



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

## **Kelompok 4 P1:**

Tiara Ayu Pertiwi

G1401201012

Raziqizzan Putrandi

G1401211040

Azzahra Adelia Putri

G1401211045

Windi Gunawan

G1401211082

Ubaidillah Al Hakim

G1401211086



# POKOK BAHASAN

.....



**Latar Belakang**

**Tujuan**

**Tinjauan Pustaka**

**Metodologi**

**Hasil dan Pembahasan**

**Kesimpulan**

## Latar Belakang



Analisis gerombol (*cluster*) digunakan untuk mengelompokkan objek berdasarkan algoritma tertentu yang memiliki kesamaan. Metode analisis gerombol diantaranya AHC, K-Mean, dan Autoencoder.

AHC

***Agglomerative hierarchical cluster*** tidak memerlukan partisi awal, tidak menghasilkan *noise*, mudah, deskripsi dan visualisasi informatif, kompleksitas tinggi.

Algoritma gerombol yang berdasarkan kepadatan data yaitu DBSCAN, unggul menduga gerombol acak, mendeteksi dan mengabaikan *noise*.



Pada data yang kepadatannya tidak seragam, menghasilkan homogenitas dalam kelompok serta perbedaan kelompok yang tinggi. Kelemahannya waktu relatif lama untuk data besar.

## Latar Belakang

Kesehatan merupakan unsur penting negara, didukung PBB (dalam program **SDGs**) salah satunya kesehatan merata.

Perawat  
**20%**

Dr. Gigi  
**24,4%**

Laporan Kementerian Kesehatan (2022), Jawa Barat kekurangan 20% perawat dan 24,4% dokter gigi. Memiliki infrastruktur, tenaga medis, dan fasilitas penunjang yang berada di atas rata-rata.



Pemerataan kesehatan dipengaruhi oleh wilayah tinggal. Oleh karena itu, penggerombolan perlu dilakukan untuk mengetahui daerah yang memerlukan peningkatan bidang kesehatan.

## Tujuan dan Data



### Tujuan

Membandingkan dua metode penggerombolan *unsupervised learning* yaitu antara *Agglomerative Hierarchical Cluster* (AHC) dan *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN), serta melakukan interpretasi terhadap gerombol yang telah ada.

## 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).

## Agglomerative Hierarchical Clustering

Agglomerative hierarchical cluster adalah sebuah metode pengelompokan hierarki 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).

## 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).

## Koefisien Silhouette

Koefisien Silhouette merupakan metode evaluasi yang digunakan untuk menilai kualitas klaster yang dihasilkan oleh algoritma clustering. Ini mengukur seberapa baik setiap data sesuai dengan klaster tempatnya dibandingkan dengan klaster lain. Rentang nilai Silhouette adalah dari -1 hingga +1, dimana nilai yang lebih tinggi menandakan kesesuaian data dengan klasternya sendiri dan ketidaksesuaian dengan klaster lain. Nilai Silhouette mendekati +1 menunjukkan pembentukan klaster yang baik, sedangkan nilai negatif menandakan kemungkinan kesalahan penempatan data dalam klaster. Nilai Silhouette sekitar 0 menunjukkan posisi data yang berada di antara dua klaster. Metrik ini membantu dalam menilai seberapa baik klaster telah terbentuk dan sejauh mana batas antara klaster yang jelas (Dalimunthe 2021).

# METODOLOGI - 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.

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)

# METODOLOGI - BAHAN DAN DATA

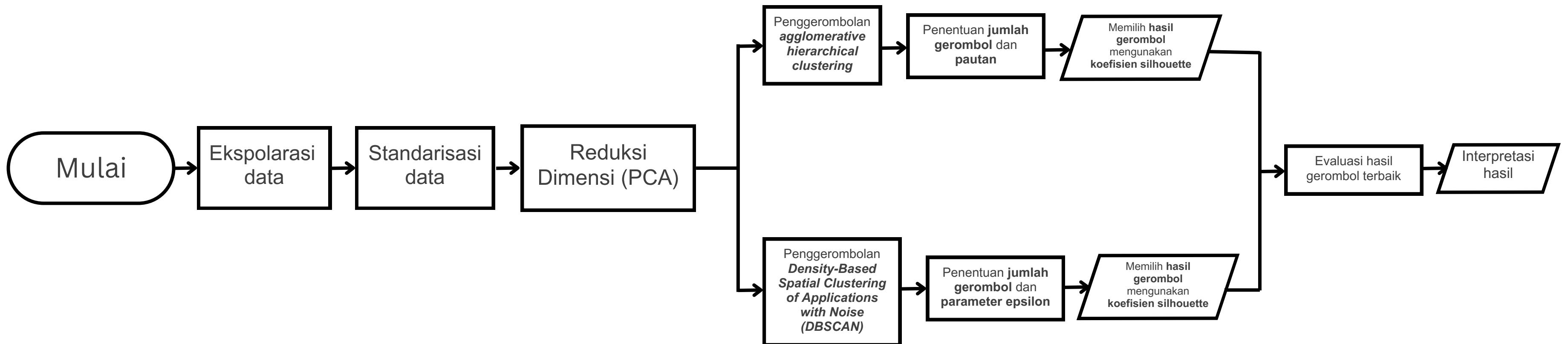


Peubah	Keterangan	Tipe peubah	Sumber
x6	Jumlah Bidan	Numerik	Kemenkes (2022)
x7	Jumlah Balita Ditimbang	Numerik	Kemenkes (2022)
x8	Persentase Penerima BPJS Kesehatan Penerima Bantuan Iuran	Numerik	Kemenkes (2022)
x9	Persentase Penerima BPJS Kesehatan Non-Penerima Bantuan Iuran	Numerik	Kemenkes (2022)
x10	Persentase Penerima Jamkesda, Asuransi Swasta, Perusahaan	Numerik	Kemenkes (2022)

# METODOLOGI - PROSEDUR PENELITIAN

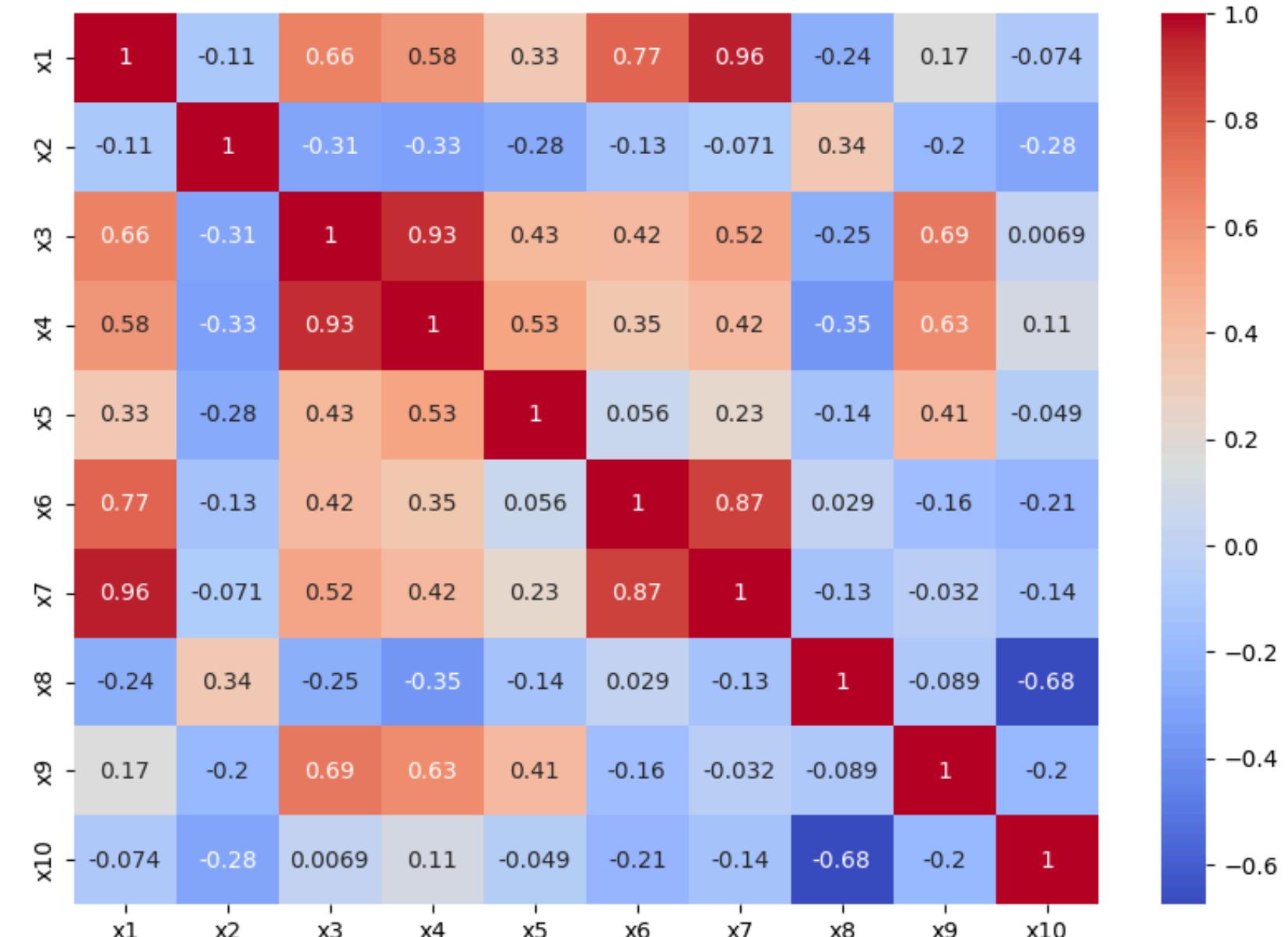
.....

## Diagram alir

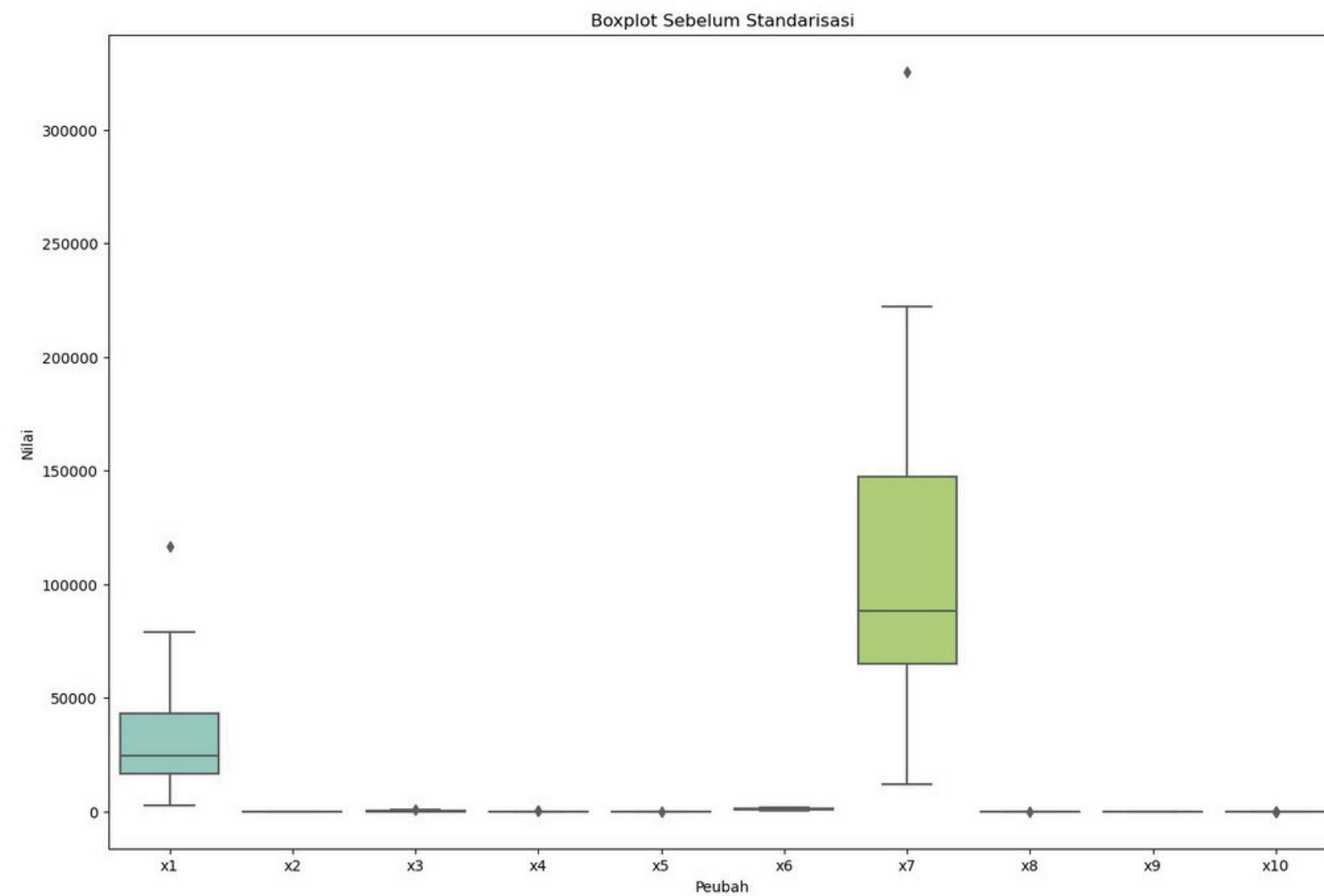


# HASIL DAN PEMBAHASAN- EKSPLORASI

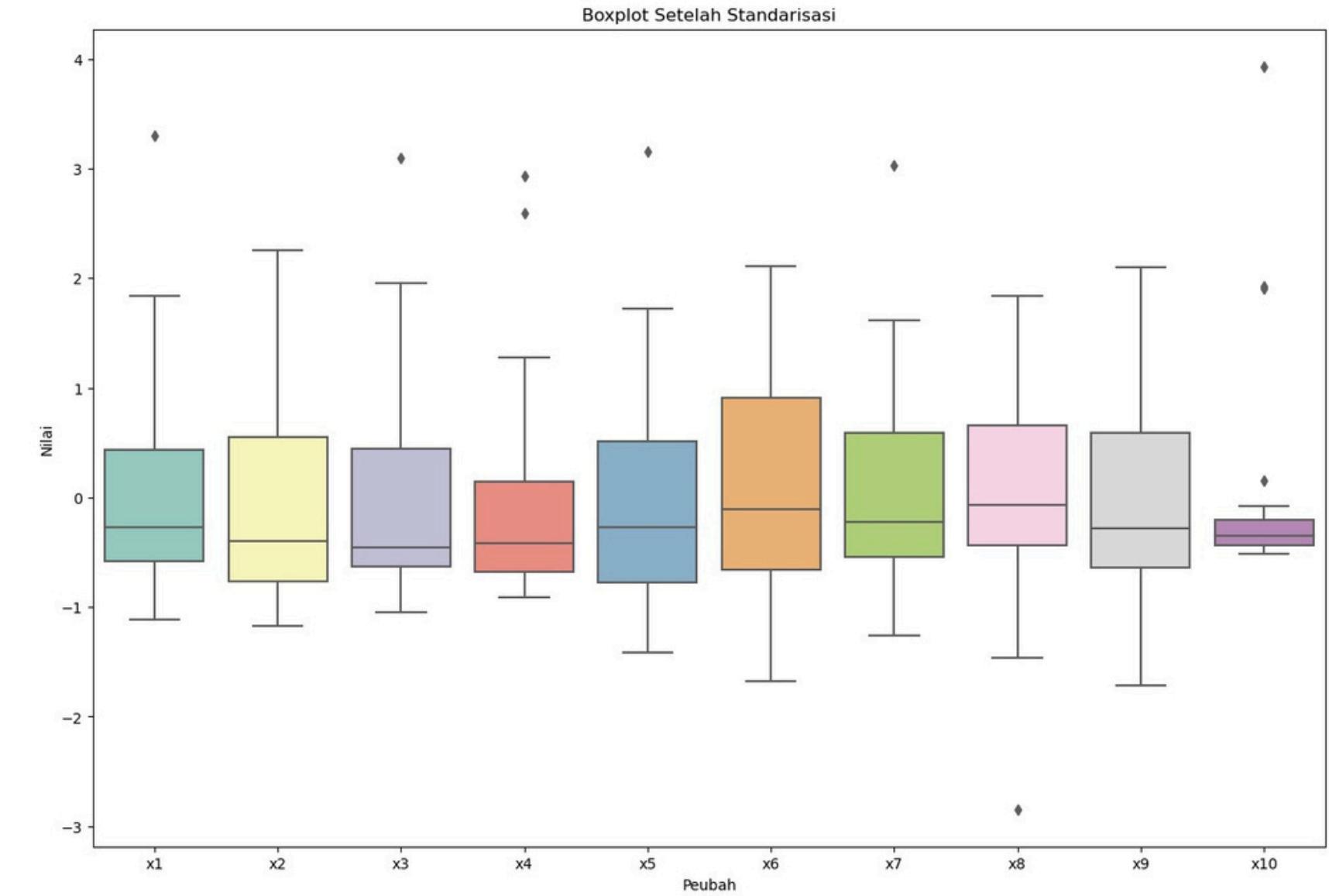
Beberapa peubah memiliki korelasi yang tinggi antara satu sama lainnya. Dua pasang peubah memiliki korelasi mendekati satu, yaitu  $x_1$  dengan  $x_7$  serta  $x_3$  dengan  $x_4$ . Sehingga diperlukan pereduksian dimensi untuk mengatasi permasalahan tersebut.



# HASIL DAN PEMBAHASAN-EKSPLORASI



## Sebelum dilakukan Standarisasi



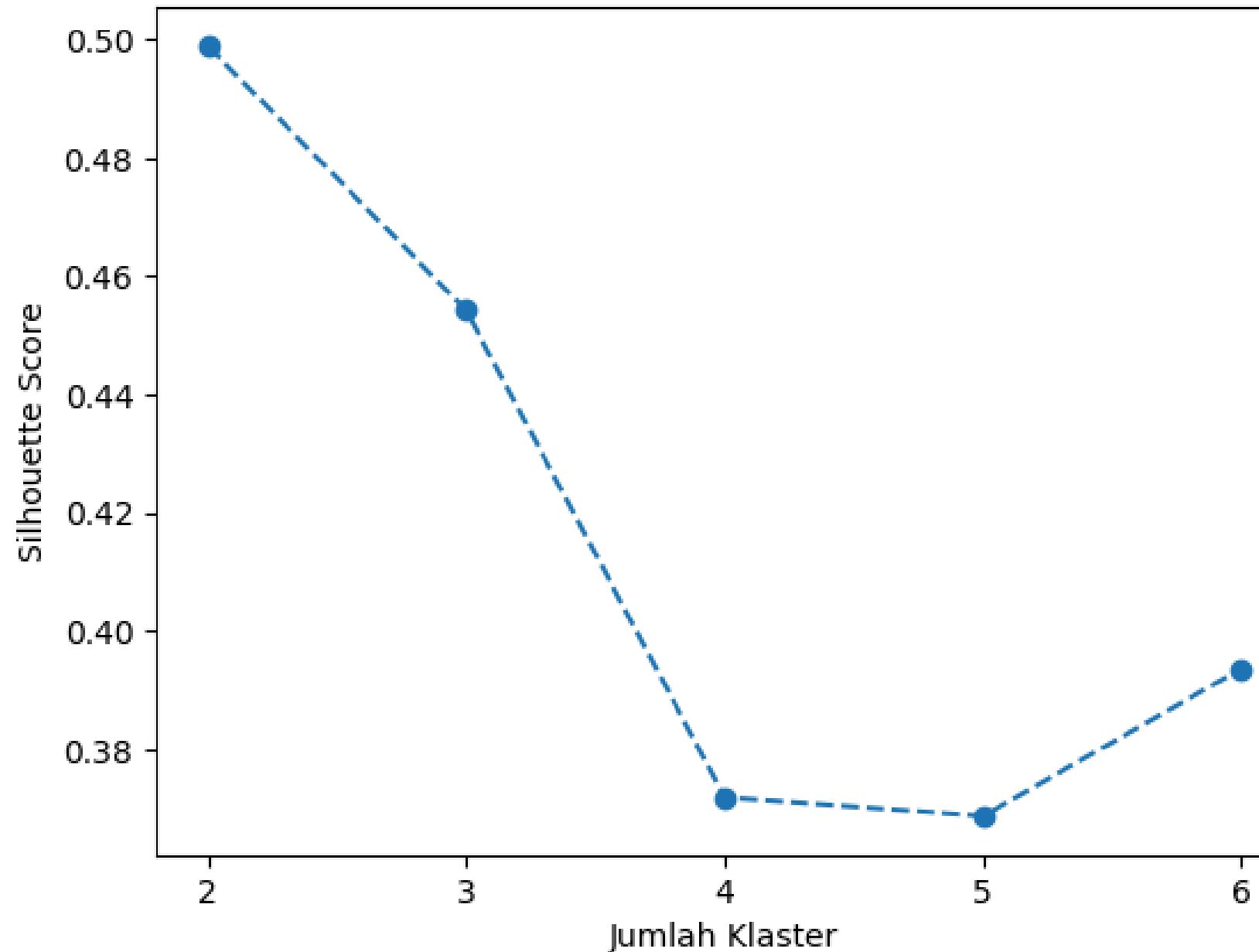
Setelah dilakukan Standarisasi

# HASIL DAN PEMBAHASAN-AHC

.....

## Elbow Plot

Elbow Plot untuk Menentukan Jumlah Klaster Optimal Average Linkage



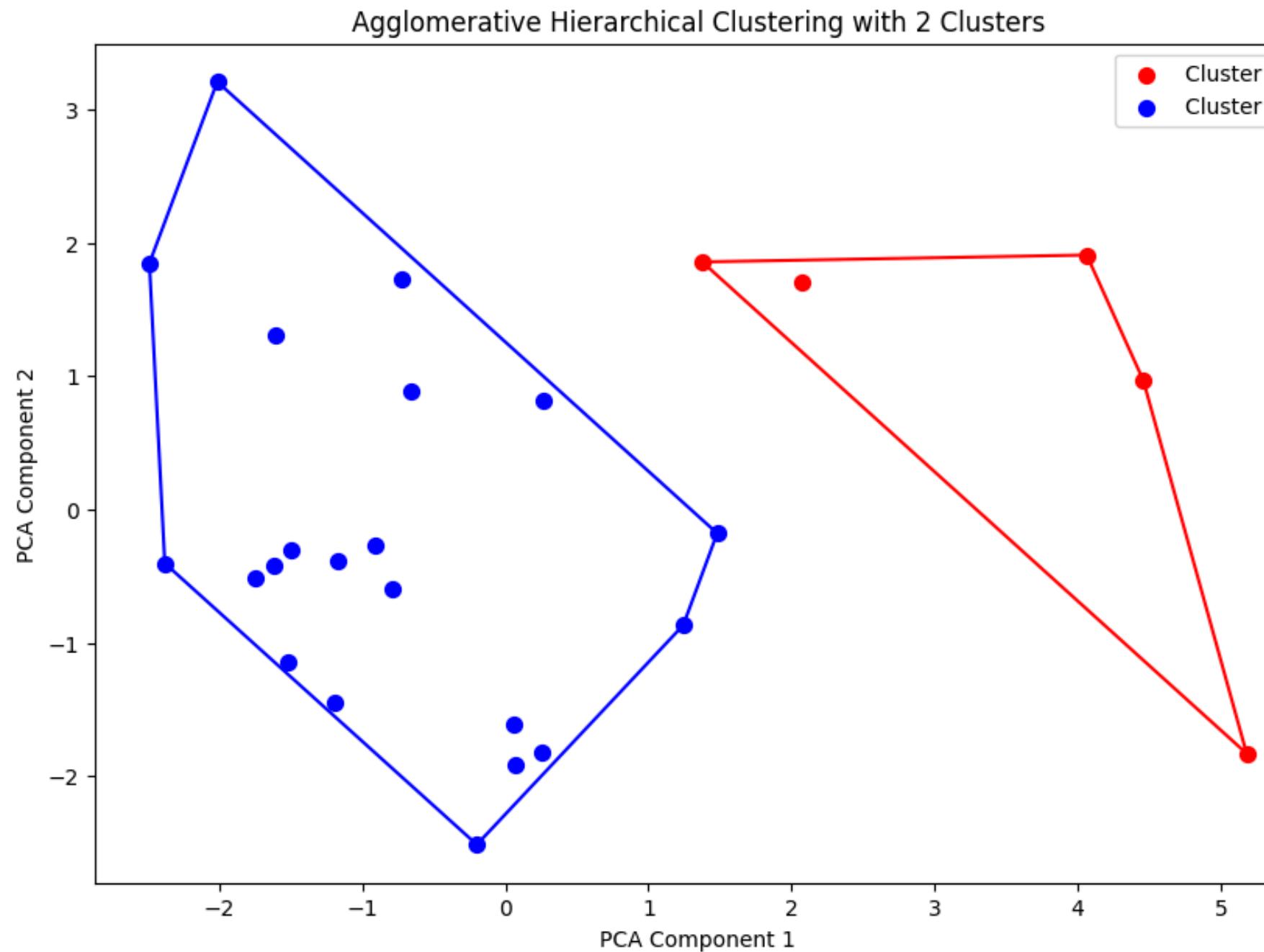
Berdasarkan Gambar di samping, terlihat bahwa jumlah cluster dengan Silhouette Score terbesar adalah 2. **Cluster optimal** yang digunakan adalah **2**.



# HASIL DAN PEMBAHASAN-AHC

.....

## *Cluster Plot*



Gambar di samping menunjukkan 2 cluster yang terbentuk dengan jelas. **Cluster pertama** terdiri dari **6 kabupaten/kota**, sedangkan cluster kedua terdiri dari **21 kabupaten/kota**.



# HASIL DAN PEMBAHASAN-AHC

.....

## 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



# HASIL DAN PEMBAHASAN-AHC

.....

## Rata Rata Nilai Tiap Cluster

<i>Cluster</i>	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

- *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.



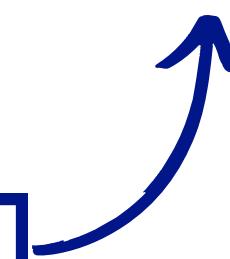
# HASIL DAN PEMBAHASAN-DBSCAN

.....

## Tuning Parameter (epsilon)

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

Optimal



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**.

# HASIL DAN PEMBAHASAN-DBSCAN

.....

## 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



# HASIL DAN PEMBAHASAN-DBSCAN

.....

## 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

### Cluster 1

Kabupaten/kota  
yang memiliki  
indikator **kesehatan**  
**yang baik**

### Cluster 2

Kabupaten/kota  
yang memiliki  
indikator **kesehatan**  
**yang kurang baik**

### Noise

Kabupaten/kota  
yang memiliki  
indikator **kesehatan**  
**yang paling beda**



# HASIL DAN PEMBAHASAN-DBSCAN

.....

## Evaluasi Model

### *Agglomerative Hierarchical Clustering*

Silhouette score sebesar

**0,4990**

### *DBSCAN*

Silhouette score sebesar

**0,4822**

Metode Agglomerative Hierarchical Clustering lebih baik dibanding DBSCAN

Selisih **0,0168**



# KESIMPULAN

.....

## AHC

- Cluster 1 : 6 kab/kota
- Cluster 2 : 21 kab/kota

## DBSCAN

- Cluster 1 : 2 kab/kota
- Cluster 2 : 23 kab/kota
- 1 noise (pencilan)

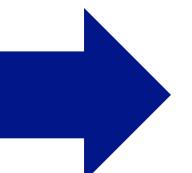
cluster  
1

Lebih tinggi : x1, x3, x4, x5, x6, x7, x9, dan x10.  
Karakteristik sosial ekonomi yang lebih kuat,  
menggambarkan tingkat kesehatan yang baik

cluster  
2

Lebih tinggi pada indikator x2 dan x8. Banyak  
penduduk dengan tingkat sosial ekonomi yang lebih  
rendah, menggambarkan tingkat kesehatan  
yang kurang baik dan membutuhkan bantuan

*Silhouette Score*  
AHC : 0,4990  
DBSCAN : 0,4882



**metode analisis penggerombolan  
AHC lebih unggul dibandingkan  
dengan 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. IJIIS: International Journal of Informatics and Information Systems, 4(1), 13–20. <https://doi.org/10.47738/ijiis.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.



# DAFTAR PUSTAKA

.....

- 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.
- 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>.
- Nevil, S. 2023. How to Calculate Z-Score and Its Meaning. <https://www.investopedia.com/terms/z/zscore.asp>.
- Suhirman, Wintolo H. 2019. SYSTEM FOR DETERMINING PUBLIC HEALTH LEVEL USING THE AGGLOMERATIVE HIERARCHICAL CLUSTERING METHOD. *Compiler*. 8(1): 95-104.
- 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.



# TERIMA KASIH



**IPB University**  
— Bogor Indonesia —

Department of Statistics  
Jl. Meranti W22 L4  
Kampus IPB Dramaga Bogor 16680  
Telp.: 0251-8624535  
E-mail: [statistika@apps.ipb.ac.id](mailto:statistika@apps.ipb.ac.id)



# IPB University

— Bogor Indonesia —

Inspiring Innovation with Integrity  
in Agriculture, Ocean and Biosciences for a Sustainable World