dilakukan evaluasi kualitas hasil pengelompokan dengan *Silhouette Coefficient*, untuk menentukan jumlah cluster optimal.
7. Interpretasi Hasil
Menganalisis karakteristik tiap cluster berdasarkan rata-rata variabel, serta memetakan kesenjangan sosial dan pembangunan antar wilayah.
8. Kesimpulan dan Saran
Menyimpulkan hasil pengelompokan dan memberikan rekomendasi untuk perumusan kebijakan pendidikan dan pembangunan daerah.

C. Hasil Penelitian dan Pembahasan

Pendekatan *Hierarchical Agglomerative Clustering* (HAC) digunakan sebagai metode analisis untuk mengelompokkan kabupaten dan kota di Provinsi Jawa Barat berdasarkan kesamaan karakteristik dalam hal jumlah anak putus sekolah dan Indeks Pembangunan Manusia (IPM) tahun 2024. Dalam proses analisis, digunakan beberapa metode *linkage*, yaitu *Single Linkage*, *Complete*

Linkage, dan *Average Linkage*. Lalu akan dilakukan pemilihan metode terbaik berdasarkan evaluasi koefisien korelasi cophenetic.

Statistika Deskriptif

Sebelum melakukan analisis *cluster*, penting untuk terlebih dahulu meninjau statistik deskriptif dari data. Berikut merupakan statistik deskriptif pada data tercantum pada Tabel 1 berikut:

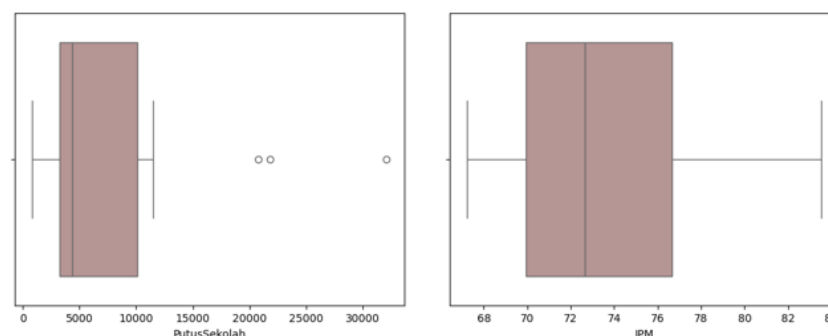
Tabel 2. Statistika Deskriptif

Variabel	Mean	Max	Min	Standar Deviasi	25%	50%	75%
Putus Sekolah	7.424	32.095	846	7.321	2.211	4.328	10.089
IPM	73,81	83,54	67,24	4,68	69,94	72,65	76,64

Pada Tabel 2 terlihat bahwa rata-rata angka putus sekolah di Provinsi Jawa Barat adalah 7.424 siswa dengan variasi yang besar yaitu 7.321 dimana jumlah minimum siswa putus sekolah sebanyak 846 siswa dan maksimum sebanyak 32.095 siswa yang mengindikasikan terdapat perbedaan yang jauh antar wilayah. Sementara untuk rata-rata Indeks Pembangunan Manusia (IPM) berada pada angka 7,81 yang termasuk pada kategori tinggi dengan IPM maksimum 83,54 dan minimum 67,24, dan variasi yang relatif kecil yaitu sebesar 4,64.

Preprocessing Data

Preprocessing data dilakukan untuk memastikan kualitas data sebelum analisis. Hasil pengecekan menunjukkan bahwa tidak terdapat data hilang maupun duplikat pada variabel yang digunakan, yaitu Putus Sekolah dan Indeks Pembangunan Manusia (IPM). *Outlier* ditemukan pada variabel Putus Sekolah berdasarkan visualisasi *boxplot*, sementara IPM tidak memiliki *outlier*. Karena kedua variabel memiliki skala yang berbeda, data distandarisasi menggunakan metode *Z-score*. Berikut visualisasi *boxplot outlier* pada Gambar 1 dan hasil standarisasi pada Tabel 2.



Gambar 1. Boxplot Outlier

Tabel 3. Standardisasi data dengan *Z-score*

Putus Sekolah	IPM
3,4341	-0,3101
2,002	-1,0547
:	:
-0,6077	-0,2668
-0,9156	-0,0380

Uji Normalitas dan korelasi

1. Normalitas

Uji normalitas dilakukan menggunakan *Shapiro-Wilk*, diperoleh hasil untuk variabel putus sekolah memiliki nilai $p < 0,0001$ dan IPM sebesar 0,0319. Karena kedua p -value lebih kecil dari taraf signifikansi 5%, maka dapat disimpulkan bahwa data tidak berdistribusi normal.

2. Korelasi *Spearman*

Data tidak berdistribusi normal maka analisis dilanjutkan menggunakan korelasi *Spearman* dan diperoleh hasil sebagai berikut:

Tabel 4. Korelasi *Spearman*

	Putus Sekolah	IPM
Putus Sekolah	1	-0,3816
IPM	-0,3816	1

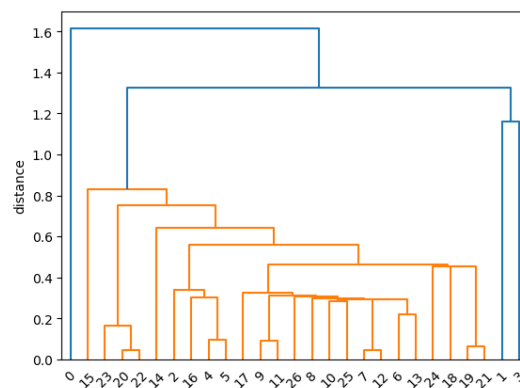
Tabel 4 menunjukkan bahwa hasil korelasi *Spearman* sebesar -0,38, yang mengindikasikan hubungan yang lemah atau tidak ada hubungan yang jelas antara variabel Putus Sekolah dan IPM (Indeks Pembangunan Manusia).

Hierarchical Agglomerative Clustering

Berikut merupakan hasil clustering yang disajikan dalam dendrogram dengan jumlah subjek berupa 27 Kabupaten/Kota di Jawa Barat.

1. Single linkage

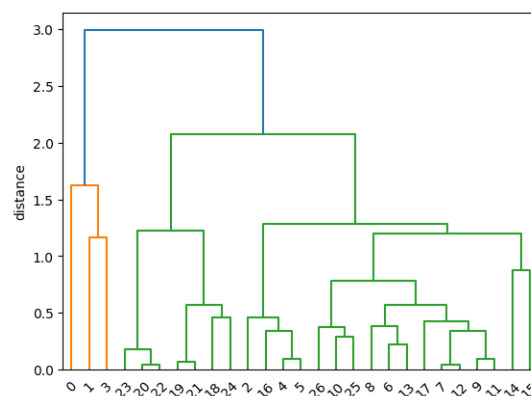
Proses menggabungkan kluster berdasarkan jarak terdekat antar objek dari masing-masing kluster. Hasil dari proses ini divisualisasikan dalam bentuk dendrogram pada Gambar 2.



Gambar 2. Dendrogram *Single Linkage*

2. Average linkage

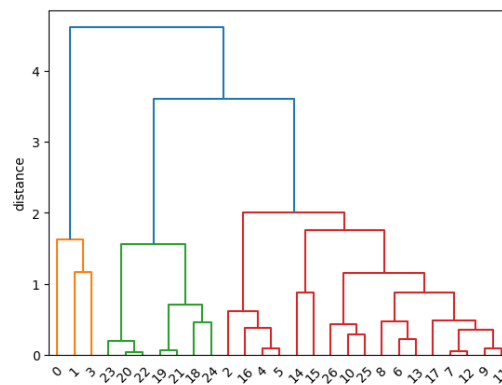
Proses menggabungkan kluster berdasarkan rata-rata jarak antar seluruh pasangan objek dari masing-masing kluster. Hasil dari proses ini divisualisasikan dalam bentuk dendrogram pada Gambar 3.



Gambar 3. Dendrogram *Average Linkage*

3. Complete linkage

Proses menggabungkan kluster berdasarkan jarak terjauh antara objek di masing-masing kluster. Hasil dari proses ini divisualisasikan dalam bentuk dendrogram pada Gambar 4.



Gambar 4. Dendrogram *Complete Linkage*

Uji Korelasi *Cophenetic*

Penelitian ini menggunakan metode *Hierarchical Agglomerative Clustering* dengan tiga pendekatan *linkage Single*, *Average*, dan *Complete*. Perbandingan efektivitas metode dilakukan melalui analisis nilai koefisien korelasi *cophenetic*, sebagaimana ditunjukkan pada Tabel 6.

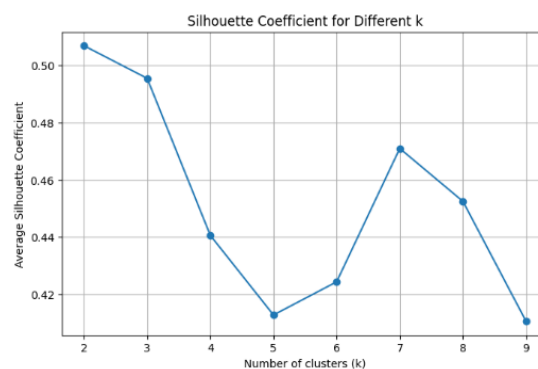
Tabel 5. Korelasi *Cophenetic*

	Korelasi <i>Cophenetic</i>
<i>Single Linkage</i>	0,7556
<i>Average Linkage</i>	0,8086
<i>Complete Linkage</i>	0,7983

Berdasarkan Tabel 5 diperoleh bahwa metode *average linkage* memiliki korelasi *cophenetic* lebih besar dibanding metode lain, sehingga pengelompokan menggunakan *Hierarchical Agglomerative* dengan *average linkage* digunakan pada penelitian ini.

Penentuan Nilai *K* dengan *Silhouette Coefficient*

Penentuan jumlah kluster terbaik dilakukan setelah metode *Average Linkage* dipilih berdasarkan nilai *cophenetic correlation* tertinggi. Validasi dilakukan menggunakan *Silhouette Coefficient*, di mana nilai indeks tertinggi menunjukkan jumlah kluster yang optimal. Diperoleh hasil *silhouette coefficient* sebagai berikut:

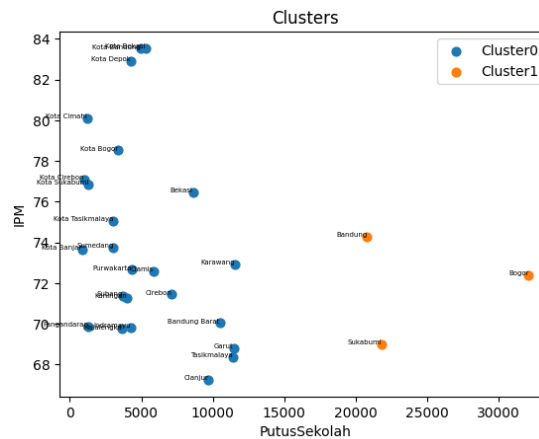


Gambar 5. Nilai *Silhouette Coefficient*

Berdasarkan Gambar 5 diatas menunjukkan bahwa nilai *Silhouette Coefficient* tertinggi terjadi pada jumlah *cluster* (k) = 2, dengan rata-rata sekitar 0,5. Setelah itu, nilai terus menurun seiring bertambahnya jumlah *cluster*, menunjukkan bahwa kualitas pengelompokan menjadi kurang baik. Dengan demikian, pengelompokan terbaik untuk data ini adalah dengan 2 *cluster*.

Clustering dengan K Optimal

Berikut merupakan *scatterplot* hasil *clustering* dengan menggunakan metode *average linkage* sebagaimana Gambar 6. di bawah ini:



Gambar 6. Plot Visualisasi *Clustering*

Berdasarkan hasil analisis dengan metode *Average Linkage*, terbentuk dua kluster utama. Kluster 0 terdiri dari mayoritas kabupaten/kota dengan rata-rata angka putus sekolah sebesar 5.240 dan nilai IPM yang relatif tinggi, yaitu 74,06. Sementara itu, Kluster 1 mencakup wilayah dengan angka putus sekolah jauh lebih tinggi (rata-rata 24.894), dan IPM yang lebih rendah, yaitu 71,88. Pola ini menunjukkan bahwa daerah dengan tingkat putus sekolah yang tinggi cenderung memiliki kualitas pembangunan manusia yang lebih rendah, yang menandakan adanya kesenjangan dalam pemerataan akses pendidikan dan pembangunan di wilayah tersebut.

Klaster 1 terdiri dari Kabupaten Bogor, Kabupaten Sukabumi, dan Kabupaten Bandung. Ketiga wilayah ini memiliki angka putus sekolah yang sangat tinggi dibandingkan wilayah lain, serta nilai IPM yang relatif lebih rendah. Meskipun bukan merupakan kota besar, ketiga kabupaten ini memiliki cakupan wilayah yang luas dan jumlah penduduk yang besar, yang dapat menimbulkan tantangan tersendiri dalam penyediaan akses pendidikan dan layanan dasar. Hal ini kemungkinan menjadi salah satu alasan mengapa ketiganya membentuk klaster tersendiri dalam hasil analisis clustering ini.

Tabel 6. Anggota Cluster Average Linkage

Cluster	Anggota	Jumlah Anggota
0	Cianjur, Garut, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, Karawang, Bekasi, Bandung Barat, Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar	24
1	Bogor, Sukabumi, Bandung	3

D. Kesimpulan

Penelitian ini berhasil mengelompokkan 27 kabupaten/kota di Provinsi Jawa Barat berdasarkan variabel angka putus sekolah dan Indeks Pembangunan Manusia (IPM) tahun 2024 menggunakan metode *Agglomerative Hierarchical Clustering*. Metode *average linkage* dipilih sebagai pendekatan terbaik karena menghasilkan nilai koefisien cophenetic tertinggi. Validasi menggunakan *silhouette coefficient* menunjukkan bahwa dua cluster merupakan jumlah optimal. Hasil pengelompokan menunjukkan bahwa wilayah dengan angka putus sekolah yang tinggi cenderung memiliki IPM yang lebih rendah. Temuan ini menekankan pentingnya intervensi kebijakan yang lebih terfokus dalam meningkatkan akses pendidikan dan pemerataan pembangunan manusia antarwilayah.

Ucapan Terimakasih

Puji syukur penulis pandatkan ke hadirat Allah Swt atas limpahan rahmat dan karuniannya. Ucapan terima kasih juga disampaikan kepada dosen pembimbing, laboran dan staf pengajar yang telah membantu dalam pengolahan data, serta kepada teman-teman yang selalu memberikan semangat dan dukungan selama penelitian berlangsung.

Daftar Pustaka

- Adi Hidayat. (2025). *Jumlah Anak Putus Sekolah di Indonesia Periode 2016-2024*. Databoks. <https://databoks.katadata.co.id/pendidikan/statistik/68341c9809724/jumlah-anak-putus-sekolah-di-indonesia-periode-2016-2024#:~:text=Menurut data Kementerian Pendidikan Dasar,SMK 9%2C3 ribu orang>.
- Khoirunnisa', A., Wibowo, F. A. S., & Kismiantini, K. (2023). Perbandingan Analisis Agglomerative Hierarchical Clustering Berdasarkan Indikator Pendidikan di Provinsi Jawa Barat. *Prosiding Seminar Pendidikan Matematika Dan Matematika*, 7(March). <https://doi.org/10.21831/pspmm.v7i1.273>
- Novian, W., & Herlina, M. (2022). Faktor-faktor yang Mempengaruhi Perilaku Masyarakat terhadap Penggunaan Internet pada Desa Digital Sukaraja. *Jurnal Riset Statistika*, 161–168. <https://doi.org/10.29313/jrs.v2i2.1466>
- Alfi Fadliana. (2015). *Penerapan Metode Agglomerative Hierarchical Clustering Untuka Klasifikasi Kabupaten/Kota di Provinsi Jawa Timur Berdsarkan Kualitas Pelayanan Keluarga Berencana*. 16(2), 39–55.
- Hair, J.F.,Black,W.C.,Babin,B.J., dan Anderson, R.E..2010. *Multivariate Data Analysis*, 7th Edition. New jersey: Pearson Prentice Hall.
- Indra, Nur, N., Iqram, M., & Inayah, N. (2023). Perbandingan K-Means dan Hierarchical Clustering dalam Pengelompokan Daerah Beresiko Stunting. *INOVTEK Polbeng - Seri Informatika*, 8(2), 356. <https://doi.org/10.35314/isi.v8i2.3612>
- Pratikto, R. O., & Damastuti, N. (2021). Klasterisasi Menggunakan Agglomerative Hierarchical Clustering Untuk Memodelkan Wilayah Banjir. *JOINTECS (Journal of Information Technology and Computer Science)*, 6(1), 13. <https://doi.org/10.31328/jointecs.v6i1.1473>
- Sofia, A. D., & Kudus, A. (2023). Pengelompokan Kabupaten/Kota berdasarkan Indikator Indeks Pembangunan Manusia 2022 Menggunakan K-Harmonic Means Clustering. *Jurnal Riset Statistika*, 3(2), 163–172. <https://doi.org/10.29313/jrs.v3i2.3130>
- Rahmawati1, I., Wilandari, Y., & Kartikasari, P. (2024). Analisis perbandingan Silhouette Coefficient dan Metode Elbow pada Pengelompokkan Provinsi di Indonesia Berdasarkan Indikator IPM dengan K-Medoids 1,2,3. *JURNAL GAUSSIAN*, 13, 13–24. <https://doi.org/10.14710/j.gauss.13.1.13-24>
- Suliadi, & Marliana, R. R. (2024). *Machine Learning*. Universitas Islam Bandung.

- Todaro, M. P., & Smith, S. C. (2020). Economic Development. Thirteenth Edition. In *Pearson* (Issue 13th Edition).
- Triayudi, A., Pinastawa, I. W. R., Pratama, J. D., Ningsih, S., Ilham, Wibowo, A., Judijanto, L., Sabilirasyad, I., Matondang, N., Hasnita, Desnelita, Y., & Gustientiedina. (2024). Clustering. In Tim PT Penamuda Media & Enbookdesign (Eds.), *Sustainability (Switzerland)* (Cetakan Pertama). PT Penamuda Media
- Viqri, Z. S. C., & Kurniati, E. (2023). Perbandingan Penerapan Metode Fuzzy Time Series Model Chen-Hsu dan Model Lee dalam Memprediksi Kurs Rupiah terhadap Dolar Amerika. *DataMath: Journal of Statistics and Mathematics*, 1(1), 19–26. <https://doi.org/10.29313/datamath.v1i1.12>