



Analisis Perbandingan Metode *Hierarchical Clustering* dan K-Means pada Indikator Pendidikan di Jawa Barat Tahun 2024

Putri Nabilah*

Universitas Singaperbangsa Karawang, *Penulis Korespondensi: 2310631050047@student.unsika.ac.id

Karunia Eka Lestari

Universitas Singaperbangsa Karawang

ABSTRAK

Penelitian ini bertujuan menganalisis perbandingan antara dua metode klusterisasi, yaitu *hierarchical clustering* dan *k-means*, dalam mengelompokkan pada indikator pendidikan di Jawa Barat pada tahun 20224. Data penelitian bersumber dari Badan Pusat Statistika (BPS) dan mencakup sepuluh variabel, yaitu angka partisipasi kasar (APK) dan angka partisipasi murni (APM) pada jenjang SD, SMP, dan SMA, rata-rata lama sekolah, harapan lama sekolah, anka melek huruf penduduk usia 15 tahun ke atas, serta Indeks Pembangunan Manusia (IPM). Metode *hierarchical clustering* memungkinkan pembentukan kluster secara bertingkat tanpa harus menentukan jumlah kluster sejak awal, sedangkan *k-means* digunakan untuk mengelompokkan data kedalam sejumlah kluster yang telah ditetapkan sebelumnya. Hasil analisis menunjukkan bahwa *hierarchical clustering* lebih efektif dalam mengungkap pola data yang kompleks, sedangkan *k-means* lebih unggul dalam kecepatan dan efisiensi. Kedua metode menghasilkan empat kluster utama dengan struktur yang relatif serupa, dimana wilayah perkotaan seperti Kota Bandung, Kota Bekasi, dan Kota Depok memiliki capaian pendidikan lebih tinggi dibandingkan kabupaten di bagian selatan Jawa Barat, sehingga temuan ini menunjukkan adanya ketimpangan pendidikan antarwilayah dan dapat menjadi dasar perumusan kebijakan pemerataan pendidikan yang berbasis data.

Kata kunci: Analisis kluster, *k-means*, *hierarchical clustering*, indikator pendidikan

PENDAHULUAN

Pendidikan merupakan salah satu faktor kunci dalam upaya peningkatan kualitas sumber daya manusia serta pembangunan nasional. Berdasarkan Undang-Undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, pendidikan berfungsi mengembangkan kemampuan individu sekaligus membentuk karakter dan peradaban bangsa yang bermartabat, dengan tujuan mencerdaskan kehidupan bangsa. Secara ideal, seluruh wilayah di Indonesia diharapkan memiliki capaian indikator pendidikan yang merata, meliputi angka melek huruf, rata-rata lama sekolah, dan harapan lama sekolah. Prinsip keadilan dalam pemerataan pendidikan menjadi salah satu fokus utama pembangunan manusia, sebagaimana tercantum dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024 yang menekankan peningkatan mutu dan pemerataan akses pendidikan di seluruh wilayah Indonesia (Bappenas, 2020).

Namun, kondisi ideal tersebut belum sepenuhnya tercapai, terutama pada tingkat provinsi. Berdasarkan data Badan Pusat Statistika (BPS, 2024), Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Barat masih menunjukkan ketimpangan antar kabupaten/kota khususnya pada indikator pendidikan. Misalnya, Kota Bandung dan Kota Bekasi memiliki rata-rata lama sekolah di atas 11 tahun, sedangkan beberapa kabupaten seperti Indramayu dan Subang masih berada di bawah rata-rata Provinsi, yaitu sekitar 8-9 tahun. Ketimpangan ini menunjukkan adanya perbedaan signifikan dalam akses, kualitas, dan pemerataan fasilitas pendidikan antarwilayah, yang berpotensi memperlebar kesenjangan sosial dan ekonomi jika tidak ditangani secara sistematis dan berbasis data. Penelitian terdahulu menunjukkan bahwa pendekatan klusterisasi berbasis indikator pendidikan mampu memberikan gambaran yang jelas mengenai perbedaan karakteristik

antar wilayah, sehingga relevan digunakan dalam analisis perbandingan metode *hierarchical clustering* dan *k-means* (Widodo dkk., 2025). Selain itu, penelitian sebelumnya juga mengungkapkan bahwa metode *agglomerative hierarchical clustering* efektif dalam menampilkan struktur pengelompokan wilayah berdasarkan indikator pendidikan secara hierarki (Khoirunnisa dkk., 2023).

Meskipun berbagai penelitian telah menerapkan metode klasterisasi dalam analisis indikator pendidikan di Provinsi Jawa Barat, sebagian besar penelitian masih berfokus pada penggunaan satu metode klasterisasi atau menggunakan data pada periode tertentu. Hingga saat ini, kajian yang secara khusus membandingkan kinerja metode *hierarchical clustering* dan *k-means* menggunakan data indikator pendidikan terbaru tahun 2024 masih terbatas. Oleh karena itu, penelitian ini bertujuan untuk melakukan analisis perbandingan kedua metode tersebut guna mengidentifikasi perbedaan hasil pengelompokan wilayah serta memperoleh gambaran yang lebih komprehensif mengenai kondisi pendidikan di Jawa Barat sebagai dasar perumusan kebijakan pendidikan yang lebih tepat sasaran dan berbasis data.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menganalisis dan membandingkan metode *hierarchical clustering* dan *k-means* dalam mengidentifikasi serta mengelompokkan kabupaten/kota di Provinsi Jawa Barat berdasarkan indikator tahun 2024. Melalui pendekatan komparatif ini, diharapkan dapat diperoleh gambaran yang lebih jelas mengenai tingkat pemerataan pendidikan di Jawa Barat, sehingga hasil penelitian dapat dijadikan sebagai dasar dalam perumusan kebijakan dan strategi peningkatan mutu pendidikan yang lebih terarah, efisien, dan berbasis bukti empiris.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode deskriptif. Data yang digunakan merupakan data sekunder pada indikator pendidikan di Jawa Barat tahun 2024 dikumpulkan sebagai dasar analisis. Selanjutnya, dilakukan teknik analisis data dengan dua pendekatan klasterisasi. Pada proses *Hierarchical Clustering*, digunakan pendekatan *complete linkage* untuk mengukur jarak terjauh antara dua klaster. Sementara itu, pada proses *K-Means Clustering*, jumlah klaster optimal ditentukan menggunakan metode *Elbow* dan *Silhouette Indeks* untuk memastikan hasil klasterisasi yang paling representatif. Untuk *Hierarchical Clustering*, dendrogram dipakai untuk mempermudah analisis hubungan hierarkis antar data. Sedangkan untuk *K-Means*, *scatter plot* digunakan untuk menggambarkan distribusi data dalam klaster.

Proses dan tahapan penelitian yang digunakan yaitu menggunakan tahapan *Knowledge Discovery in Database* (KDD). *Knowledge Discovery in Database* (KDD) merupakan keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (pattern) dalam data, Dimana pola yang ditemukan bersifat *valid*, baru, dapat bermanfaat dan dapat dipahami. Kerangka *Knowledge Discovery in Database* (KDD) digunakan dalam penelitian ini karena mampu mengakomodasi tahapan pengolahan data secara sistematis, mulai dari seleksi data hingga interpretasi hasil klasterisasi menggunakan metode *Hierarchical Clustering* dan *K-means*. Berikut merupakan tahapan dari *Knowledge Discovery in Database* (KDD).

1. Data Collection

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari publikasi Badan Pusat Statistika Provinsi Jawa Barat. Sejumlah 27 kabupaten/kota di Provinsi Jawa Barat digunakan dalam penelitian ini. Variabel yang digunakan dalam penelitian ini adalah variabel yang berfungsi pada pengelompokan kabupaten/kota berdasarkan indikator Pendidikan di Provinsi Jawa Barat. Adapun indikator-indikatornya dapat dilihat pada Tabel 1 dengan total variabel penelitian sejumlah 10 variabel.

Tabel 1. Variabel Penelitian

Variabel	Indikator
X1	Angka Partisipasi Kasar SD
X2	Angka Partisipasi Kasar SMP
X3	Angka Partisipasi Kasar SMA
X4	Angka Partisipasi Murni SD
X5	Angka Partisipasi Murni SMP
X6	Angka Partisipasi Murni SMA
X7	Rata-rata Lama Sekolah
X8	Harapan Lama Sekolah
X9	Angka Melek Aksara Usia >15 Tahun
X10	Indeks Pembangunan Manusia

2. Data Cleaning

Tahapan ini dilakukan untuk menjamin kualitas data melalui pemeriksaan duplikasi, kesalahan penginputan, dan penanganan nilai kosong (*missing value*) agar hasil klusterisasi tetap akurat.

3. Data Integration

Pada tahap ini, seluruh data indikator Pendidikan digabungkan ke dalam satu dataset berdasarkan wilayah kabupaten/kota. Proses integrasi bertujuan untuk membentuk struktur data yang konsisten sehingga siap digunakan dalam analisis clustering.

4. Data Selection

Tahap seleksi data dilakukan dengan memilih sepuluh variabel indikator Pendidikan yang relevan dengan tujuan penelitian. Variabel-variabel ini dipilih karena mampu merepresentasikan tingkat partisipasi Pendidikan, capaian Pendidikan, serta kualitas Pembangunan manusia di setiap wilayah.

5. Data Mining

Tahap ini merupakan tahap utama analysis, yaitu penerapan metode *Hierarchical Clustering* dan *K-Means Clustering* menggunakan perangkat lunak *R Studio*. Sebelum proses klusterisasi, data dinormalisasi menggunakan *Standard Scaler* untuk menyamakan skala antar variabel dengan rumus:

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

dengan x adalah nilai data, μ adalah nilai rata-rata, dan σ adalah simpangan baku.

a. Hierarchical Clustering

Metode *Hierarchical Clustering* dilakukan menggunakan pendekatan *Agglomerative* dengan *complete linkage*, yaitu penggabungan klaster berdasarkan jarak terjauh antar objek. Jarak data dihitung menggunakan *Euclidean Distance*:

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

Hasil klasterisasi ditampilkan dalam bentuk dendrogram untuk mempermudah analisis hubungan hierarkis antarwilayah.

b. *K-Means*

Pada metode *K-Means*, jumlah klaster optimal ditentukan menggunakan metode *Elbow* dan *Silhouette Index*. Proses klasterisasi dilakukan dengan meminimalkan jarak antara data dan klaster (*centroid*). Pusat klaster dihitung menggunakan rumus:

$$\mu_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i$$

Dimana μ_k adalah *centroid* klaster ke- k dan n_k adalah jumlah data dalam klaster tersebut.

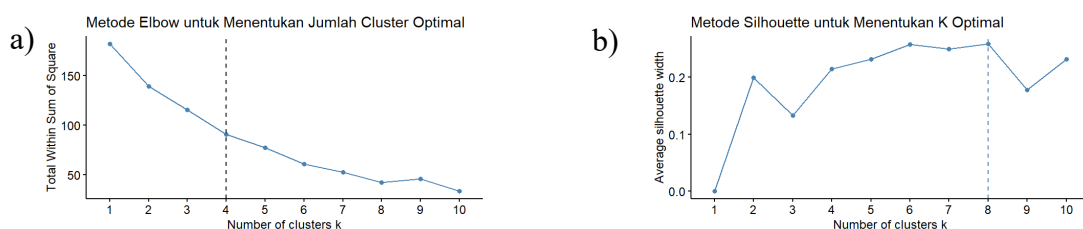
6. Evaluasi dan Perbandingan Metode

Evaluasi hasil klasterisasi dilakukan dengan membandingkan jumlah klaster, struktur pengelompokan wilayah, serta kemudahan interpretasi hasil dari kedua metode. Perbandingan ini menggambarkan pemerataan indikator Pendidikan di Provinsi Jawa Barat tahun 2024.

HASIL DAN PEMBAHASAN

Analisis perbandingan antara metode *hierarchical* dan *k-means clustering* diterapkan pada data indikator Pendidikan di Provinsi Jawa Barat tahun 2024 yang diperoleh dari Badan Pusat Statistik (BPS). Variabel yang digunakan dalam analisis mencakup angka partisipasi kasar (APK), angka partisipasi murni (APM), pada jenjang SD, SMP, dan SMA, serta angka melek huruf (AMH), rata-rata lama sekolah, harapan lama sekolah, dan IPM.

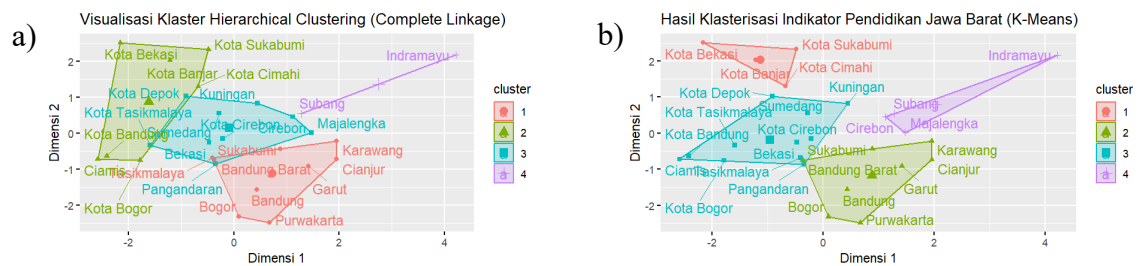
Penentuan jumlah klaster optimal pada metode *K-Means Clustering* dilakukan menggunakan metode *Elbow* dan metode *Silhouette Index*. Berdasarkan Gambar 1(a), terlihat bahwa nilai *Total Within Sum Of Square (WSS)* mengalami penurunan yang signifikan hingga $k = 4$, kemudian penurunannya mulai melandai. Kondisi ini menunjukkan adanya titik siku (*elbow point*) pada empat kluster, yang menandakan bahwa penambahan klaster setelah titik tersebut tidak memberikan peningkatan kualitas klaster yang signifikan.



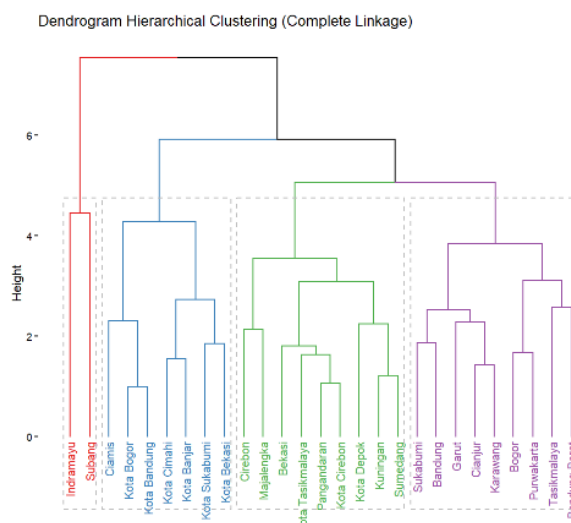
Gambar 1. Penentuan Nilai K Optimal Menggunakan Metode *Elbow*(a) dan *Silhouette*(b).

Selanjutnya, hasil metode *silhouette* pada Gambar 1(b) memperlihatkan bahwa nilai *average silhouette width* mencapai nilai relative tinggi dan stabil pada $k = 4$. Nilai *silhouette* yang lebih tinggi menunjukkan bahwa objek dalam klaster memiliki tingkat kesamaan yang tinggi terhadap klasternya sendiri serta memiliki perbedaan yang jelas dengan klaster lain. Berdasarkan kedua metode tersebut, dapat disimpulkan bahwa jumlah kluster optimal adalah 4 kluster, sehingga nilai ini digunakan dalam proses klasterisasi *K-Means*.

Hasil klusterisasi menggunakan *Hierarchical Clustering* dengan pendekatan *complete linkage* disajikan dalam bentuk *scatter plot* dan *dendrogram*, sebagaimana ditampilkan pada Gambar 2.



Berdasarkan Gambar 2, visualisasi kluster menunjukkan bahwa 27 kabupaten/kota di Provinsi Jawa Barat terbagi ke dalam empat kluster utama. Metode *complete linkage* mengelompokkan wilayah berdasarkan jarak terjauh antar anggota kluster, sehingga kluster yang terbentuk cenderung memiliki tingkat homogenitas internal yang tinggi. Secara umum, karakteristik kluster yang terbentuk dapat dijelaskan sebagai berikut. Kluster 1 didominasi oleh wilayah dengan indikator Pendidikan relative tinggi, yang umumnya merupakan daerah perkotaan, dengan nilai APK, APM, dan AMH yang lebih baik dibandingkan kluster lainnya. Kluster 2 terdiri atas wilayah dengan indikator Pendidikan pada kategori menengah, yang menunjukkan capaian Pendidikan cukup baik namun belum merata pada seluruh jenjang Pendidikan. Kluster 3 mencakup wilayah dengan indikator Pendidikan relatif lebih rendah, khususnya pada jenjang Pendidikan menengah, yang mengindikasikan masih adanya tantangan dalam partisipasi dan keberlanjutan Pendidikan. Sementara itu, Kluster 4 menunjukkan karakteristik yang berbeda secara signifikan dibandingkan kluster lain, baik karena nilai indikator yang sangat tinggi maupun sangat rendah, sehingga membentuk kelompok tersendiri dalam struktur klusterisasi. Hubungan hierarki antar kabupaten/kota divisualisasikan lebih lanjut melalui dendrogram pada Gambar 3, yang memperlihatkan proses penggabungan kluster secara bertahap berdasarkan jarak proses penggabungan kluster secara bertahap berdasarkan jarak antar data. Pemotongan dendrogram pada tingkat tertentu menghasilkan empat kluster, yang konsisten dengan hasil penentuan kluster optimal pada metode *K-Means*.



Gambar 3. Dendrogram Hasil *Hierarchical Clustering* dengan *Complete Linkage*

Hasil klasterisasi menggunakan *K-Means Clustering* menunjukkan pola pengelompokan yang relatif konsisten dengan metode *Hierarchical Clustering*, meskipun *K-Means* lebih menekankan pada kedekatan data terhadap *centroid* klaster, sedangkan *Hierarchical Clustering* menonjolkan hubungan hierarki antarwilayah. Dengan demikian, penggunaan kedua metode ini secara bersamaan memberikan gambaran yang lebih komprehensif mengenai pola pemerataan indikator pendidikan antar kabupaten/kota di Provinsi Jawa Barat tahun 2024.

Perbandingan kedua metode menunjukkan bahwa baik *Hierarchical Clustering* (*complete linkage*) maupun *K-Means Clustering* sama-sama menghasilkan empat kluster utama, yang menandakan adanya konsistensi pola pengelompokan wilayah. Namun demikian, terdapat perbedaan karakteristik sebagai berikut: *Hierarchical Clustering* lebih unggul dalam menampilkan struktur hubungan hierarki antar wilayah, sehingga memudahkan analisis kedekatan dan perbedaan secara bertahap melalui dendrogram. *K-Means Clustering* lebih efektif dalam menunjukkan pemisahan kluster secara visual dan *interpretative*, terutama dalam melihat posisi relatif wilayah terhadap *centroid* kluster.

Secara keseluruhan, kombinasi kedua metode memberikan Gambaran yang lebih komprehensif mengenai pemerataan dan ketimpangan indikator Pendidikan antar kabupaten/kota di Provinsi Jawa Barat tahun 2024. Hasil klasterisasi ini dapat menjadi dasar dalam perumusan kebijakan Pendidikan yang lebih terarah dan berbasis karakteristik wilayah.

SIMPULAN

Hasil dari penelitian ini menunjukkan bahwa metode *Hierarchical Clustering* dan *K-Means* secara efektif mampu mengelompokkan kabupaten/kota di Provinsi Jawa Barat tahun 2024 berdasarkan kesamaan karakteristik pada indikator pendidikan, khususnya Angka Partisipasi Kasar (APK) dan Angka Partisipasi Murni (APM). Penerapan *complete linkage* pada *Hierarchical Clustering* memberikan keunggulan dalam mengungkap struktur hubungan hierarki antar wilayah, sementara metode *K-Means* menghasilkan pembagian klaster yang lebih efisien dengan tingkat homogenitas internal yang relatif tinggi.

Hasil Perbandingan menunjukkan adanya konsistensi struktur klaster antara kedua metode, meskipun terdapat perbedaan pada penempatan beberapa wilayah. Keempat klaster tersebut merepresentasikan wilayah dengan capaian pendidikan sangat tinggi, tinggi, sedang, dan rendah, di mana klaster dengan capaian pendidikan tinggi hingga sangat tinggi didominasi oleh wilayah perkotaan, sementara klaster dengan capaian pendidikan rendah umumnya terdiri dari wilayah kabupaten. Temuan ini menegaskan bahwa penggunaan kedua metode secara komplementer mampu menghasilkan pemetaan kondisi pendidikan yang robust, reliabel, dan informatif.

Meskipun demikian, penelitian ini masih memiliki keterbatasan, terutama pada penggunaan data sekunder, keterbatasan jumlah indikator pendidikan, serta belum mempertimbangkan dimensi spasial dan faktor kualitatif pendidikan. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas variabel analisis dan mengintegrasikan pendekatan lain guna memperoleh gambaran ketimpangan pendidikan yang lebih komprehensif sebagai dasar perumusan kebijakan pendidikan berbasis data.

DAFTAR PUSTAKA

Aselnino, P., & Wijayanto, A. W. (2024). Analisis Perbandingan Metode Hierarchical dan Non-Hierarchical dalam Pembentukan Cluster Provinsi di Indonesia Berdasarkan Indikator *Women Empowerment*. *Indonesian Journal of Applied Statistics*, 6(1), 57-68. <https://doi.org/10.13057/ijas.v6i1.68876>

- Amrullah, A., Purnamasari, I., Sari, B. N., Garno, & Voutama, A. (n.d.). Analisis cluster faktor penunjang Pendidikan menggunakan *algoritma k-means* (studi kasus: Kabupaten Karawang). Program Studi Teknik Informasi, Universitas Singaperbangsa Karawang.
- Azkannaila, A., & Lestari, K. E. (2025). Klasterisasi Siswa Sekolah Menengah Atas Berdasarkan Kemampuan Literasi Matematis dan Kemampuan Numerasi. *Polinomial: Jurnal Pendidikan Matematika*, 4(3), 498-506.
- Badan Perencanaan Pembangunan Nasional. 2019. Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024. Kementrian Perencanaan Pembangunan Nasional/Bapepenas.
- Badan Pusat Statistik Indonesia (15 Mei 2025). [Metode Baru] Indeks Pembangunan Manusia Menurut Provinsi, 2024. <https://www.bps.go.id/id/statistics-table/2/NDk0IzI=-metode-baru-indeks-pembangunan-manusia-menurut-provinsi.html>
- Badan Pusat Statistik Indonesia (15 November 2024). [Metode Baru] Rata-rata Lama Sekolah (Tahun), 2024. <https://www.bps.go.id/id/statistics-table/2/NDE1IzI=-metode-baru-rata-rata-lama-sekolah.html>.
- Badan Pusat Statistik Indonesia (15 November 2024). [Metode Baru] Harapan Lama Sekolah (Tahun), 2024. <https://www.bps.go.id/id/statistics-table/2/NDE3IzI=-metode-baru--harapan-lama-sekolah--tahun-.html>
- Badan Pusat Statistik Indonesia (2 Desember 2024). Angka Melek Aksara Penduduk 15 Tahun ke Atas Menurut Provinsi (Persen), 2024. <https://www.bps.go.id/id/statistics-table/2/MTQ1OCMy/angka-melek-aksara-penduduk-15-tahun-ke-atas-menurut-provinsi.html>
- Badan Pusat Statistik Indonesia (2 Desember 2024). Angka Partisipasi Kasar (APK) Menurut Provinsi dan Jenjang Pendidikan, 2024. <https://www.bps.go.id/id/statistics-table/2/MzAzIzI=/angka-partisipasi-kasar-apk-menurut-provinsi-dan-jenjang-pendidikan.html>
- Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and. *Techniques*, Waltham: Morgan Kaufmann Publishers.
- Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.
- Kementerian Pendidikan Dan Kebudayaan. 2024. Indikator Pendidikan di Indonesia Tahun 2020-2024
- Khoirunnisa, A., Wibowo, F. A. S., & Kismiantini. (2023). Perbandingan Analisis *Agglomerative Hierarchical Clustering* Berdasarkan Indikator Pendidikan di Provinsi Jawa barat. 10.21831/pspm.v7i1.273.
- Putra, I. P., & Fadhillah, A. (2025). Perbandingan metode *k-means* dan *hierarchical clustering* dalam pengelompokan data penduduk miskin di Kabupaten Cianjur. DOI: 10.35870/ljit.v3i1.4028
- Sugiyono. (2019). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Bandung: Alfabeta.
- Syafiyah, U., Asrafi, I., Wicaksono, B., Puspitasari, D. P., & Sirait, F. M. (2022, November). Analisis Perbandingan Hierarchical dan Non-Hierarchical Clustering Pada Data Indikator Ketenagakerjaan di Jawa Barat Tahun 2020. In *Seminar Nasional Official Statistics* (Vol. 1, pp. 803-812).
- Widiastuti, D., & Wakit, M. (2022). Analisis Pengelompokan Data Pendidikan Menggunakan Metode *Elbow* dan *Silhouette* pada *Algoritma K-Means*. *Jurnal Teknologi Informasi dan Komputer*, 8(2), 120–131.
- Widodo, E., Mulkiyah, A. R., Anggaraini, D.T., Nur Anisah, H. A., Wardani, D. A., & Astuti, M. R. (2025)

- Yudhanegara, M. R., & Lestari, K. E. (2019, November). Clustering for multi-dimensional data set: a case study on educational data. In *Journal of Physics: Conference Series* (Vol. 1280, No. 4, p. 042025). IOP Publishing. DOI 10.1088/1742-6596/1280/4/042025
- Yudhanegara, M. R., & Lestari, K. E. (2020). The graduate views on the contribution of higher education to future careers. *International Journal of Innovation, Creativity, and Change*, 14(11), 1140-1153. https://www.ijicc.net/images/Vol_14/Iss_11/141114_Yudhanegara_2020_E1_R.pdf
- Yudhanegara, M. R., & Lestari, K. E. (2024). Pelatihan penggunaan algoritma K-Means clustering untuk mengidentifikasi karakteristik siswa. *JALIYE: Jurnal Abdimas, Loyalitas, dan Edukasi*, 3(1), 19–25. <https://doi.org/10.47662/jaliye.v3i1.710>.