

Evaluasi *K-Means* dan *Hierarchical Clustering* dalam Segmentasi Wilayah Penerimaan Bantuan Sosial Pangan di Provinsi X

DOI: <http://dx.doi.org/10.35889/jutisi.v14i2.2672>

Creative Commons License 4.0 (CC BY – NC)



Fathoni¹, Annisa Khairani^{2*}, Risma Nur'Aini³, Gina Destia Gultom⁴,
 Intan Aidita Alfitriah⁵, Ali Ibrahim⁶

Sistem Informasi, Universitas Sriwijaya, Palembang, Indonesia
 *e-mail Corresponding Author: 09031282227133@student.unsri.ac.id

Abstract

The allocation of food social assistance in Indonesia faces challenges in targeting accuracy and distribution effectiveness. To improve distribution efficiency, this study explores regional segmentation using two clustering methods K-Means and Hierarchical Clustering based on social and economic characteristics. The analysis uses 324 secondary records from the Satu Data Indonesia portal, categorized by regency and city in Province X. Clustering performance was evaluated using the Silhouette Coefficient and Davies-Bouldin Index. Results show that K-Means Clustering outperforms Hierarchical Clustering, achieving a Silhouette Coefficient of 0.5371 and a Davies-Bouldin Index of 0.7173 with five clusters. In contrast, Hierarchical Clustering produced a Silhouette Coefficient of 0.4976 and a Davies-Bouldin Index of 0.7607. Based on these findings, K-Means is recommended for more effective regional segmentation in the distribution of food social assistance

Keywords: Regional Segmentation; Food Social Assistance; K-Means; Hierarchical Clustering

Abstrak

Alokasi bantuan sosial pangan di Indonesia masih menghadapi kendala ketepatan sasaran dan efektivitas penyaluran. Salah satu pendekatan untuk meningkatkan efisiensi penyaluran adalah dengan melakukan segmentasi wilayah penerima bantuan sosial pangan berdasarkan atribut sosial dan ekonomi. Penelitian ini bertujuan untuk mengkaji dua teknik klasterisasi, yaitu *K-Means* dan *Hierarchical Clustering*, untuk melakukan segmentasi wilayah penerima bantuan sosial pangan di Provinsi X. Data yang digunakan adalah data sekunder yang bersumber dari portal Satu Data Indonesia sebanyak 324 *record* yang dikelompokkan berdasarkan kabupaten dan kota. Evaluasi kinerja klasterisasi dilakukan dengan menggunakan dua metrik, yaitu *Silhouette Coefficient* dan *Davies-Bouldin Index*. Hasil penelitian menunjukkan bahwa pendekatan *K-Means Clustering* menghasilkan segmentasi wilayah yang unggul, ditunjukkan dengan nilai *Silhouette Coefficient* sebesar 0,5371 dan *Davies-Bouldin Index* sebesar 0,7173 untuk lima klaster. Pendekatan *Hierarchical Clustering* menghasilkan nilai *Silhouette Coefficient* sebesar 0,4976 dan *Davies-Bouldin Index* sebesar 0,7607. Dengan demikian, metode *K-Means* direkomendasikan untuk menggambarkan wilayah dalam distribusi bantuan sosial pangan.

Kata kunci: Segmentasi Wilayah; Bantuan Sosial Pangan; K-Means; Hierarchical Clustering

1. Pendahuluan

Teknik *clustering* semakin banyak diterapkan dalam analisis data sosial akhir-akhir ini, terutama dalam hal distribusi sosial. Tren ini muncul seiring meningkatnya tuntutan untuk menyalurkan bantuan secara lebih tepat dan efisien, khususnya di wilayah yang memiliki tingkat kemiskinan tinggi. Salah satu metode yang sering dimanfaatkan adalah metode *K-Means Clustering* karena mampu mengenali pola penyebaran penerima Bantuan Sosial Tunai (BST) dengan mempertimbangkan letak geografis dari setiap keluarga penerima [1]. Di sisi lain, pemilihan algoritma dan metrik yang tepat dalam *Hierarchical Clustering* juga memegang

peranan penting dalam meningkatkan kualitas pengelompokan. Dengan pendekatan yang sesuai, struktur dan hubungan antar kelompok data dapat dipahami secara lebih mendalam, sehingga hasil analisis dapat dimanfaatkan dalam pengambilan keputusan berbasis wilayah secara lebih akurat [2]. Oleh karena itu, penggunaan *K-Means* dan *Hierarchical Clustering* dalam menganalisis distribusi bantuan sosial pangan menjadi semakin relevan untuk meningkatkan efektivitas program bantuan pemerintah.

Meskipun metode *clustering* tersebar luas di berbagai bidang, tantangan terbesar dalam mengelompokkan penerima bantuan sosial adalah menentukan metode *clustering* yang paling akurat dan efisien. Permasalahan yang sering muncul adalah rendahnya akurasi pengelompokan ketika menangani data dari wilayah dengan jumlah penerima bantuan yang sangat beragam. Kecenderungan *K-Means Clustering* untuk menghasilkan kluster berukuran seragam menjadikannya kurang optimal saat digunakan pada data dengan distribusi yang tidak merata [3]. Sebaliknya, *Hierarchical Clustering* lebih mampu mengidentifikasi hubungan antar wilayah dengan lebih baik, tetapi memiliki keterbatasan dalam hal efisiensi komputasi saat diterapkan pada data skala besar [4].

Sebagian besar penelitian sebelumnya berfokus pada penggunaan satu metode *clustering* dalam menganalisis distribusi bantuan sosial. Dalam praktiknya, *K-Means Clustering* kerap digunakan tanpa memperhitungkan kompleksitasnya hubungan antar penerima bantuan [5]. Sebagai alternatif, analisis ini mengombinasikan dua pendekatan pengelompokan, yaitu *K-Means* dan *Hierarchical Clustering*, guna memperoleh pemahaman yang lebih mendalam terhadap pola distribusi bantuan sosial pangan. Selain itu, teknik penilaian berbasis *Silhouette Score* yang jarang digunakan dalam beberapa penelitian sebelumnya untuk menilai efektivitas hasil pengelompokan [6].

Penelitian ini bertujuan untuk menerapkan serta membandingkan penggunaan metode *K-Means* dan *Hierarchical Clustering* dalam proses pengelompokan wilayah kabupaten/kota, dengan mengacu pada jumlah penerima bantuan sosial pangan di Provinsi X. Dengan mengevaluasi dua metode *clustering* tersebut, penelitian ini diharapkan menemukan metode yang lebih akurat dan efektif untuk mendukung distribusi bantuan sosial berdasarkan analisis data. Selain itu, hasil penelitian ini dapat berupa referensi bagi lembaga pemerintah maupun organisasi sosial dalam merumuskan strategi distribusi yang lebih optimal dan tepat sasaran, sehingga berkontribusi dalam meningkatkan kualitas kebijakan sosial dalam manajemen dan distribusi bantuan sosial pangan terutama di tingkat daerah.

2. Tinjauan Pustaka

Bagian ini membahas penelitian-penelitian sebelumnya yang berfokus pada penerapan algoritma *K-Means* dan *Hierarchical Clustering* dalam segmentasi wilayah penerima bantuan sosial pangan. Sejumlah penelitian sebelumnya telah meneliti pemanfaatan metode *K-Means* untuk melakukan klasifikasi terhadap daerah-daerah penerima bantuan sosial. Salah satunya adalah penelitian yang menerapkan algoritma *K-Means* dalam pengelompokan penerima bantuan di wilayah Kota dan Kabupaten Cirebon [7]. Pengelompokan dilakukan dengan mempertimbangkan parameter seperti jarak distribusi, jumlah kelurahan, dan usia penerima yang menghasilkan tiga kluster. Hasil pengelompokan ini memiliki skor DBI senilai 0,505 yang menunjukkan kualitas segmentasi cukup baik. Salah satu kluster mencerminkan wilayah dengan distribusi bantuan lebih cepat dan kepadatan penduduk tinggi. Selanjutnya pendekatan berbasis *RapidMiner* digunakan untuk mengklasifikasikan prioritas penerima dana desa [8]. Segmentasi dilakukan dengan mempertimbangkan kondisi ekonomi dan sosial yang menunjukkan bahwa metode *clustering* berhasil mengurangi kesalahan dalam penentuan penerima bantuan. Penelitian ini menunjukkan kontribusi nyata dari *data mining* terhadap pengambilan keputusan berbasis data.

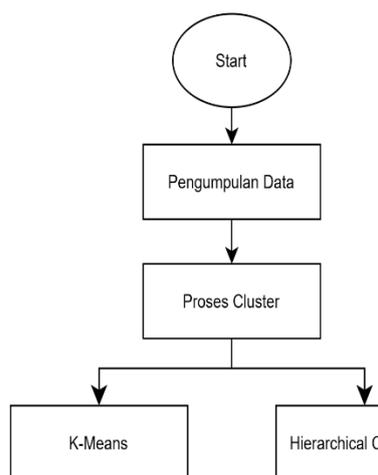
Selain penggunaan *K-Means*, beberapa peneliti juga melakukan evaluasi performa dengan metode lain. Untuk mengelompokkan wilayah yang berisiko mengalami stunting, dilakukan analisis menggunakan parameter seperti tingkat imunisasi, proporsi balita stunting, cakupan pemberian ASI eksklusif, serta akses terhadap fasilitas sanitasi dalam mengevaluasi performa metode *K-Means* dan *Hierarchical Clustering*. Metode *K-Means* menghasilkan skor *Silhouette Coefficient* sebesar 0,48, yang menunjukkan performa yang lebih baik dibandingkan dengan metode *Hierarchical Clustering* [9]. Selanjutnya, segmentasi penerima Bantuan Langsung Tunai (BLT) Dana Desa dengan memanfaatkan algoritma *K-Means*, yang membagi kelompok penerima bantuan ke dalam tiga klasifikasi utama berdasarkan kondisi ekonomi dan

tingkat kesejahteraan, di mana performa terbaik diperoleh saat menggunakan $K=3$ dengan skor *Davies-Bouldin Index* sebesar 0,738 [10]. Sementara itu, perbandingan antara *K-Means* dan *Hierarchical Clustering* untuk klasifikasi penerima bantuan sosial di 34 provinsi menunjukkan bahwa *K-Means* lebih efektif dalam membagi provinsi menjadi dua kluster: provinsi dengan tingkat kebutuhan tinggi dan provinsi dengan tingkat kebutuhan rendah [11].

Secara komprehensif tinjauan pustaka diatas menunjukkan bahwa *K-Means* banyak digunakan dan cenderung memberikan performa lebih stabil dibandingkan *Hierarchical Clustering*, terutama dalam konteks klasifikasi berbasis parameter sosial dan ekonomi. Namun demikian, sebagian besar penelitian masih berfokus pada hasil akhir kluster tanpa mengeksplorasi secara mendalam bagaimana algoritma berinteraksi dengan kondisi data yang kompleks dan variabel dinamis. Selain itu, evaluasi terhadap efektivitas algoritma dalam konteks distribusi bantuan sosial berskala besar, seperti lintas provinsi atau berbasis sistem informasi masih terbatas. Oleh karena itu, penelitian ini mengusulkan pendekatan yang lebih mendalam dengan mengevaluasi performa kedua metode *clustering* dalam kondisi data yang lebih dinamis dan variabel yang beragam. Selain itu, penelitian ini akan menganalisis bagaimana algoritma *clustering* dapat berkontribusi dalam meningkatkan efisiensi dan ketepatan distribusi bantuan sosial di Provinsi X. Dengan demikian, penelitian ini tidak hanya berkontribusi pada pengembangan metode segmentasi berbasis *clustering*, tetapi juga memberikan wawasan baru bagi pengambil kebijakan dalam meningkatkan efektivitas program bantuan sosial.

3. Metodologi

Penelitian ini dilakukan melalui tiga tahapan utama sebagaimana ditunjukkan pada Gambar 1, yakni pengumpulan data, pelaksanaan proses klusterisasi diikuti menggunakan *K-Means* dan *Hierarchical Clustering*, serta tahap akhir berupa evaluasi hasil kluster yang dihasilkan.



Gambar 1. Tahapan Penelitian

Adapun untuk penelitian ini dimulai dengan mengumpulkan data sekunder yang diperoleh dari portal Satu Data Indonesia, yaitu dataset "Jumlah Penerima Bantuan Sosial Pangan Berdasarkan Kabupaten/Kota di Provinsi X", yang dihimpun dari beberapa instansi pemerintah sebagai penyedia data. Parameter yang dianalisis mencakup jumlah penerima bantuan sosial pangan di berbagai wilayah, cakupan bantuan yang diterima, serta sejumlah faktor pendukung lainnya yang memengaruhi distribusi bantuan sosial pangan. Setelah data berhasil dikumpulkan, proses pengelompokan wilayah dilakukan menggunakan dua pendekatan klusterisasi, yakni *K-Means* dan *Hierarchical Clustering*, berdasarkan karakteristik penerima bantuan sosial. Kualitas segmentasi hasil klusterisasi kemudian dianalisis menggunakan dua indikator evaluasi, yaitu *Silhouette Coefficient*, dan *Davies-Bouldin Index*.

1) *K-Means Clustering*

Metode klusterisasi *K-Means Clustering* cocok digunakan dalam situasi dengan jumlah item yang besar. Algoritma *K-Means* merupakan teknik klusterisasi yang bekerja secara iteratif,

mudah diimplementasikan, dan populer karena menggunakan konsep jarak sebagai dasar pengukurannya [12].

Langkah-langkah dalam menerapkan algoritma *K-Means* meliputi beberapa tahapan sebagai berikut [13]:

- a. Menetapkan jumlah kluster (k) yang akan dibentuk.
- b. Menentukan secara acak titik pusat awal (centroid) untuk tiap kluster.
- c. Mengelompokkan data ke kluster terdekat berdasarkan jarak ke centroid. Perhitungan jarak tersebut dapat dilakukan menggunakan metode Euclidean, yang dirumuskan sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan:

$d(x, y)$ = jarak antara data x dan pusat kluster y

x_i = nilai data x pada observasi ke- i

y_i = nilai pusat kluster y pada observasi ke- i

n = jumlah total observasi

- d. Mengklasifikasikan setiap data ke dalam kluster yang memiliki jarak paling dekat dengan data tersebut.
- e. Menentukan nilai centroid dengan menghitung nilai rata-rata kluster dari setiap anggota kluster dengan persamaan:

$$\text{Centroid} = \sum \frac{a_i}{n} \quad (2)$$

Keterangan:

a_i = Nilai keanggotaan data pada kluster

n = Jumlah total anggota dalam kluster

2) Hierarchical Clustering

Hierarchical Clustering merupakan teknik pengelompokan yang Menyusun struktur kluster secara berjenjang, menggunakan pendekatan *agglomerative (bottom-up)* maupun *divisive (top-down)*.

Agglomerative Hierarchical Clustering merupakan salah satu pendekatan yang paling umum digunakan, di mana prosesnya dimulai dengan menganggap setiap titik data sebagai kluster terpisah. Selanjutnya, kluster-kluster yang paling mirip akan digabungkan secara bertahap hingga seluruh data membentuk satu kluster besar.

Langkah-langkah dalam *Agglomerative Hierarchical Clustering* [14]:

- a. Hitung matriks jarak antar titik data, misalnya dengan *Euclidean Distance*.
- b. Gabungkan dua kluster terdekat berdasarkan metode linkage yang dipilih:
 - *Single Linkage*: Jarak minimum antar titik dari dua kluster.
 - *Complete Linkage*: Jarak maksimum antar titik dari dua kluster.
 - *Average Linkage*: Rata-rata jarak antar titik dari dua kluster.
 - *Ward's Method*: Menggabungkan kluster dengan meminimalkan variansi dalam kluster.
- c. Perbarui matriks jarak setelah penggabungan kluster.
- d. Lanjutkan proses penggabungan secara bertahap hingga seluruh data tergabung dalam satu kluster besar, atau hingga jumlah kluster yang diinginkan tercapai.
- e. Gunakan dendrogram untuk mengidentifikasi jumlah kluster yang optimal dengan mengamati titik pemotongan yang paling sesuai secara visual.

3) Silhouette Coefficient

Silhouette Coefficient mengukur sejauh mana hasil pengelompokan data membentuk kluster yang kohesif dan terpisah dengan baik. Indikator ini menunjukkan tingkat kemiripan suatu titik data terhadap kluster tempatnya berada dibandingkan dengan kluster lain di sekitarnya [15]. Rumusnya adalah:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

Keterangan:

$S(i)$ = Nilai *Silhouette Coefficient* untuk titik data- i

$a(i)$ = Rata-rata jarak antara titik data ke- i terhadap titik lainnya dalam klaster yang sama (*intra-cluster distance*)

$b(i)$ = Rata-rata jarak antara titik data ke- i terhadap titik-titik pada klaster terdekat yang berbeda (*inter-cluster distance*)

Nilai *Silhouette Coefficient* berkisar antara -1 hingga 1:

- Mendekati 1 → Jika nilainya mendekati 1, maka klaster memiliki struktur yang baik karena titik lebih dekat ke klaster asalnya.
- Mendekati 0 → Jika nilainya mendekati 0, titik data berada di antara dua klaster, menunjukkan ambiguitas.
- Mendekati -1 → Jika nilainya mendekati -1, titik lebih mirip dengan klaster lain, menandakan pengelompokan yang kurang baik.

4) *Davies-Bouldin Index*

Davies-Bouldin Index digunakan untuk menilai kinerja model klasterisasi dengan melihat tingkat kemiripan antar anggota klaster dan perbedaan antara klaster-klaster tersebut. Nilai *Davies-Bouldin Index* yang paling optimal adalah nilai terkecil [16]. DBI dihitung dengan menggunakan rasio yang diperoleh melalui perhitungan yang dijelaskan dalam persamaan 4 berikut.

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{ij}) \quad (4)$$

Keterangan:

DBI = *Davies-Bouldin Index*

k = Jumlah klaster

R_{ij} = Rasio antar klaster i dan j

4. Hasil dan Pembahasan

1) Dataset

Penelitian ini memanfaatkan sebanyak 324 data *record* yang mencerminkan jumlah penerima bantuan sosial pangan pada level kabupaten/kota X. Setiap *record* data memuat informasi mengenai identitas wilayah seperti id, kode provinsi, kode serta nama kabupaten/kota, status data (apakah masih rencana atau telah direalisasikan) dan jumlah penerima bantuan. Rincian atribut yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Atribut Data

No	Kode	Atribut
1	XYZ ₁	ID
2	XYZ ₂	Kode_Provinsi
3	XYZ ₃	Nama_Provinsi
4	XYZ ₄	Kode_Kabupaten_Kota
5	XYZ ₅	Nama_Kabupaten_Kota
6	XYZ ₆	Penerima_Bantuan
7	XYZ ₇	Jumlah_Penerima_Manfaat
8	XYZ ₈	Satuan
9	XYZ ₉	Tahun

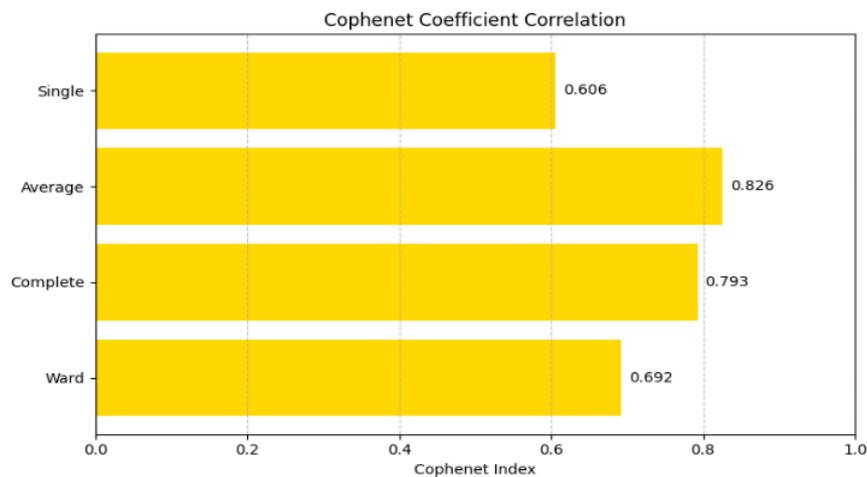
Untuk memberikan gambaran mengenai data yang dianalisis, sampel data disajikan pada Tabel 2.

Tabel 2. Sampel Data

XYZ ₁	XYZ ₂	XYZ ₃	XYZ ₄	XYZ ₅	XYZ ₆	XYZ ₇	XYZ ₈	XYZ ₉
1	32	Provinsi X	3201	Bogor	Rencana	171483	KPM	2018
2	32	Provinsi X	3202	Sukabumi	Rencana	163547	KPM	2018
3	32	Provinsi X	3203	Cianjur	Rencana	189959	KPM	2018
4	32	Provinsi X	3204	Bandung	Rencana	168246	KPM	2018
5	32	Provinsi X	3205	Garut	Rencana	171870	KPM	2018
6	32	Provinsi X	3206	Tasikmalaya	Rencana	125555	KPM	2018
7	32	Provinsi X	3207	Ciamis	Rencana	84940	KPM	2018
8	32	Provinsi X	3208	Kuningan	Rencana	82082	KPM	2018
9	32	Provinsi X	3209	Cirebon	Rencana	165213	KPM	2018
...								

2) Hasil Hierarchical Clustering

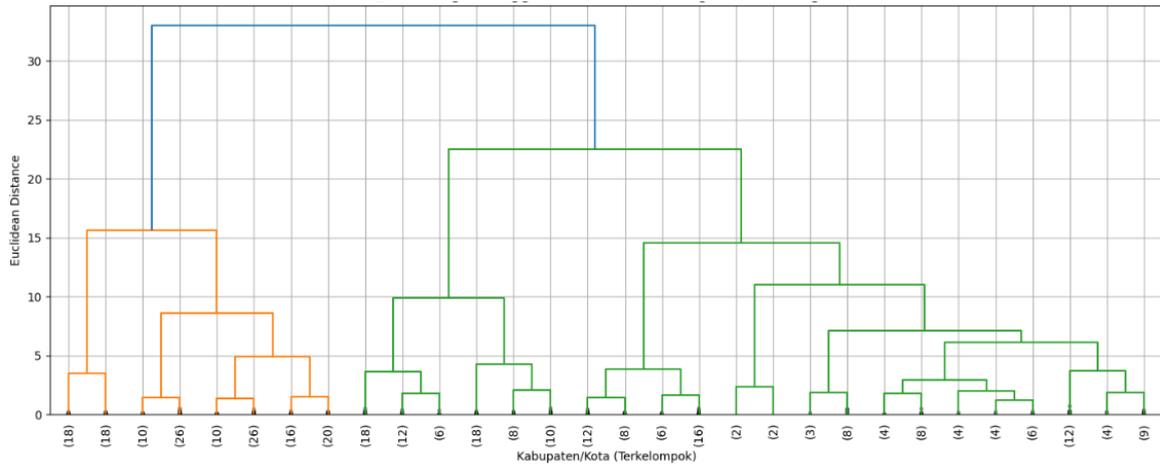
Dalam penelitian ini, proses pengelompokan terhadap penerima bantuan sosial pada Tingkat kabupaten/kota dilakukan menggunakan metode *Hierarchical Clustering* dengan pendekatan *Agglomerative Clustering*. Sebelum pembuatan dendogram dilakukan, terlebih dahulu dilakukan pemilihan jenis *linkage* yang paling sesuai. Nilai *cophenet coefficient correlation* dihitung untuk masing-masing jenis *linkage*, yaitu *single*, *average*, *complete*, dan *ward* yang hasilnya dijelaskan pada bagian berikut.



Gambar 2. Perbandingan Hasil Metode *Linkage*

Gambar 2 memperlihatkan hasil evaluasi terhadap beberapa jenis metode *linkage* yang diterapkan pada algoritma *Hierarchical Clustering*. Metode yang diuji meliputi *single linkage*, *complete linkage*, *average linkage*, *ward method*. Berdasarkan nilai *cophenet coefficient correlation*, metode *Average linkage* terbukti paling efektif dalam membentuk dendogram dengan nilai tertinggi yaitu 0.826.

Visualisasi dendogram pada Gambar 3 dihasilkan dari proses *Hierarchical Clustering* dengan pendekatan *average linkage*.



Gambar 3. Visualisasi Dendrogram *Hierarchical Clustering*

Tabel 3. DBI Hierarchical Clustering

Jumlah Kluster	<i>Davies-Bouldin Index</i>	<i>Silhouette Coefficient</i>
2	1.0293	0.4084
3	0.8743	0.4658
4	0.7607	0.4976
5	0.6945	0.4967

Tabel 3 menunjukkan bahwa jumlah kluster paling optimal adalah empat kluster, yang ditentukan melalui evaluasi nilai *Silhouette Coefficient*. Semakin tinggi nilai tersebut, maka semakin baik kualitas pembentukan kluster. Pembagian empat kluster tersebut dijelaskan lebih lanjut pada uraian berikut.

a. Kluster 1

Kluster ini mencakup wilayah dengan tingkat jumlah penerima bantuan yang relatif rendah. Wilayah-wilayah ini umumnya tersebar di daerah yang memiliki kepadatan penduduk yang rendah, serta cakupan bantuan yang merata namun tidak terlalu besar. Adapun daerah yang termasuk dalam kluster ini adalah Kabupaten Majalengka, Kabupaten Bandung Barat, Kabupaten Subang, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Sumedang.

b. Kluster 2

Kluster ini mencakup wilayah-wilayah dengan jumlah penerima bantuan menengah. Kluster ini mencakup kabupaten-kabupaten dengan jumlah penduduk dan luas wilayah yang sedang, sehingga distribusi bantuannya berada pada tingkat menengah. Adapun daerah yang termasuk dalam kluster ini antara lain: Kabupaten Karawang, Kabupaten Indramayu, Kabupaten Bogor, Kabupaten Garut.

c. Kluster 3

Kluster ini mencakup wilayah dengan tingkat penerima bantuan yang tinggi. Wilayah-wilayah ini cenderung memiliki populasi padat dan kebutuhan bantuan yang cukup tinggi, namun tidak mendominasi seperti kluster sebelumnya. Adapun daerah yang termasuk dalam kluster ini adalah Kabupaten Cianjur, Kota Tasikmalaya, Kota Sukabumi.

d. Kluster 4

Kluster ini mencakup wilayah yang memiliki jumlah penerima bantuan yang sangat rendah, yang disebabkan oleh kondisi ekonomi daerah yang relatif kuat. Adapun daerah yang termasuk dalam kluster ini adalah Kota Bekasi, Kota Depok, dan Kota Bandung.

3) Hasil *K-Means Clustering*

Tabel 4. DBI *K-Means Clustering*

Jumlah Klaster	<i>Davies-Bouldin Index</i>	<i>Silhouette Coefficient</i>
2	1.0308	0.4282
3	0.8338	0.4459
4	0.7318	0.5180
5	0.7173	0.5371

- a. Klaster 1
Klaster ini mencakup wilayah dengan jumlah penerima bantuan yang rendah. Daerah-daerah ini umumnya memiliki kepadatan penduduk rendah atau distribusi program bantuan yang merata namun tidak besar. Adapun daerah yang termasuk dalam klaster ini adalah Kabupaten Majalengka, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Pangandaran, Kabupaten Bandung Barat, Kabupaten Subang.
- b. Klaster 2
Klaster ini terdiri atas wilayah-wilayah dengan jumlah penerima bantuan menengah, mencerminkan wilayah dengan kepadatan sedang dan kebutuhan bantuan tingkat menengah. Kabupaten Karawang, Kabupaten Indramayu, Kabupaten Bogor, Kabupaten Garut.
- c. Klaster 3
Klaster ini mencakup wilayah dengan tingkat penerima bantuan yang tinggi. Wilayah-wilayah ini cenderung memiliki populasi padat dan kebutuhan bantuan yang cukup tinggi, namun tidak mendominasi seperti klaster sebelumnya. Adapun daerah yang termasuk dalam klaster ini adalah Kabupaten Cianjur, Kota Tasikmalaya, Kota Sukabumi.
- d. Klaster 4
Klaster ini mencerminkan wilayah dengan jumlah penerima bantuan sangat tinggi, biasanya merupakan kota metropolitan atau pusat urbanisasi di mana tingkat permintaan bantuan cukup besar. Adapun daerah yang termasuk dalam klaster ini adalah Kota Banjar, Kota Cirebon.
- e. Klaster 5
Klaster ini mencakup wilayah yang memiliki jumlah penerima bantuan yang sangat rendah, yang disebabkan oleh kondisi ekonomi daerah yang relatif kuat. Adapun daerah yang termasuk dalam klaster ini adalah Kota Bekasi, Kota Depok, dan Kota Bandung.

4) Evaluasi

Evaluasi hasil klasterisasi dilakukan menggunakan perhitungan *Silhouette Coefficient*. Penelitian ini membandingkan hasil klasterisasi yang diperoleh dari dua metode berbeda, yaitu *K-Means Clustering* dan *Hierarchical Clustering*. Berikut disajikan perbandingan nilai *Silhouette Coefficient* dan *Davies-Bouldin Index* yang diperoleh dari kedua metode tersebut.

Tabel 5. Hasil Perbandingan Algoritma

No	Algoritma	<i>Silhouette Coefficient</i>	<i>Davies-Bouldin Index</i>	Jumlah Klaster
1	<i>Hierarchical Clustering</i>	0.4976	0.7607	4
2	<i>K Means Clustering</i>	0.5371	0.7173	5

Perbandingan nilai *Silhouette Coefficient* antara metode *K-Means Clustering* dan *Hierarchical Clustering* mengindikasikan bahwa *K-Means* lebih unggul dalam mengelompokkan wilayah penerima bantuan sosial pangan. Nilai *Silhouette Coefficient* sebesar 0,5371 pada *K-Means* menunjukkan kualitas klaster yang lebih baik dibandingkan *Hierarchical Clustering*, yang hanya mencapai 0,4976. Temuan ini membuktikan bahwa *K-Means* mampu membentuk kelompok data yang homogen dengan batas klaster yang lebih tegas. Hal ini selaras dengan penelitian sebelumnya [9], yang menunjukkan bahwa algoritma *K-Means* menghasilkan nilai

Silhouette Coefficient lebih tinggi dibandingkan *Hierarchical Clustering* (0,48 dibandingkan 0,47). Nilai koefisien yang lebih tinggi mencerminkan pemisahan antar kluster lebih baik dan kohesi internal yang lebih kuat. Struktur kluster yang dihasilkan oleh *K-Means* memiliki tingkat kemiripan data dalam satu kluster yang tinggi sekaligus perbedaan yang tegas antara kluster yang berbeda. Sebaliknya, nilai yang rendah pada *Hierarchical Clustering* menunjukkan batas antar kluster yang kurang jelas dan keterkaitan antar kluster yang cukup tinggi. Selisih 0,0395 antara kedua metode ini secara statistik membuktikan bahwa *K-Means* mampu menghasilkan segmentasi wilayah penerima bantuan sosial pangan yang lebih presisi dan terstruktur dengan baik, sehingga lebih sesuai untuk diterapkan dalam analisis kebijakan distribusi bantuan yang membutuhkan ketepatan sasaran. Keunggulan ini diperkuat oleh temuan [9], yang menyatakan bahwa *K-Means* lebih efektif dalam mengelompokkan wilayah karena mampu menghasilkan pemisahan kluster yang lebih jelas dibandingkan *Hierarchical Clustering*.

Berdasarkan nilai *Davies-Bouldin Index*, metode *K-Means* menunjukkan performa yang lebih unggul dengan nilai 0,7173, lebih rendah dibandingkan 0,7607 pada *Hierarchical Clustering*. Nilai DBI yang lebih rendah ini mengonfirmasi bahwa kluster yang terbentuk oleh *K-Means* memiliki kekompakan internal yang lebih tinggi serta pemisahan antar kluster yang lebih jelas. Keunggulan ini juga tercermin dalam kemampuan *K-Means* membentuk lima kluster yang lebih terperinci dibanding empat kluster pada *Hierarchical Clustering*, sehingga menghasilkan identifikasi pola distribusi yang lebih detail. Temuan ini didukung oleh hasil penelitian sebelumnya [9], yang menjelaskan bahwa semakin kecil nilai DBI, semakin baik kualitas kluster karena objek dalam satu kluster menjadi lebih seragam dan jarak antar kluster semakin besar. Pembentukan lima kluster oleh *K-Means* ini memfasilitasi analisis yang lebih akurat dalam penentuan wilayah prioritas dibandingkan dengan empat kluster pada *Hierarchical Clustering*.

Evaluasi keseluruhan menggunakan metrik validasi internal memperlihatkan bahwa algoritma *K-Means Clustering* memiliki kinerja yang lebih unggul dibandingkan *Hierarchical Clustering*. Nilai akurasi keseluruhan *K-Means* tercatat sebesar 52,6% yang diperoleh melalui rata-rata konversi dua metrik validasi, yaitu *Silhouette Coefficient* sebesar 0,5371 (setara 76,9%) dan *Davies-Bouldin Index* sebesar 0,7173 (setara 28,3% setelah dikonversi ke skala kualitas 100%). Sebagai perbandingan, *Hierarchical Clustering* menghasilkan akurasi sebesar 49,4% dengan nilai *Silhouette Coefficient* sebesar 0,4976 dan DBI sebesar 0,7607. Temuan ini sejalan dengan penelitian [17], yang menyatakan bahwa nilai *Silhouette Coefficient* yang lebih besar dari 0,5 menunjukkan kualitas klustering yang baik, sedangkan nilai *Davies-Bouldin Index* yang kurang dari 0,75 mencerminkan kualitas klustering yang dapat diterima. Perbedaan performa sebesar 3,2% ini dianggap signifikan secara praktis, menegaskan bahwa *K-Means* lebih efektif dalam menghasilkan kluster yang terpisah secara jelas dan konsisten. Validasi dilakukan melalui tiga pendekatan:

- 1) Evaluasi metrik internal (*Silhouette Coefficient* dan *Davies-Bouldin Index*) untuk menilai kepadatan dan pemisahan antar kluster.
- 2) Visualisasi distribusi data menggunakan *Principal Component Analysis* (PCA) untuk memverifikasi sebaran titik data antar kluster secara visual.
- 3) Uji konsistensi kluster melalui replikasi proses klustering pada beberapa subset acak dari data.

Hasil keseluruhan menunjukkan bahwa *K-Means* menghasilkan struktur kluster yang lebih stabil, terutama pada konfigurasi lima kluster. Oleh karena itu, *K-Means* direkomendasikan untuk analisis berbasis wilayah yang memerlukan ketepatan sasaran dalam pengambilan keputusan. Sementara itu, *Hierarchical Clustering* tetap relevan untuk mengidentifikasi struktur hierarki antar wilayah yang saling terkait.

5. Simpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma *K-Means Clustering* menunjukkan performa lebih unggul daripada *Hierarchical Clustering* dalam mengelompokkan wilayah penerima bantuan sosial pangan di Provinsi X. Hal ini terlihat dari nilai *Silhouette Coefficient* *K-Means* yang mencapai 0,5371 dan *Davies-Bouldin Index* sebesar 0,7173 lebih baik dibanding *Hierarchical Clustering* dengan nilai masing-masing 0,4976 dan 0,7607. *K-Means* berhasil membentuk 5 kluster yang terdefinisi dengan jelas, sementara *Hierarchical Clustering* menghasilkan 4 kluster dengan dendrogram terbaik menggunakan *average linkage* (*Cophenet Coefficient* 0,826). Pembagian kluster ini mampu mengungkap pola distribusi bantuan, mencakup area dengan jumlah penerima rendah (kepadatan penduduk rendah dan kebutuhan relatif kecil),

sedang (kepadatan penduduk sedang dan kebutuhan tingkat menengah), hingga area dengan penerima tinggi (daerah padat penduduk dan kebutuhan tinggi). Temuan ini memberikan implikasi penting bagi pembuat kebijakan dalam mengoptimalkan alokasi bantuan sosial, khususnya untuk memprioritaskan distribusi ke wilayah klaster dengan kebutuhan tinggi. Untuk penelitian selanjutnya, disarankan untuk mengintegrasikan metode lain seperti DBSCAN, menambahkan variabel sosial-ekonomi yang lebih komprehensif, serta melakukan optimasi komputasi khususnya untuk *Hierarchical Clustering* pada dataset berskala besar. Secara keseluruhan, penelitian ini tidak hanya berhasil membandingkan efektivitas dua metode *clustering*, tetapi juga memberikan dasar ilmiah bagi pemerintah dalam meningkatkan presisi dan efisiensi program bantuan sosial pangan.

Daftar Referensi

- [1] I. D. Anjani and A. Bahtiar, "Penerapan Algoritma K-Means Clustering Untuk Mengelompokkan Penerima Bantuan Sosial Tunai (Bst) Di Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 2743–2747, 2024, doi: 10.36040/jati.v8i3.8974.
- [2] R. Behzadidoost and H. Izadkhah, "Identifying Effective Algorithms and Measures for Enhanced Clustering Quality: A Comprehensive Examination of Arbitrary Decisions in Hierarchical Clustering Algorithms," *J. Classif.*, vol. 42, pp. 1–31, 2025, doi: 10.1007/s00357-025-09506-5.
- [3] Z. Muhammad, "Analisis Perbandingan Algoritma K-Means dan Hierarchical Clustering untuk Pengelompokan Data Penduduk Indeks Pembangunan Manusia pada Kecamatan Percut Sei Tuan," Skripsi, Program Studi Teknik Informatika, Universitas Medan Area, Medan, 2024.
- [4] R. Alwisyah *et al.*, "Penerapan Clustering dalam Distribusi Bantuan Sosial Menggunakan K-Means," *Jurnal Teknologi Informasi dan Komputer.*, vol. 9, no. 3, pp. 112–125, 2021.
- [5] T. Gunawan *et al.*, "Evaluasi Kinerja Algoritma Clustering Menggunakan Silhouette Score," *Jurnal Sains dan Teknologi Komputer.*, vol. 10, no. 1, pp. 33–45, 2021.
- [6] A. Suhendra *et al.*, "Perbandingan Algoritma Clustering dalam Analisis Data Sosial," *Jurnal Artificial Intelligence dan Data Science.*, vol. 7, no. 2, pp. 55–68, 2022.
- [7] N. A. Hidayatullah and W. Prihartono, "Clustering Algoritma K-Means Untuk Meningkatkan Efektivitas Program Sosial Di Kota / Kabupaten Cirebon," vol. 13, no. 1, pp. 629–636, 2025.
- [8] L. Rahmawati, A. Siregar, and B. Nugroho, "Implementasi K-Means dalam Clustering Data Penerima Bantuan Sosial," *Jurnal Data Mining Indonesia.*, vol. 8, no. 4, pp. 54–67, 2021.
- [9] I. Indra, N. Nur, M. Iqram, and N. Inayah, "Perbandingan K-Means dan Hierarchical Clustering dalam Pengelompokan Daerah Beresiko Stunting," *INOVTEK Polbeng - Seri Inform.*, vol. 8, no. 2, p. 356, 2023, doi: 10.35314/isi.v8i2.3612.
- [10] Y. Filki, "Algoritma K-Means Clustering dalam Memprediksi Penerima Bantuan Langsung Tunai (BLT) Dana Desa," *J. Inform. Ekon. Bisnis*, vol. 4, pp. 166–171, 2022, doi: 10.37034/infob.v4i4.166.
- [11] R. A. Restu, S. Susilowati, and I. R. Rahadjeng, "Data Mining dalam Menentukan Cluster Penerima Program Bantuan dengan Metode K-Means" *REMIK (Riset dan E-Jurnal Manajemen Informatika Komputer)*, vol. 5, no. 4, pp. 200–215, 2023.
- [12] Y. K. Saheed, M. O. Arowolo, and A. U. Tosho, "An Efficient Hybridization of K-Means and Genetic Algorithm Based on Support Vector Machine for Cyber Intrusion Detection System," *Int. J. Electr. Eng. Informatics*, vol. 14, no. 2, pp. 426–442, 2022, doi: 10.15676/ijeei.2022.14.2.11.
- [13] N. L. W. S. R. Ginantra, H. H. Wijaya, and F. N. Arifah *et al.*, *Data Mining dan Penerapan Algoritma*. 1st ed. Yogyakarta: Yayasan Kita Menulis, 2021.
- [14] The Pennsylvania State University, "Example: Agglomerative hierarchical clustering." Accessed: Feb. 10, 2025. [Online]. Available: <https://online.stat.psu.edu/stat555/node/86/>
- [15] B. M. Metisen and H. L. Sari, "Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.
- [16] A. Fadilah, M. N. Pangestu, S. Lumbanbatu, and S. Defiyanti, "Pengelompokan Kabupaten/Kota Di Indonesia Berdasarkan Faktor Penyebab Stunting Pada Balita Menggunakan Algoritma K-Means," *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 2, p. 223, 2022, doi: 10.26798/jiko.v6i2.581.

- [17] G. Erda, C. Gunawan, and Z. Erda, "Grouping of Poverty in Indonesia Using K-Means With Silhouette Coefficient," *Param. J. Stat.*, vol. 3, no. 1, pp. 1–6, 2023, doi: 10.22487/27765660.2023.v3.i1.16435.