

Segmentasi Wilayah Berdasarkan Indikator Kesehatan Lingkungan dan Akses Pelayanan Dasar di Provinsi Jawa Timur

Indah Rahma Abdillah^{1*}, Diana Novitasari², Amellia Harmaimun Hidayah³,
Shindi Shella May Wara⁴, Wahyu Syaifullah Jauharis Saputra⁵

^{1,2,3,4,5} Program Studi Sains Data, Universitas Pembangunan Nasional “Veteran” Jawa Timur, Jl. Rungkut Madya, Gunung. Anyar, Kecamatan. Gunung. Anyar, Surabaya, Jawa Timur 60294, Indonesia



P-ISSN: 2986-4178
E-ISSN: 2988-4004

Riwayat Artikel

Dikirim: 07 Juni 2025
Direvisi: 25 Juli 2025
Diterima: 01 Oktober 2025

ABSTRAK

Upaya peningkatan kesehatan lingkungan dan pelayanan dasar memerlukan pemahaman yang mendalam terhadap karakteristik wilayah. Provinsi Jawa Timur, dengan keragaman kondisi antar Kabupaten/Kota, menjadi contoh penting dalam analisis ini. Pengelompokan wilayah dilakukan berdasarkan tujuh indikator, yaitu akses air minum layak, akses sanitasi layak, kepemilikan jamban, kasus diare, keluhan kesehatan, kepadatan penduduk, dan jumlah puskesmas. Metode *Hierarchical Agglomerative Clustering* (HAC) dan *K-Means Clustering* diterapkan untuk membentuk kluster wilayah yang homogen. Evaluasi performa klusterisasi menggunakan Silhouette Score, Calinski-Harabasz Index, dan Dunn Index menunjukkan bahwa HAC menghasilkan segmentasi yang lebih optimal. Analisis menghasilkan lima kluster wilayah dengan karakteristik berbeda: (1) kabupaten dengan kepadatan sedang, sanitasi terbaik, namun kasus diare dan keluhan kesehatan tinggi; (2) kabupaten dengan sanitasi rendah namun kasus diare dan keluhan kesehatan rendah; (3) kota dengan kepadatan sangat tinggi, sanitasi baik, namun fasilitas kesehatan terbatas; (4) kawasan metropolitan dengan kasus diare sangat tinggi akibat sanitasi buruk; (5) kabupaten dengan kepadatan rendah, akses air minum rendah, dan sanitasi cukup baik. Temuan ini memberikan dasar bagi pengembangan strategi intervensi kesehatan lingkungan yang lebih tepat sasaran.

Kata Kunci: Agglomerative Hierarchical Clustering, Jawa Timur, K-Means Clustering, Klusterisasi, Sanitasi

ABSTRACT

Efforts to improve environmental health and basic public services require a thorough understanding of regional characteristics. East Java Province, with its diverse conditions across regencies and cities, provides a valuable case for such analysis. This study applied clustering techniques based on seven indicators: access to safe drinking water, access to improved sanitation, toilet ownership, incidence of diarrhea, reported health complaints, population density, and the number of public health centers (Puskesmas). Both Hierarchical Agglomerative Clustering (HAC) and K-Means Clustering were employed to identify homogeneous regional clusters. Cluster quality was evaluated using the Silhouette Score, Calinski-Harabasz Index, and Dunn Index, with HAC demonstrating superior segmentation performance. The analysis produced five distinct clusters: (1) districts with medium population density, the best sanitation, but high incidence of diarrhea and health complaints; (2) districts with poor sanitation but low incidence of diarrhea and health complaints; (3) cities with very high population density, good sanitation, but limited healthcare facilities; (4) metropolitan areas with very high incidence of diarrhea due to poor sanitation; (5) districts with low population density, low access to drinking water, and fairly good sanitation. These insights provide a foundation for developing more targeted and effective environmental health interventions.

Keywords: Clustering, Sanitation, Agglomerative Hierarchical Clustering, K-Means Clustering, East Java

1. Pendahuluan

Isu kesehatan lingkungan dan pelayanan dasar masih menjadi tantangan, terutama di daerah-daerah di Indonesia. Ketimpangan akses air bersih dan sanitasi masih terjadi di berbagai wilayah. Menurut *World Health Organization* (WHO), kurangnya akses terhadap air minum dan sanitasi yang layak dapat menyebabkan berbagai penyakit menular seperti diare, kolera, dan infeksi saluran pencernaan yang menjadi penyebab utama kematian pada anak-anak [1]. Sementara itu, UNICEF mencatat bahwa lebih dari 25 juta orang di Indonesia masih tidak memiliki akses terhadap sanitasi yang memadai [2].

Provinsi Jawa Timur, sebagai provinsi dengan penduduk terbanyak ke-2 di Indonesia, menunjukkan ketimpangan kualitas kesehatan masyarakat antar Kabupaten/Kota yang tercermin dari perbedaan signifikan pada indikator seperti akses air minum layak, sanitasi, kepemilikan jamban, serta fasilitas kesehatan. Penelitian oleh Setyan et al. (2024) menyatakan bahwa ketimpangan ini berdampak langsung pada kualitas kesehatan masyarakat, khususnya dalam hal akses layanan dasar seperti air bersih dan sanitasi. Daerah dengan akses yang terbatas cenderung memiliki angka kejadian penyakit yang lebih tinggi, seperti diare dan infeksi saluran pernapasan [3]. Selain itu, hasil analisis pengelompokan wilayah di Jawa Timur berdasarkan indikator kesehatan menunjukkan bahwa intervensi yang tepat sasaran sangat dibutuhkan agar distribusi layanan kesehatan lebih merata [4].

Sebagai bagian dari upaya pemerataan pembangunan kesehatan, penting untuk melakukan analisis spasial untuk mengelompokkan wilayah berdasarkan kemiripan karakteristik indikator kesehatan dan pelayanan dasar. Salah satu metode yang umum digunakan untuk tujuan ini adalah metode klasterisasi. Metode *K-Means Clustering* dan *Hierarchical Agglomerative Clustering* (HAC) dipilih karena kemampuannya dalam menangani data numerik dan menyederhanakan pola kompleks menjadi kelompok yang dapat dianalisis secara lebih mendalam. K-Means merupakan algoritma *unsupervised learning* yang membagi data ke dalam sejumlah kelompok berdasarkan kemiripan fitur dengan tujuan meminimalkan varians dalam tiap kluster dan memaksimalkan varians antar kluster, sehingga memudahkan pengenalan pola dalam data [5]. Sementara itu, HAC bekerja dengan membentuk struktur hirarki dari data melalui pendekatan *bottom-up*, di mana setiap data awalnya diperlakukan sebagai kluster tersendiri yang kemudian digabungkan secara bertahap berdasarkan kemiripan hingga membentuk kluster yang lebih besar. Pendekatan ini memungkinkan identifikasi pola-pola tersembunyi dalam data numerik dan dapat divisualisasikan melalui dendrogram [6].

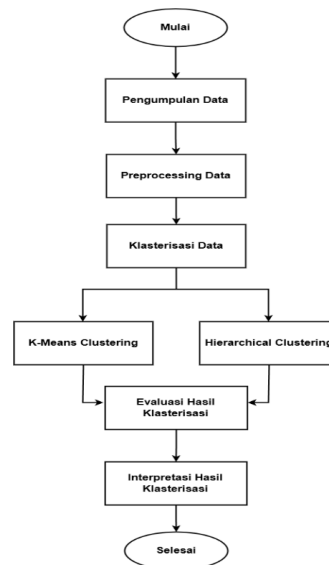
Dengan memahami karakteristik tiap kluster, intervensi kebijakan dapat dirancang secara lebih terfokus dan tepat sasaran. Selain itu, hasil klasterisasi juga dapat digunakan dalam pemetaan sumber daya, perencanaan anggaran, dan pengembangan program berbasis kebutuhan lokal. Penelitian ini akan menerapkan metode *K-Means* dan HAC untuk mengelompokkan Kabupaten/Kota di Provinsi Jawa Timur berdasarkan indikator-indikator terkait kesehatan lingkungan dan pelayanan dasar. Selanjutnya, hasil dari kedua metode akan dibandingkan untuk mengetahui mana yang memberikan segmentasi wilayah yang lebih informatif.

2. Metodologi Penelitian

Penelitian ini menggunakan pendekatan berbasis data kuantitatif untuk mengevaluasi kondisi sanitasi di seluruh kabupaten dan kota yang ada di Provinsi Jawa Timur. Analisis dilakukan melalui penerapan dua metode klasterisasi, yakni *Agglomerative Hierarchical* dan *K-Means Clustering*, dengan input berupa data numerik dari sejumlah indikator sanitasi. Tujuan dari proses ini adalah untuk mengelompokkan wilayah-wilayah berdasarkan kesamaan karakteristik sanitasi yang dimilikinya. Hasil pengelompokan tersebut diharapkan dapat memberikan wawasan mengenai persebaran kondisi sanitasi antar daerah, serta mendukung perumusan kebijakan pembangunan sanitasi yang lebih efektif dan sesuai kebutuhan wilayah.

2.1 Tahapan Penelitian

Seluruh tahapan penelitian dirancang secara sistematis untuk memastikan proses analisis dapat berjalan dengan efektif dan hasil yang diperoleh dapat diinterpretasikan secara akurat. Untuk memberikan gambaran yang lebih sistematis mengenai tahapan analisis yang dilakukan, berikut disajikan diagram alur (*flowchart*) metodologi penelitian ini.



Gambar 1. Tahapan Penelitian

Adapun penjelasan dari masing-masing tahapan dalam *flowchart* tersebut dijabarkan sebagai berikut:

1. Pengumpulan Data

Tahap pengumpulan data dalam penelitian ini dilakukan dengan mengakses informasi dari lembaga resmi yang memiliki otoritas di bidangnya, yaitu Badan Pusat Statistik (BPS) Provinsi Jawa Timur. Data tersebut mencakup 38 kabupaten/kota di Provinsi Jawa Timur dan menggambarkan kondisi sanitasi wilayah pada tahun 2023.

Tabel 1. Atribut Dataset

Variabel	Atribut
x_1	Persentase Rumah Tangga dengan Akses Air Minum Layak
x_2	Persentase Rumah Tangga dengan Akses Sanitasi Layak
x_3	Persentase Rumah Tangga yang Memiliki Jamban Sendiri
x_4	Jumlah Kasus Diare
x_5	Persentase Penduduk yang Mengalami Keluhan Kesehatan
x_6	Angka Kepadatan Penduduk
x_7	Jumlah Puskesmas

2. Tahap *Preprocessing*

Tahap ini diawali dengan pemeriksaan keberadaan missing value dalam dataset. Selanjutnya, data dinormalisasi menggunakan metode *Z-Score* melalui fungsi *StandardScaler* pada *Python* agar skala variabel menjadi seragam dan menghindari bias dalam proses klasterisasi.

3. Implementasi Algoritma

Tahapan ini mencakup penerapan dua teknik klasterisasi, yakni *Agglomerative Hierarchical Clustering* dan *K-Means*, guna mengelompokkan data sanitasi berdasarkan kesamaan karakteristik antar wilayah.

4. Evaluasi Hasil Klasterisasi

Evaluasi hasil klasterisasi dilakukan menggunakan metrik *Silhouette Score*, *Calinski-Harabasz Index*, dan *Dunn Index* untuk menilai kualitas pemisahan dan kohesi antar kluster.

Berdasarkan hasil evaluasi tersebut, model dengan performa terbaik dipilih untuk ditampilkan secara visual dan dianalisis lebih lanjut.

5. Implementasi Hasil Klasterisasi

Hasil klasterisasi dari model terbaik dianalisis dengan mengamati profil variabel sanitasi dominan pada setiap klaster. Interpretasi ini bertujuan untuk memberikan gambaran kondisi sanitasi yang spesifik dan relevan di masing-masing wilayah.

2.2 Metode *Elbow*

Penentuan jumlah klaster optimal dilakukan dengan menggunakan metode *Elbow*. Metode ini didasarkan pada perhitungan nilai *Within-Cluster Sum of Squares* (WCSS), yaitu jumlah kuadrat jarak antara setiap data dengan centroid klasternya. Secara matematis, WCSS dapat dituliskan sebagai:

$$WCSS = \sum_{k=1}^K \sum_{i=1}^{n_k} \|x_i - C_k\|^2 \quad (1)$$

Di mana x_i Adalah data ke-i, C_k Adalah centroid klaster ke-k, dan n_k Adalah jumlah data dalam klaster tersebut. Semakin kecil nilai WCSS, semakin baik data dikelompokkan dalam klaster yang bersangkutan.

Metode *Elbow* bekerja dengan menghitung nilai WCSS untuk berbagai pilihan jumlah klaster k. Selanjutnya, nilai WCSS diplot terhadap jumlah klaster. Grafik yang dihasilkan biasanya menunjukkan penurunan tajam pada awalnya, kemudian melandai. Titik “siku” (*elbow point*) pada grafik tersebut dianggap sebagai jumlah klaster optimal, karena pada titik tersebut penambahan jumlah klaster tidak lagi memberikan pengurangan WCSS yang signifikan.

2.3 Algoritma *Hierarchical Clustering*

Metode *Hierarchical Clustering* merupakan salah satu teknik klasterisasi yang digunakan dalam penelitian ini untuk mengelompokkan kabupaten/kota di Provinsi Jawa Timur berdasarkan indikator sanitasi. Pendekatan yang digunakan adalah *agglomerative (bottom-up)*, di mana setiap objek data awalnya dianggap sebagai klaster terpisah, kemudian digabung secara bertahap hingga membentuk satu klaster besar [7]. Proses penggabungan antar klaster dilakukan berdasarkan metode Ward (*Ward linkage*), yaitu metode yang bertujuan meminimalkan ragam (variansi) dalam klaster dengan menghitung peningkatan *Sum of Squares Error* (SSE) akibat penggabungan dua klaster [8]. Jarak antar objek dihitung menggunakan metrik Euclidean yang dirumuskan sebagai berikut:

$$d_{Euclidean} = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (2)$$

Di mana a_i dan b_i merupakan nilai dari data ke-i pada dua objek yang dibandingkan. Sementara itu, peningkatan SSE akibat penggabungan klaster dihitung berdasarkan jarak kuadrat *Euclidean* sebagai berikut:

$$d_{ik}^2 = \sum_{j=1}^p (x_{ij} - x_{kj})^2 \quad (3)$$

Dengan x_{ij} dan x_{kj} masing-masing adalah nilai fitur ke-j dari objek i dan k, serta p adalah jumlah variabel.

2.4 Algoritma *K-Means*

Metode *K-Means Clustering* merupakan algoritma klasterisasi berbasis partisi yang digunakan dalam penelitian ini untuk mengelompokkan daerah berdasarkan kesamaan karakteristik sanitasi. *K-Means* bertujuan untuk membagi data ke dalam

sejumlah kluster k yang telah ditentukan sebelumnya, di mana setiap data akan dimasukkan ke dalam kluster dengan centroid terdekat [9]. K-Means bersifat iteratif dan berfokus pada minimisasi jarak antara objek dengan pusat klasternya menggunakan jarak Euclidean, yang dirumuskan sebagai berikut [10]:

$$dEuclidean = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (4)$$

Di mana a_i dan b_i merupakan nilai dari data ke- i pada dua objek yang dibandingkan. Posisi centroid baru dihitung berdasarkan rata-rata seluruh data dalam masing-masing kluster, dengan rumus:

$$C_k = \frac{1}{n_k} \sum d_i \quad (5)$$

Di mana n_k merupakan jumlah data dalam kluster k , sedangkan d_i merupakan jumlah dari nilai jarak yang masuk dalam masing-masing kluster.

2.5 Silhouette Score

Silhouette Score merupakan salah satu metrik evaluasi yang digunakan untuk mengukur kualitas pemisahan dan kekompakan kluster. Nilai ini menunjukkan seberapa baik suatu objek cocok dengan klasternya sendiri dibandingkan dengan kluster lain. *Silhouette Score* bernilai antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa objek sangat cocok dengan klasternya dan jauh dari kluster lainnya [11]. Adapun rumus *Silhouette Score* untuk data ke- i adalah:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (6)$$

Di mana $s(i)$ merupakan nilai *silhouette* untuk data ke- i . Nilai $a(i)$ Adalah rata-rata jarak antara data ke- i dengan seluruh data lain dalam kluster yang sama, sedangkan $b(i)$ Adalah rata-rata jarak antara data ke- i dengan data pada kluster terdekat (kluster selain tempat data tersebut berada).

2.6 Calinski-Harabasz Index

Calinski-Harabasz Index, atau disebut juga Variance Ratio Criterion, mengukur rasio antara variasi antar kluster (*inter-cluster dispersion*) dan varians dalam kluster (*intra-cluster dispersion*). Semakin besar nilai indeks ini, maka semakin baik hasil pemisahan klasternya [12]. Rumus *Calinski-Harabasz Index* dinyatakan sebagai berikut:

$$s = \frac{tr(B_K)}{tr(W_K)} \times \frac{n - k}{k - 1} \quad (7)$$

Di mana $tr(B_K)$ dan $tr(W_K)$ adalah trace dari matriks disperse antar kluster

2.7 Dunn Index

Dunn Index merupakan metode validasi internal untuk mengukur kualitas klusterisasi. Indeks ini dirancang untuk menemukan kluster yang saling terpisah dengan baik (*inter-cluster distance* besar) dan memiliki penyebaran dalam kluster yang kecil (*intra-cluster distance* kecil). Semakin tinggi nilai *Dunn Index*, semakin baik hasil klusterisasi [13].

$$DI = \frac{d(c_p, c_q)}{diam(c_r)} \quad (8)$$

Di mana $d(c_p, c_q)$ adalah jarak antara dua kluster yang berbeda, yaitu kluster c_p dan c_q . $diam(c_r)$ adalah diameter dari kluster (c_r), yaitu jarak maksimum antar dua titik dalam satu kluster yang sama (maximum intra-cluster distance). $d(c_p, c_q)$ adalah nilai minimum dari seluruh jarak antar kluster yang menunjukkan kedekatan terdekat antar kluster berbeda. $diam(c_r)$ adalah nilai maksimum dari diameter seluruh kluster, yang menunjukkan seberapa tersebar data dalam kluster tertentu.

3. Hasil dan Pembahasan

3.1 Analisis Deskriptif

Data Analisis yang dipakai dalam penelitian ini adalah data tahun 2023 kondisi sanitasi lingkungan di Jawa Timur dengan melibatkan sejumlah data sebanyak 38 data yang berasal dari Badan Pusat Statistik (BPS) Jawa Timur. Pengelolaan data ini dengan menggunakan Program *Python*.

Variabel	count	mean	std	min	max	25%	50%	75%
x_1	38	96.12	4.81	79.26	100.00	95.69	97.61	98.84
x_2	38	27.77	5.70	15.72	37.95	24.46	28.20	32.28
x_3	38	76.10	16.2	32.10	98.93	67.88	78.48	88.83
x_4	38	4823	4659	328	22327	1179	3486	6874
x_5	38	27.7	5.70	15.72	37.95	24.46	28.20	32.28
x_6	38	1938	2262	410	8667	645	855.50	1205
x_7	38	25.58	13.0	3.00	63.00	20.25	25.00	32.75

Tabel 2. Statistik Data Secara Umum

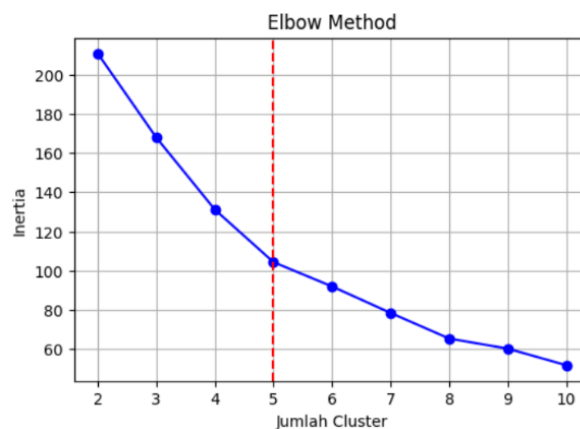
Tabel di atas memaparkan hasil dari statistika deskriptif dari variabel yang ada dalam dataset sanitasi.

3.2 Preprocessing

Sebelum dilakukan modelling dalam clustering, hal yang harus dilakukan terlebih dahulu ada *pre-processing*. Dalam tahap ini, bertujuan untuk menghasilkan kualitas data yang baik dan optimal. Berikut ini tahapan yang dilakukan dalam penelitian ini, antara lain:

1. Penanganan *Missing Value*
2. Standarisasi Data

3.3 Menentukan Jumlah Kluster



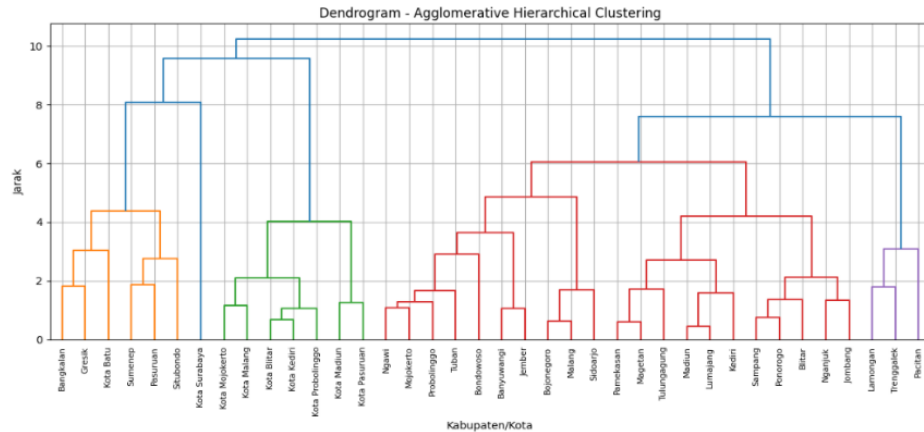
Gambar 2. Kluster Optimal Berdasarkan Metode Elbow

Penentuan jumlah kluster optimal dalam penelitian ini dilakukan menggunakan metode Elbow. Titik siku (*elbow*) yang jelas pada $k=5$ menjadi dasar pemilihan jumlah kluster optimal. Oleh karena itu, jumlah kluster $k=5$ diterapkan baik pada metode *K-Means* maupun *Hierarchical Agglomerative Clustering* (HAC) dalam penelitian ini.

3.4 Penerapan Metode Clustering

3.4.1 Hierarchical Agglomerative Clustering (HAC)

Pada penelitian ini, jumlah kluster optimal ditentukan sebanyak $k=5$ berdasarkan hasil analisis metode Elbow sebelumnya. Hasil pengelompokan divisualisasikan pada Gambar 3.



Gambar 3. Dendrogram HAC

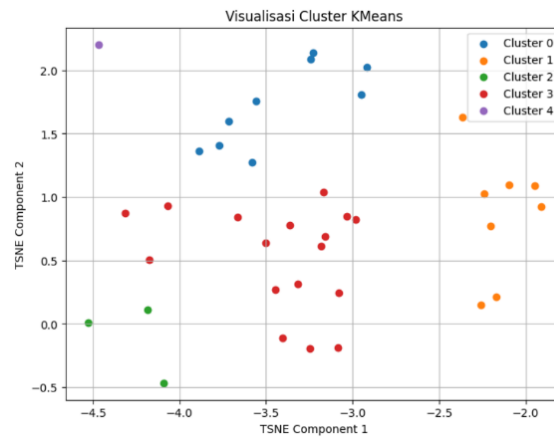
Hasil klasterisasi menunjukkan bahwa Kabupaten/Kota di Provinsi Jawa Timur terbagi ke dalam 5 kluster dengan karakteristik yang berbeda-beda. Tabel berikut menyajikan anggota masing-masing kluster berdasarkan metode HAC:

Kluster	Jumlah Anggota	Anggota
0	21	Ponorogo, Tulungagung, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Probolinggo, Sidoarjo, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Sampang, Pamekasan
1	6	Situbondo, Pasuruan, Gresik, Bangkalan, Sumenep, Kota Batu
2	7	Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun
3	1	Kota Surabaya
4	3	Pacitan, Trenggalek, Lamongan

Tabel 3. Anggota Kluster HAC

3.4.2 K-Means Clustering

Dalam penelitian ini, jumlah kluster optimal ditentukan sebanyak $k=5$, berdasarkan hasil analisis menggunakan metode Elbow. Visualisasi hasil klasterisasi disajikan pada Gambar 4.



Gambar 4. Hasil Klasterisasi *K-Means*

Hasil klasterisasi menunjukkan bahwa Kabupaten/Kota di Provinsi Jawa Timur terbagi ke dalam 5 kluster. Tabel berikut menyajikan anggota masing-masing kluster berdasarkan metode *K-Means*:

Kluster	Jumlah Anggota	Anggota
0	9	Malang, Situbondo, Pasuruan, Sidoarjo, Bojonegoro, Tuban, Gresik, Bangkalan, Sumenep
1	8	Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Batu
2	3	Pacitan, Trenggalek, Lamongan
3	17	Ponorogo, Tulungagung, Blitar, Kediri, Lumajang, Jember, Banyuwangi, Bondowoso, Probolinggo, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Sampang, Pamekasan
4	1	Kota Surabaya

Tabel 4. Anggota Kluster *K-Means*

3.5 Evaluasi

Evaluasi performa metode klasterisasi dilakukan untuk menentukan metode yang paling sesuai dalam menggambarkan segmentasi wilayah berdasarkan indikator kesehatan lingkungan dan pelayanan dasar. Evaluasi dilakukan menggunakan tiga metrik, yaitu *Silhouette Score*, *Calinski-Harabasz Index*, dan *Dunn Index*. Hasil evaluasi metode klasifikasi ditampilkan pada Tabel 5.

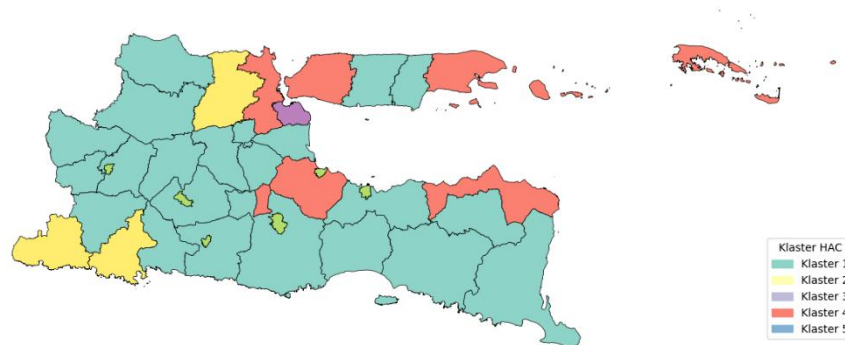
Metode	<i>Hierarchical Agglomerative</i>	<i>K-Means</i>
<i>Silhouette Score</i>	0.30	0.29
<i>Calinski-Harabasz Index</i>	12.41	12.79
<i>Dunn Index</i>	0.40	0.30

Tabel 5. Hasil Evaluasi

Berdasarkan hasil evaluasi, metode Agglomerative Hierarchical Clustering (HAC) menunjukkan performa yang lebih baik secara keseluruhan, dengan nilai Silhouette Score dan Dunn Index yang lebih tinggi dibandingkan K-Means. Meskipun nilai Calinski-Harabasz Index K-Means sedikit lebih tinggi, perbedaan tersebut tidak signifikan. Dengan mempertimbangkan stabilitas struktur klaster dan kualitas pemisahan klaster, HAC dipilih sebagai metode terbaik untuk interpretasi lebih lanjut.

3.6 Karakteristik Klaster dari Metode Terbaik

Berdasarkan hasil evaluasi yang telah dilakukan pada subbab sebelumnya, metode *Hierarchical Agglomerative Clustering* (HAC) dipilih sebagai metode terbaik untuk pengelompokan Kabupaten/Kota di Provinsi Jawa Timur. Pemetaan hasil klasterisasi ditampilkan pada Gambar 5. Setiap warna pada peta mewakili masing-masing klaster, sehingga memudahkan identifikasi sebaran spasial tiap klaster di wilayah Jawa Timur.



Gambar 5. Peta Hasil Klasterisasi

Selanjutnya, karakteristik masing-masing klaster dianalisis berdasarkan nilai rata-rata dari sejumlah indikator utama yang digunakan dalam proses klasterisasi. Rangkuman karakteristik tiap klaster disajikan pada Tabel 6 berikut:

Klaster	x_1	x_2	x_3	x_4	x_5	x_6	x_7
0	96.81	30.32	72.92	5086.14	30.32	901.71	30.24
1	96.86	18.61	72.37	4722	18.61	853.33	23.67
2	99.41	28.13	84.64	1337.86	28.13	5590.57	7.71
3	98.15	21.43	88.35	22327	21.43	8667	63
4	81.44	29.50	81.83	5627.67	29.50	594	26

Tabel 6. Karakteristik Tiap Klaster

Dari tabel di atas, dapat disimpulkan beberapa karakteristik penting dari masing-masing klaster sebagai berikut:

1. Klaster 0

Wilayah dalam klaster ini memiliki akses air minum yang sangat tinggi (96.8%) dan tingkat sanitasi paling baik di antara klaster lain (30.3%), meskipun secara umum masih tergolong rendah. Kepemilikan jamban cukup baik (72.9%). Kepadatan penduduk berada pada tingkat sedang dengan 901 orang per km² dan terdapat 30 puskesmas. Meskipun layanan kesehatan cukup memadai, jumlah kasus diare (5068 kasus) dan keluhan kesehatan (30.3%) masih tergolong tinggi, mengindikasikan bahwa peningkatan sanitasi sangat diperlukan untuk menurunkan beban penyakit.

2. Klaster 1

Wilayah dalam klaster ini memiliki akses air minum yang sangat tinggi (96.86%) namun tingkat sanitasi tergolong rendah (18.6%), yang merupakan salah satu terendah di antara klaster. Kepemilikan jamban cukup baik (72.3%). Kepadatan penduduk berada pada tingkat sedang dengan 853 orang per km² dan terdapat 23 puskesmas. Jumlah kasus diare (4722 kasus) dan keluhan kesehatan (18.6%) relatif rendah, menunjukkan kemungkinan adanya faktor lain yang membantu menjaga kesehatan masyarakat meskipun sanitasi masih kurang.

3. Klaster 2

Wilayah dalam klaster ini memiliki akses air minum tertinggi (99.41%) dengan tingkat sanitasi yang baik (28.1%) dan kepemilikan jamban sangat tinggi (84.6%). Kepadatan penduduk sangat tinggi, mencapai 5590 orang per km², namun jumlah puskesmas sangat rendah (7.7). Kasus diare di klaster ini sangat rendah (1337 kasus) dan keluhan kesehatan juga relatif rendah (28.1%), menunjukkan kondisi lingkungan dan perilaku sehat yang baik meskipun beban layanan kesehatan berpotensi tinggi karena sedikitnya fasilitas kesehatan.

4. Klaster 3

Wilayah dalam klaster ini memiliki akses air minum yang sangat tinggi (98.15%) dan kepemilikan jamban tertinggi di antara klaster (88.3%). Namun, tingkat sanitasi tergolong rendah (21.4%). Kepadatan penduduk paling tinggi di antara seluruh klaster, yakni 8667 orang per km², dengan jumlah puskesmas terbanyak (63). Jumlah kasus diare sangat tinggi (22.327 kasus), meskipun keluhan kesehatan tergolong rendah (21.4%), mengindikasikan tantangan sanitasi dan kepadatan penduduk yang signifikan di wilayah ini.

5. Klaster 4

Wilayah dalam klaster ini memiliki akses air minum terendah di antara klaster (81.44%), namun tingkat sanitasi (29.5%) dan kepemilikan jamban (81.82%) tergolong baik. Kepadatan penduduk rendah, yakni 594 orang per km², dengan jumlah puskesmas sedang (26). Kasus diare (5627 kasus) dan keluhan kesehatan (29.5%) cukup tinggi, yang menunjukkan perlunya peningkatan akses air minum dan edukasi kesehatan di wilayah ini.

4. Kesimpulan

Penelitian ini berhasil mengelompokkan Kabupaten/Kota di Provinsi Jawa Timur berdasarkan indikator kesehatan lingkungan dan pelayanan dasar menggunakan dua metode klusterisasi, yaitu Hierarchical Agglomerative Clustering (HAC) dan K-Means Clustering. Berdasarkan evaluasi performa model menggunakan Silhouette Score, Calinski-Harabasz Index, dan Dunn Index, metode HAC dipilih sebagai metode terbaik untuk menggambarkan segmentasi wilayah. Hasil klusterisasi menunjukkan bahwa terdapat 5 klaster wilayah dengan karakteristik yang berbeda-beda, di antaranya:

1. Wilayah kabupaten dengan kepadatan sedang, sanitasi terbaik di antara klaster, akses air minum cukup tinggi, namun kasus diare dan keluhan kesehatan masih relatif tinggi.
2. Wilayah kabupaten dan sebagian daerah pedesaan dengan akses air minum sangat baik, sanitasi rendah, serta kasus diare dan keluhan kesehatan yang relatif rendah.
3. Wilayah perkotaan padat dengan akses air minum dan kepemilikan jamban terbaik, sanitasi baik, namun fasilitas kesehatan terbatas dan potensi beban layanan tinggi meskipun kasus diare rendah.
4. Kawasan metropolitan dengan kepadatan penduduk sangat tinggi dan fasilitas kesehatan banyak, namun menghadapi tantangan kasus diare sangat tinggi akibat sanitasi buruk dan kepadatan penduduk yang tinggi.
5. Wilayah kabupaten dengan kepadatan rendah, akses air minum terendah di antara klaster, sanitasi cukup baik, dan kasus diare serta keluhan kesehatan cukup tinggi, menandakan kebutuhan peningkatan akses air bersih dan edukasi kesehatan.

Hasil tersebut memberikan gambaran bahwa ketimpangan akses sanitasi dan fasilitas kesehatan masih terjadi antar wilayah di Jawa Timur. Dengan memahami karakteristik tiap klaster, intervensi kebijakan dapat dirancang lebih terarah. Dengan demikian, hasil penelitian ini dapat menjadi bahan pertimbangan bagi pemerintah daerah dalam perencanaan pembangunan kesehatan lingkungan yang lebih efektif.

5. Daftar Pustaka

- [1] WHO, "Drinking Water," 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/drinking-water>
- [2] UNICEF Indonesia, "Water, Sanitation and Hygiene," UNICEF Indonesia. [Online]. Available: <https://www.unicef.org/indonesia/water-sanitation-and-hygiene>
- [3] A. P. Setyan, M. Jamal, Widodo, and Sanyoto, "Analisis Dampak Persentase Sanitasi Dan Sumber Air Minum Terhadap Penyakit Di Jawa Timur," *J. Apl. Sistern dan Tek. Inform. Pomos.*, vol. 02, no. 01, pp. 49–56, 2024, [Online]. Available: <https://jurnal.stt-pomosda.ac.id/index.php/jastip/article/view/143>
- [4] Y. Liana, V. Ratnasari, and A. S. Deskriptif, "Pengelompokan Kabupaten / Kota di Jawa Timur dengan Pendekatan Metode Ensemble ROCK," *J. SAINS DAN SENI ITS*, vol. 11, no. 2, 2022.
- [5] E. U. Oti, S. I. Onyeagu, C. H. Nwankwo, W. K. Alvan, and G. A. Osuji, "NEW K-MEANS CLUSTERING METHODS THAT MINIMIZES THE TOTAL INTRA-CLUSTER VARIANCE," *African J. Math. Stat. Stud.*, vol. 3, no. 5, p. 42, 2020, [Online]. Available: www.abjournals.org
- [6] T. Märzinger, J. Kotik, and C. Pfeifer, "Application of hierarchical agglomerative clustering (Hac) for systemic classification of pop-up housing (puh) environments," *Appl. Sci.*, vol. 11, no. 23, 2021, doi: 10.3390/app112311122.
- [7] L. Rahmawati, S. Widya Sihwi, and E. Suryani, "Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus : Dokumen Skripsi Jurusan Kimia, Fmipa, Universitas Sebelas Maret)," *J. Teknol. Inf. ITSmart*, vol. 3, no. 2, p. 66, 2016, doi: 10.20961/its.v3i2.654.
- [8] N. Akbar, "Analisis Hierarchical Clustering pada Karakteristik Penduduk Bekerja Jurnal Informatika Dan Rekayasa Komputer di Provinsi Jambi," *J. Inform. Dan Rekayasa Komputer(JAKAKOM)*, vol. 4, no. September, pp. 1040–1050, 2024, doi: 10.33998/jakakom.v4i2.
- [9] J. Khomsiyah, A. Ramdhani, A. Damayanti, and D. Rohman, "PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN WILAYAH BANJIR," *J. Ilm. Betrik*, vol. 12. No.03, no. 03, pp. 249–253, 2021.

- [10] M. Herviany, S. Putri Delima, T. Nurhidayah, and Kasini, "Comparison of K-Means and K-Medoids Algorithms for Grouping Landslide Prone Areas in West Java Province," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. April, pp. 34–40, 2021.
- [11] A. Shoolihah, M. T. Furqon, and A. W. Widodo, "Implementasi Metode Improved K-Means untuk Mengelompokkan Titik Panas Bumi," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 11, pp. 1270–1276, 2017, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/453>
- [12] A. M. Sikana and A. W. Wijayanto, "Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering," *J. Ilmu Komput.*, vol. 14, no. 2, pp. 66–78, 2021.
- [13] R. A. Sary, N. Satyahadewi, and W. Andani, "Application of K-Means++ With Dunn Index Validation of Grouping West Kalimantan Region Based on Crime Vulnerability," *BAREKENG J. Math. Its Appl.*, vol. 18, no. 4, pp. 2283–2292, 2024, doi: 10.30598/barekengvol18iss4pp2283-2292.