

Segmentasi Provinsi Berdasarkan Indikator Sanitasi Total Berbasis Masyarakat di Indonesia: Pendekatan *Hierarchical Clustering Agglomerative*

Segmentation of Provinces Based on Community-Based Total Sanitation Indicators in Indonesia: A Hierarchical Agglomerative Clustering Approach

Nabbila Dyah Syahrani¹, Gita Rahmawati², Nasyiatul Izzah³, Safril Ahmadi Sanmas⁴,
Alfidha Rahmah⁵, Fatkhurokhman Fauzi⁶

^{1,2,3,4,5,6} Universitas Muhammadiyah Semarang, Semarang

Corresponding author : nabbilads@gmail.com

Abstrak

Sanitasi yang buruk menjadi salah satu faktor penyebab beban penyakit global, dengan sekitar 1,4 juta kematian setiap tahunnya disebabkan oleh faktor lingkungan, termasuk sanitasi yang tidak aman. Pada tahun 2023, sebanyak 4,2 miliar orang di dunia tidak memiliki akses sanitasi yang dikelola dengan aman, yang dapat menyebabkan berbagai masalah kesehatan seperti malnutrisi dan kematian. Di Indonesia, program Sanitasi Total Berbasis Masyarakat (STBM) dikembangkan oleh Kementerian Kesehatan untuk meningkatkan perilaku higienis melalui pemberdayaan masyarakat. STBM terdiri dari lima pilar: Stop Buang Air Besar Sembarangan (SBS), Cuci Tangan Pakai Sabun (CTPS), Pengelolaan Sampah Rumah Tangga (PSRT), Pengelolaan Air Minum dan Makanan Rumah Tangga (PAMMRT), serta Pengelolaan Air Limbah Domestik Rumah Tangga (PALDRT). Namun, implementasi program ini di berbagai daerah di Indonesia masih bervariasi, dengan beberapa daerah belum menerapkan STBM secara optimal. Oleh karena itu, diperlukan pemetaan daerah-daerah yang membutuhkan intervensi lebih lanjut. Penelitian ini menggunakan metode *Hierarchical Clustering* untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan data capaian STBM tahun 2023. Penelitian ini menghasilkan empat *cluster* dengan menggunakan *Agglomerative Average Linkage* sebagai metode terbaik. Hasil *clustering* menunjukkan *Cluster 1* memiliki kondisi sanitasi terburuk, *Cluster 2* memiliki pencapaian terbaik, sementara *Cluster 3* dan *Cluster 4* berada di antara keduanya, dengan beberapa aspek memerlukan perbaikan. Diperoleh nilai evaluasi hasil *clustering* menggunakan *silhouette coefficient* sebesar 0,5027.

Kata Kunci : *Hierarchical Clustering*, STBM, Akses Sanitasi

Abstract

Poor sanitation is one of the contributing factors to the global disease burden, causing approximately 1.4 million deaths annually due to environmental factors, including unsafe sanitation. In 2023, about 4.2 billion people worldwide lacked access to safely managed sanitation, leading to various health issues such as malnutrition and death. In Indonesia, the Community-Based Total Sanitation (STBM) program was developed by the Ministry of Health to promote hygienic behavior through community empowerment. STBM consists of five pillars: Stop Open Defecation (SBS), Handwashing with Soap (CTPS), Household Waste Management (PSRT), Household Drinking Water and Food Management (PAMMRT), and Domestic Wastewater Management (PALDRT). However, the implementation of this program varies across different regions in Indonesia, with some areas not yet optimally applying STBM. Therefore, mapping areas that require further intervention is necessary. This study employs the Hierarchical Clustering method to group provinces in Indonesia based on STBM achievement data from 2023. The study resulted in four clusters, using Agglomerative Average Linkage as the best method. The clustering results show that Cluster 1 has the worst sanitation conditions, Cluster 2 has achieved the best performance, while Clusters 3 and 4 fall in between, with some aspects requiring improvement. The evaluation result of the clustering, using the silhouette coefficient, yielded a value of 0.5027.

Keywords : *Hierarchical Clustering*, STBM, Sanitation Access

PENDAHULUAN

Kualitas lingkungan yang sehat dapat mencegah sekitar seperempat dari beban penyakit global (Kemenkes RI, 2024). Menurut laporan WASH, sebanyak 1,4 Juta orang meninggal tiap tahunnya disebabkan oleh faktor lingkungan, salah satunya adalah sanitasi. Pada tahun 2023, sebanyak 4,2 miliar orang didunia hidup tanpa akses sanitasi yang dikelola dengan aman. Sanitasi yang tidak aman menyebabkan sebanyak 564 kematian dan juga dapat menyebabkan malnutrisi (WHO, 2024). Upaya global untuk meningkatkan akses sanitasi sangat penting untuk mengurangi beban penyakit serta memperbaiki kualitas hidup bagi jutaan orang di seluruh dunia.

Menurut WHO, sanitasi didefinisikan sebagai penyediaan sarana dan layanan yang aman untuk membuang urin dan feses manusia (WHO & UNICEF, 2023). Sanitasi adalah upaya yang dilakukan secara sengaja untuk membiasakan perilaku hidup bersih, bertujuan menghindari kontak langsung manusia dengan kotoran dan limbah berbahaya lainnya. Langkah ini diharapkan dapat melindungi serta meningkatkan kesehatan manusia (Sa'ban et al., 2020).

Sanitasi telah menjadi perhatian di Indonesia sejak tahun 1993, yang ditunjukkan oleh berbagai studi yang mengkaji akses pasokan air dari pedesaan maupun perkotaan. Seiring berjalannya waktu melalui WHO dan World Bank, dunia mulai menyoroti pengelolaan limbah cair pada tahun 2000 (FK UNUSA, 2023). Saat ini, sanitasi berkelanjutan untuk semua telah menjadi target dari Tujuan Pembangunan Berkelanjutan (SGDs) nomor 6 (United Nations, n.d.).

Sebagai upaya dalam mencapai target tersebut, Kementerian Kesehatan Indonesia mengembangkan program Sanitasi Total Berbasis Masyarakat (STBM) sebagai pendekatan mengubah perilaku higienis dan saniter melalui pemberdayaan masyarakat dengan cara pemukiman. STBM memiliki 5 Pilar diantaranya : Stop Buang Air Besar Sembarangan (SBS), Cuci Tangan Pakai Sabun (CTPS), Pengelolaan Sampah Rumah Tangga (PSRT), Pengelolaan Air Minum dan Makanan Rumah Tangga (PAMMRT), Pengelolaan Air Limbah Domestik Rumah Tangga (PALDRT) (Kemenkes RI, n.d.).

Pelaksanaan Program Sanitasi Total Berbasis Masyarakat (STBM) di Indonesia cukup bervariasi, dimana terdapat beberapa daerah yang belum menerapkan program ini secara optimal (Stiawati, 2021). Berdasarkan hal tersebut, masih terdapat tantangan dalam memastikan keberlanjutan sanitasi yang baik di seluruh wilayah di Indonesia. Oleh karena itu, penting untuk memahami karakteristik implementasi STBM di masing-masing daerah agar dapat mengidentifikasi daerah-daerah yang memerlukan intervensi lebih lanjut. Salah satu metode yang dapat digunakan adalah dengan pendekatan *clustering*.

Clustering merupakan model multivariat yang bertujuan melakukan pengelompokan data berdasarkan karakteristik data tersebut. Analisis *cluster* terbagi menjadi *Hierarchical* dan *Non-Hierarchical* (Alamtaha et al., 2023). Dalam penelitian ini, akan digunakan metode *Hierarchical Clustering* untuk melakukan pengelompokan. *Hierarchical Clustering* merupakan metode yang didasarkan pada struktur seperti dendogram, dimana data dibagi dan digabungkan seperti cabang pohon dengan mengelompokkan data dengan kesamaan terdekat (Singh & Srivastava, 2020). Metode ini dapat dilakukan dengan pendekatan *agglomerative* dimana terdiri dari beberapa jenis metode diantaranya yaitu, *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan *Centroid Linkage*.

Metode *Hierarchical Clustering Agglomerative* telah banyak diterapkan pada penelitian-penelitian terdahulu. Penelitian yang dilakukan oleh Krisman Pratama dan Ulfa Khaira (2021) mengelompokkan data titik panas di Provinsi Jambi menggunakan metode *Agglomerative Hierarchical Clustering*. Hasil penelitian menunjukkan bahwa jumlah cluster terbaik adalah 2, dengan nilai *Silhouette Coefficient* sebesar 0,5856 (Krisman

Pratama Simanjuntak, 2021). Selanjutnya dilakukan oleh Dessy Iyan Yulianti, dkk (2023) menunjukkan bahwa pengelompokan data menghasilkan 3 kluster, dengan pendonor terbanyak berada di Kecamatan Purwakarta, selanjutnya Kecamatan Bungursari. Evaluasi dilakukan dengan melihat Silhouette Coefficient yang menunjukkan tingkat akurasi sebesar 0,7133 (Iyan Yulianti et al., 2023). Penelitian yang lain dilakukan oleh Fachrian Bimantoro Putra, dkk (2023) dalam Pengelompokan Kabupaten/Kota Di Papua Berdasarkan Indikator Kemiskinan, hasil penelitian mengindikasikan bahwa jumlah kluster yang optimal adalah 3 kluster dengan nilai *silhouette score* sebesar 0,48 (Bimantoro Putra et al., 2023).

Dengan demikian, penelitian ini akan menerapkan metode *Hierarchical Clustering Agglomerative* dalam mengelompokkan provinsi-provinsi di Indonesia berdasarkan data capaian STBM tahun 2023. Penelitian ini bertujuan untuk melihat provinsi-provinsi yang telah berhasil menerapkan STBM secara optimal serta provinsi yang masih memerlukan perhatian lebih lanjut mengenai pelaksanaan program tersebut. Dengan demikian, hasil penelitian ini diharapkan dapat membantu dalam merumuskan strategi intervensi dalam meningkatkan kualitas sanitasi khususnya di wilayah yang paling membutuhkan.

METODE

1. Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari publikasi Kementerian Kesehatan Indonesia. Data yang digunakan adalah persentase capaian Sanitasi Total Berbasis Masyarakat per provinsi di Indonesia tahun 2023. Variabel yang digunakan meliputi Stop Buang Air Besar Sembarangan (SBSS), Cuci Tangan Pakai Sabun (CTPS), Pengelolaan Air Limbah Domestik Rumah Tangga (PALDRT), Pengelolaan Sampah Rumah Tangga (PSRT), dan Pengelolaan Air Minum dan Makanan Rumah Tangga (PAMMRT).

2. Analisis Cluster

Menurut Santoso (2017), analisis *cluster* akan menghasilkan beberapa *cluster*. Sementara menurut Supranto (2004), analisis *cluster* adalah teknik untuk mengelompokkan obyek pada *cluster* yang cukup homogen disebut *cluster*, sehingga objek yang berbeda akan dikelompokkan dalam *cluster* yang lain. Hirarki *cluster* merupakan pengembangan dari suatu hirarki atau struktur mirip pohon bercabang. Pengelompokan hirarki memiliki algoritma dengan melakukan pembuatan *m cluster* yang terdiri atas 1 obyek dari suatu data. Selanjutnya antara satu *cluster* digabung dengan *cluster* lain yang memiliki jarak terdekat (Widyawati et al., 2020)

3. Multikolinieritas

Multikolinieritas terjadi ketika terdapat hubungan linear sempurna atau pasti antara beberapa atau seluruh variabel. Keberadaan multikolinieritas dianggap sebagai pelanggaran dalam analisis kluster karena dapat memengaruhi hasil pengelompokan. Salah satu metode untuk mendeteksi multikolinieritas adalah dengan menghitung nilai *Variance Inflation Factor* (VIF) (Al Farizi et al., 2023). Dimana nilai VIF kurang dari 10 maka dapat disimpulkan tidak terdapat multikolinieritas begitupun sebaliknya (Mahmudan, 2020). Persamaan nilai VIF dapat dilihat pada persamaan (1) (Al Farizi et al., 2023).

$$VIF = \frac{1}{1 - R^2} \quad (1)$$

4. Uji KMO dan Bartlett

Kaiser Meyer Olkin (KMO) adalah uji kecukupan sampel yang berfungsi untuk membandingkan antara koefisien korelasi yang terlihat dengan koefisien korelasi parsial.

Kecukupan sampel adalah asumsi yang harus terpenuhi dalam analisis *cluster* yang dapat diuji dengan KMO. Adapun persamaan KMO dapat dilihat pada persamaan (2) (Maghfiro et al., 2024).

$$KMO = \frac{\sum_{i=1}^n \sum_{j \neq i}^n r_{ij}^2}{\sum_{i=1}^n \sum_{j \neq i}^n r_{ij}^2 + \sum_{i=1}^n \sum_{j \neq 1}^n r_{x_i x_k - x_j}^2} \quad (2)$$

Dimana:

$i : 1, 2, \dots, p$

$j : 1, 2, \dots, p$

$k : 1, 2, \dots, p$

5. Cluster Hierarki Agglomerative

Penelitian ini menerapkan analisis kluster tipe *agglomerative* dengan lima metode berbeda, yang dikenal sebagai "*Agglomerative Hierarchical Clustering*". Analisis ini dibagi menjadi empat jenis metode berdasarkan pendekatan dalam mengelompokkan objek, yaitu *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan *Ward*. Kelima metode tersebut bertujuan untuk mengelompokkan objek yang memiliki karakteristik serupa (Dewi & Ahadiyah, 2022).

a. Cophenetic

Koefisien korelasi *cophenetic* merupakan koefisien korelasi antara elemen-elemen asli matriks ketidakmiripan (*dissimilarity distance*) dan elemen-elemen yang dihasilkan oleh dendogram (matriks *cophenetic*) (A. Dani et al., 2019). Perhitungan koefisien ini dilakukan menggunakan rumus (A. T. R. Dani et al., 2020) :

$$r_{coph} = \frac{\sum_{m < n} (d_{mn} - \bar{d})(d_{coph \sim mn} - \bar{d}_{coph})}{\sqrt{[\sum_{m < n} (d_{mn} - \bar{d})^2][\sum_{m < n} (d_{coph \sim mn} - \bar{d}_{coph})^2]}} \quad (3)$$

dengan:

r_{coph} : koefisien korelasi *cophenetic*

d_{mn} : jarak asli antara objek ke- m dan ke- n

\bar{d} : rata-rata d_{mn}

$d_{coph \sim mn}$: jarak *cophenetic* objek ke- m dan ke- n

\bar{d}_{coph} : rata-rata $d_{coph \sim mn}$

Korelasi *cophenetic* berfungsi untuk menilai sejauh mana hasil pengelompokan hierarki mencerminkan struktur data aslinya. Dengan membandingkan jarak antar objek dalam matriks ketidakmiripan asli dengan jarak pada dendogram, korelasi ini memberikan indikasi seberapa baik dendogram menggambarkan hubungan antar objek. Semakin mendekati 1 nilai korelasinya, semakin akurat pengelompokan tersebut merepresentasikan data asli (Tri et al., 2020).

b. Silhouette

Metode *Silhouette* adalah pendekatan untuk mengevaluasi kluster yang menggabungkan kohesi dan separasi. Kohesi dihitung dari semua objek dalam kluster, sedangkan separasi diukur melalui jarak rata-rata antara objek dalam kluster dan kluster terdekat (Muningsih & Kiswati, 2018). Jarak antar data dihitung menggunakan rumus *Euclidean*. *Silhouette* dapat dihitung untuk setiap kluster atau keseluruhan kluster yang dihasilkan oleh algoritma klustering. Nilai *silhouette* untuk seluruh data dengan jumlah kluster k , yang dikenal sebagai $sil(c)$,

didefinisikan sebagai rata-rata nilai *silhouette* dari semua kluster. *r* (Paembonan et al., 2021), seperti yang dijelaskan berikut:

$$sil(c) = sil(k) \frac{1}{|k|} \sum_{i=1}^k sil(C_i) \quad (4)$$

dengan :

$sil(c)$: nilai *silhouette* semua *cluster*
 $|k|$: banyaknya *cluster* k
 $il(C_i)$: rata-rata nilai *silhouette*

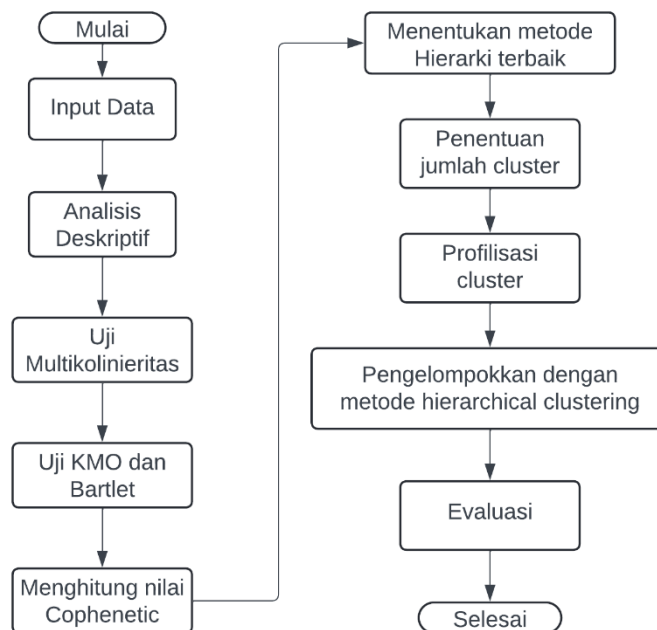
Setelah nilai koefisien siluet dihitung menggunakan rumus di atas, kategori evaluasi kualitas pengelompokan dapat dilihat melalui rentang nilai koefisien siluet. Tabel 1 menunjukkan kategori pengelompokan berdasarkan nilai siluet yang diperoleh.

Tabel 1.
Kategori Evaluasi Nilai *Silhouette*

Nilai <i>Silhouette Coefficient</i>	Kategori
$0,7 < s \leq 1$	<i>Strong structure</i>
$0,5 < s \leq 0,7$	<i>Medium structure</i>
$0,25 < s \leq 0,5$	<i>Weak structure</i>
$s \leq 0,25$	<i>No structure</i>

6. Alur Penelitian

Gambar 1:
Diagram Alir Penelitian



HASIL DAN PEMBAHASAN

1. Analisis Deskriptif

Analisis deskriptif digunakan untuk mengungkapkan informasi dari data. Proses ini mencakup perhitungan seperti rata-rata, median, standar deviasi, serta nilai maksimum dan minimum.

Tabel 2.
Nilai Analisis Deskriptif

Variabel	Min	Mean	Max	Std
Pilar 1 (SBS)	0,0000	0,8214	1,2084	0,2719
Pilar 2 (CTPS)	0,0324	0,4433	0,9740	0,2723
Pilar 3 (PSRT)	0,0000	0,4429	1,0000	0,2892
Pilar 4 (PAMMRT)	0,0000	0,3066	0,8875	0,2338
Pilar 5 (PLCRT)	0,0000	0,2690	0,8867	0,2331

Berdasarkan Tabel 2, Pilar 1 memiliki rata-rata tertinggi, yaitu 0,821 dengan simpangan baku 0,272, menunjukkan variasi yang cukup besar. Pilar 2 dan 3 memiliki rata-rata yang sama, yaitu 0,443 dengan simpangan baku 0,289, menunjukkan distribusi data yang serupa. Pilar 4 dan 5 memiliki rata-rata yang lebih rendah, masing-masing 0,307 dan 0,269, dengan variasi yang lebih kecil. Data ini menunjukkan bahwa Pilar 1 memiliki nilai tertinggi dan variasi terbesar, sementara Pilar 4 dan 5 memiliki nilai dan variasi yang lebih rendah.

2. Uji Multikolinearitas

Multikolinearitas adalah asumsi yang harus dipenuhi dalam analisis *cluster* agar tidak ada variabel yang bersifat duplikatif dalam analisis. Hasil pengujian awal pada lima variabel menunjukkan adanya masalah multikolinearitas. Menurut (Yaldi et al., 2022) salah satu solusi dari masalah multikolinearitas adalah dengan menghilangkan variabel yang memiliki korelasi tinggi. Dengan demikian, variabel PSRT tidak digunakan dalam analisis dan diperoleh hasil dari pengujian asumsi multikolinearitas yang telah memenuhi dapat dilihat pada tabel 3.

Tabel 3.
Pengujian Asumsi Multikolinearitas

Variabel	VIF
SBS	1,686
CTPS	4,913
PAMMRT	8,974
PLCRT	6,006

Berdasarkan hasil pada Tabel 3, nilai VIF semua variabel berada di bawah 10. Hal ini menunjukkan bahwa tidak terdapat masalah multikolinearitas yang signifikan di antara variabel-variabel tersebut. Oleh karena itu, semua variabel tersebut dapat digunakan dalam analisis lebih lanjut, seperti regresi atau analisis *clustering*, tanpa khawatir adanya interaksi berlebihan antara variabel.

3. Uji KMO dan Bartlett

Uji *Kaiser-Meyer-Olkin (KMO)* berfungsi untuk menilai kecukupan sampel dalam analisis faktor, memastikan apakah data layak direduksi menjadi beberapa faktor. Uji *Bartlett* menguji apakah terdapat korelasi signifikan antarvariabel dalam dataset, yang memungkinkan terbentuknya faktor-faktor.

Tabel 4.
Uji KMO dan Bartlett

Metode Uji	Nilai
------------	-------

<i>Kaisen Mayer Olkin (KMO)</i>	0,79212
<i>Barlett's Test of Sphericity</i>	0,00000

Berdasarkan pada table 4 hasil uji asumsi analisis faktor menggunakan KMO dan *Bartlett* menunjukkan hasil yang baik. Nilai KMO sebesar 0,79212 menunjukkan bahwa data cukup memadai untuk dilakukan analisis faktor. Sementara itu, uji *Bartlett* dengan nilai signifikansi sebesar 0,00000 menunjukkan adanya korelasi yang signifikan antar variabel, sehingga analisis dapat dilanjutkan.

4. Penentuan Metode Hierarki Terbaik dengan Nilai *Cophenetic*

Penentuan Metode Hierarki terbaik digunakan pertimbangan nilai korelasi *cophenetic*. Hasil perhitungan korelasi *cophenetic* ditunjukkan pada tabel 5

Tabel 5.

Metode	Nilai <i>Cophenetic</i>
<i>Single Linkage</i>	0,54859
<i>Complete Linkage</i>	0,66023
<i>Average Linkage</i>	0,75119
<i>Ward Linkage</i>	0,62972
<i>Centroid Linkage</i>	0,73889

Berdasarkan tabel 5, metode *Average Linkage* memiliki performa terbaik dalam pembentukan *cluster*. Dengan demikian, metode tersebut akan digunakan dalam membentuk *cluster*.

5. Penentuan Jumlah *Cluster*

Tabel 6 menunjukkan hasil perhitungan nilai *Calinski Harabasz Index* (CHI) untuk melihat nilai k (jumlah *cluster*) optimum. Semakin tinggi nilai CHI menunjukkan *cluster* yang lebih kompak dan terpisah secara jelas antar *cluster* lainnya.

Tabel 6.

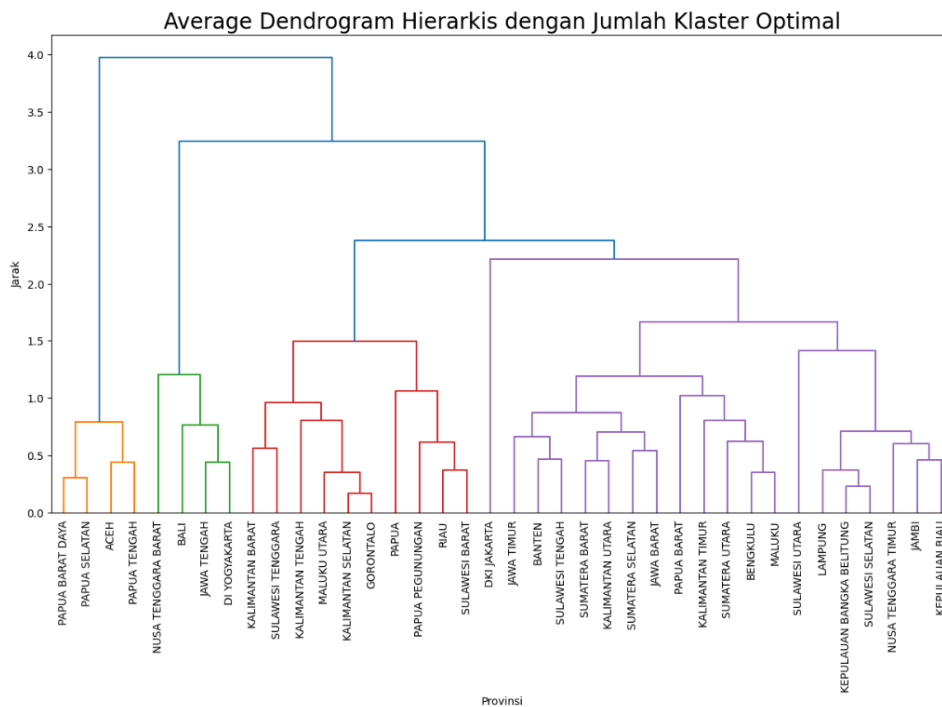
Jumlah <i>Cluster</i>	CH Indeks
2	18,809
3	23,782
4	41,497
5	35,402

Dapat dilihat pada tabel 6, jumlah *cluster* sebanyak 4 menghasilkan nilai CHI yang terbesar. Sehingga diperoleh jumlah *cluster* optimal yang dibentuk adalah sebanyak 4 *cluster*.

6. Hasil *Clustering*

Gambar 2:

Dendogram Average Linkage Hierarchical Clustering



Berdasarkan pengelompokan provinsi pada gambar 1, diperoleh 4 *cluster* provinsi yang terbentuk. Anggota *cluster* pertama berjumlah 4 provinsi, *cluster* kedua berjumlah 4 provinsi, *cluster* ketiga berjumlah 10 provinsi, dan *cluster* keempat berjumlah 20 provinsi seperti yang tertera pada tabel 7.

Tabel 7.
Anggota Cluster

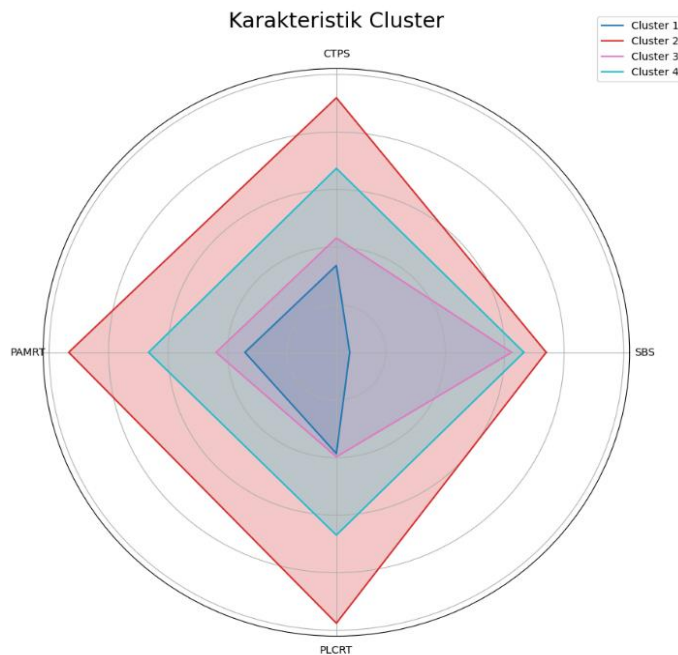
Cluster	Provinsi
Cluster 1	Aceh, Papua Barat Daya, Papua Selatan, Papua Tengah
Cluster 2	Jawa Tengah, DI Yogyakarta, Bali, Nusa Tenggara Barat
Cluster 3	Riau, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara, Papua, Papua Pegunungan
Cluster 4	Sumatera Utara, Sumatera Barat, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Timur, Banten, Nusa Tenggara Timur, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Maluku, Papua Barat

7. Profilisasi Cluster

Karakteristik *cluster* dapat dilihat dengan rata-rata nilai setiap variabel untuk setiap *cluster*. Untuk karakteristik tiap *cluster* dapat dilihat pada gambar 2.

Gambar 3:

Dendrogram Average Linkage Hierarchical Clustering



Berdasarkan hasil analisis kluster, *Cluster 1* menunjukkan nilai yang sangat rendah untuk semua variabel, menandakan performa terendah secara keseluruhan. *Cluster 2*, sebaliknya, memiliki skor yang sangat tinggi, terutama pada variabel CTPS, PAMRT, dan PLCRT, mencerminkan performa terbaik di antara kluster lainnya. *Cluster 3* memiliki nilai yang mendekati rata-rata untuk SBS, tetapi berada di bawah rata-rata untuk variabel lainnya, menunjukkan bahwa *cluster* ini relatif seimbang namun lemah pada beberapa aspek. Sementara itu, *Cluster 4* sedikit di atas rata-rata untuk semua variabel, walaupun tidak terdapat perbedaan yang mencolok pada rata-rata keseluruhan.

Berdasarkan gambar 2, setiap kelompok menunjukkan kondisi sanitasi yang berbeda. *Cluster 1* memiliki kondisi terburuk, dengan nilai sangat rendah pada semua indikator, seperti Stop Buang Air Besar Sembarangan (SBSS), Cuci Tangan Pakai Sabun (CTPS), dan pengelolaan air limbah serta sampah rumah tangga, mencerminkan sanitasi yang sangat buruk. *Cluster 2* merupakan kluster dengan sanitasi terbaik, memiliki nilai sangat tinggi di semua aspek, terutama dalam kebiasaan mencuci tangan dan pengelolaan air limbah. *Cluster 3* berada di rata-rata untuk SBSS, namun lemah dalam aspek lain, seperti kebersihan tangan dan pengelolaan limbah, menunjukkan adanya tantangan di bidang sanitasi. *Cluster 4* sedikit lebih baik dari rata-rata di semua indikator, mencerminkan kondisi sanitasi yang cukup baik, tetapi belum optimal. Perbedaan ini menggambarkan variasi tingkat kebersihan di berbagai wilayah yang memerlukan upaya peningkatan, terutama pada kluster-kluster dengan skor rendah.

8. Evaluasi

Hasil *cluster* di evaluasi menggunakan nilai *silhouette coefficient*. Nilai yang semakin mendekati 1 mengindikasikan kualitas *cluster* yang semakin baik.

Tabel 8.

Nilai <i>Silhouette Coefficient</i>	
<i>Silhouette Coefficient</i>	Kriteria Penilaian
0,5027	Medium Structure

Tabel 8 menunjukkan nilai *silhouette coefficient* dari hasil *cluster* yang diperoleh yaitu sebesar 0,5027. Nilai tersebut menunjukkan bahwa struktur *cluster* dalam data

adalah medium, artinya objek-objek cenderung cocok dengan *cluster* mereka dan kurang bertumpang tindih dengan *cluster* lainnya.

KESIMPULAN

Berdasarkan hasil analisis, diperoleh kesimpulan bahwa metode *Agglomerative Average Linkage* merupakan metode terbaik untuk pengelompokan berdasarkan data capaian STBM Indonesia 2023, ditunjukkan dengan nilai *cophenetic* tertinggi yaitu sebesar 0,75119. Selanjutnya, berdasarkan nilai *CH Index* tertinggi (41,497), diperoleh jumlah *cluster optimum* adalah sebanyak 4 *cluster*. Dari hasil pengelompokan diperoleh anggota *cluster* pertama berjumlah 4 provinsi, *cluster* kedua berjumlah 4 provinsi, *cluster* ketiga berjumlah 10 provinsi, dan *cluster* keempat berjumlah 20 provinsi.

Hasil profilisasi *cluster* menunjukkan *Cluster 1* memiliki performa terendah secara keseluruhan. *Cluster 2*, sebaliknya, memiliki skor yang sangat tinggi, terutama pada variabel CTPS, PAMRT, dan PLCRT. *Cluster 3* memiliki nilai yang mendekati rata-rata untuk SBS, tetapi berada di bawah rata-rata untuk variabel lainnya. Sementara itu, *Cluster 4* sedikit di atas rata-rata untuk semua variabel, meskipun tidak ada perbedaan yang signifikan dari rata-rata keseluruhan. Hasil evaluasi *clustering* menggunakan nilai *Silhouette* menunjukkan nilai yang cukup baik yaitu sebesar 0,5027 atau dalam kategori *medium structure*.

DAFTAR PUSTAKA

- Al Farizi, M. F., Harianto, F., Dewanti, M. S., Siburian, C. A., Mardianto, M. F. F., Amelia, D., & Ana, E. (2023). Pengelompokan Daerah di Jawa Timur Berbasis Indikator Kesejahteraan Masyarakat dengan Pendekatan Analisis Cluster Hierarki. *Inferensi*, 6(2), 141. <https://doi.org/10.12962/j27213862.v6i2.15452>
- Alamtaha, Z., Djakaria, I., & Yahya, N. I. (2023). Implementasi Algoritma Hierarchical Clustering dan Non-Hierarchical Clustering untuk Pengelompokan Pengguna Media Sosial. *Estimasi: Journal of Statistics and Its Application*, 4(1), 2721–379. <https://doi.org/10.20956/ejsa.vi.24830>
- Bimantoro Putra, F., Tri Rian Dani, A., & Wigantono, S. (n.d.). *PENERAPAN ALGORITMA HIERARCHICAL CLUSTERING DALAM PENGELOMPOKKAN KABUPATEN/KOTA DI PAPUA BERDASARKAN INDIKATOR KEMISKINAN APPLICATION OF THE HIERARCHICAL CLUSTERING ALGORITHM IN GROUPING DISTRICTS/CITIES IN PAPUA BASED ON POVERTY INDICATORS*.
- Dani, A. T. R., Wahyuningsih, S., & Rizki, N. A. (2020). Pengelompokan Data Runtun Waktu menggunakan Analisis Cluster (Studi Kasus: Nilai Ekspor Komoditi Migas dan Nonmigas Provinsi Kalimantan Timur Periode Januari 2000-Desember 2016). *Jurnal EKSPONENSIAL*, 11(1), 29–38.
- Dani, A., Wahyuningsih, S., & Rizki, N. (2019). Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu. *Jambura Journal of Mathematics*, 1, 64–78. <https://doi.org/10.34312/jjom.v1i2.2354>
- Dewi, A. F., & Ahadiyah, K. (2022). Agglomerative Hierarchy Clustering Pada Penentuan Kelompok Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pendidikan. *Zeta - Math Journal*, 7(2), 57–63. <https://doi.org/10.31102/zeta.2022.7.2.57-63>
- FK UNUSA. (2023). *Situasi Sanitasi di Indonesia*. <https://fkes.unusa.ac.id/2023/03/23/situasi-sanitasi-di-indonesia/>.
- Iyan Yulianti, D., Iman Hermanto, T., & Defriani, M. (2023). RESOLUSI: Rekayasa Teknik Informatika dan Informasi Analisis Clustering Donor Darah dengan Metode

- Agglomerative Hierarchical Clustering. *Media Online*, 3(6), 308.
<https://djournals.com/resolusi>
- Kemkes RI. (n.d.). *Wujudkan Sanitasi Indonesia yang Lebih Bersih, Lebih Sehat*. Retrieved September 26, 2024, from <https://stbm.kemkes.go.id/>
- Kemkes RI. (2024). *Laporan Kinerja Kegiatan Penyehatan Lingkungan Tahun 2023*.
- Krisman Pratama Simanjuntak, U. K. (2021). *Hotspot Clustering in Jambi Province Using Agglomerative Hierarchical Clustering Algorithm*. 1, 7–16.
- Maghfiro, M., Wardhani, N. W. S., & Iriany, A. (2024). Ensemble Cluster Method For Clustering Cabbage Production In East Java. *Inferensi*, 7(2), 129.
<https://doi.org/10.12962/j27213862.v7i2.20378>
- Mahmudan, A. (2020). Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering. *Jurnal Matematika, Statistika Dan Komputasi*, 17(1), 1–13. <https://doi.org/10.20956/jmsk.v17i1.10727>
- Muningsih, E., & Kiswati, S. (2018). SISTEM APLIKASI BERBASIS OPTIMASI METODE ELBOW UNTUK PENENTUAN CLUSTERING PELANGGAN. In *JOUTICA* (Vol. 3, Issue 1).
- Paembonan, S., Abduh, H., & Kunci, K. (2021). *Penerapan Metode Silhouette Coefficient Untuk Evaluasi Clustering Obat Clustering; K-means; Silhouette coefficient* (Vol. 6, Issue 2). <https://ojs.unanda.ac.id/index.php/jiit/index>
- Sa'ban, L. M. A., Sadat, A., & Nazar, A. (2020). Meningkatkan Pengetahuan Masyarakat Dalam Perbaikan Sanitasi Lingkungan. *Dinamisia : Jurnal Pengabdian Kepada Masyarakat*, 5(1), 10–16. <https://doi.org/10.31849/dinamisia.v5i1.4365>
- Singh, S., & Srivastava, S. (2020). Review of Clustering Techniques in Control System. *Procedia Computer Science*, 173, 272–280.
<https://doi.org/10.1016/j.procs.2020.06.032>
- Stiawati, T. (2021). Program Sanitasi Total Berbasis Masyarakat (STBM) untuk Merubah Perilaku Hidup Sehat di Kelurahan Kasunyatan Kota Serang Provinsi Banten. *Sawala : Jurnal Administrasi Negara*, 9(2), 179–191.
<https://doi.org/10.30656/sawala.v9i2.3607>
- Tri, A., Dani, R., Wahyuningsih, S., Nanda, D., & Rizki, A. (2020). Pengelompokan Data Runtun Waktu menggunakan Analisis Cluster (Studi Kasus: Nilai Ekspor Komoditi Migas dan Nonmigas Provinsi Kalimantan Timur Periode Januari 2000-Desember 2016) Grouping of Time Series Data using Cluster Analysis (Case Study: Export Value of Oil and Non-oil Commodities in East Kalimantan Provinces Period. *Jurnal EKSPONENSIAL*, 11(1).
- United Nations. (n.d.). *Sustainable Development Goals*. <https://www.un.org/sustainabledevelopment/water-and-sanitation/>
- WHO. (2024). *Sanitation*. WHO. <https://www.who.int/news-room/fact-sheets/detail/sanitation>
- WHO, & UNICEF. (2023). *Progress on household drinking water, sanitation and hygiene 2000–2022: special focus on gender*.
- Widyawati, W., Saptomo, W. L. Y., & Utami, Y. R. W. (2020). Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan. *Jurnal Ilmiah SINUS*, 18(1), 75. <https://doi.org/10.30646/sinus.v18i1.448>
- Yaldi, E., Pasaribu, J. P. K., Suratno, E., Kadar, M., Gunardi, G., Naibaho, R., Hati, S. K., & Aryati, V. A. (2022). Penerapan Uji Multikolinieritas Dalam Penelitian Manajemen Sumber Daya Manusia. *Jurnal Ilmiah Manajemen Dan Kewirausahaan (JUMANAGE)*, 1(2), 94–102. <https://doi.org/10.33998/jumanage.2022.1.2.89>