

**Perbandingan Metode Agglomerative Hierarchical Cluster dan
DBSCAN dalam Pengelompokan Kabupaten/Kota di Jawa Barat
Berdasarkan Indikator Kesehatan Tahun 2022**



Kelompok 4 Paralel 1:

Tiara Ayu Pertiwi	G1401201012
Raziqizzan Putrandi	G1401211040
Azzahra Adelia Putri	G1401211045
Windi Gunawan	G1401211082
Ubaidillah Al Hakim	G1401211086

**DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**

1. Pendahuluan

Analisis cluster merupakan salah satu metode statistika yang digunakan untuk mengelompokkan sesuatu berdasarkan algoritma tertentu. Menurut Anderberg (2014) Algoritma dalam analisis cluster dapat mengumpulkan amatan ke dalam kelompok-kelompok yang sebelumnya tidak terpikirkan oleh manusia. Tujuan dari pengelompokkan ini adalah mendapatkan beberapa kelompok dengan amatan dalam satu kelompok memiliki kesamaan satu sama lain dibandingkan dengan amatan dalam kelompok lainnya.

Analisis cluster memiliki beberapa metode diantaranya adalah Agglomerative hierarchical cluster (AHC), K-Mean, dan Autoencoder. Metode AHC merupakan salah satu metode klasik yang dibandingkan dengan metode klasik seperti K-Means dan Fuzzy-C means, AHC memiliki kelebihan seperti tidak memerlukan partisi awal serta tidak menghasilkan titik-titik noise (Liu *et al.* 2021). Selain itu, AHC mudah dilakukan dan dapat memberikan deskripsi dan visualisasi yang sangat informatif untuk struktur pengelompokan data yang potensial. Meskipun begitu AHC memiliki kompleksitas yang lebih tinggi dibandingkan kedua metode tersebut.

Salah satu algoritma dari analisis cluster yang berdasarkan pada kepadatan data pada suatu area cluster. Algoritma tersebut adalah Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Keunggulan DBSCAN dibandingkan analisis cluster lainnya diantaranya adalah kemampuan untuk menduga cluster dengan bentuk acak, dapat mendeteksi dan mengabaikan noise (Hanafi dan Saadaftar 2022). Selain itu, Penelitian yang dilakukan oleh Deng (2020) menyebutkan bahwa algoritma DBSCAN ketika diberlakukan pengclusteran pada data yang kepadatannya tidak seragam menghasilkan homogenitas dalam kelompok serta perbedaan antar kelompok yang tinggi. Walaupun begitu kelemahan dari DBSCAN adalah memerlukan waktu yang relatif lama ketika diberikan data yang besar.

Kesehatan merupakan salah satu unsur yang penting dalam sebuah negara, hal ini didukung oleh Persatuan Bangsa Bangsa (PBB) dalam programnya yaitu Sustainable Development Goals (SDGs) yang salah satunya adalah kesehatan yang merata. Menurut laporan Kementerian Kesehatan (2022), menyebutkan pada Provinsi Jawa Barat terdapat kekurangan sebanyak 20% pada perawat dan 24,4% pada dokter gigi. Meskipun demikian, Provinsi Jawa Barat memiliki infrastruktur, tenaga medis, dan fasilitas penunjang yang berada di atas rata-rata.

Pemerataan kesehatan dipengaruhi oleh wilayah kabupaten/kota tempat tinggal. Hal tersebut sejalan dengan pernyataan Rakasiwi dan Kautsar (2021) yang menyatakan bahwa kebutuhan pelayanan kesehatan terpengaruhi oleh tempat tinggal. Oleh karena itu, pengclusteran perlu dilakukan untuk mengetahui kelompok daerah yang memerlukan peningkatan dalam bidang kesehatan dengan hal tersebut, diharapkan pemerintah Provinsi Jawa Barat dapat menyusun kebijakan yang lebih efisien dan tepat sasaran.

Tujuan:

1. Membandingkan dua metode pengclusteran *unsupervised learning* yaitu antara AHC dan DBSCAN.

2. Mengclusterkan wilayah kabupaten dan kota pada Provinsi Jawa Barat berdasarkan indikasi kesehatan dan melakukan interpretasi dari cluster yang didapatkan.

2. Metodologi

2.1 Bahan dan Data

Data yang digunakan dalam penelitian ini adalah data kesehatan Indonesia yang diperoleh dari laman open data jawa barat tahun 2022. Data terdiri dari 10 peubah. Objek amatan dalam penelitian ini adalah kabupaten/kota di Provinsi Jawa Barat, Indonesia yang berjumlah 27.

Tabel 1 Peubah yang digunakan dalam penelitian

Peubah	Keterangan	Tipe Peubah	Sumber
x1	Jumlah Ibu Bersalin	Numerik	Kemenkes (2022)
x2	Presentase Balita Stunting	Numerik	Kemenkes (2022)
x3	Jumlah Dokter Umum	Numerik	Kemenkes (2022)
x4	Jumlah Dokter Gigi	Numerik	Kemenkes (2022)
x5	Jumlah Laboratorium Kesehatan	Numerik	Kemenkes (2022)
x6	Jumlah Bidan	Numerik	Kemenkes (2022)
x7	Jumlah Balita Ditimbang	Numerik	Kemenkes (2022)
x8	Persentase Penerima BPJS Kesehatan Penerima Bantuan Iuran	Numerik	Kemenkes (2022)
x9	Persentase Penerima BPJS Kesehatan Non-Penerima Bantuan Iuran	Numerik	Kemenkes (2022)
x10	Persentase Penerima Jamkesda, Asuransi Swasta, Perusahaan	Numerik	Kemenkes (2022)

2.2 Standarisasi Z-score

Z-score adalah teknik normalisasi data yang penting untuk standarisasi. Ini merupakan ukuran statistik yang menghubungkan nilai tertentu dengan rata-rata dari sekumpulan nilai (Nevil 2023). Dalam penelitian oleh Agarwal (2014), Z-score diakui sangat berguna terutama saat atribut memiliki rentang nilai yang tidak diketahui atau adanya outlier signifikan yang memengaruhi normalisasi min-max. Di samping itu, Z-score juga banyak digunakan dalam berbagai bidang, salah satunya dapat digunakan dalam analisis statistik dan penelitian medis yang dilakukan oleh Kranzusch *et al.* (2020). Bahkan, metode ini terbukti efektif dalam tahap preprocessing untuk klasifikasi menggunakan

algoritma machine learning, seperti yang dilakukan oleh Henderi (2021). Hal ini menegaskan bahwa Z-score memiliki fleksibilitas dan efektivitas dalam berbagai bidang (Pratama 2023).

2.3 Agglomerative Hierarchical Cluster

Agglomerative hierarchical cluster adalah sebuah metode pengelompokan hirarki yang menggunakan pendekatan dari bawah ke atas (bottom-up). Prosesnya dimulai dengan setiap data dianggap sebagai sebuah kelompok, kemudian kelompok-kelompok potensial ditemukan berdasarkan jarak antar data untuk digabungkan menjadi kelompok yang lebih besar. Proses ini terus berulang sehingga membentuk struktur hierarki yang tampak bergerak ke atas (Agglomerative), membentuk jenjang (Hierarki). Hierarchical clustering adalah teknik pengelompokan yang membentuk struktur pohon. Prosesnya dilakukan secara bertingkat atau bertahap, dan terdapat dua metode dalam algoritma Hierarchical Clustering, yaitu Agglomerative (bottom-up) dan Divisive (top-down) (Yulianti *et al.* 2023).

Adapun jenis linkage yang digunakan dalam algoritma Agglomerative Hierarchical Clustering adalah sebagai berikut:

a. Single Linkage (Nearest Neighbor Methods)

Single linkage clustering adalah teknik pengelompokan hierarki yang menentukan jarak antara kelompok berdasarkan titik-titik terdekat di setiap kelompok. Dalam pengelompokan single linkage, jarak antara dua kelompok didefinisikan sebagai jarak terpendek antara dua titik mana pun di dua kelompok tersebut (Jarman 2020).

$$d_{(UV)} = \min(d_{UW}, d_{VW})$$

Salah satu keuntungan utama dari pengelompokan single linkage adalah kemampuannya untuk mendeteksi kelompok yang memanjang atau kelompok dengan bentuk yang tidak teratur. Namun, metode ini sensitif terhadap noise dan outlier, yang dapat menyebabkan efek rantai di mana kelompok bergabung berdasarkan titik-titik dekat tunggal, tanpa memperhatikan struktur keseluruhan data.

b. Complete Linkage (Furthest Neighbor Methods)

Complete linkage terjadi apabila kelompok-kelompok digabung berdasarkan jarak terjauh dari satu cluster dengan cluster lainnya (Suhirman *et al.* 2019).

$$d_{(UV)} = \max(d_{UW}, d_{VW})$$

Salah satu keuntungan dari pengelompokan complete linkage adalah kemampuannya untuk menangani noise dan outlier dengan lebih baik daripada metode single linkage. Namun, metode ini juga memiliki kelemahan, di mana dalam kehadiran noise, pengelompokan dapat memberikan hasil yang tidak optimal (Jarman 2020).

c. Average Linkage (between Groups Methods)

$$d_{(UV)W} = \frac{d_{(UW)} + d_{(VW)}}{n_{(UW)} + n_{(VW)}}$$

Salah satu keunggulan dari pengelompokan average linkage adalah kemampuannya untuk mengatasi noise dan outlier lebih baik daripada metode single linkage dan complete linkage. Metode ini sering digunakan dalam analisis pengelompokan hierarkis karena cenderung memberikan hasil yang lebih stabil dan konsisten

2.4 Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

Algoritma Density-Based Spatial Clustering of Applications with Noise (DBSCAN) adalah algoritma pengelompokan berbasis kerapatan distribusi titik data. Algoritma ini umum digunakan dalam berbagai bidang, termasuk astronomi, untuk mengidentifikasi kelompok titik data. DBSCAN bekerja dengan mengelompokkan titik data yang saling berdekatan sementara menandai titik-titik yang berada di daerah kerapatan rendah sebagai outlier atau titik noise. Secara keseluruhan, DBSCAN adalah algoritma pengelompokan yang kuat yang sangat berguna dalam skenario dimana jumlah kelompok tidak diketahui, dan kelompok-kelompok tersebut mungkin memiliki bentuk dan kerapatan yang bervariasi (Zhang 2019).

Kelebihan dari DBSCAN adalah DBSCAN tidak memerlukan jumlah kelompok yang harus ditentukan sebelumnya, dapat mengidentifikasi kelompok dengan bentuk dan ukuran yang beragam, tahan terhadap outlier dan noise dalam data, tidak mengasumsikan distribusi tertentu dari titik data dalam kelompok. Langkah algoritma DBSCAN adalah sebagai berikut.

- a. Mulai dari titik yang belum dikunjungi dan temukan semua titik terdekat dalam jarak ϵ .
- b. Jika jumlah titik terdekat lebih besar dari atau sama dengan \minPts , sebuah kelompok terbentuk.
- c. Perluas kelompok dengan menambahkan titik tetangga secara rekursif.
- d. Titik-titik yang tidak termasuk dalam kelompok apapun ditandai sebagai noise.

2.5 Koefisien Silhouette

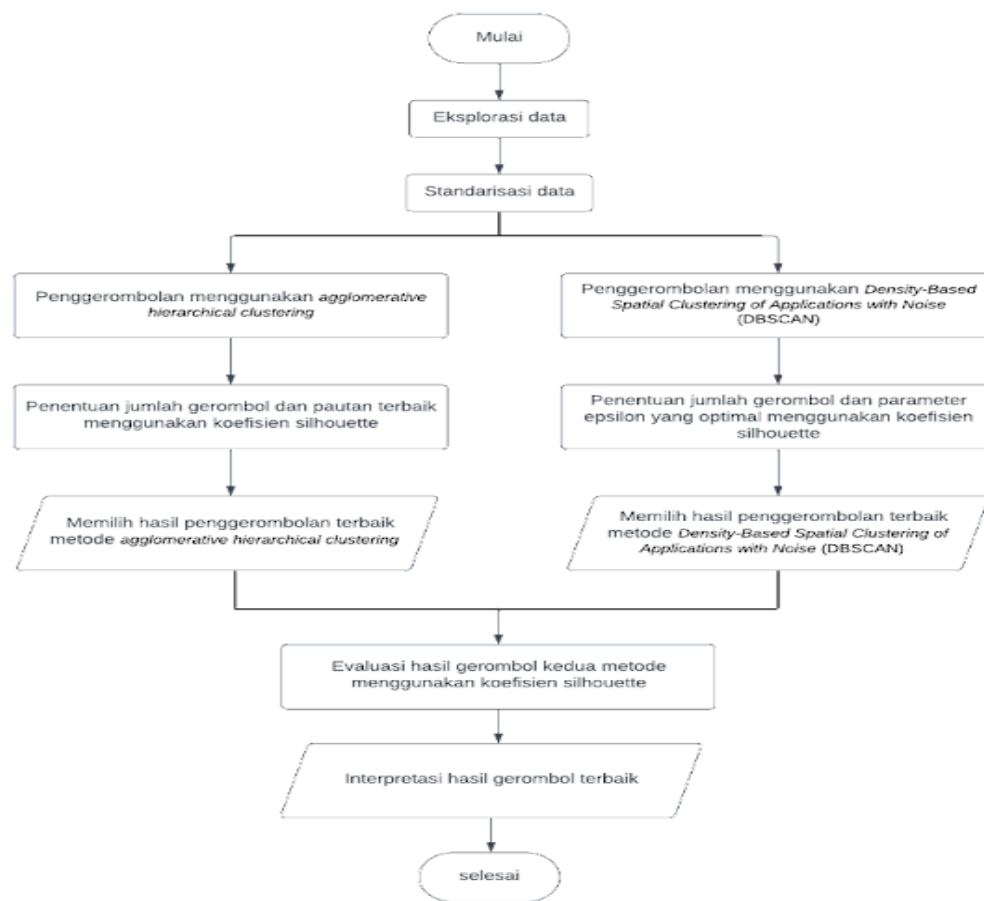
Koefisien Silhouette merupakan metode evaluasi yang digunakan untuk menilai kualitas kluster yang dihasilkan oleh algoritma clustering. Ini mengukur seberapa baik setiap data sesuai dengan kluster tempatnya dibandingkan dengan kluster lain. Rentang nilai Silhouette adalah dari -1 hingga +1, dimana nilai yang lebih tinggi menandakan

Silhouette mendekati +1 menunjukkan pembentukan kluster yang baik, sedangkan nilai negatif menandakan kemungkinan kesalahan penempatan data dalam kluster. Nilai Silhouette sekitar 0 menunjukkan posisi data yang berada di antara dua kluster. Metrik ini membantu dalam menilai seberapa baik kluster telah terbentuk dan sejauh mana batas antara kluster yang jelas (Dalimunthe 2021).

2.6 Prosedur Analisis Data

1. Melakukan eksplorasi data pada setiap peubah dengan boxplot dan matriks korelasi.
2. Melakukan standarisasi data menggunakan Z-score. Bertujuan untuk memastikan peubah yang berskala besar tidak mendominasi peubah yang berskala kecil.

3. Melakukan Pengclusteran kabupaten/kota dengan menggunakan Metode *agglomerative hierarchical clustering*.
 - a. Menentukan Jumlah cluster dan pautan terbaik
 - b. Melakukan pemilihan jumlah cluster dan pautan terbaik berdasarkan nilai koefisien silhouette.
4. Melakukan pengclusteran kabupaten/kota dengan menggunakan Metode *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN).
 - a. Menentukan epsilon (1-1000) dan jumlah cluster yang optimal.
 - b. Melakukan pemilihan jumlah cluster optimal serta parameter optimal bagi epsilon berdasarkan nilai koefisien Silhouette.
5. Mengevaluasi hasil pengclusteran terbaik menggunakan koefisien Silhouette dan indeks DBCV dari hasil pengclusteran terbaik kedua metode.
6. Memilih hasil pengclusteran terbaik dan melakukan interpretasi hasil cluster terbaik dengan melihat karakteristik dari setiap cluster.

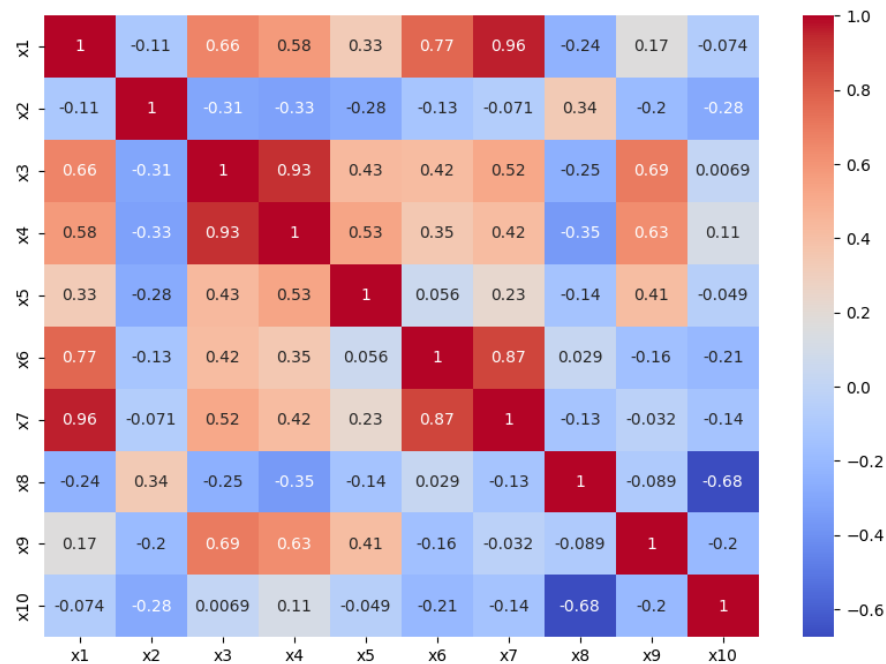


Gambar 1 Diagram Alir Penelitian

3. Hasil dan Pembahasan

3.1 Eksplorasi Data

Tahapan awal sebelum melakukan analisa data lebih lanjut dilakukan eksplorasi data, untuk melihat data secara umum. Pada penelitian kali ini, eksplorasi yang digunakan menggunakan matriks korelasi. Berdasarkan Gambar 2 terlihat beberapa peubah memiliki korelasi yang tinggi antara satu sama lainnya. Dua pasang peubah memiliki korelasi mendekati satu, yaitu x1 dengan x7 serta x3 dengan x4. x1 dan x7 memiliki korelasi yang tinggi dikarenakan jika ada balita yang ditimbang berasal dari ibu yang melahirkan pada tahun tersebut.

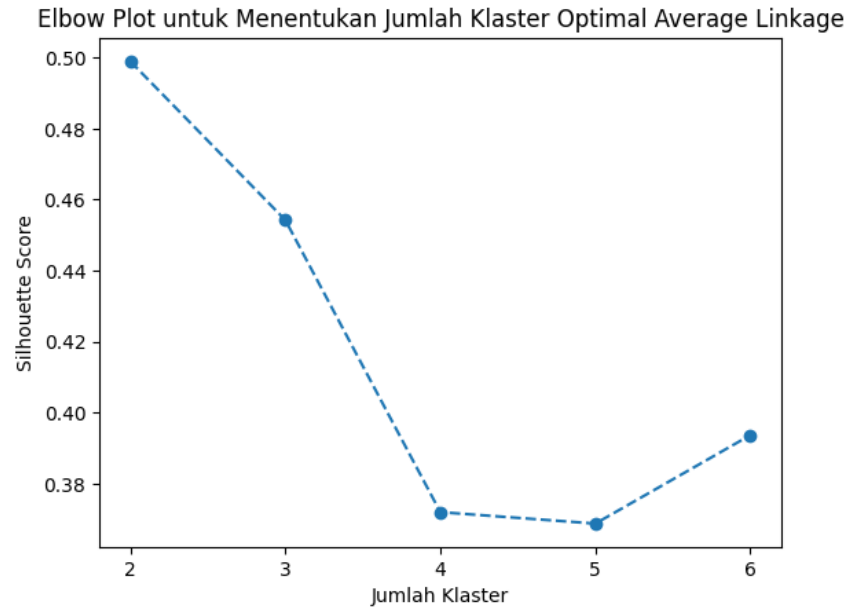


Gambar 2 Matriks korelasi antar peubah

Korelasi yang tinggi dapat mempengaruhi hasil dari clustering oleh karena itu, perlu dilakukan reduksi dimensi menggunakan *Principal Component Analysis* (PCA). Sebelum dilakukan PCA, terlebih dahulu melakukan standarisasi. Penggunaan standarisasi digunakan untuk menyeimbangkan skala dari setiap peubah. Metode yang digunakan dalam metode ini adalah *z-score*.

3.2 Agglomerative Hierarchical Clustering (AHC)

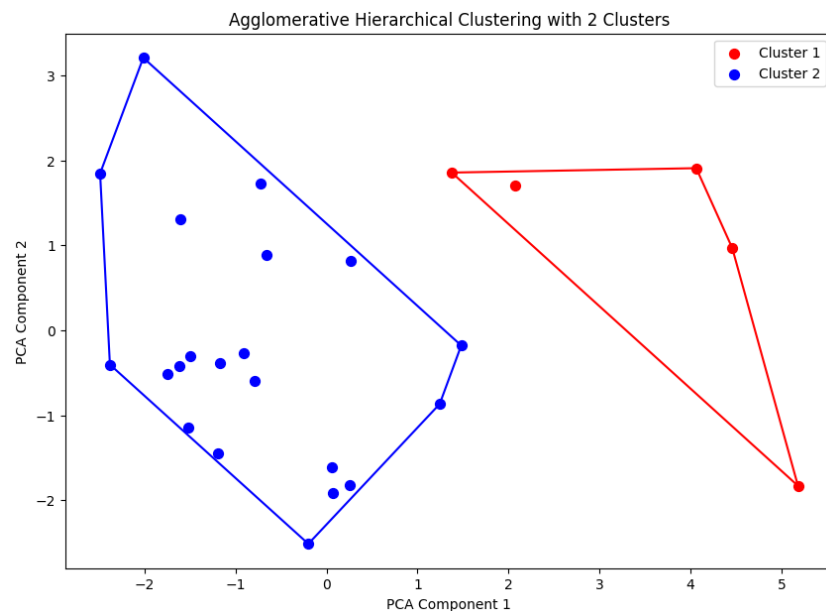
Klasterisasi menggunakan metode AHC dilakukan setelah melalui proses eksplorasi data. Langkah pertama yang dilakukan adalah menentukan *cluster* optimal menggunakan *elbow method* yang didasarkan pada nilai *Silhouette Score*. Penelitian ini menguji jumlah *cluster* mulai dari 2 hingga 6. Visualisasi *elbow plot* dapat dilihat pada Gambar 3.



Gambar 3 *Elbow plot* untuk menentukan *cluster* optimal

Berdasarkan Gambar 2 di atas, terlihat bahwa jumlah *cluster* dengan *Silhouette Score* terbesar adalah 2. Hal ini mengindikasikan bahwa 2 *cluster* adalah jumlah yang paling optimal untuk mengclusterkan kabupaten/kota di Jawa Barat berdasarkan 10 peubah yang digunakan.

Setelah menentukan bahwa jumlah *cluster* optimal adalah 2, data kemudian di-*cluster* menggunakan metode AHC dengan *linkage average*. Hasil *clustering* divisualisasikan dalam bentuk *cluster plot* untuk memudahkan interpretasi distribusi data dalam *cluster*. Visualisasi *cluster plot* dapat dilihat pada Gambar 4.



Gambar 4 *Cluster plot* AHC

Gambar 4 di atas menunjukkan 2 *cluster* yang terbentuk dengan jelas. *Cluster* pertama terdiri dari 6 kabupaten/kota, sedangkan *cluster* kedua terdiri dari 21 kabupaten/kota. Sebaran *cluster* setiap kabupaten/kota dapat dilihat pada Tabel 2.

Tabel 2 Klasterisasi Kabupaten/Kota

Kabupaten/Kota	Cluster
Kabupaten Bogor, Kabupaten Bekasi, Kota Bogor, Kota Bandung, Kota Bekasi, dan Kota Depok.	1
Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Bandung, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Sumedang, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bandung Barat, Kabupaten Pangandaran, Kota Sukabumi, Kota Cirebon, Kota Cimahi, Kota Tasikmalaya, dan Kota Banjar	2

Selanjutnya dihitung rata rata nilai untuk setiap *cluster* untuk melihat gambaran tentang kontribusi setiap peubah dalam pembentukan *cluster*. Rata rata nilai untuk setiap *cluster* dapat dilihat pada Tabel 3.

Tabel 3 Rata rata nilai tiap *cluster*

Cluster	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
1	62105.8	3.1	555.2	98.4	10.6	1231.0	165028.6	27.5	35.0	17.8
2	24388.5	6.7	134.5	17.8	4.8	905.6	89733.6	35.3	17.8	9.8

Berdasarkan Tabel 3 di atas, *cluster* 1 menunjukkan nilai yang lebih tinggi pada sebagian besar indikator kesehatan, terutama pada peubah x1, x3, x4, x5, x6, x7, x9, dan x10. Hal ini dapat mengindikasikan bahwa kabupaten/kota dalam klaster ini memiliki karakteristik ekonomi dan sosial yang lebih kuat pada peubah-peubah tersebut. *Cluster* 2 memiliki nilai yang lebih tinggi pada peubah x2 dan x8, yang mungkin mengindikasikan bahwa kabupaten/kota dalam klaster ini memiliki lebih banyak penduduk dengan status sosial ekonomi lebih rendah yang membutuhkan bantuan kesehatan.

3.3 Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

Metode selanjutnya yang digunakan untuk melakukan pengclusteran ialah DBSCAN menggunakan data yang sudah dilakukan standarisasi dan reduksi dimensi dengan PCA. Parameter yang dilakukan tuning pada metode DBSCAN ialah epsilon dari 1 s.d 5 dan parameter tetapnya ialah $minPts = 2$. Parameter yang dipilih didasarkan pada nilai koefisien silhouette tertinggi. setelah dilakukan tuning parameter jumlah cluster yang didapatkan dapat dilihat pada Tabel 4.

Tabel 4 Evaluasi hasil pengclusteran DBSCAN untuk epsilon optimal berdasarkan koefisien silhouette

Epsilon	minPts	Jumlah cluster	Noise	Koefisien Silhouette
1	2	5	1	0.36
2	2	2	1	0.48
3	2	1	1	-1
4	2	1	1	-1
5	2	1	1	-1

Cluster yang terbentuk dari epsilon yang optimal ialah 2 dengan nilai koefisien silhouette yang didapatkan sebesar 0.48. Nilai koefisien silhouette yang didapatkan sudah cukup baik untuk mengclusterkan data kesehatan tersebut. Sehingga cluster yang di gunakan pada metode DBSCAN ialah dua dengan noise satu kabupaten.

Tabel 5 Klasterisasi Kabupaten/Kota

Kabupaten/Kota	Cluster
Kota Bogor dan Kota Bekasi.	1
Kabupaten Bekasi, Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Bandung, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Sumedang, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bandung Barat, Kabupaten Pangandaran, Kota Sukabumi, Kota Depok, Kota Bandung, Kota Cirebon, Kota Cimahi, Kota Tasikmalaya, dan Kota Banjar	2
Kabupaten Bogor	Noise

Dapat dilihat pada tabel 5 bahwa pada cluster satu terdapat 2 kabupaten/kota sedangkan pada cluster kedua terdapat 23 kabupaten/kota serta terdapat 1 kabupaten/kota sebagai noise atau pencilan. Interpretasi setiap cluster dapat dilakukan dengan melihat rata-rata peubah disetiap cluster sehingga dapat dilihat sebaran atau ciri dari setiap cluster.

Tabel 6 Rata rata nilai tiap *cluster*

Cluster	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
1	65325.5	2.5	687.5	132	7.5	1405	173512.5	18.8	33.9	32.6
2	25020.6	6.4	157.8	22.8	5.6	887.5	88563.8	35.3	19.7	9.7
Noise	116562	4.7	596	79	9	1950	325557	28.4	29	6.5

Berdasarkan Tabel 6 di atas, *cluster* 1 menunjukkan nilai yang lebih tinggi pada sebagian besar indikator kesehatan, terutama pada peubah x1, x3, x4, x5, x6, x7, x9, dan x10. Hal ini dapat mengindikasikan bahwa kabupaten/kota dalam klaster ini memiliki karakteristik ekonomi dan sosial yang lebih kuat sehingga Kesehatan pada cluster ini

tergolong baik. *Cluster 2* memiliki nilai yang lebih tinggi pada peubah x2 dan x8, yang mungkin mengindikasikan bahwa kabupaten/kota dalam kluster ini memiliki lebih banyak penduduk dengan status sosial ekonomi lebih rendah yang mengindikasikan kurang baiknya Kesehatan di *cluster* ini sehingga membutuhkan bantuan Kesehatan lebih.

3.4 Evaluasi Model

Evaluasi model clustering menggunakan *silhouette score* merupakan salah satu pendekatan untuk mengukur kualitas kluster yang dihasilkan. Penelitian ini menggunakan dua metode clustering yang dibandingkan yaitu *Agglomerative Hierarchical Clustering* dan DBSCAN, dengan pembentukan masing-masing 2 kluster. Hasil evaluasi menunjukkan bahwa *silhouette score* untuk metode *Agglomerative Hierarchical Clustering* adalah sebesar 0,4990, sedangkan untuk DBSCAN adalah 0,4822. Skor tersebut mengindikasikan seberapa baik setiap titik data cocok dengan kluster tempatnya berada, dengan rentang nilai antara -1 hingga 1. Nilai *silhouette* yang semakin mendekati +1 menunjukkan bahwa kluster yang terbentuk sudah terdefinisi dengan baik, sedangkan skor mendekati 0 menandakan adanya tumpang tindih antar-kluster, dan skor negatif menunjukkan titik data yang mungkin salah diklasterkan (Sunarko *et al.*, 2023).

Evaluasi menunjukkan bahwa model *Agglomerative Hierarchical Clustering* secara moderat berhasil membentuk dua kluster yang signifikan. Hal ini menunjukkan keefektifan metode ini dalam memisahkan data berdasarkan peubah yang dianalisis. Perbandingan *silhouette score* antara kedua metode menunjukkan bahwa metode *Agglomerative Hierarchical Clustering* memiliki keunggulan tipis dibanding DBSCAN. Perbedaan *silhouette score* tersebut sebesar 0,0168. Oleh karena itu, dari kedua metode tersebut, *Agglomerative Hierarchical Clustering* dengan 2 kluster dapat dianggap lebih baik daripada DBSCAN.

4. Kesimpulan

Berdasarkan hasil yang diperoleh dalam penelitian ini, penggunaan metode analisis penggerombolan AHC maupun DBSCAN memberikan hasil yang tidak jauh berbeda pada jumlah cluster optimalnya, yaitu 2 cluster. Hanya saja pada metode DBSCAN disertai dengan 1 data noise atau pencilan. Disamping itu juga pada metode AHC, cluster 1 terdiri dari 6 kabupaten/kota dan pada cluster 2 terdiri dari 21 kabupaten/kota, sedangkan pada metode DBSCAN cluster 1 terdapat 2 kabupaten/kota dan cluster 2 terdiri dari 23 kabupaten/kota disertai dengan 1 kabupaten yang menjadi noise atau pencilan.

Hasil analisis menggunakan metode AHC dan DBSCAN, yang mana didapatkan dari rata-rata setiap peubah pada cluster menunjukkan interpretasi yang tidak berbeda. Cluster 1 pada kedua metode memiliki nilai rata-rata yang lebih tinggi pada sebagian besar peubah, terutama pada indikator jumlah ibu bersalin, dokter umum, dokter gigi, laboratorium kesehatan, bidan, dan balita ditimbang, serta persentase penerima BPJS Kesehatan non-penerima bantuan iuran dan persentase penerima Jamkesda, asuransi swasta dan perusahaan. Hal ini dapat mendefinisikan bahwa kabupaten atau kota yang berada pada cluster 1 memiliki karakteristik sosial ekonomi yang lebih kuat, yang mana menggambarkan tingkat kesehatan yang baik. Sedangkan pada cluster 2 baik menggunakan metode AHC maupun DBSCAN

menunjukkan nilai rata-rata yang lebih tinggi pada indikator persentase balita yang mengalami stunting dan persentase penerima BPJS Kesehatan dengan bantuan iuran. Hasil tersebut menunjukkan bahwa kabupaten atau kota yang berada pada cluster 2 memiliki penduduk dengan karakteristik sosial ekonomi yang terbelah lebih rendah. Sehingga dapat disimpulkan bahwa kabupaten atau kota pada cluster 2 memiliki tingkat kesehatan yang lebih rendah, dan membutuhkan bantuan pemerintah untuk meningkatkannya.

Kualitas karakter yang dihasilkan oleh metode AHC dan DBSCAN ini diukur dengan *Silhouette score*. Dimana metode AHC berhasil menghasilkan 2 cluster yang signifikan, dengan *silhouette score* terbesar yang dihasilkan adalah 0,4990. Sedangkan metode DBSCAN memiliki *silhouette score* tertinggi sebesar 0,4822. Dengan demikian dapat dikatakan bahwa metode AHC lebih unggul untuk digunakan dalam analisis penggerombolan dibandingkan dengan metode DBSCAN.

Daftar Pustaka

- Agarwal, S. 2014. Data mining: Data mining concepts and techniques. In Proceedings -2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013. <https://doi.org/10.1109/ICMIRA.2013.45>
- Anderberg, Michael R. Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks. Vol. 19. Academic press, 2014.
- Dalimunthe S., Hanafiah A. 2021. Implementation of Agglomerative Hierarchical Clustering Based on The Classification of Food Ingredients Content of Nutritional Substances. IT Journal Research and Development. 6(1): 60-69.
- Deng, D. (2020). DBSCAN Clustering Algorithm Based on Density. 2020 7th International Forum on Electrical Engineering and Automation (IFEEA). doi:10.1109/ifeea51475.2020.00199.
- Hanafi, N, Saadatfar H. (2022). A fast DBSCAN algorithm for big data based on efficient density calculation, Expert Systems with Applications, Volume 203, 117501, <https://doi.org/10.1016/j.eswa.2022.117501>.
- Henderi, H. 2021. Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. IJIS: International Journal of Informatics and Information Systems, 4(1), 13–20. <https://doi.org/10.47738/ijis.v4i1.73>
- Jarman, AM. 2020. Hierarchical cluster analysis: Comparison of single linkage, complete linkage, average linkage and centroid linkage method. Georgia Southern University, 29.
- Kementerian Kesehatan Republik Indonesia. (2022). Profil kesehatan Indonesia 2022. Kementerian Kesehatan Republik Indonesia.
- Kranzusch, R., Wiesemann, S., Zange, L., Jeuthe, S., Ferreira, T., Kuehne, T., Pieske, B., Tillmanns, C., Friedrich, M. G., Schulz-menger, J., & Messroghli, D. R. 2020. Z-score mapping for standardized analysis and reporting of cardiovascular magnetic resonance

modified Look-Locker inversion recovery (MOLLI) T1 data : Normal behavior and validation in patients with amyloidosis. *Journal of Cardiovascular Magnetic Resonance*, 7, 1–10.

Liu, N., Xu, Z., Zeng, X.-J., & Ren, P. (2021). An agglomerative hierarchical clustering algorithm for linear ordinal rankings. *Information Sciences*, 557, 170–193. doi:10.1016/j.ins.2020.12.056.

Nevil, S. 2023. How to Calculate Z-Score and Its Meaning. <https://www.investopedia.com/terms/z/zscore.asp>.

Pratama, IWP. 2023. Standarisasi Z-Score sebagai Pendekatan Alternatif dalam Evaluasi Prestasi Akademik Mahasiswa: Studi Kasus di Politeknik eLBajo Commodus. *Jurnal Penelitian Terapan Mahasiswa*. 1(2): 77-85.

Rakasiwi LS, Kautsar A. (2021). Pengaruh Faktor Demografi dan Sosial Ekonomi terhadap Status Kesehatan Individu di Indonesia. *Kajian Ekonomi Keuangan*. Vol 5. <http://dx.doi.org/10.31685/kek.V5.2.1008>.

Suhirman, Wintolo H. 2019. SYSTEM FOR DETERMINING PUBLIC HEALTH LEVEL USING THE AGGLOMERATIVE HIERARCHICAL CLUSTERING METHOD. *Compiler*. 8(1): 95-104.

Sunarko, GS., Wasino, & Sutrisno T. 2023. Klasterisasi sentimen ulasan pengguna aplikasi bca mobile pada platform google play store dengan algoritma k-means clustering. *Jurnal Ilmu Komputer Dan Sistem Informasi*, 11(1): 1-6.

Yulianti DI., Hermanto TI., Defriani M. 2023. Analisis Clustering Donor Darah dengan Metode Agglomerative Hierarchical Clustering. *Rekayasa Teknik Informatika dan Informasi*. 3(6): 285–290.

Zhang M. 2019. Use Density-Based Spatial Clustering of Applications with Noise (DBSCAN) Algorithm to Identify Galaxy Cluster Members. *IOP Publishing*. 252(4): 1755-1315.