

A Comparative Study of K-Means and Single Linkage for Regional Clustering in West Kalimantan in 2024 Based on Educational Indicators

Ridho Alamsyah^{1*}, Andina Putri Maharani¹, Dwi Vika Mutia¹, Sofiah Nur Alifah¹, Sofie Malika¹, Muhammad Riefky¹

¹Universitas Sebelas Maret, Surakarta, Indonesia

*Corresponding author: ridhoalamsyahcilik@gmail.com

Abstract

Education is a strategic investment in human resource development, as evidenced by several indicators, including the Average Years of Schooling (AYS), Expected Years of Schooling (EYS), Net Enrolment Rate (NER), and Gross Enrolment Ratio (GER). West Kalimantan Province faces the challenge of providing equal educational quality across regions. This issue has led to a low Human Development Index (HDI), ranking 30th of 34 provinces in Indonesia in 2024. These regional differences make it hard for the government to decide on the right policy actions. This study aims to group districts and cities in West Kalimantan based on their education quality indicators. It also compares the performance of two clustering methods, K-Means and Single Linkage, to find out which one produces the best results. The data used is secondary data from 2024, sourced from BPS West Kalimantan Province. The analysis included descriptive statistics, a data adequacy test, a multicollinearity test, and a cluster quality evaluation using the Elbow method based on Within-Cluster Sum of Squares (WCSS). At three clusters, K-Means produces a WCSS value of 26.6143, while Single Linkage produces a WCSS value of 33.7651. Therefore, K-Means with three clusters is selected as the optimal clustering method because it produces a lower WCSS value than Single Linkage. The K-Means clustering results consist of three groups with different characteristics based on the four education variables used in the study. These results can support local governments in formulating education policies tailored to the characteristics of each cluster.

Keywords: Clustering, Education, K-Means

Article History:

Received : 17 June 2026

Revised : 31 August 2026

Accepted : 10 September 2026

© 2026 Alamsyah et al. Published by
Program Studi Statistika Fakultas MIPA
Universitas Lambung Mangkurat.



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

How to Cite:

R. Alamsyah, A. P. Maharani, D. V. Mutia, S. N. Alifah, S. Malika, and M. Riefky, "A Comparative Study of K-Means and Single Linkage for Regional Clustering in West Kalimantan in 2024 Based on Educational Indicators," *RAGAM: Journal of Statistics and Its Application*, vol. 5, no. 2, pp. 145-159, 2026, [doi: 10.20527/ragam.v5i2.19477](https://doi.org/10.20527/ragam.v5i2.19477)

1. PENDAHULUAN

Pendidikan adalah faktor penentu kualitas sumber daya manusia dan keberhasilan pembangunan daerah. Menurut *Human Capital Theory* yang dikembangkan oleh Gary Becker dan Theodore Schultz, pendidikan dianggap sebagai investasi strategis untuk meningkatkan kemampuan dan produktivitas individu. Peningkatan kualitas pendidikan dipercaya dapat mendorong pertumbuhan ekonomi dan kemajuan wilayah. Oleh karena itu, pembangunan di sektor pendidikan tidak hanya memberikan manfaat bagi suatu individu, tetapi bisa meningkatkan daya saing daerah secara lebih luas. Secara global, komitmen terhadap pendidikan berkualitas terlihat dalam *Sustainable Development Goals* 4 (SDGs 4) dari *United Nations*. SDGs 4 menyoroti pendidikan inklusif, berkualitas, dan memberikan kesempatan belajar sepanjang hayat untuk semua orang. Agenda ini sejalan dengan Asta Cita poin ke-4 yang memprioritaskan membangun sumber daya manusia

yang unggul melalui peningkatan akses serta kualitas pendidikan. Indonesia masih menghadapi tantangan besar dalam pemerataan pendidikan di beberapa wilayah. Tantangan ini terasa di provinsi dengan wilayah yang luas dan kompleks seperti Kalimantan Barat. Sebagai salah satu provinsi terbesar, Kalimantan Barat kesulitan mendistribusikan layanan pendidikan secara merata ke seluruh kabupaten dan kota. Akibatnya, Indeks Pembangunan Manusia (IPM) di wilayah ini masih rendah. Kalimantan Barat mencatat IPM di bawah rata-rata nasional dan berada di peringkat ke-30 dari 34 provinsi di Indonesia [1]. Ketidakmerataan pendidikan di Kalimantan Barat dapat dievaluasi menggunakan empat indikator kunci yaitu Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), Angka Partisipasi Kasar (APK) dan Angka Partisipasi Sekolah (APS). RLS mencerminkan rekam pendidikan penduduk dewasa, HLS merupakan proyeksi lama waktu belajar bagi generasi muda mendatang, sedangkan APK digunakan untuk menggambarkan tingkat keikutsertaan penduduk dalam suatu jenjang pendidikan secara keseluruhan tanpa membatasi usia. Sementara, APS berperan sebagai indikator penting dalam menilai sejauh mana sistem pendidikan mampu menjangkau penduduk usia sekolah. Namun, adanya variasi dan perbedaan nilai dari empat indikator tersebut antar kab/kota seringkali menyulitkan pemerintah dalam menentukan wilayah yang perlu diprioritaskan untuk mendapatkan intervensi kebijakan secara lebih cepat.

Selama ini, proses pengambilan keputusan kerap dilakukan secara umum tanpa memperhatikan kesamaan karakteristik khusus antar wilayah. Maka dari itu, diperlukan sebuah teknik analisis yang dapat mengklasifikasikan wilayah dengan profil pendidikan yang serupa. Pengelompokan objek ini dapat dilakukan melalui analisis kluster (*clustering analysis*). Sebagai bagian dari metode statistik multivariat interdependen, analisis kluster bertujuan untuk merangkum kumpulan data dengan cara menyatukan objek penelitian ke kelompok yang didasarkan pada persamaan sifat atau karakter tertentu. Penelitian sebelumnya telah membahas pengelompokan wilayah berdasarkan indikator pendidikan menggunakan berbagai metode analisis kluster untuk mendukung pengambilan keputusan dalam pembangunan pendidikan. Penelitian yang dilakukan Basalamah et al. [2] mengungkapkan adanya tiga kluster di setiap jenjang (SD, SMP, dan SMA/ sederajat) ketika 34 provinsi di Indonesia dikategorikan menurut angka penyelesaian pendidikan dalam kurun waktu 2020 sampai 2022. Pada jenjang SD, kluster 1 (tinggi) berjumlah 29 anggota sebagai capaian tertinggi, kluster 2 (sedang) berjumlah 4 anggota, dan kluster 3 (rendah) berjumlah 1 anggota yaitu Papua sebagai capaian terendah. Pada jenjang SMP, kluster 1 (tinggi) berjumlah 20 anggota sebagai capaian tertinggi, kluster 2 (sedang) berjumlah 13 anggota, dan kluster 3 (rendah) berjumlah 1 anggota yaitu Papua sebagai capaian terendah. Pada jenjang SMA, kluster 1 (tinggi) berjumlah 7 anggota sebagai capaian tertinggi, kluster 2 (sedang) berjumlah 24 anggota, dan kluster 3 (rendah) berjumlah 3 anggota yaitu Nusa Tenggara Timur, Gorontalo, dan Papua sebagai capaian terendah [2]. Berdasarkan temuan ini, metode analisis kluster terbukti efektif dalam mengelompokkan daerah-daerah dengan profil pendidikan yang serupa. Namun, sebagian besar penelitian terdahulu hanya menggunakan satu metode *clustering* tanpa melakukan perbandingan kinerja antar metode, serta lebih banyak dilakukan pada

wilayah di luar Provinsi Kalimantan Barat. Padahal, karakteristik data pendidikan yang berbeda dapat menghasilkan struktur klaster yang berbeda pula sehingga komparasi hasil metode *clustering* yang tepat dapat menjadi aspek penting untuk memperoleh hasil pengelompokan yang optimal. Selain itu, kondisi geografis dan tingkat pemerataan pendidikan yang beragam di Kalimantan Barat memerlukan kajian khusus untuk menghasilkan informasi yang relevan bagi perumusan kebijakan daerah. Kondisi ini mengindikasikan bahwa masih minimnya kajian komparatif antara metode *K-Means* dan *Single Linkage* untuk keperluan klasterisasi kab/kota di Kalimantan Barat dengan menggunakan parameter kualitas pendidikan. Sehingga tujuan penelitian ini adalah mengelompokkan wilayah di Kalimantan Barat tahun 2024 menurut indikator pendidikan menggunakan komparasi metode *K-Means* dan *single linkage clustering* dengan optimalisasi WCSS (*Within-Cluster Sum of Squares*) dengan metode Elbow. Melalui metode komparasi tersebut, hasil penelitian ini diharapkan mampu memberikan manfaat praktis dan dapat menjadi acuan berbasis data bagi Pemerintah Kalimantan Barat serta pemangku kepentingan lainnya dalam memetakan daerah pembangunan. Dengan mengetahui karakteristik spesifik dari setiap klaster, pemerintah dapat merumuskan kebijakan pemerataan pendidikan yang lebih presisi, mengalokasikan anggaran dan sumber daya secara efisien, serta merancang program intervensi yang tepat sasaran untuk meningkatkan capaian RLS, HLS, APK, APS, agar dapat mendorong IPM.

2. TINJAUAN PUSTAKA

Penelitian Mashitawai et al. [3] memanfaatkan data sekunder BPS NTT 2023 untuk meneliti 22 wilayah kabupaten dan kota. Riset ini menggunakan enam variabel untuk mengukur kualitas pendidikan: RLS, APS, Angka Partisipasi Kasar (APK), Rasio Guru-Murid pada jenjang SMA, Angka Melek Huruf (AMH), dan jumlah SMA. Metode yang digunakan yaitu membandingkan *K-Means* dan *K-Medoids* dengan penentuan total klaster optimal dengan skor *Silhouette*. Hasil pengklasteran optimal dengan algoritma *K-Means* membentuk dua kategori. Klaster 1 berisi 8 kabupaten dengan kualitas pendidikan tinggi (Sumba Barat, Sumba Timur, Sikka, Manggarai, Rote Ndao, Sumba Tengah, Sumba Barat Daya, dan Sabu Raijua), sedangkan Klaster 2 mencakup 14 kab/kota yang memiliki kualitas pendidikan rendah [3].

Dalam penelitiannya, Thamrin et al. [4] menganalisis data sekunder indikator kesehatan masyarakat tahun 2020 yang mencakup 19 kab/kota di Sumatera Barat. Terdapat lima indikator yang diteliti: jumlah warga yang sakit/mengeluhkan kesehatannya (X1), warga berobat jalan (X2), warga yang pernah dirawat inap (X3), pemanfaatan jaminan kesehatan rawat inap (X4), serta rata-rata hari rawat inap (X5). Metode yang diterapkan yaitu analisis klaster hierarki metode *Single Linkage* dengan ukuran jarak *Manhattan (City Block Distance)*, di mana data terlebih dahulu ditransformasikan dengan standarisasi *Z-Score*, dilakukan uji korelasi *Pearson*, serta reduksi dimensi dengan *Principal Component Analysis/PCA* untuk mereduksi korelasi antar indikator. Hasil analisis tersebut membagi wilayah menjadi lima klaster. Wilayah

dengan tingkat kesehatan tinggi menempati klaster 1 (Kepulauan Mentawai) dan klaster 4 (Pariaman). Wilayah berstatus kesehatan sedang, yang mencakup 12 kab/kota, terkumpul di klaster 2. Sementara itu, wilayah dengan tingkat kesehatan rendah terbagi ke dalam klaster 3 (Solok, Tanah Datar, Lima Puluh Kota, dan Sijunjung) serta klaster 5 (Padang) [4].

Studi yang dilakukan Rozikin et al. [5] memanfaatkan data sekunder dari Dinas Pendidikan Provinsi Jawa Timur tahun 2021 yang mencatat angka putus sekolah tingkat SMA/SMK sederajat di 24 cabang dinas. Penelitian ini menganalisis 12 variabel pemicu putus sekolah, yang meliputi pernikahan, tuntutan pekerjaan, mengikuti kursus, pindah sekolah, kendala finansial, kondisi kesehatan, kehendak orang tua, hilangnya minat belajar, sanksi dari sekolah (dikeluarkan), kematian, pengangguran, serta kendala administratif. Metode yang diterapkan adalah analisis klaster hierarki menggunakan metode *Ward's* dan *Single Linkage*, menggunakan tahapan standarisasi data, pengujian kecukupan sampel menggunakan *Kaiser-Meyer-Olkin* (KMO), dan mengevaluasi metode terbaik berdasarkan perbandingan varian dalam klaster dan varian antar klaster. Berdasarkan hasil analisis, metode *Single Linkage* terbukti paling unggul dengan capaian 96.73%, mengungguli metode *Ward's* yang memperoleh angka 94.63%. Pengelompokan menggunakan *Single Linkage* menghasilkan 4 klaster, yaitu satu klaster yang memiliki karakter penyebab kursus, mutasi, sakit, permintaan orang tua, dan kesalahan administratif. satu klaster dengan penyebab yang relatif merata. Satu klaster yang didominasi masalah ekonomi. Satu klaster yang didominasi faktor rumah tangga, bekerja, tidak mau sekolah, dikeluarkan oleh sekolah, dan menganggur [5].

2.1 K-Means

K-Means merupakan metode pengelompokan yang bersifat non hierarkis, di mana objek-objek dikelompokkan ke sejumlah klaster yang telah ditentukan sebelumnya. Teknik *K-Means* merupakan strategi yang terbukti efektif untuk mengorganisasi data dan memungkinkan item dalam suatu dataset dikelompokkan berdasarkan sifat-sifat yang sama. Metode ini mengelompokkan objek dengan kemiripan tinggi pada klaster sama, sedangkan objek pada klaster berbeda memiliki kemiripan yang lebih rendah [6].

2.2 Single Linkage

Pendekatan *single linkage* yaitu metode pengelompokan hierarkis aglomeratif yang didasarkan pada jarak terpendek antar objek. Tahap awal, teknik ini mengidentifikasi objek-objek yang memiliki jarak terpendek dalam matriks. Objek yang memiliki jarak terdekat dan digabungkan menjadi satu klaster [7]. Pendekatan ini cenderung menghasilkan klaster memanjang karena penggabungan didasarkan pada anggota klaster yang memiliki jarak terdekat.

2.3 Rata-rata Lama Sekolah (RLS)

Rata-rata Lama Sekolah (RLS) merepresentasikan total waktu (dalam tahun) yang dihabiskan oleh masyarakat untuk mengikuti pendidikan formal [8]. Indikator tersebut berfungsi sebagai tolak ukur utama dalam mengevaluasi mutu dan kualitas pendidikan

penduduk di suatu kawasan tertentu. Penduduk yang menyelesaikan pendidikan SD maka diartikan menempuh pendidikan 6 tahun, tamat SMP dihitung 9 tahun, tamat SMA dihitung 12 tahun tanpa menghitung tinggal kelas atau tidak.

2.4 Harapan Lama Sekolah (HLS)

Harapan Lama Sekolah (HLS) menjadi indikator baru menggantikan angka melek huruf dalam metode baru dalam menentukan Indeks Pembangunan Manusia (IPM) [9]. HLS kini digunakan sebagai indikator utama karena lebih baik dalam menggambarkan kualitas pendidikan. HLS adalah perkiraan durasi sekolah (dalam tahun) yang diharapkan dapat ditempuh oleh seorang anak kelak, di mana perhitungannya diberlakukan untuk penduduk dengan batas usia minimal 7 tahun.

2.5 Angka Partisipasi Sekolah (APS)

Angka Partisipasi Sekolah (APS) merupakan gambaran seberapa banyak penduduk usia sekolah yang mengenyam pendidikan formal. Indikator ini menunjukkan seberapa besar persentase anak pada kelompok umur tertentu yang sedang mengenyam pendidikan di berbagai jenjang [10]. APS menjadi salah satu poin dalam cerminan keberhasilan di bidang pendidikan.

2.6 Angka Partisipasi Kasar (APK)

Angka Partisipasi Kasar (APK) merupakan indikator yang digunakan untuk melihat seberapa besar penduduk yang sedang mengikuti pendidikan pada suatu jenjang. Nilai APK dinyatakan dalam bentuk persentase dan dapat digunakan untuk menggambarkan tingkat partisipasi pendidikan di suatu wilayah. Dalam penelitian ini, APK yang digunakan meliputi APK SD, APK SMP, dan APK SMA. APK SD digunakan untuk melihat partisipasi pendidikan pada jenjang Sekolah Dasar, APK SMP untuk jenjang Sekolah Menengah Pertama, dan APK SMA untuk jenjang Sekolah Menengah Atas. Ketiga indikator tersebut digunakan karena dapat memberikan gambaran mengenai partisipasi pendidikan pada setiap jenjang di kab/kota Kalimantan Barat.

3. METODE PENELITIAN

Penelitian ini menerapkan pendekatan metode kuantitatif yang mengandalkan pengumpulan dan analisis data secara sistematis. Pendekatan ini bertujuan untuk mendeskripsikan suatu fenomena, melakukan pengujian hipotesis, dan memproyeksikan hasil yang spesifik [11]. Perbandingan antara metode *K-Means* dan *Single Linkage* menjadi fokus utama dalam penelitian ini untuk melihat mana yang lebih baik dalam mengklasterkan kab/kota menurut indikator pendidikan. Komponen data yang dipakai berupa data sekunder yang didapat dari publikasi BPS Kalimantan Barat tahun 2024 (<https://kalbar.bps.go.id/id>). Variabel yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Definisi	Satuan
X1	RLS	Rata-rata lama sekolah penduduk	Tahun
X2	HLS	Harapan lama sekolah yang akan ditempuh anak usia sekolah	Tahun
X3	APS	Persentase penduduk usia sekolah yang masih sekolah	Persen
X4	APK	Persentase penduduk yang bersekolah pada suatu jenjang pendidikan tanpa memperhatikan usia.	Persen

3.1 Tahap Penelitian

Tahapan penelitian ini dilakukan sebagai berikut.

1. Mengumpulkan data sekunder berupa RLS, HLS, APK dan APS seluruh kab/kota Kalimantan Barat tahun 2024 dari publikasi resmi BPS.
2. Deskripsi karakteristik data dengan visualisasi dan statistik deskriptif yang mencakup nilai rata-rata, nilai tengah, nilai minimum, nilai maksimum dan standar deviasi.
3. Melakukan standarisasi data menggunakan metode *Z-score*.
4. Melakukan uji asumsi kecukupan data dan multikolinearitas.
5. Menentukan jumlah kluster optimal menggunakan Elbow Method berdasarkan nilai *WCSS (Within-Cluster Sum of Squares)*.
6. Mengelompokkan wilayah di Kalimantan Barat dengan metode kluster optimum.
7. Membandingkan hasil pengelompokan dari metode K-Means dan Single Linkage serta mengidentifikasi karakteristik masing-masing kluster.
8. Kesimpulan dan saran.

3.2 K-Means

K-Means merupakan teknik yang bertujuan memisahkan sekumpulan data ke beberapa kelompok sesuai kemiripan antar objek. Metode ini termasuk dalam *unsupervised learning* karena proses pengelompokan tanpa ada data kategori sebelumnya. *K-Means* banyak digunakan karena perhitungannya relatif mudah dan mampu memberikan hasil pengelompokan yang cukup baik.

Proses pengelompokan data menggunakan *K-Means* dilakukan sebagai berikut.

1. Menentukan jumlah awal kluster (*k*).
2. Menginisialisasi *centroid* secara acak.
3. Menghitung jarak antara tiap data dengan *centroid* menggunakan metode *Euclidean*.
4. Mengelompokkan data ke dalam kluster dengan *centroid* terdekat.
5. Melakukan perhitungan ulang posisi *centroid* berdasarkan rata-rata data di setiap kluster.
6. Mengulangi langkah 3–5 hingga tidak terjadi perubahan signifikan pada anggota kluster atau *centroid* (konvergen).

Jarak *Euclidean* digunakan untuk mengukur dekatnya data dan *centroid*. Persamaan *Euclidean* dituliskan sebagai Persamaan (1) berikut

$$D(p, q) = \sqrt{(X_{1p} - X_{1q})^2 + (X_{2p} - X_{2q})^2 + \dots + (X_{np} - X_{nq})^2} \quad (1)$$

Keterangan:

$D(p, q)$: jarak antara data ke-p dan *centroid* ke-q
 X_{np} : nilai atribut ke-n dari data ke-p
 X_{nq} : nilai atribut ke-n dari data ke-q

3.3 Single Linkage

Single Linkage adalah metode *hierarchical clustering* yang bertujuan menggabungkan kluster berdasarkan jarak terdekat antar anggota kluster. Pada metode ini, dua kluster yang memiliki jarak paling kecil akan bergabung menjadi satu kluster baru. Hasil dari *single linkage* dapat divisualisasikan dalam bentuk dendrogram atau diagram pohon [12]. Proses pengelompokan data menggunakan *Single Linkage* dapat dilakukan sebagai berikut.

1. Menghitung jarak antar seluruh data
2. Menetapkan dua data atau kluster dengan jarak paling dekat
3. Menggabungkan kluster tersebut
4. Menghitung kembali jarak antar kluster
5. Mengulangi proses sampai terbentuk jumlah kluster yang diinginkan

Jarak *Euclidean Distance* digunakan untuk menghitung jarak antar data yang dapat dituliskan sebagai Persamaan (2) berikut ini.

$$d(x, y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (2)$$

Keterangan:

$D(X, Y)$: Jarak antar data
 X_i : Nilai variabel data ke-i
 y_i : Nilai variabel pembanding ke-i
 N : Jumlah variabel

Metode *single linkage* digunakan untuk menghitung jarak antar kluster berdasarkan jarak yang dapat dituliskan sebagai Persamaan (3) berikut ini.

$$D(A, B) = \min\{d(x, y) : x \in A, y \in B\} \quad (3)$$

Keterangan:

$D(A, B)$: Jarak antara kluster A dan B
 $d(x, y)$: Jarak antar objek
 $x \in A$: Anggota pada kluster A
 $x \in B$: Anggota pada kluster B
 $\min$: Nilai jarak terkecil

3.4 Standarisasi Z-Score

Standardisasi *Z-Score* dilakukan untuk menyamakan skala dari setiap variabel yang digunakan dalam penelitian. Hal ini diperlukan karena keenam indikator pendidikan memiliki nilai dan satuan yang berbeda. Dengan melakukan standardisasi, setiap variabel dapat memiliki pengaruh yang lebih seimbang dalam proses pengelompokan. Perhitungan *Z-Score* dilakukan dengan mengurangi nilai suatu data dengan rata-rata variabel, kemudian dibagi dengan standar deviasinya. Rumus *Z-Score* dapat dituliskan sebagai persamaan (4) berikut ini.

$$Z_i = \frac{X_i - \bar{X}}{S} \quad (4)$$

Keterangan :

Z_i : Nilai hasil standardisasi
 X_i : Nilai data ke-i
 $\bar{X}$: Nilai rata-rata
 S : Standar deviasi

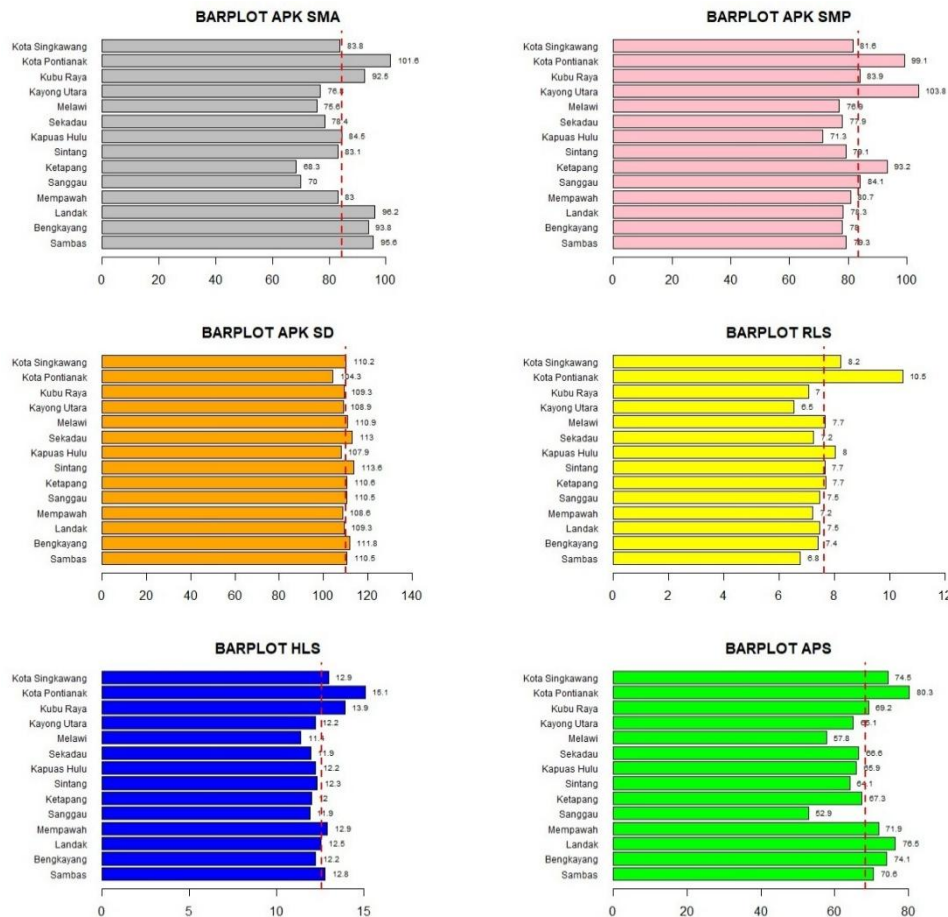
3.5 WCSS dan Elbow Method

Within-Cluster Sum of Squares (WCSS) digunakan untuk melihat seberapa dekat data yang berada dalam satu kluster dengan pusat klasternya. Semakin kecil nilai WCSS, berarti data dalam kluster tersebut semakin berdekatan dan memiliki karakteristik yang lebih mirip. Metode Elbow digunakan untuk memvisualisasikan hubungan antara jumlah kluster dan nilai WCSS, sehingga dapat membantu menentukan jumlah kluster yang paling sesuai [13]. Penentuan jumlah kluster dilakukan dengan melihat perubahan nilai WCSS pada beberapa jumlah kluster. Jumlah kluster dipilih pada titik ketika penurunan WCSS mulai tidak terlalu besar atau membentuk pola seperti siku. Dengan demikian, jumlah kluster yang digunakan tidak terlalu sedikit maupun terlalu banyak.

4. HASIL DAN PEMBAHASAN

4.1 Karakteristik Data Indikator Kualitas Pendidikan

Karakteristik data dimanfaatkan untuk merepresentasikan kondisi suatu wilayah. Berikut adalah grafik yang dapat menggambarkan kondisi kualitas pendidikan di Kalimantan Barat tahun 2024 ditinjau dari aspek RLS, HLS, APS, dan APK.



Sumber: Badan Pusat Statistik, diolah (2026)

Gambar 1. RLS, HLS, APS, APK SD SMP SMA di Kalimantan Barat Tahun 2024

Gambar 1 memuat RLS, HLS, APS, dan APK SD/SMP/SMA di Kalimantan Barat tahun 2024. RLS menunjukkan variasi capaian pendidikan antarwilayah, dengan Pontianak tertinggi (10.5 tahun) yang berarti rata-rata penduduknya menempuh pendidikan hingga SMA, sedangkan Kayong Utara terendah (6.5 tahun), mengindikasikan rata-rata pendidikan baru setingkat SD. Sebagian besar kab/kota berada pada kisaran 7–8 tahun, menandakan capaian pendidikan masih relatif rendah dan belum merata. Wilayah dengan RLS di bawah rata-rata meliputi Kubu Raya, Kayong Utara, Sekadau, Sanggau, Mempawah, Landak, Bengkayang, dan Sambas, sedangkan wilayah dengan RLS di atas atau sama dengan rata-rata meliputi Singkawang, Pontianak, Melawi, Kapuas Hulu, Sintang, dan Ketapang. HLS di Kalimantan Barat secara umum lebih tinggi dibandingkan RLS. Pontianak kembali menjadi yang tertinggi dengan HLS 15.06 tahun, menandakan peluang menempuh pendidikan hingga perguruan tinggi, sementara Melawi terendah (11.4 tahun). Seluruh wilayah memiliki HLS di atas 11 tahun, menunjukkan harapan pendidikan yang cukup baik meski masih ada perbedaan antarwilayah. Wilayah dengan HLS di bawah rata-rata meliputi Kayong Utara, Melawi, Sekadau, Kapuas Hulu, Sintang, Ketapang, Sanggau, dan Bengkayang, sedangkan wilayah dengan HLS di atas atau sama dengan rata-rata meliputi Singkawang, Pontianak, Kubu Raya, Mempawah, Landak, dan Sambas. APS menunjukkan variasi cukup besar, dengan Pontianak tertinggi (80.3%) yang mencerminkan partisipasi pendidikan tinggi, sedangkan Sanggau terendah (52.9%).

Landak dan Singkawang juga menunjukkan APS relatif tinggi, sementara wilayah lain berada pada tingkat sedang hingga rendah, mencerminkan ketimpangan akses pendidikan. Wilayah dengan APS di bawah rata-rata meliputi Kayong Utara, Melawi, Sekadau, Kapuas Hulu, Sintang, Ketapang, dan Sanggau, sedangkan wilayah dengan APS di atas atau sama dengan rata-rata meliputi Singkawang, Pontianak, Kubu Raya, Mempawah, Landak, Bengkayang, dan Sambas. APK SD juga bervariasi, dengan Sintang tertinggi (113.6%) dan Pontianak terendah (104.3%). Bengkayang dan Sekadau juga menunjukkan APK SD relatif tinggi, sedangkan wilayah lain lebih rendah, mencerminkan perbedaan capaian antarwilayah. Wilayah dengan APK SD di bawah rata-rata meliputi Pontianak, Kubu Raya, Kayong Utara, Kapuas Hulu, Mempawah, dan Landak, sedangkan wilayah dengan APK SD di atas atau sama dengan rata-rata meliputi Singkawang, Melawi, Sekadau, Sintang, Ketapang, Sanggau, Bengkayang, dan Sambas. Untuk APK SMP, Kayong Utara memiliki nilai tertinggi (103.8%), sedangkan Kapuas Hulu terendah (71.3%), dengan rata-rata 83.4%. Wilayah dengan APK SMP di bawah rata-rata meliputi Singkawang, Kubu Raya, Melawi, Sekadau, Kapuas Hulu, Sintang, Mempawah, Landak, Bengkayang, dan Sambas, sedangkan yang di atas atau sama dengan rata-rata meliputi Pontianak, Kayong Utara, Ketapang, dan Sanggau. Kondisi ini menunjukkan bahwa partisipasi pendidikan SMP masih berbeda antarkabupaten/kota di Kalimantan Barat. APK SMA tertinggi mencapai 101.6%, sedangkan Ketapang terendah (68.3%), dengan rata-rata 84.5%. Wilayah dengan APK SMA di bawah rata-rata meliputi Kayong Utara, Melawi, Sekadau, Sintang, Ketapang, Sanggau, dan Mempawah, sedangkan yang di atas atau sama dengan rata-rata meliputi Singkawang, Pontianak, Kubu Raya, Kapuas Hulu, Landak, Bengkayang, dan Sambas. Kondisi tersebut menegaskan bahwa partisipasi pendidikan SMA juga masih menunjukkan perbedaan antar kab/kota di Kalimantan Barat.

4.2 Asumsi Kecukupan Data dan Multikolinearitas

Sebelum analisis kluster dilakukan, data diperiksa terlebih dahulu melalui uji asumsi untuk memastikan bahwa data yang digunakan memenuhi syarat analisis. Pengujian ini meliputi uji kecukupan data dan multikolinearitas. Pemeriksaan kecukupan data dilakukan dengan nilai *Kaiser-Meyer-Olkin* (KMO), sedangkan multikolinearitas dilakukan dengan korelasi *pearson* untuk melihat arah hubungan antar variabel. Uji Asumsi kecukupan data menggunakan metode KMO yang dilakukan digunakan untuk melihat apakah data yang digunakan representatif sehingga layak untuk dianalisis sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Hasil KMO

Keterangan	Nilai
<i>Overall</i> MSA (KMO)	0.69

Sumber: Data diolah menggunakan R-Studio (2026)

Tabel 2 menunjukkan adanya nilai KMO sebesar 0.69 berada di atas 0.50, sehingga memenuhi uji asumsi kecukupan data dan variabel dapat dianalisis lebih lanjut.

Selanjutnya Uji multikolinearitas digunakan untuk mengidentifikasi ada atau tidaknya hubungan antar variabel prediktor [14].

Tabel 3. Hasil Asumsi Multikolinearitas

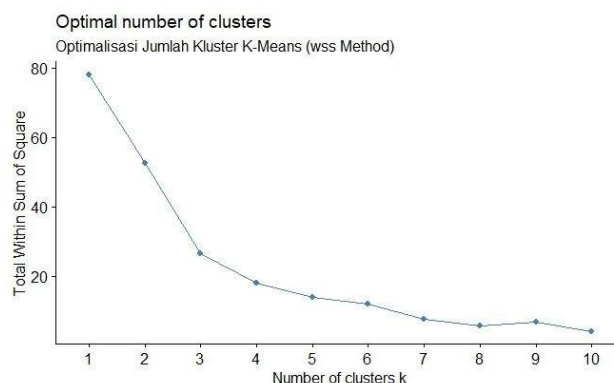
	<i>RLS</i>	<i>HLS</i>	<i>APS</i>	<i>APK SD</i>	<i>APK SMP</i>	<i>APK SMA</i>
RLS	1.0000	0.6055	0.4113	-0.6011	0.2274	0.3438
HLS	0.6055	1.0000	0.6861	-0.7038	0.3797	0.7126
APS	0.4113	0.6861	1.0000	-0.4495	0.1439	0.7779
APK SD	-0.6011	-0.7038	-0.4495	1.0000	-0.4547	-0.4229
APK SMP	0.2274	0.3797	0.1439	-0.4547	1.0000	-0.0867
APK SMA	0.3438	0.7126	0.7779	-0.4229	-0.867	1.0000

Sumber: Data diolah menggunakan R-Studio (2026)

Tabel 3 memuat hasil pengujian uji asumsi multikolinearitas melalui korelasi *Pearson* memperlihatkan bahwa sebagian besar nilai korelasi antar variabel berada di bawah 0.7. Akan tetapi, terdapat beberapa variabel dengan korelasi relatif kuat, yaitu HLS dengan APK SD sebesar -0.7038, HLS dengan APK SMA sebesar 0.7126, dan APS dengan APK SMA sebesar 0.7779. Hal ini menunjukkan adanya hubungan yang relatif kuat pada beberapa pasangan variabel yang digunakan.

4.3 Komparasi Metode *K-Means Clustering* dan *Single Linkage Hierarchical Clustering*

Metode *K-Means* yang diterapkan untuk mengelompokkan wilayah berdasarkan indikator kualitas pendidikan dengan jumlah kluster yang diuji hingga $k=10$. Penentuan jumlah kluster optimal dilakukan menggunakan metode Elbow berdasarkan nilai *Within-Cluster Sum of Squares* (WCSS). Hasil ini divisualisasikan ke dalam grafik pada Gambar 2.

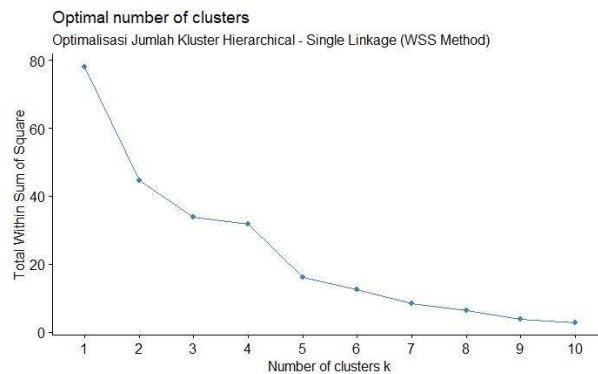


Sumber: Data diolah menggunakan R-Studio (2026)

Gambar 2. Optimalisasi Jumlah Kluster

Berdasarkan Gambar 2 nilai WCSS mengalami penurunan seiring dengan bertambahnya jumlah kluster. Penurunan nilai WCSS terlihat cukup besar dari $k=1$ hingga $k=3$, kemudian penurunan cenderung lebih landai. Berdasarkan titik siku (*elbow*) pada grafik, jumlah kluster optimal pada metode *K-Means* adalah sebanyak tiga kluster. Berdasarkan nilai WCSS metode *K-Means* mengalami penurunan dari 78.0005 pada $k=1$ menjadi 2.7829 pada $k=10$. Penurunan WCSS yang cukup besar terjadi hingga $k=3$. Hal ini mendukung penentuan jumlah kluster optimalnya adalah sebanyak tiga kluster.

Kemudian metode *Single Linkage Hierarchical Clustering* dilakukan sebagai metode pembandingan. Metode ini mengelompokkan objek berdasarkan jarak terdekat antar anggota kluster. Penentuan jumlah kluster terbaik dilakukan dengan melihat nilai WCSS pada beberapa kemungkinan jumlah kluster.



Sumber: Data diolah menggunakan R-Studio (2026)

Gambar 3. Optimalisasi Jumlah Kluster

Berdasarkan Gambar 3 nilai WCSS mengalami penurunan seiring dengan bertambahnya jumlah kluster. Penurunan nilai WCSS terlihat cukup besar dari $k=1$ hingga $k=3$, kemudian penurunan cenderung lebih landai. Berdasarkan titik siku (*elbow*) pada grafik, jumlah kluster optimal pada metode *Single Linkage* adalah sebanyak tiga kluster. Berdasarkan nilai WCSS metode *Single Linkage* mengalami penurunan dari 78.0005 pada $k=1$ menjadi 2.7829 pada $k=10$. Jumlah kluster optimal pada metode *Single Linkage* ditentukan sebanyak tiga kluster dengan nilai WCSS sebesar 33.7651.

4.4 Pengelompokkan Kab/Kota dengan Metode *Clustering* yang Optimum

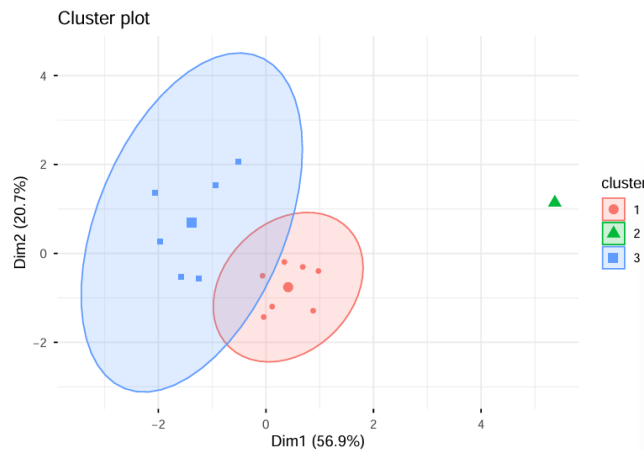
Untuk memetakan kab/kota di Kalimantan Barat menurut indikator pendidikannya, penelitian ini mengaplikasikan algoritma *K-Means* serta *Single Linkage*. Sementara itu, evaluasi untuk menentukan jumlah kelompok yang paling optimal diukur menggunakan skor WCSS. Nilai skor WCSS digunakan untuk mengukur kualitas hasil klusterisasi, di mana semakin kecil nilainya maka semakin baik kluster yang terbentuk [15].

Tabel 4. Hasil Perhitungan Nilai WCSS

	Metode	Nilai WCSS
K = 3	<i>K-Means</i>	26.6143
K = 3	<i>Single Linkage</i>	33.7651

Sumber: Data diolah menggunakan R-Studio (2026)

Berdasarkan Tabel 4, metode *K-Means* dan *Single Linkage* sama-sama menghasilkan jumlah kluster optimal sebanyak 3 kluster. Metode *K-means* dengan 3 kluster menghasilkan nilai WCSS sebesar 26.6143, sedangkan metode *Single Linkage* dengan 3 kluster menghasilkan nilai WCSS sebesar 33.7651. Nilai WCSS yang lebih kecil pada metode *K-means* menunjukkan bahwa anggota dalam kluster memiliki kemiripan karakteristik yang lebih tinggi dibandingkan metode *Single Linkage*.



Sumber: Data diolah menggunakan R-Studio (2026)

Gambar 4. Plot Hasil Klasterisasi Algoritma *K-Means*

Berdasarkan Gambar 4, visualisasi klaster menggunakan *K-Means* dengan jumlah klaster optimum sebanyak tiga klaster. Klaster pertama ditunjukkan oleh warna merah dengan simbol lingkaran, klaster kedua ditunjukkan oleh warna hijau dengan simbol segitiga, sedangkan klaster ketiga ditunjukkan dengan warna biru bersimbol persegi.

Tabel 5. Hasil Klasterisasi Terbaik

Klaster	Jumlah Anggota	Kabupaten/Kota
1	6	Sanggau, Ketapang, Sintang, Sekadau, Melawi, Kayong Utara
2	1	Kota Pontianak
3	7	Sambas, Bengkayang, Landak, Mempawah, Kapuas Hulu, Kubu Raya, Kota Singkawang

Sumber: Data diolah menggunakan R-Studio (2026)

Tabel 5 yang disajikan di atas dapat di jelaskan sebagai berikut:

- Klaster 1 :

Klaster pertama terdiri atas 6 wilayah, yaitu Sanggau, Ketapang, Sintang, Sekadau, Melawi, Kayong Utara. Kelompok ini mencatatkan performa yang relatif rendah pada berbagai indikator pendidikan. Hal ini di tandai oleh belum meratanya akses layanan pendidikan, yang dibuktikan dengan rendahnya angka RLS, HLS, APS, APK SD SMP SMA. Minimnya capaian tersebut menandakan terbatasnya peluang masyarakat untuk melanjutkan studi serta kurang optimalnya partisipasi pendidikan. Jika tidak segera diatasi, situasi ini berpotensi menghambat terwujudnya Tujuan Pembangunan Berkelanjutan (SDGs) poin ke-4 yaitu pendidikan yang inklusif dan berkualitas sekaligus menghambat tercapainya Visi Indonesia Emas 2045. Oleh sebab itu, akselerasi peningkatan mutu pendidikan di wilayah tersebut harus menjadi prioritas, khususnya guna mendongkrak indikator RLS, HLS, APS, APK SD SMP SMA.

- Klaster 2 :

Klaster kedua diisi oleh satu wilayah, yakni Pontianak, yang mencatatkan prestasi pendidikan lebih tinggi dari daerah-daerah lain. Tingginya capaian ini sangat wajar mengingat fasilitas dan akses pendidikan di wilayah perkotaan jauh lebih memadai. Hal ini terlihat dari tingginya angka RLS, HLS, dan APS yang menandakan tingginya

kesadaran dan peluang masyarakat untuk terus bersekolah. Karena wilayah dengan pendidikan rendah bisa menghambat Visi Indonesia Emas 2045 dan target pemerataan pendidikan dalam SDGs poin ke-4. Capaian positif Pontianak ini harus dijaga dan perlu terus didorong agar RLS, HLS, APS, APK SD SMP SMA terus semakin meningkat.

- **Klaster 3 :**

Klaster ketiga mencakup 7 wilayah, yaitu Sambas, Bengkayang, Landak, Mempawah, Kapuas Hulu, Kubu Raya, Kota Singkawang. Kelompok ini mencatatkan performa pendidikan pada tingkat menengah, tidak setinggi Pontianak namun juga tidak serendah kelompok ketiga. Kondisi ini ditandai oleh capaian RLS, HLS, APS, APK SD SMP SMA, yang relatif sedang, mengindikasikan bahwa akses dan partisipasi pendidikan di wilayah ini sudah berjalan, tetapi belum optimal. Untuk mendekati capaian klaster pertama serta mendukung SDGs poin ke-4 dan Visi Indonesia Emas 2045, diperlukan upaya peningkatan yang berkelanjutan pada indikator RLS, HLS, APS, APK SD SMP SMA di wilayah-wilayah ini.

5. KESIMPULAN

Berdasarkan hasil penelitian, didapatkan bahwa metode *single linkage* dan *k-means* menunjukkan jumlah klaster paling optimal yaitu tiga klaster. Berdasarkan data yang ada, *K-Means Clustering* memiliki nilai WCSS sebesar 26.6143 dan *Single Linkage* memiliki nilai WCSS sebesar 33.7651. Dengan demikian, *K-means* dengan tiga klaster memiliki nilai WCSS yang lebih kecil dibandingkan *Single Linkage* dengan tiga klaster, sehingga *K-means* dengan tiga klaster dipilih sebagai metode *clustering* yang optimum dalam penelitian ini. Pemerintah daerah khususnya Kalimantan Barat dapat memanfaatkan hasil tersebut untuk kebijakan tepat sasaran dengan memprioritaskan daerah yang masuk anggota klaster rendah tanpa mengabaikan upaya penguatan pada klaster menengah agar dapat mendekati capaian klaster tinggi. Pemerintah Kalimantan Barat dapat melakukan peningkatan akses pendidikan melalui beasiswa, pelatihan kerja, pendidikan vokasi, serta pemerataan sarana dan kualitas tenaga pendidik untuk mendukung tujuan pembangunan berkelanjutan ke-4 yaitu Pendidikan Berkualitas. Untuk penelitian selanjutnya, disarankan untuk menambahkan indikator pendidikan lainnya, seperti tingkat melek huruf, angka putus sekolah, rasio guru terhadap murid, atau IPM, sehingga karakteristik pengelompokan daerah dapat dianalisis secara lebih komprehensif. Selain itu disarankan juga untuk menggunakan metode lainnya seperti *K-Medoids* dan SOM agar dapat membandingkan hasilnya.

DAFTAR PUSTAKA

- [1] M. N. H. T. Putra, "Klasterisasi Kabupaten / Kota di Provinsi Kalimantan Barat Berdasarkan Rasio Guru dengan Murid di Tingkat SD, SMP, dan SMA Serta IPM Tahun 2022 Menggunakan Metode K-Means Clustering," *Forum Anal. Stat.*, vol. 3, no. 1, pp. 42–50, 2023.
- [2] A. T. Basalamah and R. Setyadi, "Penerapan Algoritma K-Means Clustering Pada

Tingkat Penyelesaian Pendidikan Di Provinsi Indonesia,” *J-ICOM J. Inform. dan Teknol. Komput.*, vol. 04, no. 02, pp. 114–121, 2023.

- [3] R. K. Mashitawai *et al.*, “Penerapan Algoritma K-Means dan K-Medoids untuk Klasterisasi Wilayah Berdasarkan Indikator Kualitas Pendidikan di Provinsi Nusa Tenggara Timur Tahun 2023,” *Semin. Nas. Sains dan Teknol. Seri IV*, vol. 2, no. 2, pp. 48–57, 2025.
- [4] D. R. Thamrin and D. Murni, “Analisis Cluster Hierarki Metode Single Linkage Pada Kabupaten / Kota di Provinsi Sumatera Barat Berdasarkan Indikator Kesehatan,” *J. Math. UNP*, vol. 7, no. 3, pp. 45–51, 2022.
- [5] M. R. Rozikin and W. Pramesti, “Implementasi Metode Ward ’ s dan Single Linkage dalam Pengelompokan Anak Putus Sekolah di Provinsi Jawa Timur,” *Cakrawala J. Litbang Kebijak.*, vol. 17, pp. 77–90, 2023.
- [6] M. Amelia, A. Faqih, and A. R. Rinaldi, “Penerapan Metode K-Means Clustering dalam Pemetaan Kemiskinan Kabupaten/Kota di Indonesia untuk Perencanaan Kebijakan yang Tepat,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, pp. 437–444, 2025.
- [7] N. A. Maori *et al.*, “Metode Single Linkage Pada Agglomerative Hierarchical Clustering Dalam Penentuan Tingkat Kepuasan Mahasiswa Terhadap Layanan Akademik,” *SIMETRIS*, vol. 15, no. 2, pp. 1–10, 2024.
- [8] BPS, “Rata - Rata Lama Sekolah Menurut Kab/Kota(Tahun),2024.”
- [9] BPS, “Indeks Pembangunan Manusia(IPM) Menurut Provinsi Tahun 2024.”
- [10] BPS, “Angka Partisipasi Sekolah Menurut Kab/Kota(Tahun),2024.”
- [11] H. Qudratuddarsi, E. Meivawati, and R. Saputra, “Pelatihan Penelitian Metode Kuantitatif dan Systematic literature review bagi Dosen dan Mahasiswa,” *Beru’-beru’Jurnal Pengabd. Kpd. Masy.*, vol. 3, no. 1, pp. 22–32, 2024.
- [12] S. A. Firda, H. A. Widiyanto, R. Adelisa, and E. Sunandi, “Perbandingan Single Linkage dan K-Medoids Pada Data Penghasilan Dan Demografi Area Komunitas Chicago,” *J. FOURIER*, vol. 14, no. 2, pp. 68–79, 2025, doi: 10.14421/fourier.2025.142.68-79.
- [13] L. C. Fahsyah, C. Wijaya, F. M. Bintang, J. J. Mulyono, F. Ramadhan, and F. Amsury, “Penerapan Clustering K-Means Untuk Segmentasi Pelanggan Pada Bisnis Retail,” *HOAQ (High Educ. Organ. Arch. Qual. J. Teknol. Inf.*, vol. 17, no. 1, pp. 38–52, 2026, doi: 10.52972/hoaq.vol17no1.p38-52.
- [14] T. Apriliana and E. Widodo, “Analisis Cluster Hierarki untuk Pengelompokan Provinsi di Indonesia berdasarkan Jumlah Base Transceiver Station dan Kekuatan Sinyal,” *KONSTELASI: Konvergensi Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 286–296, 2023.
- [15] M. I. Ramadhan, A. Nazir, M. Irsyad, S. Sanjaya, and F. Syafria, “Analisis Clustering Menggunakan Metode K-Means untuk Mengidentifikasi Pola Kepuasan Alumni,” *Indones. J. Mach. Learn. Comput. Sci.*, vol. 6, no. 1, pp. 31–41, 2026.