

## Pengelompokan Provinsi di Indonesia berdasarkan Ketimpangan Akses Layanan Kesehatan Tahun 2024 Menggunakan Pendekatan *Cluster Hirarki*

***Clustering of Indonesian Provinces Based on Inequality in Access to Health Services in 2024 Using a Hierarchical Clustering Approach***

**Ariza Nabilah Rahma Na’ifa<sup>1</sup>, Dewi Rohayah<sup>2</sup>, Intan Yuliati<sup>3</sup>, Nayla Tsabita Amalia Shofa<sup>4</sup>,  
Elly Pusporani<sup>5,a</sup>, Dita Amelia<sup>6</sup>**

<sup>1,2,3,4,5,6</sup>Program Studi S1 Statistika, Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga, Indonesia

<sup>a)</sup> Corresponding author: [elly.pusporani@fst.unair.ac.id](mailto:elly.pusporani@fst.unair.ac.id)

### **ABSTRACT**

*Health disparities remain a major challenge in Indonesia, particularly in terms of access to healthcare services across provinces. This study aims to classify 38 Indonesian provinces based on inequality in healthcare access in 2024 using a hierarchical clustering approach. Three key indicators were used: the number of hospitals, the number of medical personnel, and the percentage of people experiencing health complaints who opted for self-medication. The analysis identified the average linkage method as the most suitable model, supported by the highest cophenetic correlation coefficient (0,911). The results revealed two distinct clusters. The first cluster includes most provinces outside Java Island, characterized by limited healthcare infrastructure and personnel. The second cluster comprises four provinces on Java Island with advanced healthcare facilities but a high rate of self-medication. These findings suggest that healthcare access inequality is influenced not only by infrastructure but also by social and behavioral factors. Therefore, policy recommendations should be tailored accordingly: infrastructure improvement and equitable distribution of medical personnel for the first cluster, and health education interventions for the second. This study contributes to evidence-based policy design in line with the Sustainable Development Goals (SDGs), particularly the goal of ensuring equitable healthcare access for all.*

**Keywords:** healthcare access inequality, hierarchical clustering, average linkage, self-medication, indonesia, sdgs

### **1. Pendahuluan**

Kesehatan merupakan salah satu aspek fundamental dalam pembangunan manusia. Akses terhadap layanan kesehatan yang merata dan berkualitas menjadi indikator penting dalam menentukan tingkat kesejahteraan suatu wilayah. Akses ini bukan hanya soal ada atau tidaknya fasilitas kesehatan, tetapi juga mencakup apakah biayanya terjangkau, lokasinya mudah dijangkau, pelayanannya baik, dan sesuai dengan budaya masyarakat (Putri & Sari, 2021; Nugroho *et al.*, 2020). Di Indonesia, ketimpangan akses terhadap layanan kesehatan masih menjadi tantangan yang signifikan, terutama ketika melihat disparitas antarprovinsi yang cukup tinggi (Kemenko PMK, 2023). Hal ini juga diperparah oleh penyebaran tenaga medis yang tidak merata, infrastruktur yang belum memadai, serta perbedaan kondisi ekonomi (Kemenkes RI, 2023). Program Jaminan Kesehatan Nasional (JKN) memang membantu meningkatkan akses, tapi belum bisa mengatasi seluruh ketimpangan yang ada (Ramadhan & Utami, 2022). Karena itu, diperlukan kebijakan yang berdasarkan data dan fakta agar semua orang, tanpa terkecuali, bisa mendapatkan layanan kesehatan secara adil.

Ketimpangan ini tidak hanya mencerminkan perbedaan dalam ketersediaan fasilitas kesehatan, tetapi juga menunjukkan adanya kesenjangan dalam pemerataan tenaga medis dan kemampuan masyarakat untuk mengakses layanan kesehatan yang layak. Salah satu indikator ketimpangan yang dapat mencerminkan kondisi tersebut adalah jumlah rumah sakit di tiap provinsi, jumlah tenaga medis yang tersedia, serta persentase penduduk yang memiliki keluhan kesehatan namun memilih untuk mengobati sendiri. Ketiga indikator ini menggambarkan aspek struktural maupun perilaku masyarakat dalam mengakses layanan kesehatan.

Menurut data Badan Pusat Statistik (BPS), beberapa provinsi di luar Pulau Jawa masih menunjukkan angka yang relatif rendah dalam hal cakupan layanan kesehatan dasar, seperti ketersediaan puskesmas, rumah sakit, dan rasio tenaga medis terhadap jumlah penduduk (BPS, 2023). Ketimpangan ini dapat berdampak pada penurunan kualitas hidup, tingginya angka kematian ibu dan bayi, serta memperbesar disparitas kesejahteraan antardaerah.

Penelitian ini juga sejalan dengan Tujuan Pembangunan Berkelanjutan (*Sustainable Development Goals/SDGs*) poin ketiga, yaitu “Menjamin kehidupan yang sehat dan mendorong kesejahteraan bagi semua orang di segala usia.” Salah satu target dalam tujuan ini adalah memastikan akses layanan kesehatan yang merata dan inklusif, terutama untuk kelompok masyarakat yang rentan. Oleh karena itu, memahami pola ketimpangan antarprovinsi menjadi landasan penting dalam merancang kebijakan kesehatan yang berbasis bukti (*evidence-based policy*).

Namun demikian, sebagian besar penelitian sebelumnya yang membahas ketimpangan akses layanan kesehatan di Indonesia masih bersifat deskriptif dan kurang mengadopsi pendekatan statistik eksploratif seperti *cluster analysis*. Misalnya, studi oleh Riyadi & Hanum (2021) hanya mengkaji ketimpangan distribusi tenaga kesehatan tanpa mempertimbangkan faktor fasilitas dan perilaku masyarakat. Demikian pula, penelitian oleh Wulandari *et al.* (2020) lebih menyoroti persepsi masyarakat terhadap pelayanan kesehatan tanpa membandingkan perbedaan kondisi antarwilayah secara sistematis. Ketidakhadiran pendekatan kuantitatif yang komprehensif ini menjadi celah yang cukup krusial, karena tanpa segmentasi yang jelas antarprovinsi, upaya penajaman kebijakan menjadi kurang efektif. Oleh karena itu, penting dilakukan analisis *cluster* hierarki untuk mengelompokkan provinsi berdasarkan kesamaan karakteristik akses layanan kesehatan agar dapat memberikan rekomendasi kebijakan yang lebih spesifik dan kontekstual.

Dalam penelitian ini digunakan metode analisis *cluster* hierarki (*hierarchical cluster analysis*) karena mampu mengelompokkan provinsi berdasarkan tingkat kemiripan karakteristik secara objektif tanpa harus menentukan jumlah kelompok sejak awal (Aselnino dan Wijayanto, 2024). Metode ini memanfaatkan jarak antarobservasi untuk menghasilkan struktur pengelompokan yang divisualisasikan dalam bentuk dendrogram, sehingga memudahkan interpretasi hubungan antarwilayah (Liu, 2025). Pendekatan ini dianggap paling sesuai untuk topik ketimpangan akses layanan kesehatan karena data yang digunakan bersifat multivariat dan tidak selalu memiliki distribusi normal (Utari dan Hanun, 2021). Selain itu, metode hierarki juga memungkinkan peneliti untuk melihat pola kedekatan antarprovinsi secara bertahap, sehingga hasilnya dapat memberikan gambaran yang lebih mendalam tentang struktur ketimpangan antarwilayah di Indonesia.

Penelitian ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan ketimpangan akses layanan kesehatan pada tahun 2024 menggunakan pendekatan *cluster* hierarki. Dengan menggunakan tiga indikator utama, yaitu jumlah rumah sakit, jumlah tenaga medis, dan persentase penduduk yang memiliki keluhan kesehatan namun memilih mengobati sendiri, diharapkan hasil pengelompokan dapat mencerminkan realitas yang lebih holistik. Penggunaan pendekatan *cluster analysis*, khususnya metode *cluster* hierarki, memungkinkan pengelompokan wilayah secara objektif berdasarkan kemiripan data, sehingga dapat menjadi dasar bagi pengambilan keputusan yang lebih tepat sasaran dalam upaya pemerataan akses layanan kesehatan di Indonesia dan sebagai bahan evaluasi dalam upaya pencapaian target-target SDGs, khususnya dalam pemerataan akses layanan kesehatan di tahun 2024 dan seterusnya.

## 2. Tinjauan Pustaka

### 2.1 Analisis Cluster

*Cluster* adalah teknik analisis data yang bertujuan untuk mengelompokkan sejumlah objek ke dalam kelompok-kelompok atau *cluster* berdasarkan tingkat kemiripan karakteristik yang dimiliki oleh objek-objek tersebut. Salah satu pendekatan yang paling sering digunakan dalam analisis *cluster* adalah metode hierarki (*hierarchical clustering*), yang menyusun data dalam bentuk struktur bertingkat. Dalam pendekatan ini, proses pengelompokan dilakukan secara bertahap dengan menggabungkan objek-objek yang paling mirip (metode *agglomerative*) atau memisahkan dari satu kelompok besar menjadi kelompok-kelompok kecil (metode *divisive*). Hasil dari metode ini umumnya divisualisasikan dalam bentuk *dendrogram* yang menunjukkan hubungan hierarkis antar *cluster* (Bhahari & Kusnawi, 2024; Xu & Tian, 2022).

### 2.2 Pengelompokan Cluster Hirarki

Metode *cluster* hierarki adalah teknik pengelompokan data yang menyusun objek berdasarkan tingkat kemiripan dalam bentuk struktur bertingkat. Pendekatan ini terbagi menjadi dua prinsip utama, yaitu *agglomerative* (pemusatan), yang dimulai dengan setiap objek sebagai *cluster* terpisah dan secara bertahap digabungkan berdasarkan kemiripan, serta *divisive* (pemisahan), yang dimulai dari satu *cluster* besar lalu dipecah menjadi *cluster-cluster* kecil. Salah satu ukuran kemiripan yang umum digunakan dalam metode ini adalah jarak *Euclidean*, yang dihitung sebagai akar kuadrat dari jumlah kuadrat perbedaan nilai atribut antar objek (Wijuniarni *et al.*, 2022; Han *et al.*, 2022; Tan *et al.*, 2020).

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (1)$$

Hasil pengelompokan dengan metode *cluster* hierarki biasanya divisualisasikan dalam bentuk *dendrogram*, yaitu diagram pohon yang menunjukkan proses penggabungan atau pemisahan *cluster*. Pendekatan ini dapat diterapkan menggunakan metode seperti *single linkage*, *complete linkage*, *average linkage*, dan *Ward's method* (Han *et al.*, 2022; Wijuniarni *et al.*, 2022).

#### 1. Single Linkage

*Single linkage* mengukur kesamaan atau kemiripan antar *cluster* berdasarkan jarak terdekat antara objek dalam *cluster* tersebut (Han *et al.*, 2022). Dapat dihitung dengan rumus berikut ini:

$$d_{(uv)w} = \min \{d_{uw}, d_{vw}\} \quad (2)$$

dengan,

$d_{uw}$  : Jarak terdekat dari *cluster* U dan W

$d_{vw}$  : Jarak terdekat dari *cluster* V dan W

## 2. Complete Linkage

*Complete linkage* mengukur kesamaan atau kemiripan antar *cluster* berdasarkan jarak terjauh antara objek dalam *cluster* tersebut (Han *et al.*, 2022). Dapat dihitung dengan rumus berikut ini:

$$d_{(uv)w} = \max \{d_{uw}, d_{vw}\} \quad (3)$$

dengan,

$d_{uw}$  : Jarak terjauh dari *cluster* U dan W

$d_{vw}$  : Jarak terjauh dari *cluster* V dan W

## 3. Average Linkage

*Average linkage* menghitung jarak antar *cluster* berdasarkan rata-rata jarak antara semua pasangan objek dari dua *cluster* yang dibandingkan (Tan *et al.*, 2020). Dapat dihitung dengan rumus berikut ini:

$$d_{(uv)w} = \text{Average}\{d_{uw}, d_{vw}\} = \frac{\sum_{i=1}^n \sum_{j=1}^n d_{ij}}{N_{(uv)} N_w} \quad (4)$$

dengan,

$d_{ij}$  : Jarak antara objek ke-*i* pada *cluster* UV dan objek ke-*j* pada *cluster* W

$N_{(uv)} N_w$  : Jarak objek pada *cluster* UV dan W

## 4. Ward's Linkage

*Ward's linkage* mengelompokkan *cluster* dengan cara meminimalkan peningkatan total varians dalam *cluster* yang terbentuk (Han *et al.*, 2022). Dapat dihitung dengan rumus berikut ini:

$$d_{uv} = \frac{n_u \times n_v}{n_u + n_v} \times (\bar{x}_u - \bar{x}_v)^2 \quad (5)$$

Dengan,

$n_u, n_v$  : Jumlah objek dalam *cluster* U dan V

$\bar{x}_u, \bar{x}_v$  : *Centroid* (rata-rata) objek dari *cluster* U dan V

$(\bar{x}_u - \bar{x}_v)^2$  : Kuadrat jarak *Euclidean* antara *centroid* kedua *cluster*

## 2.3 Pemilihan Model dan Jumlah Cluster Terbaik

*Cluster* yang efektif ditandai oleh keseragaman internal yang tinggi (homogenitas) antar elemen dalam *cluster* tersebut, serta perbedaan yang signifikan (heterogenitas) antara *cluster* yang satu dengan yang lainnya. Untuk menilai kualitas *cluster*, digunakan berbagai indeks validitas internal dan koefisien *cophenetic*, yang mengukur sejauh mana struktur *cluster* mencerminkan tingkat kebaikan dari *cluster* (Botta-Dukát, 2023; Iglesias Vázquez & Zseby, 2023). Dalam penelitian ini, jumlah *cluster* dieksplorasi dalam rentang 2 hingga 7 untuk menyeimbangkan antara kompleksitas analisis dan kemudahan interpretasi. Rentang tersebut dianggap ideal karena mampu menggambarkan variasi antarprovinsi tanpa membuat hasil pengelompokan terlalu terfragmentasi, dengan pemilihan akhir didasarkan pada nilai indeks validitas dan koefisien *cophenetic* terbaik (Vichi *et al.*, 2022).

### 1. Koefisien Cophenetic

Koefisien *cophenetic* adalah ukuran korelasi antara jarak asli antar objek dalam data (misalnya jarak *Euclidean* kuadrat) dengan jarak yang dihasilkan dari dendrogram hasil *clusterisasi* hirarki. Koefisien ini digunakan untuk menilai seberapa baik dendrogram merepresentasikan data asli. Formula koefisien *Cophenetic* menghitung korelasi antara dua matriks jarak tersebut (Botta-Dukát, 2023; Saracı *et al.*, 2013). Dapat dihitung dengan rumus berikut ini:

$$r_{coph} = \frac{\sum_{i < j} (d_{ij} - \bar{d})(d_{cij} - \bar{d}_c)}{\sqrt{\left[ \sum_{i < j} (d_{ij} - \bar{d})^2 \right] \left[ \sum_{i < j} (d_{cij} - \bar{d}_c)^2 \right]}} \quad (6)$$

dengan,

$r_{coph}$  : Koefisien korelasi *cophenetic*

$d_{ij}$  : Jarak *squared euclidean* antara objek *i* dan *j*

$\bar{d}$  : Rata-rata jarak *squared euclidean* antara objek *i* dan *j*

$d_{cij}$  : Jarak *cophenetic* antara objek *i* dan *j*

$\bar{d}_c$  : Rata-rata jarak *cophenetic* antara objek *i* dan *j*

Nilai koefisien korelasi *cophenetic* berkisar antara -1 dan 1, jika nilai  $r_{coph}$  mendekati 1 berarti solusi yang dihasilkan dari proses *clustering* baik.

### 2. Internal Cluster Validation

Validasi internal adalah metode untuk menilai kualitas hasil *clusterisasi* berdasarkan informasi yang terkandung dalam data itu sendiri, tanpa melibatkan label eksternal. Evaluasi ini mempertimbangkan kepadatan data dalam *cluster*, hubungan antar objek yang berdekatan, serta tingkat pemisahan antar *cluster*. Beberapa indeks yang umum digunakan dalam validasi internal meliputi *Dunn Index*, *Silhouette Coefficient*, dan *Connectivity*, yang masing-masing menangkap aspek berbeda dari struktur *cluster* yang terbentuk (Bottadukat, 2023; Iglesias Vázquez & Zseby, 2023).

a. *Dunn Index*

*Dunn Index* merupakan salah satu ukuran validitas yang digunakan untuk menilai kualitas hasil *clusterisasi*, terutama dalam membandingkan metode *clustering* yang berbeda secara efektif. Indeks ini menghitung rasio antara jarak minimum antar *cluster* dengan diameter maksimum dalam *cluster*, di mana nilai yang lebih tinggi menunjukkan *cluster* yang lebih baik dalam hal pemisahan dan kekompakan. (Saxena et al., 2021). Dapat dihitung dengan rumus berikut ini:

$$D = \min_{i \leq k \leq n} \left\{ \min_{i \leq k \leq n} \left\{ \frac{d(c_i, c_j)}{\max_{i \leq k \leq n} (d'(c_k))} \right\} \right\} \quad (7)$$

dengan,

$D$  : *Index Dunn*

$d(c_i, c_j)$  : Jarak antar kelompok  $c_i$  dan  $c_j$

$d'(c_k)$  : Jarak dalam kelompok  $c_k$

b. *Silhouette Coefficient*

Indeks *silhouette* mengukur sejauh mana suatu objek cocok dengan *clusternya* sendiri dibandingkan dengan *cluster* lain berdasarkan perbedaan rata-rata jarak antar objek. Hasil perhitungan nilai silhouette berada pada range -1 sampai 1. Semakin besar nilai silhouette akan semakin baik kualitas suatu kelompok. (Iglesias Vázquez & Zseby, 2023). Perhitungan indeks *silhouette* didefinisikan sebagai rata – rata  $s(i)$  yaitu:

$$SC = \frac{1}{n} \sum_{i=1}^n s(i) \quad (8)$$

dengan,

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}, \min d(i, C), a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j)$$

$a(i)$  : Rata – rata jarak objek ke- $i$  dengan semua objek yang berada di dalam satu cluster  $A$

$b(i)$  : Nilai minimum dari jarak rata – rata objek  $i$  dengan semua objek pada cluster lain  $C$

c. *Connectivity*

Indeks *connectivity* mengukur tingkat keterhubungan antar objek dalam cluster dengan menghitung jumlah tetangga terdekat yang berbeda *cluster*. Nilai indeks yang lebih kecil menunjukkan hasil *cluster* yang lebih baik atau optimal (Halim & Widodo, 2017). Dapat dihitung dengan rumus berikut ini:

$$conn = \sum_{i=1}^N \sum_{j=i}^L x_{i, nn_{i(j)}} \quad (9)$$

dengan

$conn$  : Indeks *connectivity*

$nn_{i(j)}$  : pengamatan tetangga terdekat (*nearest neighbour*) dari data ke- $j$  ke data ke- $i$ , jika dalam satu *cluster* bernilai 0 dan jika berbeda, bernilai  $1/j$ .

$N$ : Banyak pengamatan

$L$  : Banyak *cluster*

### 3. Metodologi

#### 3.1 Data dan Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) berjudul *Statistik Indonesia 2025*, yang dirilis pada 28 Februari 2025. Data ini mencakup seluruh 38 provinsi di Indonesia yang ditampilkan dalam Tabel 1 berikut.

**Tabel 1.** Unit Observasi Provinsi di Indonesia

No	Provinsi	No	Provinsi	No	Provinsi
1	Aceh	11	DKI Jakarta	21	Kalimantan Tengah
2	Sumatera Utara	12	Jawa Barat	22	Kalimantan Selatan
3	Sumatera Barat	13	Jawa Tengah	23	Kalimantan Timur

No	Provinsi	No	Provinsi	No	Provinsi
4	Riau	14	DI Yogyakarta	24	Kalimantan Utara
5	Jambi	15	Jawa Timur	25	Sulawesi Utara
6	Sumatera Selatan	16	Banten	26	Sulawesi Tengah
7	Bengkulu	17	Bali	27	Sulawesi Selatan
8	Lampung	18	Nusa Tenggara Barat	28	Sulawesi Tenggara
9	Bangka Belitung	19	Nusa Tenggara Timur	29	Gorontalo
10	Kepulauan Riau	20	Kalimantan Barat	30	Sulawesi Barat
31	Maluku	34	Papua Barat	37	Papua Pegunungan
32	Maluku Utara	35	Papua Selatan	38	Papua Barat Daya
33	Papua	36	Papua Tengah		

Dalam penelitian ini menggunakan 3 indikator ketimpangan akses kesehatan di setiap provinsi di Indonesia yang ditampilkan dalam Tabel 2. sebagai berikut.

**Tabel 2.** Variabel Indikator dalam Analisis *Cluster*

Variabel Akses Kesehatan	Keterangan
Jumlah RS Umum	Jumlah total rumah sakit umum yang tersedia di setiap provinsi.
Jumlah Tenaga Medis	Jumlah total tenaga medis, termasuk dokter, perawat, dan tenaga kesehatan lain.
Persentase Penduduk Memilih Mengobati Sendiri	Persentase penduduk yang memiliki keluhan kesehatan namun tidak pergi ke fasilitas kesehatan dan memilih pengobatan mandiri.

### 3.2 Langkah Analisis

Penelitian menggunakan metode analisis *cluster* hirarki untuk menganalisis data. Prosedur untuk melakukan analisis guna mencapai tujuan penelitian adalah sebagai berikut:

1. Mengumpulkan data observasi
2. Menggambarkan karakteristik variabel penelitian secara deskriptif
3. Melakukan pengelompokan dengan langkah-langkah sebagai berikut:
  - a. Melakukan pengelompokan hirarki menggunakan pengukuran jarak Euclidean yang diterapkan pada model *single linkage*, *complete linkage*, *average linkage* dan *ward's linkage*
  - b. Perbandingan dalam analisis ini dilakukan dengan menggunakan *Internal Clustering Validation* terhadap variasi jumlah *cluster* untuk menentukan jumlah *cluster* optimal dan Koefisien *Cophenetic* untuk menentukan satu model *cluster* hirarki terbaik
4. Mendeskripsikan karakteristik tiap *cluster* yang telah berbentuk
5. Merumuskan rekomendasi kebijakan untuk pemerintah guna meningkatkan kualitas akses kesehatan di setiap provinsi di Indonesia berdasarkan hasil analisis *cluster*

### 4. Results and Discussion

#### 4.1 Statistika Deskriptif

Analisis deskriptif digunakan untuk memberikan gambaran umum mengenai karakteristik masing-masing variabel dalam penelitian ini, melalui penyajian nilai rata-rata, minimum, maksimum, serta simpangan baku, sebagaimana ditampilkan pada Tabel 3.

**Tabel 3.** Statistika Deskriptif

Indikator Ketimpangan Akses Kesehatan	N	Mean	Median	Minimum	Maksimum	StDev
Jumlah RS Umum	38	71,3158	41,5	8 (Papua Tengah)	373 (Jawa Barat)	89,4799
Jumlah Tenaga Medis	38	5451,95	2449	262 (Papua Barat Daya)	31627 (Jawa Barat)	7594,74

Persentase Penduduk Memilih Mengobati Sendiri	38	61,3463	61,995	29,11 (Papua Pegunungan)	76,64 (Gorontalo)	8,91765
---	----	---------	--------	-----------------------------	----------------------	---------

Tabel 3 menunjukkan adanya ketimpangan akses kesehatan yang signifikan antarwilayah di Indonesia. Pada indikator jumlah rumah sakit umum, rata-rata tiap wilayah memiliki 71,32 rumah sakit dengan median 41,5, menunjukkan distribusi yang tidak merata karena beberapa wilayah seperti Jawa Barat memiliki jumlah sangat tinggi (373 RS), sementara Papua Tengah hanya memiliki 8 RS. Hal serupa terlihat pada jumlah tenaga medis, di mana Jawa Barat mendominasi dengan 31.627 tenaga medis, sedangkan Papua Barat Daya memiliki jumlah yang sangat sedikit. Nilai rata-rata tenaga medis (5.452) yang jauh lebih tinggi daripada median (2.449) mengindikasikan adanya ketimpangan ekstrem, dengan beberapa wilayah memiliki sumber daya kesehatan yang jauh lebih memadai dibandingkan wilayah lainnya. Sementara itu, persentase penduduk yang memilih mengobati sendiri menunjukkan variasi yang tidak terlalu ekstrem, dengan rata-rata 61,35% dan median 62,00%. Namun, tetap terlihat perbedaan antara wilayah seperti Gorontalo (76,64%) dan Papua Pegunungan (29,11%), yang mungkin mencerminkan perbedaan aksesibilitas atau budaya dalam pengobatan. Secara keseluruhan, data ini mengungkapkan disparitas yang besar dalam fasilitas dan sumber daya kesehatan, dengan wilayah Papua secara konsisten berada di posisi terbawah dalam semua indikator, sementara Jawa Barat mendominasi dalam hal ketersediaan layanan kesehatan.

#### 4.2 Pengelompokan Hirarki

Analisis hirarki dalam penelitian ini menerapkan pengukuran jarak Euclidean dengan empat pendekatan linkage yang berbeda, yaitu *single linkage*, *complete linkage*, *average linkage*, dan *ward's linkage*. Untuk menentukan jumlah *cluster* yang optimal, dilakukan evaluasi menggunakan *Internal Clustering Validation* dengan memvariasikan jumlah *cluster* dari 2 hingga 7 *cluster* pada setiap model hirarki. Pendekatan ini memungkinkan identifikasi konfigurasi *cluster* terbaik berdasarkan *Internal Clustering Validation*. Data yang digunakan untuk pengelompokan hirarki ini menggunakan data yang telah distandarisasi.

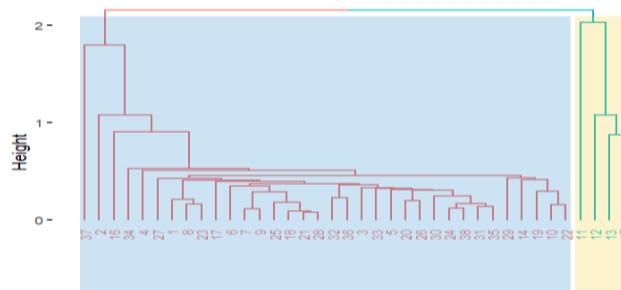
##### 1. Single Linkage

Pada model ini, Untuk menentukan jumlah *cluster* yang optimal, dilakukan evaluasi menggunakan *Internal Clustering Validation* dengan memvariasikan jumlah cluster dari 2 hingga 7 *cluster* pada setiap model hirarki dilihat dari nilai *dunn*, *connectivity*, dan *silhouette* yang terdapat pada Tabel 4.

**Tabel 4.** Uji Validitas Internal untuk *Single Linkage*

Indeks Validitas	Jumlah Cluster					
	2	3	4	5	6	7
<i>Connectivity</i>	4,771	7,216	10,145	12,645	16,074	18,614
<i>Dunn</i>	0,405	0,381	0,509	0,305	0,305	0,257
<i>Silhouette</i>	0,663	0,556	0,530	0,510	0,375	0,329

Hasil uji *Internal Clustering Validation* pada model ini ditunjukkan oleh Tabel 4, nilai *Connectivity* terendah diperoleh pada dua *cluster* (4,771) yang menunjukkan keterikatan internal paling kuat antaranggota dalam *cluster*. Nilai *Dunn Index* tertinggi muncul pada empat *cluster* (0,509), menandakan pemisahan antarcluster yang relatif baik, namun perbedaannya tidak terlalu jauh dibandingkan dua *cluster* (0,405). Sementara itu, *Silhouette Coefficient* tertinggi juga ditunjukkan oleh dua *cluster* (0,663), yang menandakan bahwa struktur ini memiliki keseragaman internal paling tinggi dan pemisahan antarcluster paling jelas. Berdasarkan kombinasi ketiga indeks tersebut, jumlah *cluster* optimal yang dipilih adalah dua, karena memberikan hasil pengelompokan yang paling kompak, stabil, dan mudah diinterpretasikan. Hasil Dendogram dari 2 *cluster* terlihat pada Gambar 1.



**Gambar 1.** Dendogram *Single Linkage*

Dua cluster optimal yang terbentuk pada model *single linkage* disajikan dalam bentuk dendogram pada Gambar 1 yang menunjukkan masing-masing berisi 34 dan 4 anggota.

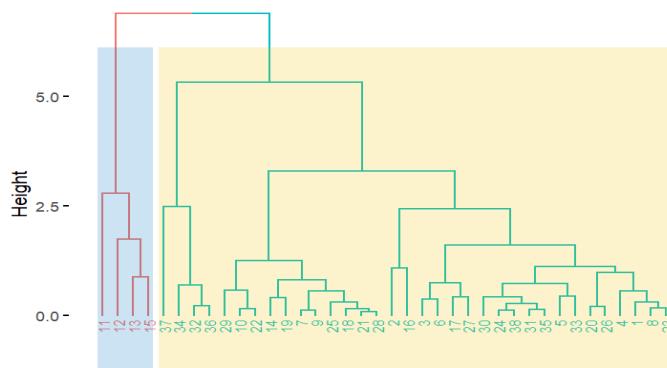
## 2. Complete Linkage

Pada model ini, Untuk menentukan jumlah *cluster* yang optimal, dilakukan evaluasi menggunakan *Internal Clustering Validation* dengan memvariasikan jumlah *cluster* dari 2 hingga 7 *cluster* pada setiap model hierarki dilihat dari nilai *dunn*, *connectivity*, dan *silhouette* yang terdapat pada Tabel 5.

**Tabel 5.** Uji Validitas Internal untuk *Complete Linkage*

Indeks Validitas	Jumlah Cluster					
	2	3	4	5	6	7
Connectivity	4,771	11,630	19,138	21,582	23,415	27,773
Dunn	0,405	0,109	0,125	0,140	0,143	0,200
Silhouette	0,663	0,480	0,388	0,368	0,337	0,379

Hasil uji *Internal Clustering Validation* pada model ini ditunjukkan oleh Tabel 5, bahwa indeks *dunn* tertinggi ada pada 2 *cluster*, sementara indeks *connectivity* terendah dan *silhouette* tertinggi menunjuk pada 2 *cluster*, sehingga pada model ini pengelompokan *cluster* optimal terdapat pada *cluster* berjumlah 2. Hasil Dendogram dari 2 *cluster* terlihat pada Gambar 2.



**Gambar 2.** Dendogram *Complete Linkage*

Dua cluster optimal yang terbentuk pada model *complete linkage* disajikan dalam bentuk dendrogram pada Gambar 2 yang menunjukkan masing-masing berisi 34 dan 4 anggota.

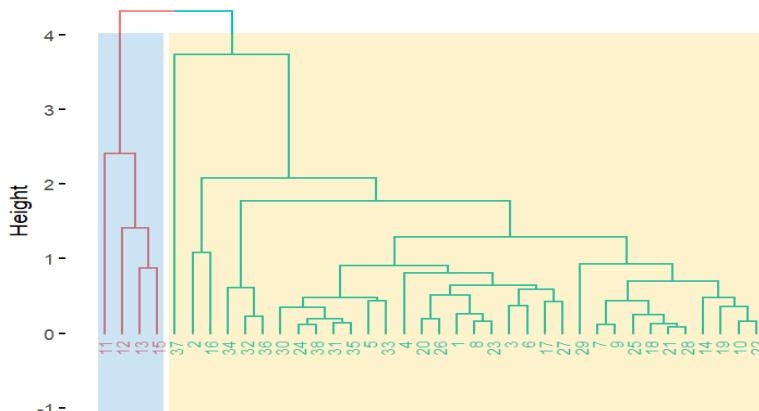
### **3. Average Linkage**

Pada model ini, Untuk menentukan jumlah *cluster* yang optimal, dilakukan evaluasi menggunakan *Internal Clustering Validation* dengan memvariasikan jumlah cluster dari 2 hingga 7 *cluster* pada setiap model hierarki dilihat dari nilai *dunn*, *connectivity*, dan *silhouette* yang terdapat pada Tabel 6.

**Tabel 6.** Uji Validitas Internal untuk *Average Linkage*

Indeks Validitas	Jumlah Cluster					
	2	3	4	5	6	7
Connectivity	4,771	7,700	10,145	14,613	20,266	22,766
Dunn	0,405	0,509	0,509	0,257	0,141	0,141
Silhouette	0,663	0,601	0,530	0,414	0,372	0,352

Hasil uji *Internal Clustering Validation* pada model ini ditunjukkan oleh Tabel 6, Nilai *Connectivity*, *Dunn Index*, dan *Silhouette Coefficient* digunakan untuk menilai kualitas hasil pengelompokan pada metode *Average Linkage*. Berdasarkan Tabel 6, nilai *Connectivity* terendah diperoleh pada dua *cluster* (4,771), menunjukkan keterikatan internal yang paling kuat antaranggota dalam *cluster*. Nilai *Dunn Index* tertinggi terdapat pada tiga dan empat *cluster* (0,509), menandakan pemisahan antarcluster yang baik, namun perbedaannya relatif kecil dibandingkan dua cluster (0,405). Sementara itu, *Silhouette Coefficient* tertinggi juga ditemukan pada dua *cluster* (0,663), yang menunjukkan tingkat keseragaman internal paling tinggi dan pemisahan antarcluster paling jelas. Berdasarkan kombinasi ketiga indeks tersebut, jumlah *cluster* optimal yang dipilih adalah dua, karena memberikan hasil pengelompokan yang paling stabil, kompak, dan mudah diinterpretasikan dalam konteks ketimpangan akses layanan kesehatan antarprovinsi di Indonesia. Hasil Dendogram dari 2 *cluster* terlihat pada Gambar 3.



Gambar 3. Dendogram Average Linkage

Dua *cluster* optimal yang terbentuk pada model *average linkage* disajikan dalam bentuk dendrogram pada Gambar 3 yang menunjukkan masing-masing berisi 34 dan 4 anggota.

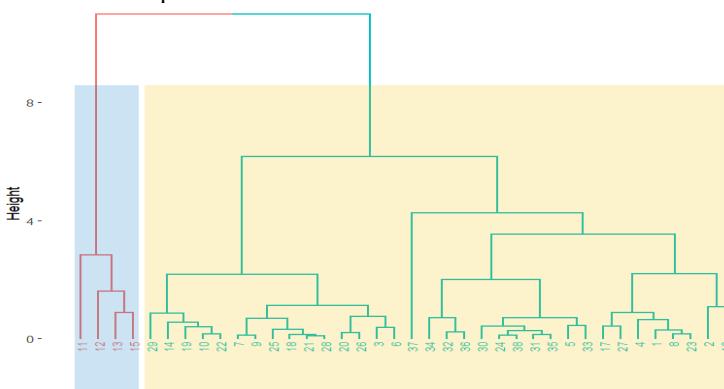
#### 4. Ward's Linkage

Pada model ini, Untuk menentukan jumlah *cluster* yang optimal, dilakukan evaluasi menggunakan *Internal Clustering Validation* dengan memvariasikan jumlah *cluster* dari 2 hingga 7 *cluster* pada setiap model hirarki dilihat dari nilai *dunn*, *connectivity*, dan *silhouette* yang terdapat pada Tabel 7.

Tabel 7. Uji Validitas Internal untuk Ward's Linkage

Indeks Validitas	Jumlah Cluster					
	2	3	4	5	6	7
Connectivity	4,771	14,746	17,675	22,236	24,680	28,477
Dunn	0,405	0,079	0,111	0,111	0,172	0,175
Silhouette	0,663	0,379	0,386	0,386	0,406	0,361

Hasil uji *Internal Clustering Validation* pada model ini ditunjukkan oleh Tabel 7, bahwa indeks *dunn* tertinggi ada pada 2 *cluster*, sementara indeks *connectivity* terendah dan *silhouette* tertinggi menunjuk pada 2 *cluster*, sehingga pada model ini pengelompokan *cluster* optimal terdapat pada *cluster* berjumlah 2. Hasil Dendrogram dari 2 *cluster* terlihat pada Gambar 4.



Gambar 4. Dendrogram Ward's Linkage

Dua *cluster* optimal yang terbentuk pada model *average linkage* disajikan dalam bentuk dendrogram pada Gambar 4 yang menunjukkan masing-masing berisi 34 dan 4 anggota.

#### 4.3 Perbandingan Model Clustering Terbaik

Keempat metode pengelompokan hirarki, yaitu *single linkage*, *complete linkage*, *average linkage*, dan *Ward's linkage*, seluruhnya menghasilkan struktur pengelompokan yang konsisten ketika jumlah *cluster* ditentukan sebanyak dua. Hal ini menunjukkan bahwa jika jumlah kelompok ditetapkan sebanyak dua, maka keempat metode tersebut sama-sama mampu membentuk hasil pengelompokan yang stabil dan dapat digunakan untuk mengidentifikasi perbedaan akses layanan kesehatan antarprovinsi. Oleh karena itu, untuk memastikan ketepatan struktur pengelompokan, pemilihan model terbaik selanjutnya dipertajam dengan menggunakan ukuran validitas internal, yaitu koefisien korelasi *cophenetic*. Ukuran ini digunakan untuk

menilai sejauh mana hasil pengelompokan mencerminkan jarak sebenarnya antarprovinsi dalam data. Analisis ini dilakukan dengan perbandingan jarak pengukuran kemiripan menggunakan koefisien korelasi *cophenetic* yang didapat dari jarak objek dendogram yang telah terbentuk untuk menentukan salah satu model pengelompokan terbaik.

**Tabel 8. Koefisien Korelasi Cophenetic**

<i>Single linkage</i>	<i>Complete linkage</i>	<i>Average linkage</i>	<i>Ward's linkage</i>
0,888	0,821	0,911	0,789

Berdasarkan Tabel 8 koefisien korelasi *cophenetic* tertinggi didapatkan 0,911 sehingga model terbaik dalam metode *clustering* hirarki pada penelitian ini adalah *Average linkage*.

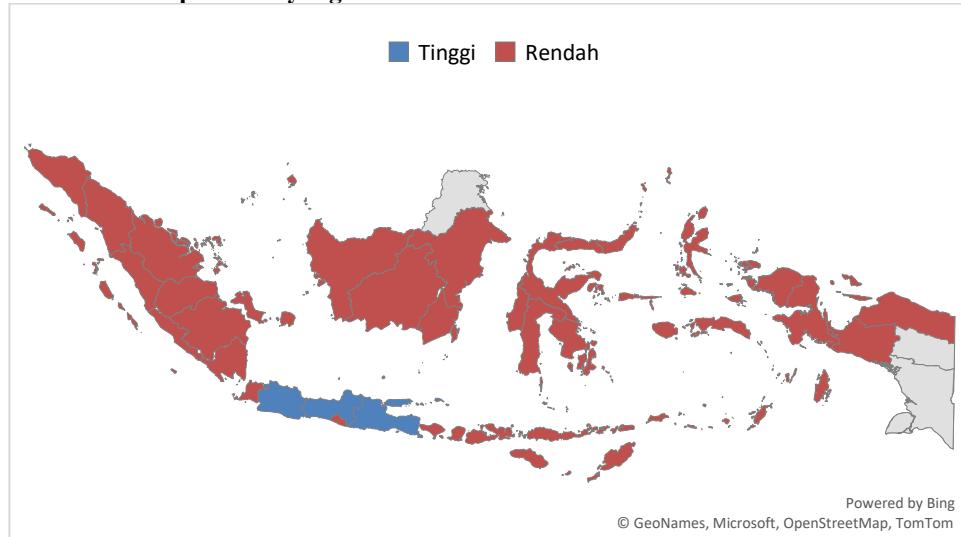
#### 4.4 Evaluasi dan Hasil Pengelompokan

Berdasarkan analisis uji *Internal Clustering Validation* serta koefisien korelasi *cophenetic* yang dilakukan terhadap keempat model pengelompokan, jumlah *cluster* optimal untuk mengelompokan provinsi di Indonesia berdasarkan akses layanan kesehatan adalah 2 *cluster* yang diperoleh dari perbandingan indeks *connectivity* dan indeks *silhouette* terhadap masing-masing model dimana hasil uji sama-sama menunjukkan optimal pada 2 *cluster*, sedangkan berdasarkan koefisien korelasi *cophenetic* diperoleh model terbaik dalam metode *clustering* hirarki pada penelitian ini adalah *Average Linkage*. Hasil *Clustering* dapat dilihat pada Tabel 9.

**Tabel 9. Hasil Clustering berdasarkan Average Linkage**

Provinsi di Indonesia	Cluster	
	1	2
	Acех, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur,	DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur.
Provinsi	Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua, Papua Barat, Papua Selatan, Papua Tengah, Papua Pegunungan, Papua Barat Daya.	

#### 4.5 Karakteristik Tiap Cluster yang Terbentuk

**Gambar 5. Pengelompokan Provinsi di Indonesia berdasarkan Akses Layanan Kesehatan**

Pemetaan hasil pengelompokan pada Gambar 5 menunjukkan terdapat 2 warna yang digunakan untuk menunjukkan hasil pengelompokan provinsi di Indonesia berdasarkan akses layanan kesehatan menggunakan model *Average linkage*. Wilayah berwarna biru menunjukkan *cluster* 1 yang terbentuk yaitu sebanyak 4 provinsi dan warna merah untuk *cluster* 2 sebanyak 38 provinsi. Untuk mengetahui pembagian pola karakteristik menggunakan data asli penelitian pada Tabel 10.

**Tabel 10.** Profilisasi *Cluster*

<b>Cluster</b>	<b>Kategori Rata-Rata Indikator</b>	<b>Jumlah Rumah Sakit Umum</b>	<b>Jumlah Tenaga Medis</b>	<b>Persentase Penduduk yang Mempunyai Keluhan Kesehatan tetapi Memilih Mengobati Sendiri</b>
1	Rendah	45,147	3095,235	60,724
2	Tinggi	293,750	25484,000	66,633

Berdasarkan hasil pengelompokan provinsi di Indonesia menggunakan metode *clustering Average Linkage*, dapat dilihat bahwa pola karakteristik untuk masing-masing *cluster* yang telah terbentuk pada Tabel 11 berikut.

**Tabel 11.** Karakteristik Setiap *Cluster*

<b>Cluster</b>	<b>Provinsi di Indonesia</b>	<b>Karakteristik</b>
1	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua, Papua Barat, Papua Selatan, Papua Tengah, Papua Pegunungan, Papua Barat Daya.	<ul style="list-style-type: none"> <li>- Jumlah rumah sakit umum yang tergolong rendah dengan rata-rata 45,147 dan nilai ini lebih rendah dari rata-rata nasional.</li> <li>- Jumlah tenaga medis yang rendah dengan rata-rata 3095,235 dan nilai ini lebih rendah dari rata-rata nasional.</li> <li>- Persentase penduduk yang mengobati sendiri tergolong rendah dengan rata-rata 60,724. Meskipun proporsi penduduk yang memilih mengobati sendiri sedikit lebih rendah, hal ini bisa saja disebabkan oleh kurangnya pilihan layanan kesehatan formal daripada kesadaran kesehatan yang tinggi.</li> <li>- Umumnya merupakan provinsi dengan infrastruktur kesehatan terbatas, kemungkinan berada di luar Pulau Jawa atau daerah 3T (tertinggal, terdepan, dan terluar).</li> <li>- Perlu perhatian lebih dalam distribusi fasilitas dan tenaga kesehatan oleh pemerintah pusat.</li> </ul>
2	DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur.	<ul style="list-style-type: none"> <li>- Jumlah rumah sakit umum yang tergolong tinggi dengan rata-rata 293,750.</li> <li>- Jumlah tenaga medis yang tinggi dengan rata-rata 25484,000.</li> <li>- Persentase penduduk yang mengobati sendiri tergolong tinggi dengan rata-rata 66,633. Meskipun ketersediaan rumah sakit dan tenaga medis sangat baik, tingginya persentase penduduk yang tetap memilih mengobati sendiri menunjukkan adanya faktor non-fasilitas yang berperan, seperti: preferensi budaya, aksesibilitas lokal, biaya, atau persepsi terhadap pelayanan kesehatan.</li> <li>- Perlu strategi non-fisik, seperti edukasi kesehatan dan peningkatan kualitas pelayanan, untuk menurunkan perilaku <i>self-medication</i>.</li> </ul>

*Clustering* ini dapat memberikan pemahaman yang lebih mendalam mengenai pola ketimpangan akses layanan kesehatan di berbagai provinsi di Indonesia. Hasil pengelompokan dapat membantu pemerintah dalam merancang kebijakan kesehatan yang lebih tepat sasaran dan efektif, sesuai dengan karakteristik masing-masing *cluster*. Provinsi yang tergolong dalam *Cluster 2*, yang memiliki jumlah rumah sakit dan tenaga medis tinggi serta persentase pengobatan mandiri yang juga tinggi, dapat difokuskan pada kebijakan peningkatan kualitas layanan kesehatan dan edukasi masyarakat mengenai pentingnya akses ke fasilitas kesehatan formal. Sementara itu, *Cluster 1* yang menunjukkan rendahnya ketersediaan fasilitas dan tenaga kesehatan dapat menjadi prioritas dalam distribusi sumber daya, pembangunan infrastruktur kesehatan, serta peningkatan jumlah tenaga medis. Pemerataan akses layanan kesehatan menjadi krusial agar seluruh penduduk, tanpa

memandang wilayah tempat tinggalnya, dapat memperoleh pelayanan yang layak dan berkualitas. Dengan mempertimbangkan perbedaan kebutuhan antar *cluster*, kebijakan yang dirancang akan lebih efisien dan berkontribusi pada pengurangan ketimpangan layanan kesehatan di Indonesia secara menyeluruh.

## 5. Kesimpulan

Penelitian ini berhasil mengelompokkan 38 provinsi di Indonesia ke dalam dua *cluster* berdasarkan ketimpangan akses layanan kesehatan menggunakan metode *cluster* hirarki, dengan pendekatan *average linkage* sebagai model terbaik berdasarkan nilai koefisien *cophenetic* tertinggi (0,911). *Cluster* pertama mencakup sebagian besar provinsi di luar Pulau Jawa dengan fasilitas dan tenaga medis yang terbatas, sementara *cluster* kedua terdiri dari empat provinsi di Pulau Jawa yang memiliki fasilitas kesehatan tinggi namun tingkat pengobatan mandiri juga tinggi. Temuan ini menunjukkan bahwa ketimpangan akses tidak hanya dipengaruhi oleh infrastruktur, tetapi juga oleh faktor sosial dan perilaku masyarakat. Oleh karena itu, diperlukan kebijakan yang disesuaikan seperti peningkatan infrastruktur dan distribusi tenaga medis untuk *cluster* pertama, serta intervensi non-struktural seperti edukasi kesehatan untuk *cluster* kedua, guna mendukung tercapainya pemerataan akses kesehatan sesuai target SDGs.

## 6. Referensi

- 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>
- Badan Pusat Statistik. (2023). *Statistik Kesehatan Indonesia 2023*. Jakarta: BPS.
- Bhahari, R. H., & Kusnawi. (2024). *Clustering Analysis of Socio-Economic Districts/Cities in East Java Province Using PCA and Hierarchical Clustering Methods*. Sinkron: Jurnal dan Penelitian Teknik Informatika, 8(4). <https://doi.org/10.33395/sinkron.v8i4.14078>
- Botta-Dukát, Z. (2023). *A new approach for evaluating internal cluster validation indices*. arXiv. <https://arxiv.org/abs/2308.03894>
- Halim, N. N., & Widodo, E. (2017). Clustering Dampak Gempa Bumi di Indonesia Menggunakan Kohonen Self Organizing Maps (SOM). Prosiding SI MaNIs (Seminar Nasional Integrasi Matematika Dan Nilai-Nilai Islami), 1(1): 188–194.
- Han, J., Pei, J., & Kamber, M. (2022). *Data Mining: Concepts and Techniques* (4th ed.). Morgan Kaufmann.
- Iglesias Vázquez, F., & Zseby, T. (2023). *Temporal silhouette: validation of stream clustering robust to concept drift*. *Machine Learning*, 113(11), 2067–2091. <https://doi.org/10.1007/s10994-023-06462-2>
- Kementerian Kesehatan Republik Indonesia. (2023). *Profil Kesehatan Indonesia Tahun 2022*. Jakarta: Kemenkes RI.
- Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan Republik Indonesia. (2023). *Menko PMK serahkan penghargaan Swasti Saba untuk kabupaten/kota sehat 2023*. URL: <https://www.kemenkopmk.go.id/menko-pmk-serahkan-penghargaan-swasti-saba-untuk-kabupatenkota-sehat-2023>. Diakses tanggal 22 Juni 2025.
- Liu, X., Cao, J., Ge, R., & Jiang, O. (2025). Evaluating medical service performance of hospitals in Sichuan Province, China: exploratory factor analysis and hierarchical clustering analysis based on diagnosis-related groups. *Frontiers in Public Health*, 13, 1617945.
- Nugroho, A., Suryanto, S., & Utomo, B. (2020). Disparitas akses pelayanan kesehatan di Indonesia: analisis data Riskedas. *Jurnal Kebijakan Kesehatan Indonesia*, 9(2), 87–96. <https://doi.org/10.7454/jkki.v9i2.3021>
- Putri, A. N., & Sari, K. P. (2021). Dimensi akses pelayanan kesehatan dan pengaruhnya terhadap pemanfaatan fasilitas kesehatan. *Jurnal Kesehatan Masyarakat Andalas*, 15(1), 45–52. <https://doi.org/10.24893/jkma.v15i1.396>
- Ramadhan, A. F., & Utami, S. R. (2022). Efektivitas program JKN dalam memperluas akses layanan kesehatan: Studi kasus di wilayah pedesaan. *Media Kesehatan Masyarakat Indonesia*, 21(4), 232–240. <https://doi.org/10.30597/mkmi.v21i4.16150>
- Riyadi, B., & Hanum, N. (2021). Ketimpangan Distribusi Tenaga Kesehatan di Indonesia: Analisis Deskriptif Berdasarkan Data Nasional. *Jurnal Kebijakan Kesehatan Indonesia*, 10(2), 85–95.
- Saraçlı, S., Doğan, N., & Doğan, İ. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation. *Journal of Inequalities and Applications*, 2013(1), 203. <https://doi.org/10.1186/1029-242X-2013-203>
- Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., ... & Lin, C. T. (2021). Parallel and scalable Dunn Index for the validation of big data clustering. *Pattern Recognition Letters*, 143, 1-7. <https://doi.org/10.1016/j.patrec.2021.01.011>

- Tan, P.-N., Steinbach, M., & Kumar, V. (2020). *Introduction to Data Mining* (2nd ed.). Pearson.
- Utari, D. T., & Hanun, D. S. (2021). Hierarchical Clustering Approach for Region Analysis of Contraceptive Users. *EKSAKTA: Journal of Sciences and Data Analysis*, 99-108. <https://doi.org/10.20885/EKSAKTA.vol2.iss2.art3>
- Vichi, M., Cavicchia, C., & Groenen, P. J. (2022). Hierarchical means clustering. *Journal of Classification*, 39(3), 553-577.
- Wijuniamurti, S., Nugroho, S., & Rachmawati, R. (2022). Agglomerative Nesting (AGNES) Method and Divisive Analysis (DIANA) Method for Hierarchical Clustering on Some Distance Measurement Concepts. *Journal of Statistics and Data Science*, 1(1). <https://doi.org/10.33369/jsds.v1i1.21009>
- Wulandari, R. D., Kurniasari, I. D., & Sari, D. K. (2020). Persepsi Masyarakat Terhadap Akses Pelayanan Kesehatan di Indonesia. *Jurnal Administrasi Kesehatan Indonesia*, 8(1), 34–42.
- Xu, D., & Tian, Y. (2022). A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science*, 9, 1–36. <https://doi.org/10.1007/s40745-020-00233-9>