E-ISSN : 2988-1986
https://ejournal.cahayailmubangsa.institute/index.php/kohesi

Kohesi: Jurnal Multidisiplin Saintek
Volume 9 No. 2 Tahun 2025

SEGMENTASI WILAYAH JAWA TIMUR BERDASARKAN JUMLAH TENAGA KESEHATAN MENGGUNAKAN *K-MEANS* DAN *AGGLOMERATIVE*

Mujtahida Azmi Fitri[1], Intan Nur'aulia[2], Findy Neyshella Sari[3], Aviolla Terza Damaliana[4], Shindi Shella May Wara[5]
Program Studi Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur
E-*mail*: 23083010036@student.upnjatim.ac.id[1], 23083010087@student.upnjatim.ac.id[2]
23083010095@student.upnjatim.ac.id[3], aviolla.terza.sada@upnjatim.ac.id[4]
shindi.shella.fasilkom@upnjatim.ac.id[5]

ABSTRACT

The unequal distribution of healthcare workers in East Java Province poses a major challenge to delivering quality health services. This study aims to analyze regional segmentation based on the number of healthcare workers using two clustering methods: K-Means and Agglomerative Hierarchical Clustering. Secondary data from East Java's Central Bureau of Statistics (BPS) for the year 2024 was used, covering various categories of healthcare personnel. After normalization using the Standard Scaler, the optimal number of clusters was determined through the Elbow Method, which indicated three as the most suitable number of clusters. Evaluation using the Silhouette Score revealed that Agglomerative Clustering (score 0.575) outperformed K-Means (score 0.397) in generating well-defined and representative clusters. The clustering results showed that Surabaya City formed a separate cluster due to its significantly higher number of healthcare workers, while other regions were grouped into medium and low-level clusters. Visualizations using PCA and dendrogram further supported the established cluster structure. These findings offer valuable insights for formulating more equitable and data-driven healthcare workforce distribution policies, with the goal of improving access to and the quality of healthcare services across East Java.

Keywords: Health workforce distribution, clustering, K-Means, Agglomerative Clustering, East Java.

ABSTRAK

Distribusi tenaga kesehatan yang tidak merata di Provinsi Jawa Timur menjadi tantangan utama dalam penyediaan layanan kesehatan yang berkualitas. Penelitian ini bertujuan untuk menganalisis segmentasi wilayah berdasarkan jumlah tenaga kesehatan menggunakan dua metode klasterisasi: K-Means dan Agglomerative Hierarchical Clustering. Data sekunder dari BPS Jawa Timur tahun 2024 digunakan, mencakup berbagai kategori tenaga kesehatan. Setelah proses normalisasi dengan Standard Scaler, jumlah klaster optimal ditentukan menggunakan metode Elbow, yang menunjukkan tiga klaster sebagai pilihan terbaik.

Article History
Received: Juli 2025
Reviewed: Juli 2025
Published: Juli 2025

Plagiarism Checker No 235
Prefix DOI : 10.8734/Kohesi.v1i2.365
Copyright : Author
Publish by : Kohesi

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Evaluasi menggunakan Silhouette Score menunjukkan bahwa Agglomerative Clustering (skor 0,575) lebih unggul dibandingkan K-Means (skor 0,397) dalam menghasilkan klaster yang jelas dan representatif. Hasil klasterisasi menunjukkan bahwa Kota Surabaya membentuk klaster tersendiri karena jumlah tenaga kesehatan yang sangat tinggi, sedangkan wilayah lain terbagi ke dalam klaster menengah dan rendah. Visualisasi melalui PCA dan dendrogram memperkuat struktur klaster yang terbentuk. Temuan ini memberikan wawasan penting bagi perumusan kebijakan distribusi tenaga kesehatan yang lebih merata dan berbasis data, guna meningkatkan akses dan kualitas layanan kesehatan di seluruh wilayah Jawa Timur.

Kata Kunci: Distribusi tenaga kesehatan, klasterisasi, K-Means, Agglomerative Clustering, Jawa Timur.

1. PENDAHULUAN

Tenaga kesehatan adalah elemen utama dalam sistem pelayanan kesehatan yang berkelanjutan. Ketersediaan dan distribusi tenaga kesehatan yang seimbang menjadi faktor penting dalam memastikan kualitas pelayanan kesehatan untuk masyarakat. Ketidakmerataan dalam distribusi tenaga kesehatan di berbagai daerah dapat menyebabkan perbedaan yang besar dalam akses dan kualitas layanan kesehatan. Masalah ini menjadi fokus utama, terutama di Provinsi Jawa Timur, yang memiliki luas area dan jumlah penduduk yang besar, serta berbagai karakteristik sosial dan ekonomi di setiap kabupaten atau kota.

Masalah mengenai penyebaran tenaga kesehatan sudah ada sejak lama. Informasi dari Dinas Kesehatan Provinsi dan juga laporan tingkat nasional menunjukkan terdapat perbedaan jumlah tenaga kesehatan di berbagai kabupaten atau kota, di mana beberapa daerah memiliki lebih banyak tenaga kesehatan, sedangkan daerah lainnya kekurangan. Ketidakseimbangan ini bisa memiliki pengaruh langsung pada layanan publik, keberhasilan program kesehatan, serta pencapaian tujuan pembangunan di sektor kesehatan.

Salah satu metode yang dapat diterapkan untuk secara objektif menganalisis pola distribusi ini adalah analisis clustering. Teknik ini efektif untuk mengelompokkan area yang memiliki karakteristik yang serupa, dalam konteks ini, dilihat dari jumlah tenaga kesehatan. Dua teknik pengelompokan yang sering digunakan adalah K-Means Clustering dan Agglomerative Hierarchical Clustering. K-Means adalah metode pemisahan yang mengelompokkan data ke dalam sejumlah kluster berdasarkan kesamaan data dengan pusat kluster (centroid), sementara Agglomerative merupakan metode hierarkis yang secara bertahap membentuk kluster dari data individual hingga tercipta satu hierarki [1], [2].

Beberapa penelitian sebelumnya telah mengimplementasikan metode clustering dalam konteks data kesehatan. Maryani et al. [1] mengelompokkan indikator keluarga sehat di Indonesia menggunakan K-Means Clustering. Sementara itu, Barekeng et al. [2] memanfaatkan kombinasi metode Agglomerative dan K-Means untuk mengelompokkan indeks kesehatan lingkungan di provinsi-provinsi di Indonesia. Fahmiyah & Ningrum [3] juga menggunakan metode K-Means untuk segmentasi provinsi berdasarkan Indeks Pembangunan Manusia (IPM), dan Wahyuni [4] memanfaatkan K-Means untuk pengelompokan wilayah tertinggal berdasarkan indikator kemiskinan.

Penelitian yang secara khusus menganalisis distribusi tenaga kesehatan di Indonesia dilakukan oleh Pahlevi et al. [5], yang menggunakan metode K-Means untuk mengelompokkan wilayah berdasarkan jumlah tenaga kesehatan. Penelitian ini menekankan pentingnya klasifikasi wilayah sebagai dasar pengambilan kebijakan distribusi tenaga kesehatan yang lebih merata dan efektif.

Namun demikian, penelitian tentang segmentasi daerah kabupaten atau kota di Provinsi Jawa Timur dengan dasar jumlah tenaga kesehatan yang menggunakan metode K-Means dan Agglomerative Clustering masih sangat jarang. Oleh karena itu, tujuan dari penelitian ini adalah untuk menganalisis dan mengelompokkan kabupaten atau kota di Jawa Timur berdasarkan jumlah tenaga kesehatan pada tahun 2024. Hasil dari analisis ini diharapkan dapat memberikan wawasan kepada para pembuat kebijakan dalam menetapkan prioritas dan strategi untuk meratakan distribusi tenaga kesehatan di tingkat daerah.

2. METODE PENELITIAN

Penelitian ini menggunakan data sekunder yang diperoleh dari statistik resmi terkait industri di Jawa Timur RM MANOVA diterapkan untuk menguji perbedaan variabel jumlah perusahaan, tenaga kerja, investasi dan nilai produksi antara dua tahun berturut-turut (2017-2018). Analisis Data yang digunakan dalam penelitian ini diperoleh dari Badan Pusat Statistiik (BPS) Jawa Timur, yang menyediakan data statistik Jumlah Tenaga kerja Industri termasuk jumlah perusahaan, tenaga kerja, investasi, nilai produksi. Variabel dalam penelitian ini terdiri dari satu variabel independen dan empat variabel dependen. Tujuan untuk mengetahui jumlah tenaga kerja industri berdasarkan tahun.

Analisis dilakukan menggunakan Uji Manova satu arah untuk menguji apakah perbedaan tidak signifikan dalam variabel Tahun Jumlah perusahaan antara dua data Tahun 2017-2018. Beberapa pengujian statistik yang dilakukan dalam penelitian ini meliputi :

2.1 Langkah Analisis

Langkah awal dalam analisis segmentasi wilayah dilakukan dengan mengumpulkan dan mempersiapkan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Timur tahun 2024. Data ini mencakup informasi jumlah tenaga kesehatan untuk berbagai kategori profesi yang tercantum dalam Tabel 1.

Tabel 1. Variabel Dataset dan Tipe Data

Nama Variabel	Tipe Data
Perawat	int64
Bidan	float64
Tenaga Kefarmasian	int64
Tenaga Kesehatan Masyarakat	int64
Tenaga Kesehatan Lingkungan	int64
Tenaga Gizi	int64
Jumlah Tenaga Medis	int64
Jumlah Tenaga Kesehatan Psikologi Klinis	int64
Jumlah Tenaga Keterapian Fisik	int64
Jumlah Tenaga Keteknisan Medis	int64
Jumlah Tenaga Teknik Biomedika	int64

Setelah dikumpulkan, data diolah menggunakan perangkat lunak Python. Tahapan analisis diawali dengan eksplorasi data secara deskriptif untuk memahami karakteristik distribusi tenaga kesehatan di setiap wilayah. Selanjutnya dilakukan proses pra-pemrosesan seperti normalisasi data menggunakan Standard Scaler. Proses segmentasi wilayah dilakukan menggunakan dua metode klasterisasi, yaitu K-Means dan Agglomerative Hierarchical Clustering. Evaluasi terhadap hasil klasterisasi dilakukan menggunakan metrik seperti Silhouette Score untuk menilai kualitas pemisahan antar klaster.

2.2 Normalisasi Data

Tahap awal sebelum dilakukannya analisis klasterisasi adalah proses normalisasi data, guna memastikan bahwa seluruh variabel berada dalam skala yang sebanding. Proses ini penting karena metode klasterisasi yang digunakan bersifat berbasis jarak, sehingga perbedaan skala antar variabel dapat menyebabkan distorsi dalam pengelompokan data. Oleh karena itu, digunakan metode Standard Scaler berbasis Z-score sebagai teknik normalisasi.

Metode Z-score mengubah nilai asli (X) menjadi nilai standar (Z) dengan rumus sebagai berikut:

$$Z = \frac{X - \mu}{\sigma}$$

dengan :

Z = nilai yang telah dinormalisasi (Z-score),

X = nilai asli variabel,

μ = rata-rata dari variabel,

σ = standar deviasi dari variabel.

Transformasi ini menghasilkan data dengan distribusi baru yang memiliki rata-rata nol dan standar deviasi satu, sehingga semua fitur memiliki bobot yang setara dalam perhitungan jarak. Dengan demikian, tidak ada variabel yang mendominasi proses klasterisasi hanya karena memiliki nilai absolut yang lebih besar.

Penggunaan normalisasi berbasis Z-score juga telah terbukti secara empiris dapat meningkatkan stabilitas algoritma klasterisasi serta memperbaiki akurasi hasil pengelompokan pada data numerik berskala berbeda. Studi sebelumnya menunjukkan bahwa standardisasi merupakan salah satu langkah krusial dalam proses prapemrosesan data untuk analisis berbasis pembelajaran tanpa pengawasan [7].

2.3 Penentuan Jumlah Klaster Optimal

Penentuan jumlah klaster yang optimal merupakan langkah penting dalam analisis klasterisasi untuk memastikan bahwa struktur pengelompokan yang dihasilkan mencerminkan pola alami dalam data. Pada penelitian ini digunakan metode Elbow, yang merupakan pendekatan visual berbasis metrik Within-Cluster Sum of Squares (WCSS).

WCSS adalah total jumlah kuadrat jarak antara setiap titik data dengan centroid klasternya. Nilai WCSS dihitung untuk berbagai jumlah klaster (k), kemudian dipetakan ke dalam grafik untuk menemukan titik optimal. Secara matematis, WCSS dapat dirumuskan sebagai:

$$WCSS = \sum_{j=1}^{k} \sum_{i=1}^{n_j} ||x_i^{(j)} - \mu_j||^2$$

Dimana :

k = Jumlah klaster,

n_j = Jumlah data pada klaster ke-j,

$x_i^{(j)}$ = Titik data ke-i dalam klaster ke-j,

μ_j = Centroid dari klaster ke-j,

$||x_i^{(j)} - \mu_j||^2$ = Jarak Euclidean kuadrat antara titik data ke-iii pada klaster ke-j dengan centroid klaster tersebut.

Dalam metode Elbow, nilai WCSS diplot terhadap jumlah klaster (k). Jumlah klaster optimal ditentukan pada titik di mana penurunan WCSS mulai melambat secara signifikan, membentuk pola seperti "siku" pada grafik. Titik tersebut dianggap sebagai kompromi terbaik antara kompleksitas model (jumlah klaster) dan kemampuan representasi terhadap struktur data. Pendekatan ini telah banyak digunakan dalam studi segmentasi wilayah dan terbukti efektif dalam mengidentifikasi jumlah klaster yang efisien dan interpretatif [8].

2.4 Metode Klasterisasi

Setelah jumlah klaster optimal ditentukan, proses segmentasi wilayah dilakukan menggunakan dua pendekatan algoritma klasterisasi, yaitu K-Means Clustering dan Agglomerative Hierarchical Clustering. K-Means Clustering adalah algoritma partisi yang membagi data ke dalam k klaster berdasarkan jarak minimum antara data dan centroid klaster. Tujuannya adalah meminimalkan variasi dalam klaster (intra-cluster) dan memaksimalkan perbedaan antar klaster (inter-cluster).

$$J = \sum_{j=1}^{k} \sum_{x_i \epsilon C_j} \left|\left|x_i - \mu_j\right|\right|^2$$

Dimana :

J = Total jumlah kuadrat jarak antara data dan pusat klaster,

k = Jumlah klaster,

Cj = Klaster ke-j,

xi = Data ke-i dalam klaster Cj,

μ_j = Centroid (titik pusat) dari klaster Cj,

$\left|\left|x_i - \mu_j\right|\right|$ = Jarak Euclidean antara data ke-i dan centroid ke-j.

K-Means dikenal luas karena kesederhanaan dan efisiensi komputasinya, serta cocok digunakan untuk data berukuran besar dan berdimensi tinggi.

Agglomerative Hierarchical Clustering merupakan algoritma klasterisasi bottom-up yang memulai pengelompokan dari setiap data sebagai klaster tunggal. Selanjutnya, klaster-klaster tersebut digabungkan secara bertahap berdasarkan ukuran kemiripan (linkage criteria), seperti single linkage, complete linkage, atau average linkage, hingga seluruh data tergabung dalam

satu klaster. Jarak antar klaster dapat dihitung menggunakan metode average linkage sebagai contoh:

$$D(A,B) = \frac{1}{|A||B|}\sum_{x\in A}\sum_{y\in B}||x-y||$$

Dimana :

A dan B = dua klaster yang dibandingkan,

$|A|, |B|$ = jumlah data dalam masing-masing klaster,

$||x-y||$ = jarak Euclidean antara titik x dan y.

Hasil dari metode ini divisualisasikan dalam bentuk dendrogram, yang merepresentasikan struktur hierarki dan tingkat kemiripan antar klaster. Metode ini efektif untuk mengungkap hubungan bertingkat antar wilayah berdasarkan kemiripan distribusi tenaga kesehatan.

2.5 *Visualisasi Klaster*

Untuk metode klasterisasi hierarkis, digunakan dendrogram sebagai alat visualisasi. Dendrogram menggambarkan proses penggabungan antar klaster secara bertingkat, serta menunjukkan jarak antar data sebelum digabungkan. Visualisasi ini memudahkan identifikasi struktur klaster dan penentuan jumlah klaster potensial melalui pemotongan pada tingkat tertentu.

Penggunaan dendrogram dalam penelitian ini bertujuan untuk mengevaluasi pola distribusi tenaga kesehatan antar wilayah secara lebih intuitif dan mendalam, serta mendukung interpretasi hasil dari metode Agglomerative Hierarchical Clustering.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Deskriptif

Analisis deskriptif dilakukan untuk memahami karakteristik awal dari setiap variabel tenaga kesehatan di wilayah Provinsi Jawa Timur. Hasil statistik deskriptif disajikan dalam Tabel 2, yang mencakup nilai rata-rata (mean), simpangan baku (standar deviasi), nilai minimum, dan maksimum dari setiap jenis tenaga kesehatan.

Tabel 2. Hasil Statistika Deskriptif

Variabel	Mean	Standar Deviasi	Min	Max
Perawat	761.66	1209.12	185	7446
Bidan	5.79	11.88	1	69
Tenaga Kefarmasian	1949.39	1834.35	568	11570
Tenaga Kesehatan Masyarakat	873.08	433.79	154	2047
Tenaga Kesehatan Lingkungan	473.92	570.08	107	3482
Tenaga Gizi	96.63	82.30	26	532
Jumlah Tenaga Medis	48.16	34.71	11	228
Jumlah Tenaga Kesehatan Psikologi Klinis	93.37	70.35	28	437
Jumlah Tenaga Keterapian Fisik	43.97	69.83	4	433
Jumlah Tenaga Keteknisan Medis	180.63	194.36	57	1209
Jumlah Tenaga Teknik Biomedika	269.24	357.82	79	2280

Berdasarkan hasil analisis deskriptif, variabel Tenaga Kefarmasian memiliki rata-rata tertinggi sebesar 1.949,39 dengan standar deviasi sebesar 1.834,35, yang mencerminkan variasi yang besar antar wilayah. Variabel Perawat juga menonjol dengan rata-rata 761,66 dan standar deviasi 1.209,12, menunjukkan ketimpangan distribusi yang cukup signifikan, terutama dengan nilai maksimum mencapai 7.446. Sementara itu, variabel dengan rata-rata terendah adalah Jumlah Tenaga Keterapian Fisik (43,97) dan Bidan (5,79), dengan standar deviasi bidan yang relatif besar terhadap nilai rata-ratanya, menandakan adanya outlier atau wilayah ekstrem.

Secara keseluruhan, sebagian besar variabel menunjukkan sebaran nilai yang luas, menandakan distribusi tenaga kesehatan yang tidak merata antar kabupaten/kota. Untuk itu, sebelum dilakukan klasterisasi menggunakan metode K-Means dan Agglomerative Clustering, data perlu dinormalisasi guna menghilangkan pengaruh skala antar variabel. Normalisasi ini penting agar setiap variabel berkontribusi secara seimbang terhadap pembentukan klaster, mengingat kedua metode sangat bergantung pada perhitungan jarak antar data.

3.2 Normalisasi Data

Proses normalisasi dilakukan sebelum tahap klasterisasi guna menyetarakan skala antar variabel dan menghilangkan bias akibat perbedaan satuan atau rentang nilai. Pada penelitian ini, metode standarisasi Z-score digunakan karena mampu mengubah distribusi data menjadi

memiliki rata-rata nol dan simpangan baku satu.

Tabel 3. Hasil Normalisasi

Perawat	Bidan	Tenaga Kefarmasian	Tenaga Kesehatan Masyarakat	....	Jumlah Tenaga Teknik Biomedika
-0.483327	-0.255805	-0.668707	-1.086515	....	-0.470816
-0.265407	-0.164049	-0.064857	-0.133348	....	-0.173435
-0.443096	0.000000	-0.489154	-0.855232	....	-0.357528
.....	.....	.....	.....	....	.....
-0.384425	-0.164049	-0.737213	-1.679909	....	-0.535957

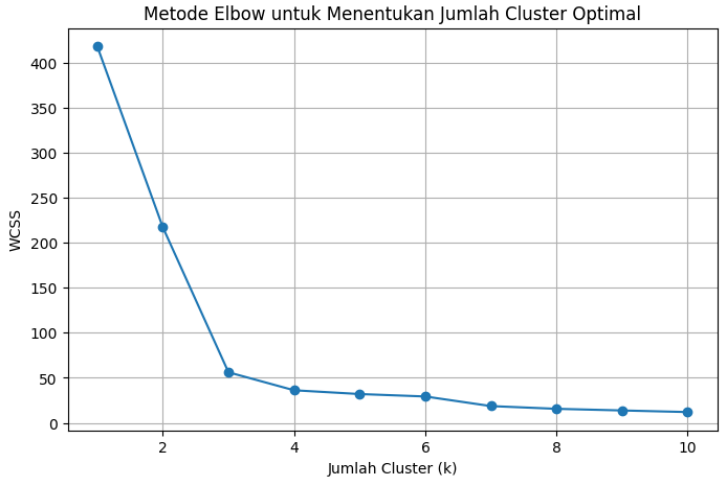
Hasil normalisasi menunjukkan bahwa seluruh variabel seperti “Perawat”, “Tenaga Kefarmasian”, hingga “Jumlah Tenaga Teknik Biomedika” telah berada dalam skala yang seragam, dengan nilai-nilai yang tersebar di kisaran -1 hingga +1. Misalnya, nilai variabel “Perawat” pada beberapa wilayah berada di rentang -0.48 hingga -0.14, yang menunjukkan telah berhasil dinormalisasi dari skala awalnya yang sangat besar.

Dengan skala yang telah diseragamkan, setiap variabel memiliki kontribusi yang seimbang dalam proses penghitungan jarak antar objek data, yang sangat penting dalam algoritma K-Means dan Agglomerative Clustering. Hal ini memastikan bahwa hasil pengelompokan tidak dipengaruhi oleh besarnya nilai suatu variabel semata, melainkan berdasarkan pola distribusi data secara keseluruhan. Setelah tahap normalisasi ini, proses selanjutnya adalah menentukan jumlah klaster optimal yang akan digunakan. Penentuan ini dilakukan dengan menggunakan metode evaluasi seperti Elbow Method dan Silhouette Score, yang dapat mengidentifikasi jumlah klaster paling representatif terhadap struktur data.

3.3 Penentuan Jumlah Klaster Optimal

Untuk menentukan jumlah cluster yang optimal digunakan metode Elbow yang memvisualisasikan hubungan antara jumlah klaster (k) dengan nilai Within-Cluster Sum of Squares (WCSS).

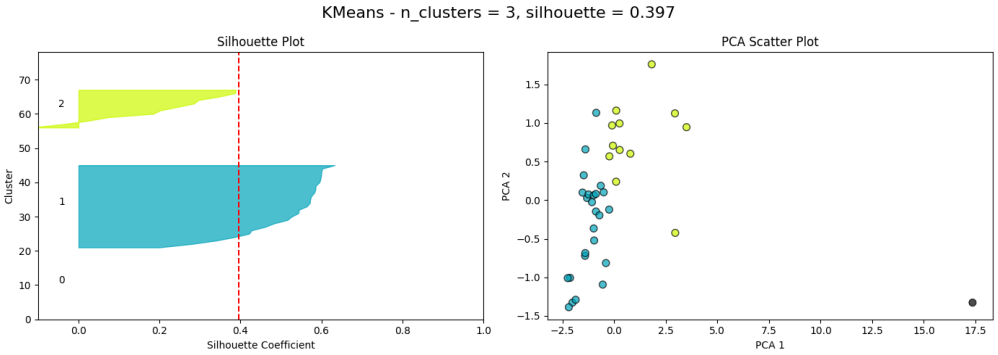
Gambar 1. Grafik Elbow



Gambar 1 di atas menampilkan hasil visualisasi Elbow Method yang digunakan untuk menentukan jumlah cluster optimal dalam analisis segmentasi wilayah berdasarkan jumlah tenaga kesehatan di Provinsi Jawa Timur. Grafik menunjukkan hubungan antara jumlah klaster (k) dan nilai Within-Cluster Sum of Squares (WCSS), dengan rentang k dari 1 hingga 10. Terlihat bahwa nilai WCSS mengalami penurunan tajam dari klaster 1 ke 3, dan setelahnya penurunannya cenderung melandai. Titik siku atau "elbow" pada grafik ini terlihat jelas pada k = 3, yang menandakan bahwa pada titik tersebut terjadi keseimbangan antara jumlah klaster dan variabilitas data di dalam klaster. Oleh karena itu, jumlah klaster optimal yang digunakan dalam penelitian ini adalah tiga. Pemilihan jumlah klaster ini menjadi dasar dalam penerapan algoritma K-Means dan Agglomerative Clustering pada tahap berikutnya, guna menghasilkan segmentasi wilayah yang lebih representatif terhadap distribusi tenaga kesehatan.

3.4 Hasil Klasterisasi

Gambar 2. Visualisasi KMeans, Silhouette Plot (Kiri), Scatter Plot Variabel (Kanan)

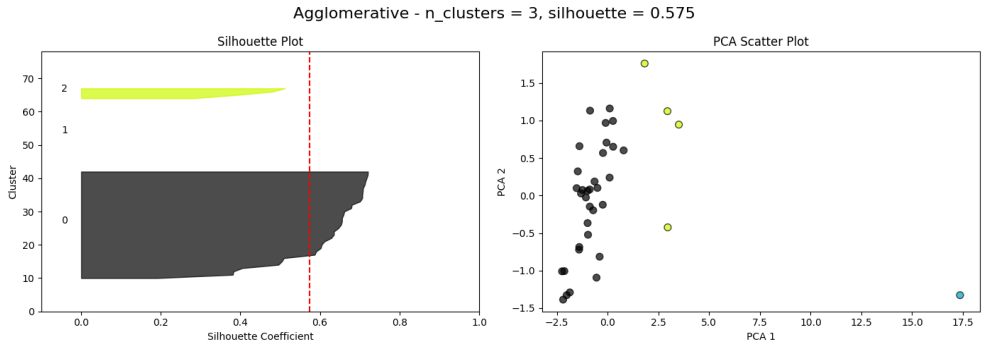


Pada visualisasi gambar 2, hasil pengelompokan menggunakan metode KMeans clustering dengan tiga klaster (k = 3) menunjukkan nilai silhouette score sebesar 0,397. Nilai ini mengindikasikan bahwa hasil klasterisasi berada dalam kategori cukup baik, meskipun masih terdapat beberapa tumpang tindih antar klaster. Plot silhouette memperlihatkan bahwa klaster 1 mendominasi dengan jumlah anggota terbanyak, diikuti oleh klaster 2, sementara klaster 0

hanya terdiri dari satu anggota. Visualisasi melalui scatter plot PCA menunjukkan bahwa sebagian besar data terkonsentrasi dalam dua kelompok utama, sedangkan satu titik berada jauh terpisah sebagai outlier, yaitu Kota Surabaya dalam klaster 0. Hal ini menunjukkan bahwa Surabaya memiliki karakteristik jumlah tenaga kesehatan yang sangat berbeda (jauh lebih tinggi) dibanding wilayah lainnya, sesuai dengan posisinya sebagai pusat layanan kesehatan dan kota metropolitan.

Karakteristik wilayah dalam masing-masing klaster memperkuat hasil tersebut. Klaster 0, yang hanya beranggotakan Kota Surabaya, merepresentasikan wilayah dengan jumlah tenaga kesehatan yang sangat tinggi dan berbeda secara signifikan dari daerah lain. Klaster 1 terdiri dari 25 wilayah, yang sebagian besar merupakan kabupaten dan kota kecil seperti Pacitan, Ponorogo, Kota Blitar, dan Kota Mojokerto, dengan karakteristik jumlah tenaga kesehatan yang rendah hingga sedang. Sementara itu, Klaster 2 beranggotakan 12 wilayah, termasuk Kota Malang, Sidoarjo, Gresik, dan Banyuwangi, yang mencerminkan wilayah dengan jumlah tenaga kesehatan yang relatif tinggi, namun masih berada dalam rentang umum dan tidak ekstrem seperti Surabaya.

Gambar 3. Visualisasi Agglomerative, Silhouette Plot (Kiri), Scatter Plot Variabel (Kanan)



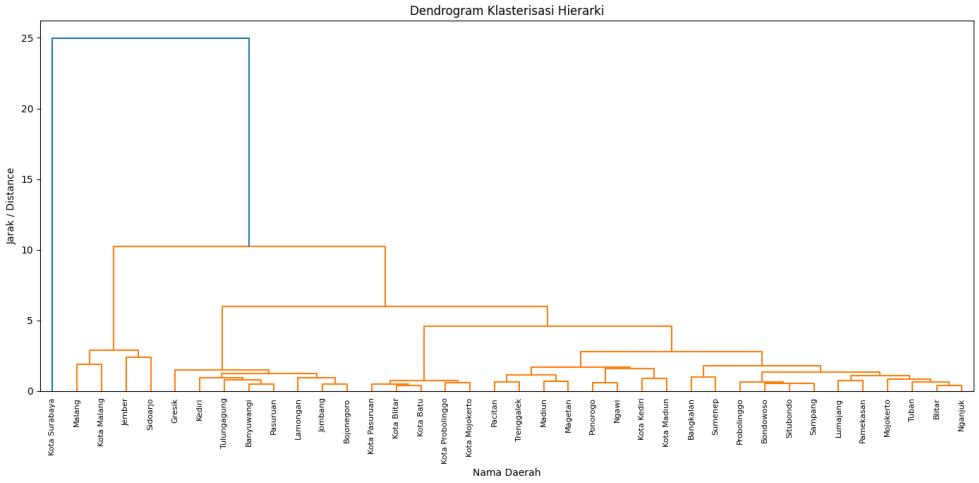
Gambar 2 di atas menunjukkan hasil klasterisasi menggunakan metode Agglomerative Clustering dengan tiga klaster (n_clusters = 3), dengan nilai koefisien siluet sebesar 0,575. Nilai ini mengindikasikan bahwa pemisahan antar klaster cukup baik dan struktur klaster relatif jelas. Plot siluet memperlihatkan sebagian besar objek memiliki nilai positif, menandakan kecocokan terhadap klasternya masing-masing. Sementara itu, visualisasi PCA menunjukkan bahwa ketiga klaster cukup terpisah, meskipun terdapat sedikit tumpang tindih.

Karakteristik masing-masing klaster memperkuat hasil ini. Klaster 1 diisi oleh wilayah dengan jumlah tenaga kesehatan sangat tinggi, seperti Surabaya, yang berperan sebagai pusat layanan kesehatan. Klaster 0 mencakup daerah dengan ketersediaan tenaga kesehatan rendah, umumnya kabupaten non-perkotaan. Klaster 2 berada di tingkat sedang, terdiri dari

kota/kabupaten berkembang seperti Malang, Jember, dan Sidoarjo. Segmentasi ini menunjukkan bahwa klasterisasi tidak hanya efektif secara visual, tetapi juga mencerminkan kondisi nyata distribusi tenaga kesehatan.

3.5 Hasil Visualisasi Klaster

Gambar 4. Grafik Dendogram



Pada gambar 4 menunjukkan hasil klasterisasi hirarki menggunakan metode agglomerative dengan pendekatan linkage tertentu (misalnya, average linkage). Pada dendrogram tersebut, terlihat bahwa wilayah-wilayah di Jawa Timur membentuk tiga klaster utama yang dipisahkan berdasarkan jarak atau tingkat kemiripan antar daerah. Klaster pertama secara mencolok dihuni oleh Kota Surabaya, yang berdiri sendiri karena memiliki karakteristik jumlah tenaga kesehatan yang jauh lebih tinggi dibanding wilayah lainnya. Klaster kedua mencakup beberapa wilayah seperti Malang, Kota Malang, Jember, dan Sidoarjo, yang cenderung memiliki jumlah tenaga kesehatan cukup tinggi namun tidak se-ekstrem Surabaya. Sementara itu, klaster ketiga berisi sebagian besar kabupaten lainnya yang memiliki jumlah tenaga kesehatan relatif rendah dan saling memiliki kemiripan. Jarak atau ketinggian pada dendrogram menunjukkan tingkat disimilaritas antar klaster, di mana semakin tinggi garis penggabungan, semakin besar perbedaan antar kelompok. Struktur dendrogram ini mencerminkan segmentasi yang alami dalam data distribusi tenaga kesehatan di wilayah Jawa Timur dan mendukung validitas pengelompokan hasil klasterisasi.

3.6 Analisis Perbandingan

Hasil analisis menunjukkan bahwa metode K-Means dan Agglomerative Clustering memberikan hasil klasterisasi yang berbeda dalam mengelompokkan wilayah berdasarkan distribusi tenaga kesehatan. Metode Agglomerative Clustering menghasilkan nilai koefisien siluet sebesar 0,575, lebih tinggi dibandingkan K-Means yang hanya mencapai 0,397, sehingga

mengindikasikan kualitas pemisahan klaster yang lebih baik. Plot siluet menunjukkan bahwa sebagian besar wilayah dalam hasil klasterisasi Agglomerative memiliki nilai siluet positif, menandakan kecocokan tinggi antar objek dalam satu klaster. Sebaliknya, hasil K-Means menunjukkan adanya tumpang tindih antar klaster, ditunjukkan oleh nilai siluet yang lebih rendah dan cenderung negatif pada beberapa wilayah. Visualisasi PCA juga memperkuat temuan ini, di mana klaster hasil Agglomerative lebih terpisah secara spasial dibandingkan K-Means.

Adapun karakteristik masing-masing klaster yang terbentuk melalui metode Agglomerative menunjukkan variasi signifikan dalam jumlah tenaga kesehatan.

Tabel 4. Hasil Klaster Agglomerative

Klaster ke-	Wilayah	Keterangan
0	Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Lumajang, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Gresik, Bangkalan, Sampang, Pamekasan, Sumenep, Kota Kediri, Kota Blitar, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Batu	Jumlah tenaga kesehatan relatif rendah (rata-rata 444 perawat, 39 tenaga medis). Mencerminkan keterbatasan akses layanan kesehatan.
1	Kota Surabaya	Jumlah tenaga kesehatan sangat tinggi (rata-rata 7.446 perawat, 11.570 tenaga kefarmasian). Menjadi pusat layanan kesehatan provinsi.
2	Malang, Jember, Sidoarjo, Kota Malang	Jumlah tenaga kesehatan menengah (rata-rata 1.708 perawat, 3.783 tenaga kefarmasian). Menopang layanan kesehatan regional.

Segmentasi ini menunjukkan bahwa hasil klasterisasi tidak hanya membentuk kelompok yang terpisah secara visual pada dendrogram, tetapi juga mencerminkan kondisi nyata distribusi tenaga kesehatan di berbagai wilayah di Jawa Timur.

KESIMPULAN

Penelitian ini bertujuan untuk melakukan segmentasi wilayah Kabupaten/Kota di Provinsi Jawa Timur berdasarkan jumlah tenaga kesehatan dengan menggunakan dua metode

klasterisasi, yaitu K-Means dan Agglomerative Clustering. Analisis menunjukkan bahwa metode Agglomerative Clustering menghasilkan klasterisasi yang lebih baik dengan nilai koefisien siluet 0,575, lebih tinggi dibandingkan K-Means yang menghasilkan 0,397. Temuan ini menunjukkan bahwa struktur klaster yang dihasilkan oleh metode agglomerative lebih efektif dalam menggambarkan kesamaan antar wilayah berdasarkan jumlah tenaga kesehatan.

Proses klasterisasi menghasilkan tiga klaster utama, yaitu klaster dengan jumlah tenaga kesehatan sangat tinggi (Kota Surabaya), klaster dengan tingkat menengah (Malang, Sidoarjo, Jember), dan klaster dengan jumlah tenaga kesehatan relatif rendah yang mencakup sebagian besar kabupaten lainnya. Segmentasi ini mencerminkan adanya ketimpangan distribusi tenaga kesehatan antarwilayah serta memberikan gambaran kondisi aktual yang lebih terstruktur.

Namun demikian, untuk mendukung pemerataan distribusi tenaga kesehatan yang optimal, hasil segmentasi ini perlu diintegrasikan dengan faktor-faktor lain seperti jumlah penduduk di masing-masing wilayah, tingkat beban penyakit, kondisi geografis, serta aksesibilitas layanan kesehatan. Tidak semua daerah dengan jumlah penduduk rendah memerlukan jumlah tenaga kesehatan yang sama seperti daerah padat penduduk atau dengan tingkat penyakit tinggi. Oleh karena itu, kebijakan pemerataan tenaga kesehatan perlu disesuaikan dengan kebutuhan riil masing-masing wilayah, sehingga dapat mengoptimalkan pelayanan kesehatan dan meningkatkan efisiensi distribusi tenaga kesehatan di Provinsi Jawa Timur secara berkelanjutan.

DAFTAR PUSTAKA

[1] H. Maryani, A. Rizkianti, dan N. Izza, “Classification of Healthy Family Indicators in Indonesia Based on a K-means Cluster Analysis,” *J. Prev. Med. Public Health*, vol. 57, no. 3, pp. 234–241, May 2024, doi: 10.3961/jpmph.23.497.

[2] U. Mahmudah dan M. S. Lola, “A Two-Step Cluster for Classifying Provinces in Indonesia Based on Environmental Quality,” *BAREKENG: J. Math. & App.*, vol. 17, no. 3, pp. 1685–1694, Sept. 2023, doi: 10.30598/barekengvol17iss3pp1685-1694.

[3] I. Fahmiyah dan R. A. Ningrum, “Human Development Clustering in Indonesia: Using K-Means Method and Based on Human Development Index Categories,” *J. Adv. Technol. Multidiscip. (JATM)*, vol. 2, no. 1, pp. 27–33, 2023.

[4] R. Wahyuni, “K-Means Clustering for Grouping Indonesia Underdeveloped Regions in 2020 Based on Poverty Indicators,” *J. Parameter*, vol. 2, no. 1, pp. 8–15, 2021. [Online]. Available: https://bestjournal.untad.ac.id/index.php/parameter/article/view/15675.

[5] A. Setiyaji dan H. D. Purnomo, "Analysis of Health Worker Distribution in Indonesia Using K-Means," *J. Teknol. dan Sist. Inf. (JATISI)*, vol. 11, no. 3, 2024, doi: 10.35957/jatisi.v11i3.8367.

[6] Badan Pusat Statistik Provinsi Jawa Timur, *Statistik Kesehatan Provinsi Jawa Timur 2024* [Online]. Surabaya: BPS Jatim, 2024. Available: https://jatim.bps.go.id

[7] C. Wongoutong, W. Petcharakorn, and T. Kerdprasop, "*The impact of neglecting feature scaling in k-means clustering*," PLOS ONE, vol. 19, no. 4, Apr. 2024. [Online]. Available: https://doi.org/10.1371/journal.pone.0310839

[8] A. Rajsya, P. Purnawansyah, and A. R. Manga', "Rancang Bangun Penerapan Metode Elbow Pada K-Means Untuk Clustering Data Persediaan Barang," LINIER: Jurnal Teknik Informatika, vol. 1, no. 4, 2024. [Online]. Available: https://jurnal.fikom.umi.ac.id/index.php/LINIER/article/view/2539

[9] I. Indra, N. Nur, M. Iqram, dan N. Inayah, "Perbandingan K-Means dan Hierarchical Clustering dalam Pengelompokan Daerah Beresiko Stunting," INOVTEK Polbeng – Seri Informatika, vol. 8, no. 2, pp. 356–366, 2023. [Online]. Available: https://ejournal.polbeng.ac.id/index.php/ISI/article/viewFile/3612/1607