

Comparison of K-Means and Ward Methods in Clustering Indonesian Provinces Based on Household Basic Service Access

Nurul Mulya Syahwa, Fajri Juli Rahman Nur Zentrato, Muhammad Arief Rivano ,
Zamahsary martha, dan Tessy Octavia Mukhti*

Departemen Statistika, Universitas Negeri Padang, Kota Padang, Indonesia

*Corresponding author: tessyoctaviam@fmipa.unp.ac.id

Submitted : 12 Januari 2026

Revised : 29 Januari 2026

Accepted : 03 Februari 2026

ABSTRACT

Disparities in household basic service access across provinces in Indonesia remain a key issue in regional development. Basic services such as access to improved drinking water, proper sanitation, electricity, and adequate housing are essential indicators of household welfare, making regional classification necessary to identify similarities and disparities among provinces. This study aims to cluster Indonesian provinces based on household basic service access indicators and to compare the performance of the K-Means method and Hierarchical Clustering using the Ward approach. The analysis was conducted using numerical data with Euclidean distance as a measure of similarity. The optimal number of clusters was determined using the Silhouette plot and further validated using the Silhouette Coefficient. The results indicate that both K-Means and Ward methods produce two optimal clusters representing provinces with relatively high and relatively low levels of household basic service access. Centroid analysis reveals clear differences between clusters across all indicators, particularly in electricity access and sanitation. Furthermore, the evaluation of clustering quality shows that the Ward method yields a higher Silhouette Coefficient than the K-Means method, indicating more compact clusters and better separation between clusters. Therefore, the Ward method is considered more effective in mapping patterns of household basic service access across provinces. The findings of this study can support regional planning by providing a clearer understanding of disparities in household basic service access in Indonesia.

Keywords: Cluster analysis; K-Means; Ward method; Household basic services; Indonesian provinces



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Akses terhadap layanan dasar rumah tangga seperti air minum layak, sanitasi layak, dan listrik merupakan komponen penting dalam meningkatkan kualitas hidup masyarakat serta kesejahteraan wilayah. Layanan dasar tersebut berperan dalam mendukung kesehatan masyarakat, aktivitas ekonomi rumah tangga, dan produktivitas tenaga kerja. Penelitian empiris menunjukkan bahwa peningkatan akses air minum layak dan sanitasi layak berpengaruh signifikan terhadap penurunan tingkat kemiskinan suatu wilayah, karena perbaikan akses layanan dasar berkaitan dengan penurunan beban kesehatan, peningkatan kapasitas kerja, serta efisiensi pengeluaran rumah tangga (Putri & Samsuddin, 2025). Selain itu, akses terhadap infrastruktur dasar seperti listrik, air minum, dan sanitasi juga memiliki hubungan positif dengan indikator kesejahteraan makro seperti *Human Development Index*, yang menegaskan bahwa layanan dasar merupakan bagian penting dalam pembangunan sosial ekonomi wilayah (Chrisnahutama, 2023). Meskipun peran layanan dasar terhadap kesejahteraan telah dibuktikan secara empiris, distribusi akses layanan dasar antarwilayah belum sepenuhnya merata. Perbedaan tingkat akses terhadap air minum layak, sanitasi layak, dan listrik antarprovinsi mencerminkan adanya variasi kondisi kesejahteraan wilayah. Distribusi akses layanan dasar rumah tangga antarprovinsi di Indonesia menunjukkan variasi yang cukup besar. Sebagai contoh, Provinsi Sumatera Utara memiliki persentase akses air minum layak sebesar 93,24%, sanitasi layak 87,47%, akses listrik 76,89%, dan hunian layak 99,28%, yang mencerminkan tingkat akses layanan dasar rumah tangga yang relatif tinggi. Sebaliknya, Provinsi Papua Pegunungan menunjukkan kondisi yang jauh lebih rendah dengan persentase akses air minum layak sebesar 32,89%, sanitasi layak 16,34%, akses listrik 9,53%, dan hunian layak 20,78%. Perbedaan yang sangat kontras

antarprovinsi tersebut mengindikasikan adanya ketimpangan akses layanan dasar rumah tangga yang signifikan. Kondisi ini menunjukkan perlunya pendekatan analisis yang mampu memetakan perbedaan karakteristik wilayah secara objektif dan sistematis agar pola ketimpangan kesejahteraan dapat diidentifikasi secara lebih jelas dan berbasis data (Ramadhan et al., 2023).

Salah satu pendekatan statistik yang banyak digunakan untuk tujuan tersebut adalah analisis kluster. Analisis kluster bertujuan mengelompokkan objek ke dalam beberapa kelompok berdasarkan tingkat kemiripan karakteristik tertentu. Dalam konteks data layanan dasar rumah tangga, analisis kluster memungkinkan penyederhanaan data multivariat menjadi kelompok provinsi yang relatif homogen, sehingga wilayah dengan karakteristik akses layanan dasar yang serupa dapat dikelompokkan dalam satu kluster. Pendekatan ini memudahkan interpretasi pola kesejahteraan wilayah serta mendukung analisis perbedaan antarwilayah secara komprehensif (Ramadhan et al., 2023). Berbagai metode klusterisasi dapat digunakan dalam analisis wilayah, di antaranya metode K-Means dan *hierarchical clustering* dengan pendekatan Ward. Metode K-Means merupakan metode klusterisasi non-hierarki yang banyak digunakan karena kesederhanaan proses perhitungan dan kemampuannya dalam mengelompokkan data numerik. Penelitian terdahulu menunjukkan bahwa K-Means mampu menghasilkan kluster wilayah yang jelas dan mudah diinterpretasikan ketika diterapkan pada indikator sosial ekonomi dan kesejahteraan wilayah, sehingga metode ini dinilai efektif dan efisien untuk pemetaan kesejahteraan berbasis data kuantitatif (Yusuf et al., 2025).

Selain K-Means, metode *hierarchical clustering* dengan pendekatan Ward juga dinyatakan layak dan efektif digunakan dalam klusterisasi wilayah. Metode Ward membentuk kluster dengan meminimalkan peningkatan variasi di dalam kluster, sehingga menghasilkan kelompok wilayah yang lebih kompak dan stabil. Studi terdahulu pada pengelompokan data sosial ekonomi wilayah menunjukkan bahwa metode Ward mampu menggambarkan struktur pengelompokan data secara lebih mendalam melalui dendrogram dan menghasilkan kluster yang representatif terhadap kondisi wilayah, khususnya pada tahap eksplorasi struktur data (Iklima & Pujiyanta, 2023). Perbedaan karakteristik antara metode K-Means dan Ward menunjukkan bahwa hasil klusterisasi sangat dipengaruhi oleh metode yang digunakan. Penelitian yang membandingkan kedua metode tersebut menyatakan bahwa penerapan metode klusterisasi yang berbeda pada data yang sama dapat menghasilkan struktur kluster yang berbeda pula, sehingga perbandingan metode menjadi penting untuk memperoleh hasil pengelompokan wilayah yang lebih sesuai dengan karakteristik data dan tujuan analisis (Iklima & Pujiyanta, 2023).

Berdasarkan uraian tersebut, penelitian ini secara khusus bertujuan untuk membandingkan metode K-Means dan metode Ward dalam klusterisasi provinsi di Indonesia berdasarkan indikator layanan dasar rumah tangga, yaitu persentase rumah tangga dengan akses air minum layak, sanitasi layak, listrik PLN, dan hunian layak. Perbandingan ini dilakukan untuk menilai perbedaan hasil klusterisasi yang dihasilkan oleh kedua metode serta menentukan metode yang paling sesuai dalam memetakan kesejahteraan wilayah secara objektif dan informatif. Hasil penelitian diharapkan dapat memberikan kontribusi dalam pengembangan analisis kluster wilayah dan menjadi dasar pendukung perumusan kebijakan pembangunan berbasis data.

II. METODE PENELITIAN

A. Data Penelitian

Penelitian ini menggunakan data sekunder berupa persentase rumah tangga dengan akses layanan dasar pada tingkat provinsi di Indonesia. Variabel yang digunakan dalam penelitian ini terdiri atas empat indikator, yaitu akses terhadap air minum layak, sanitasi layak, sumber listrik dari PLN, serta kondisi hunian yang layak. Data diperoleh dari publikasi resmi Badan Pusat Statistik melalui Survei Sosial Ekonomi Nasional (Susenas) dan dapat diakses pada <https://www.bps.go.id>. Data disusun dalam bentuk tabel lintas provinsi dan digunakan untuk membandingkan hasil klusterisasi provinsi menggunakan metode K-Means dan *hierarchical clustering* dengan pendekatan Ward, guna mengelompokkan provinsi berdasarkan tingkat kemiripan akses layanan dasar rumah tangga.

B. Teknik Analisis data

Teknik analisis data yang digunakan dalam penelitian ini adalah analisis kluster dengan metode *K-Means Clustering* dan *Hierarchical Clustering* menggunakan pendekatan *Ward*. Kedua metode diterapkan untuk mengelompokkan provinsi di Indonesia berdasarkan tingkat kemiripan akses layanan dasar rumah tangga. Seluruh variabel penelitian dinyatakan dalam bentuk persentase, sehingga memiliki satuan dan rentang skala yang sama (0–100%). Oleh karena itu, tidak dilakukan proses standarisasi data, karena perbedaan skala antar variabel tidak berpotensi mendistorsi perhitungan jarak antar objek. Penggunaan data dalam skala yang seragam dinilai telah memenuhi syarat penerapan analisis kluster berbasis jarak (Izzuddin & Wijayanto, 2024).

1) Uji Asumsi Analisis Klaster

(a) Uji Multikolinearitas

Terjadinya multikolinearitas merupakan salah satu pelanggaran dalam analisis cluster. Multikolinearitas merupakan adanya hubungan yang kuat antara variabel-variabel. Multikolinearitas adalah permasalahan yang harus diwaspadai dalam analisis cluster, karena dapat menyulitkan dalam mengidentifikasi pengaruh/efek masing-masing variabel dan memengaruhi hasil akhir pengelompokan (Salsabila et al., 2022).

Multikolinearitas dapat diidentifikasi menggunakan Variance Inflation Factor (VIF). Masalah multikolinearitas ditunjukkan apabila nilai VIF lebih besar dari 10, sedangkan nilai VIF kurang dari atau sama dengan 10 menunjukkan tidak adanya multikolinearitas yang signifikan (Matdoan et al., 2023). Nilai VIF dihitung menggunakan Persamaan :

$$VIF_j = \frac{1}{1-R_j^2} \quad (1)$$

Di mana R_j^2 merupakan koefisien determinasi dari variabel ke-j terhadap seluruh variabel lain.

(b) Uji KMO

Dugaan bahwa sampel cukup mewakili populasi dapat dinilai dengan menerapkan uji Kaiser-Meyer-Olkin (KMO). Uji ini menilai kecukupan sampling secara keseluruhan serta kecukupan untuk setiap indikator individu. Uji ini sering digunakan untuk menentukan apakah data cocok untuk analisis cluster dan apakah terdapat struktur laten atau faktor mendasar yang menjelaskan korelasi antarvariabel. Statistik KMO dihitung menggunakan rumus yang ditunjukkan dalam Persamaan 2.

$$KMO = \frac{\sum_{j=1}^p \sum_{l=1, l \neq j}^p r_{jl}^2}{\sum_{j=1}^p \sum_{l=1, l \neq j}^p r_{jl}^2 + \sum_{j=1}^p \sum_{l=1, l \neq j}^p \rho_{jl}^2} \quad (2)$$

di mana:

p = jumlah variabel

r_{ij} = koefisien korelasi antara variabel i dan j

ρ_{ij} = koefisien korelasi parsial antara variabel i dan j

Nilai Kaiser–Meyer–Olkin (KMO) digunakan untuk menilai kecukupan sampel sebelum analisis lanjutan. Dalam konteks *Exploratory Factor Analysis* (EFA) yaitu metode statistik yang digunakan untuk mengidentifikasi struktur laten atau faktor-faktor yang mendasari hubungan antarvariabel yang diamati, tanpa menetapkan struktur faktor sebelumnya. Nilai KMO yang mendekati atau lebih besar dari 0,5 menunjukkan bahwa sampel memadai untuk dilanjutkan ke tahap analisis berikutnya ($KMO \geq 0,5$) dan bahwa data cocok untuk dianalisis lebih lanjut (Bernard et al., 2020).

2) K-Means Clustering

Metode *K-Means* merupakan metode pengelompokan data yang membagi objek ke dalam beberapa klaster berdasarkan tingkat kemiripan karakteristik, sehingga objek dengan karakteristik yang serupa berada dalam klaster yang sama dan objek yang berbeda berada dalam klaster yang berbeda. Proses pengelompokan dilakukan dengan menentukan pusat klaster dan mengelompokkan data berdasarkan kedekatan terhadap pusat klaster tersebut. Tahapan ini dilakukan secara berulang hingga diperoleh pembagian klaster yang stabil dan optimal. Proses *K-Means* dilakukan dengan langkah-langkah sebagai berikut (Amalina et al., 2022) :

- Proses algoritma K-Means diawali dengan menentukan jumlah klaster (K) yang akan dibentuk. Untuk menentukan jumlah *cluster* dilihat dari grafik silhouette diperoleh dengan menghitung nilai Silhouette rata-rata untuk setiap jumlah klaster yang diuji. Penentuan jumlah klaster optimal dilakukan dengan mengamati pola grafik, di mana jumlah klaster terbaik ditunjukkan oleh nilai k yang menghasilkan puncak awal sebelum terjadi penurunan nilai Silhouette secara konsisten (Kuswardana et al., 2025).
- Selanjutnya, ditetapkan pusat klaster awal (centroid) secara acak sebanyak jumlah klaster yang ditentukan.
- Setelah itu, jarak antara setiap data dan masing-masing centroid dihitung menggunakan jarak Euclidean, kemudian setiap data dikelompokkan ke dalam klaster dengan jarak terdekat terhadap centroid. persamaan *Euclidean Distance* antara lain :

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \quad (3)$$

Keterangan :

x_i : data kriteria

μ_j : *centroid* pada cluster ke- j

- (d) Setiap data selanjutnya diklasifikasikan ke dalam kluster berdasarkan jarak terdekat terhadap pusat kluster. Setelah proses pengelompokan dilakukan, nilai pusat kluster diperbarui dengan menghitung rata-rata data yang tergabung dalam kluster tersebut dengan rumus :

$$C_k = \frac{1}{n_k} \sum d_i \quad (4)$$

Keterangan :

n_k : Jumlah data dalam *cluster* k

d_i : Jumlah dari nilai jarak yang masuk dalam masing-masing *cluster*.

- (e) Proses pengelompokan dan pembaruan pusat kluster dilakukan secara berulang hingga tidak terjadi perubahan keanggotaan kluster.
(f) Pada tahap akhir, nilai pusat kluster yang diperoleh pada iterasi terakhir digunakan sebagai dasar dalam penentuan klasifikasi data.

3) Metode Ward

Metode Ward merupakan pendekatan hierarki berbasis dispersi yang banyak diterapkan dalam analisis kluster. Tujuan utamanya adalah untuk meminimalkan variansi dalam kluster. Metode ini menerapkan prinsip analisis variansi untuk menghitung jarak antar kluster yang membedakannya dengan metode hierarki lainnya. Keunggulan metode *Ward* terletak pada kemampuannya yang lebih efisien dan cenderung membentuk kluster berukuran kecil (Matdoan et al., 2023). Tahapan untuk metode *Ward* adalah sebagai berikut :

- (a) Pendefinisian k (jumlah kluster yang akan dibentuk dengan *Silhouette Coefficient*).
(b) Menghitung jarak antar kluster (misalnya dengan Euclidean Distance), dimana masing-masing objek dianggap sebagai satu kluster sehingga $n = N$. Penghitungan Euclidean distance adalah sebagai berikut :

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (5)$$

dimana, D_e = jarak Euclidean; i = banyak objek; (x,y) = koordinat objek; dan (s,t) = koordinat *centroid*.

- (c) Menentukan 2 dari N kluster melalui nilai SSE (Sum of Square Error) terkecil sebagai kluster pertama. SSE diperoleh melalui persamaan :

$$SSE = \sum_{i=1}^n (x_i - \bar{x})'(x_i - \bar{x}) \quad (6)$$

dimana, x_i = vektor kolom objek i ; $\bar{x}$ = vektor kolom berisi rata rata objek dalam kluster; dan n = banyaknya objek.

- (d) Tinjau kembali $N-1$ himpunan agar dapat meminimalkan heterogenitas. Sehingga, N kluster dikurangi $N - 1$.
(e) Ulangi langkah b dan c, hingga didapatkan semua objek yang telah dikelompokkan ke dalam satu atau beberapa kluster.

4) Memilih Metode Terbaik dengan Melihat *Silhouette Coefficient*

Kualitas hasil klusterisasi dievaluasi menggunakan *Silhouette Coefficient*, yang mengukur tingkat kohesi dan separasi kluster. Nilai *silhouette* digunakan sebagai dasar pemilihan kluster terbaik, di mana nilai yang lebih tinggi menunjukkan struktur kluster yang lebih baik (Vania & Nurina Sari, 2023).

Nilai *silhouette* untuk setiap objek ke- i dirumuskan sebagai berikut:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (7)$$

Keterangan:

- $s(i)$: nilai *silhouette* untuk objek ke- i
- $a(i)$: rata-rata jarak objek ke- i dengan seluruh objek lain yang berada dalam kluster yang sama (kohesi kluster)

- $b(i)$: nilai rata-rata jarak terkecil antara objek ke- i dengan seluruh objek pada kluster terdekat lainnya (separasi kluster)

Kategori nilai *sillhouette coefficient* ditunjukkan pada Tabel 1.

Tabel 1. Kategori Evaluasi Nilai *Sillhouette Coefficient*

Rentang Nilai <i>Sillhouette Coefficient</i>	Kategori
0,71 - 1	Struktur kuat
0,51 - 0,7	Struktur baik
0,26 - 0,5	Struktur lemah
$\leq 0,25$	Tanpa struktur

Berdasarkan Tabel 1, nilai *sillhouette coefficient* digunakan untuk mengevaluasi kualitas hasil pengelompokan.

III. HASIL DAN PEMBAHASAN

A. Statistika Deskriptif

Statistika deskriptif digunakan untuk memberikan gambaran umum mengenai karakteristik data penelitian yang meliputi 38 provinsi di Indonesia. Berikut statistika deskriptif dari variabel penelitian yang dapat dilihat pada Tabel 2.

Tabel 2. Statistika Deskriptif Variabel Penelitian

Variabel	Min	Maks	Mean	Median
Persentase akses air minum layak (X1)	32,89	99,98	87,75	90,01
Persentase sanitasi layak (X2)	16,34	98,20	83,02	85,98
Persentase listrik dari PLN (X3)	20,78	100,00	93,36	98,46
Persentase hunian layak (X4)	9,53	88,31	65,13	68,42

Statistika deskriptif pada Tabel 2 menunjukkan bahwa seluruh variabel memiliki rentang nilai yang cukup lebar dengan nilai rata-rata yang relatif tinggi. Perbedaan nilai rata-rata antarvariabel menunjukkan bahwa tingkat pencapaian akses layanan dasar rumah tangga tidak merata antarindikator dan antarprovinsi. Hal ini mencerminkan adanya variasi kondisi pemenuhan layanan dasar, di mana beberapa aspek telah tercapai dengan baik di sebagian besar wilayah, sementara aspek lainnya masih relatif tertinggal pada sejumlah provinsi. Rentang nilai yang lebar pada masing-masing variabel menunjukkan adanya ketimpangan akses layanan dasar rumah tangga antarprovinsi, sehingga analisis kluster diperlukan untuk mengelompokkan provinsi berdasarkan kemiripan karakteristik akses layanan dasar rumah tangga.

B. Teknik Analisis data

1) Uji Asumsi Analisis Kluster

Berikut hasil uji asumsi multikolinearitas dan KMO pada penelitian ini.

(a) Uji Multikolinearitas

Nilai koefisien korelasi variabel dapat dilihat pada Tabel 2.

Tabel 3. Nilai VIF

Variabel	Nilai VIF
X1	2.06
X2	1.83
X3	2.06
X4	1.68

Berdasarkan Tabel 2, nilai Variance Inflation Factor (VIF) untuk seluruh variabel berada di bawah batas kritis 10. Nilai VIF yang relatif rendah ini menunjukkan bahwa tidak terdapat masalah multikolinearitas yang signifikan antar variabel penelitian. Dengan demikian, masing-masing variabel dapat digunakan secara bersama-sama dalam analisis kluster tanpa adanya pengaruh hubungan linear yang kuat antar variabel.

(b) Uji KMO (*Kaiser Mayer Olkin*)

Pada Tabel 3 dapat dilihat nilai KMO (*Kaiser Mayer Olkin*) untuk masing-masing variabel yang digunakan pada penelitian ini :

Tabel 4. Nilai KMO

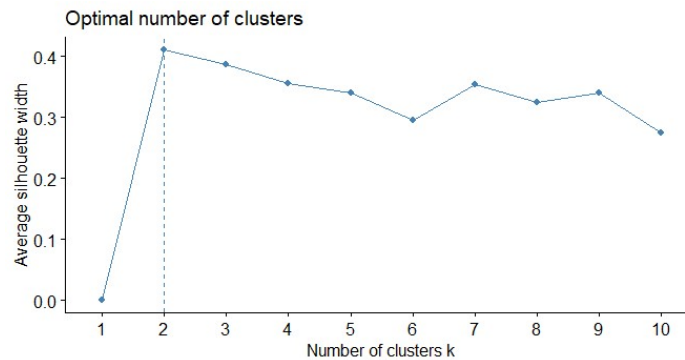
Variabel	Nilai KMO
X1	0.80
X2	0.76
X3	0.81
X4	0.75

Berdasarkan Tabel 3, nilai Kaiser–Meyer–Olkin (KMO) untuk seluruh variabel berada di atas 0,5. Nilai KMO tersebut menunjukkan bahwa tingkat kecukupan sampel dan korelasi antar variabel tergolong baik, sehingga data layak digunakan dalam analisis multivariat, termasuk analisis kluster. Nilai KMO yang berada pada kategori sedang hingga baik mengindikasikan bahwa struktur data memadai untuk mengidentifikasi pola pengelompokan antar provinsi.

2) K-Means Clustering

(a) Menentukan jumlah kluster menggunakan metode grafik *Silhouette*.

Grafik metode *Silhouette* ditampilkan pada Gambar 1.



Gambar 1. Grafik Metode *Silhouette* untuk Metode *K-Means*

Dari Gambar 1 yang menampilkan grafik *Silhouette* dengan jumlah *cluster* yang ideal adalah 2 *cluster*. Berdasarkan titik tertinggi dan garis yang mengarah ke kluster 2. Dengan menggunakan metode *silhouette coefficient*, kita dapat mengevaluasi nilai dari setiap *cluster* yang sudah terbentuk, ditampilkan pada Tabel 4.

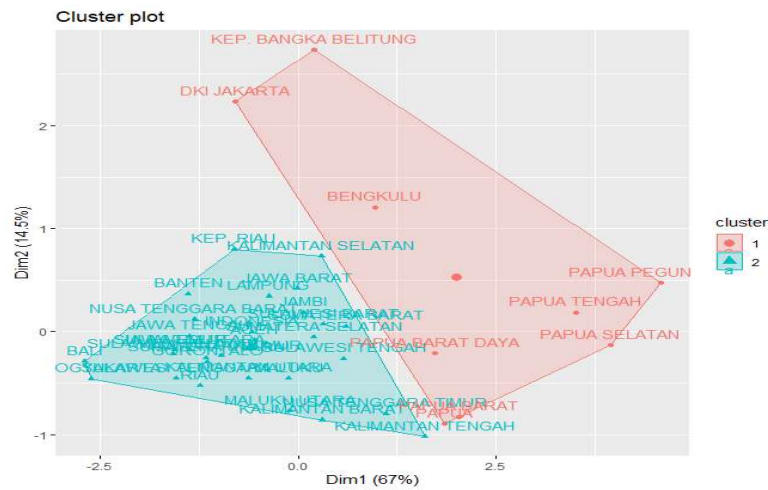
Tabel 4. Nilai *silhouette* untuk Metode *K-Means*

No	Cluster	Nilai <i>silhouette</i>
1	K2	0,47
2	K3	0,39
3	K4	0,43
4	K5	0,34

Berdasarkan Tabel 4 dapat dilihat bahwa nilai *silhouette* tertinggi berada di K2 dengan nilai *silhouette* 0,47 yang dikategorikan strukturnya lemah. Seiring bertambahnya jumlah *cluster*, nilai *silhouette* kembali menurun. Penurunan terendah terjadi ketika jumlah *cluster* mencapai 5 dengan nilai *silhouette* 0,34. Berdasarkan grafik dan nilai *silhouette* dapat disimpulkan bahwa *cluster* yang terbentuk sebanyak 2 buah *cluster*.

(b) Menetapkan hasil klasterisasi akhir

Hasil pengelompokan merujuk pada proses pengelompokan data ke dalam 2 *cluster* berdasarkan karakteristik yang serupa. Proses ini bertujuan untuk mengidentifikasi pola dan struktur dalam data, serta mengelompokkan objek yang mirip satu sama lain. Dengan memanfaatkan software R, kedua *cluster* yang dihasilkan dari analisis *cluster k-means* ini terlihat pada Gambar 2.



Gambar 2. Visualisasi Kelompok Kluster Menggunakan Metode *K-Means*

Berdasarkan gambar kluster yang terbentuk sebanyak 2 buah. Berikut anggota kluster yang terbentuk yang dapat dilihat pada Tabel 5.

Tabel 5. Anggota kluster

Cluster	Anggota
Cluster 1	Kep. Bangka Belitung, DKI Jakarta, Bengkulu, Papua Tengah, Papua Selatan, Papua Barat Daya, Papua, Papua Barat, dan Papua Pegunungan.
Cluster 2	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Lampung, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Gorontalo, Sulawesi Barat, Sulawesi Tenggara, Maluku, dan Maluku Utara

(c) Profiling dan Interpretasi Kluster.

Ringkasan data statistik dan ciri-ciri dari kedua cluster yang telah terbentuk dapat dilihat pada Tabel 6.

Tabel 6. Centroid Cluster

	Air minum	Sanitasi	Listrik	Hunian layak
Rata-rata	87,75	83,02	65,13	93,36
<i>Cluster 1</i>	76,96	70,24	43,16	77,69
<i>Cluster 2</i>	90,91	86,91	71,83	98,05

Berdasarkan Tabel 6, nilai centroid menunjukkan rata-rata masing-masing indikator akses layanan dasar rumah tangga pada setiap kluster. Secara umum, Klaster 2 memiliki nilai centroid yang lebih tinggi dibandingkan Klaster 1 pada seluruh indikator, yang mengindikasikan tingkat akses layanan dasar rumah tangga yang lebih baik. Pada indikator air minum layak, Klaster 2 memiliki nilai rata-rata sebesar 90,91%, lebih tinggi dibandingkan Klaster 1 yang sebesar 76,96%. Hal ini menunjukkan bahwa provinsi-provinsi yang tergolong dalam Klaster 2 memiliki cakupan akses air minum layak yang lebih luas. Perbedaan yang cukup mencolok juga terlihat pada indikator sanitasi layak, di mana Klaster 2 mencatat nilai rata-rata 86,91%, sedangkan Klaster 1 hanya 70,24%, yang mengindikasikan masih terbatasnya akses sanitasi layak pada wilayah-wilayah di Klaster 1.

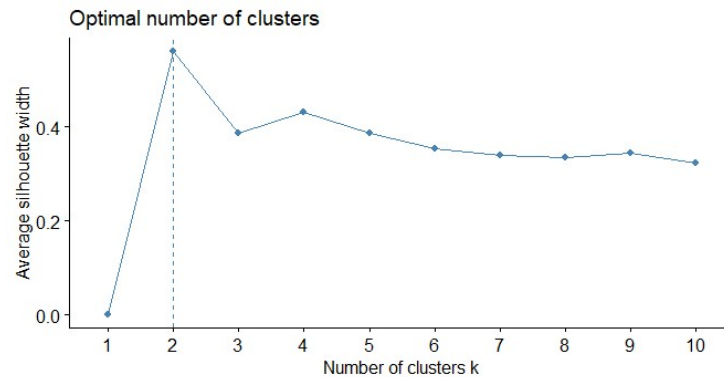
Pada indikator akses listrik, Klaster 2 menunjukkan nilai rata-rata 71,83%, jauh lebih tinggi dibandingkan Klaster 1 yang hanya 43,16%. Perbedaan ini mencerminkan adanya ketimpangan yang signifikan dalam akses listrik antar kluster. Sementara itu, pada indikator hunian layak, Klaster 2 kembali menunjukkan kondisi yang lebih baik dengan nilai rata-rata 98,05%, dibandingkan Klaster 1 sebesar 77,69%.

Secara keseluruhan, perbedaan nilai centroid antar kluster menunjukkan bahwa Klaster 2 merepresentasikan kelompok provinsi dengan tingkat akses layanan dasar rumah tangga yang relatif tinggi, sedangkan Klaster 1

mencerminkan kelompok provinsi dengan tingkat akses layanan dasar yang relatif lebih rendah. Hasil ini menegaskan adanya ketimpangan akses layanan dasar antar provinsi dan menjadi dasar penting untuk penyusunan kebijakan pembangunan yang lebih terfokus pada wilayah-wilayah yang tergolong dalam klaster.

3) Metode Ward

Tahap awal metode Ward dimulai dengan penentuan jumlah klaster yang optimum dilakukan dengan menggunakan metode Elbow. Jumlah klaster optimum yang terbentuk diamati pada grafik berikut :



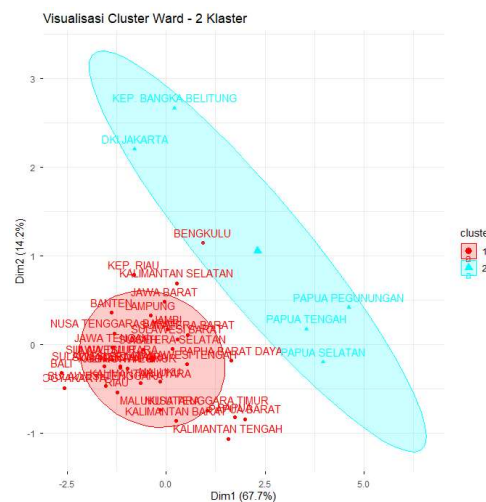
Gambar 3. Grafik Metode *Silhouette* untuk Metode Ward

Dari Gambar 3 yang menampilkan grafik *Silhouette* dengan jumlah *cluster* yang ideal adalah 2 *cluster*. Berdasarkan titik tertinggi dan garis yang mengarah ke klaster 2. Kemudian dengan menggunakan metode silhouette coefficient, kita dapat mengevaluasi nilai dari setiap cluster yang sudah terbentuk, ditampilkan pada Tabel 5.

Tabel 5. Nilai silhouette untuk Metode Ward

No	Cluster	Nilai <i>silhouette</i>
1	K2	0,74
2	K3	0,66
3	K4	0,34
4	K5	0,38

Berdasarkan Tabel 5 dapat dilihat bahwa Nilai silhouette tertinggi diperoleh pada K2 sebesar 0,74, sehingga jumlah klaster optimal pada metode Ward adalah 2 klaster. Visualisasi hasil metode Ward dengan 2 klaster dapat dilihat sebagai berikut :



Gambar 4. Visualisasi Kelompok Klaster Menggunakan Metode Ward

Berdasarkan gambar klaster yang terbentuk sebanyak 2 buah. Berikut anggota klaster yang terbentuk yang dapat dilihat pada Tabel 6.

Tabel 6. Anggota klaster

Cluster	Anggota
Cluster 1	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Lampung, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Gorontalo, Sulawesi Barat, Sulawesi Tenggara, Maluku, Maluku Utara, Bengkulu, Papua Barat Daya, Papua, dan Papua Barat.
Cluster 2	Kep. Bangka Belitung, DKI Jakarta, Papua Tengah, Papua Selatan, dan Papua Pegunungan.

Ringkasan data statistik dan ciri-ciri dari kedua cluster yang telah terbentuk dapat dilihat pada Tabel 7.

Tabel 7. *Centroid Cluster*

	Air minum	Sanitasi	Listrik	Hunian layak
Rata-rata	87,75	83,02	65,13	93,36
<i>Cluster 1</i>	89,64	86,09	70,19	97,37
<i>Cluster 2</i>	74,18	62,31	31,04	65,90

Berdasarkan Tabel 7, nilai centroid menunjukkan rata-rata masing-masing indikator akses layanan dasar rumah tangga pada setiap klaster hasil analisis menggunakan metode Ward. Secara umum, Klaster 1 memiliki nilai centroid yang lebih tinggi dibandingkan Klaster 2 pada seluruh indikator, yang mengindikasikan tingkat akses layanan dasar rumah tangga yang relatif lebih baik.

Pada indikator air minum layak, Klaster 1 mencatat nilai rata-rata sebesar 89,64, sedangkan Klaster 2 memiliki nilai rata-rata yang lebih rendah, yaitu 74,18. Hal ini menunjukkan bahwa provinsi-provinsi dalam Klaster 1 memiliki cakupan akses air minum layak yang lebih luas dibandingkan Klaster 2. Pola yang sama juga terlihat pada indikator sanitasi layak, di mana Klaster 1 memiliki nilai rata-rata 86,09, jauh lebih tinggi dibandingkan Klaster 2 yang hanya 62,31.

Perbedaan yang cukup mencolok juga tampak pada indikator akses listrik. Klaster 1 menunjukkan nilai rata-rata sebesar 70,19, sedangkan Klaster 2 hanya 31,04, yang mengindikasikan adanya ketimpangan akses listrik yang signifikan antar klaster. Selain itu, pada indikator hunian layak, Klaster 1 kembali menunjukkan kondisi yang lebih baik dengan nilai rata-rata 97,37, dibandingkan Klaster 2 yang sebesar 65,90.

Secara keseluruhan, hasil centroid klaster metode Ward menunjukkan bahwa Klaster 1 merepresentasikan kelompok provinsi dengan tingkat akses layanan dasar rumah tangga yang relatif tinggi, sedangkan Klaster 2 mencerminkan kelompok provinsi dengan keterbatasan akses layanan dasar rumah tangga yang lebih besar. Temuan ini menegaskan adanya ketimpangan akses layanan dasar antar provinsi dan menjadi dasar penting dalam penyusunan kebijakan pembangunan yang lebih terarah, khususnya bagi provinsi-provinsi yang tergolong dalam Klaster 2.

4) Perbandingan Kedua Metode

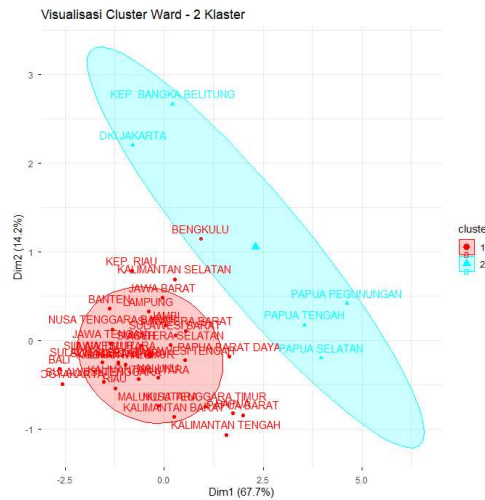
Setelah penentuan jumlah dan pembagian klaster masing-masing kabupaten/kota, dilakukan uji validasi pada setiap metode untuk menentukan klaster terbaik. Uji yang dilakukan adalah *Silhouette*. Hasil pengujian disajikan pada tabel di bawah:

Tabel 8. Hasil Perbandingan Kedua Metode

Metode	Klaster	Nilai <i>silhouette</i>
<i>K-Means</i>	2	0,47
Ward	2	0,74

Dari hasil pengujian validasi klaster diatas, diperoleh metode Ward memiliki nilai koefisien *Silhouette* yang lebih tinggi dibandingkan metode *K-Means*. Hal ini menunjukkan bahwa metode Ward dengan 2 klaster menjadi metode yang terbaik untuk data akses layanan dasar rumah tangga ini. Berdasarkan nilai centroid pada Tabel 6 (*K-Means*) dan Tabel 7 (Ward), terlihat bahwa kedua metode menghasilkan pola pengelompokan yang relatif serupa, yaitu memisahkan wilayah ke dalam dua klaster dengan karakteristik akses layanan dasar yang berbeda. Perbedaan

utama antara kedua metode terletak pada tingkat kontras antar *cluster*, di mana metode Ward menunjukkan selisih nilai *centroid* yang lebih tegas, terutama pada variabel listrik dan hunian layak. Hal ini mengindikasikan bahwa metode Ward mampu menghasilkan pemisahan klaster yang lebih jelas dibandingkan K-Means, sehingga lebih efektif dalam membedakan provinsi berdasarkan tingkat akses layanan dasar rumah tangga. Berikut grafik metode ward yang terpilih dilihat pada Gambar 5.



Gambar 5. Hasil *Clustering* Metode *Ward*

Visualisasi kluster Ward dengan dua kluster menunjukkan pemisahan provinsi yang cukup jelas berdasarkan dua dimensi utama, di mana Dim1 sebagai dimensi dengan kontribusi variasi terbesar menjadi pembeda utama antar kluster. Kluster 1 didominasi oleh sebagian besar provinsi yang posisinya relatif berdekatan, mencerminkan karakteristik akses layanan dasar rumah tangga yang relatif serupa dan berada pada tingkat sedang hingga rendah. Sebaliknya, Kluster 2 terdiri dari beberapa provinsi yang terpisah cukup jauh dari kluster pertama, seperti wilayah Papua dan Kepulauan Bangka Belitung, yang menunjukkan karakteristik akses layanan dasar yang berbeda secara signifikan. Pola sebaran ini mengindikasikan bahwa metode Ward mampu menghasilkan pemisahan kluster yang tegas dan stabil, sehingga efektif dalam membedakan provinsi berdasarkan perbedaan karakteristik akses layanan dasar rumah tangga.

IV. KESIMPULAN

Penelitian ini bertujuan untuk mengelompokkan provinsi di Indonesia berdasarkan indikator akses layanan dasar rumah tangga serta membandingkan kinerja metode K-Means dan Hierarchical Clustering metode Ward. Hasil penelitian menunjukkan bahwa kedua metode menghasilkan dua kluster sebagai jumlah kluster optimal, yang merepresentasikan kelompok provinsi dengan tingkat akses layanan dasar rumah tangga yang relatif tinggi dan relatif rendah. Berdasarkan nilai centroid, kedua metode secara konsisten menunjukkan adanya perbedaan yang cukup jelas antar kluster pada seluruh indikator, khususnya pada akses listrik dan sanitasi layak, sehingga mengindikasikan adanya ketimpangan akses layanan dasar antar provinsi. Namun demikian, hasil evaluasi kualitas kluster menunjukkan bahwa metode Ward memiliki nilai koefisien Silhouette yang lebih tinggi dibandingkan metode K-Means, yang menandakan struktur kluster yang lebih kompak dan pemisahan antar kluster yang lebih baik. Oleh karena itu, metode Ward dinilai lebih unggul dan lebih representatif dalam memetakan pola akses layanan dasar rumah tangga antar provinsi pada penelitian ini. Sebagai saran, penelitian selanjutnya dapat mempertimbangkan penambahan indikator sosial ekonomi lainnya agar pemetaan kesejahteraan wilayah menjadi lebih komprehensif. Selain itu, penggunaan metode klusterisasi lain atau pendekatan validasi tambahan juga dapat dilakukan untuk memperkuat hasil pengelompokan dan memberikan perbandingan yang lebih luas dalam analisis kluster wilayah.

DAFTAR PUSTAKA

- Amalina, T., Bima, D., Pramana, A., & Sari, B. N. (2022). Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food. *Jurnal Ilmiah Wahana Pendidikan*, 8(15), 574–583. <https://doi.org/10.5281/zenodo.7052276>
- Bernard, D. T. K., Munasinghe, A. A. S. N., Premarathne, I. K. R. J., & Wijayarathne, S. (2020). The Influence Of Facebook Marketing On Consumer Buying Intention Of Clothing: Evidence From Young Adults. *Archives of Business Research*, 8(9), 37–51. <https://doi.org/10.14738/abr.89.8992>
- Chrisnahutama, A. (2023). Akses listrik, air minum, dan sanitasi serta modal manusia di Indonesia. *Journal of Economics Research and Policy Studies*, 3(3), 185–199. <https://doi.org/10.53088/jerps.v3i3.770>
- Iklima, T., & Pujiyanta, A. (2023). Perbandingan Metode K-Means Clustering Dan Metode Ward Dalam Mengelompokkan Pelanggan Mall. *Jurnal Fasilkom*, 13(3), 349–357. <https://doi.org/https://doi.org/10.37859/jf.v13i3>
- Izzuddin, K. H., & Wijayanto, A. W. (2024). Pemodelan Clustering Ward, K-Means, Diana, dan PAM dengan PCA untuk Karakterisasi Kemiskinan Indonesia Tahun 2021. *Komputika : Jurnal Sistem Komputer*, 13(1), 41–53. <https://doi.org/10.34010/komputika.v13i1.10803>
- Kuswardana, D. A., Prasetya, D. A., Trimono, T., & Diyasa, I. G. S. M. (2025). Comparison of Elbow and Silhouette Methods in Optimizing K-Prototype Clustering for Customer Transactions. *Jurnal Ilmiah Edutic : Pendidikan Dan Informatika*, 12(1), 43–48. <https://doi.org/10.21107/edutic.v12i1.29744>
- Matdoan, M. Y., Balami, A. M., & Talakua, W. (2023). Implementasi metode Ward clustering untuk klasterisasi penduduk yang memiliki jaminan kesehatan pada kabupaten dan kota di Provinsi Maluku (Application of the Ward clustering method for clustering of population who have health insurance in districts/cities in Maluku Province). In *Majalah Ilmiah Matematika dan Statistika* (Vol. 23, Number 1). <https://jurnal.unej.ac.id/index.php/MIMS/index>
- Putri, D. B., & Samsuddin, M. A. (2025). Pengaruh Akses Air Minum Layak dan Akses Sanitasi Layak terhadap Tingkat Kemiskinan: Studi Panel Kabupaten di Jawa Barat Tahun 2020-2024. *Jurnal Ekonomi Dan Pembangunan Indonesia*, 3(3). <https://doi.org/10.61132/jepi.v3i4.1611>
- Ramadhan, H., Rizal Abdan Kamaludin, M., Alfian Nasrullah, M., & Rolliawati, D. (2023). Comparison of Hierarchical, K-Means and DBSCAN Clustering Methods for Credit Card Customer Segmentation Analysis Based on Expenditure Level. *Journal of Applied*

Informatics and Computing (JAIC), 7(2), 246–251.
<https://doi.org/https://doi.org/10.30871/jaic.v7i2.5790>

Salsabila, A., Widiharah, T., & Sudarno. (2022). *Metode K-Harmonic Means Clustering Dengan Validasi Silhouette Coefficient (Studi Kasus : Empat Faktor Utama Penyebab Stunting 34 Provinsi di Indonesia Tahun 2018)*. 11(1), 11–20.
<https://doi.org/https://ejournal3.undip.ac.id/index.php/gaussian/>

Vania, P., & Nurina Sari, B. (2023). Perbandingan Metode Elbow dan Silhouette untuk Penentuan Jumlah Klaster yang Optimal pada Clustering Produksi Padi menggunakan Algoritma K-Means. *Jurnal Ilmiah Wahana Pendidikan*, 9(21), 547–558.
<https://doi.org/10.5281/zenodo.10081332>

Yusuf, M. D., Munandar, T. A., & Ramdhania, F. K. (2025). Comparative Analysis of K-Means and Hierarchical Clustering for Regional Welfare Disparity Identification in West Java Province. *International Journal of Information Technology and Computer Science Applications*. <https://doi.org/https://doi.org/10.58776/ijitcsa.v3i3.213>