

Perbandingan Proses Klasterisasi Data Menggunakan K-Means Clustering dan Agglomerative Hierarchical Clustering

Budi Hartono*, Veronica Lusiana, Imam Husni Al Amin

Fakultas Teknologi Informasi dan Industri, Prodi Teknik Informatika, Universitas Stikubank, Semarang, Indonesia

Email: ¹budihartono@edu.unisbank.ac.id, ²vero@edu.unisbank.ac.id, ³imam@edu.unisbank.ac.id

Email Penulis Korespondensi: budihartono@edu.unisbank.ac.id

Submitted 18-06-2025; Accepted 01-08-2025; Published 30-08-2025

Abstrak

Data dalam jumlah besar membutuhkan pengolahan dan analisis yang baik. Salah satu teknik analisis data adalah klasterisasi data, yaitu mengelompokkan data menjadi beberapa kelompok atau klaster data berdasarkan kemiripan karakteristik data. Penelitian ini melakukan pengamatan pada proses penyusunan klaster dan hasil klaster menggunakan algoritma atau metode klaster K-Means dan Agglomerative Hierarchical Clustering (AHC). Penyusunan klaster dilakukan menggunakan tiga macam jumlah data yang berbeda yaitu 10 (data A10), 30 (data B30), dan 60 (data C60), dengan pilihan dua, tiga, dan empat klaster. Hasil percobaan diperoleh yaitu klaster data A10 adalah sama, namun pada data C60 adalah berbeda. Kedua metode tersebut memberikan hasil klaster yang sama yaitu dalam jumlah anggota klaster dan nomor datanya, sebaliknya hasil klaster yang berbeda yaitu apabila terdapat perbedaan jumlah anggota klaster. Hasil klaster data B30 untuk tiga klaster adalah sama, sedangkan pada dua dan empat klaster adalah berbeda. Hasil penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik pada proses penyusunan klaster data dan dapat menjadi dasar untuk memilih metode klaster yang lebih sesuai.

Kata Kunci: Klaster data; Klaster K-Means; Agglomerative Hierarchical Clustering; Centroid

Abstract

Large amounts of data require good processing and analysis. One of the data analysis techniques is data clustering, which is grouping data into several groups or data clusters based on the similarity of data characteristics. This study observed the clustering process and cluster results using the K-Means and Agglomerative Hierarchical Clustering (AHC) algorithms or methods. Clustering was carried out using three different amounts of data, namely 10 (A10 data), 30 (B30 data), and 60 (C60 data), with choices of two, three, and four clusters. The experimental results obtained were that the A10 data cluster was the same, but the C60 data was different. Both methods provide the same cluster results, namely in the number of cluster members and their data numbers; conversely, different cluster results are obtained if there are differences in the number of cluster members. The B30 data cluster results for three clusters are the same, while for two and four clusters they are different. The results of this study are expected to provide a better understanding of the data clustering process and can be a basis for selecting a more appropriate clustering method.

Keywords: Data cluster; K-Means Cluster; Agglomerative Hierarchical Clustering; Centroid

1. PENDAHULUAN

Data dalam jumlah yang besar membutuhkan pengolahan dan analisis yang baik. Salah satu teknik analisis data adalah proses klasterisasi data, yaitu proses pengelompokan data menjadi beberapa kelompok atau klaster data berdasarkan kemiripan karakteristik data. Pada data dengan jumlah besar, maka proses klasterisasi dapat dihasilkan menggunakan algoritma tertentu karena pengelompokan data tidak dapat dikerjakan secara mudah hanya dengan membaca data satu-persatu. Hasil dari proses ini dapat digunakan untuk berbagai tujuan, antara lain segmentasi pasar, mengetahui tren perilaku konsumen, pemetaan wilayah populasi, dan pengenalan pola [1]. Contoh, klaster data digunakan untuk membagi kelompok pelanggan berdasarkan usia, tingkat pendapatan, atau domisili. Algoritma klaster data yang telah banyak diterapkan yaitu *K-Means Clustering* (klaster K-Means) dan *Agglomerative Hierarchical Clustering* (AHC).

Penelitian ini melakukan pengamatan pada proses penyusunan klaster dan hasil klaster menggunakan algoritma atau metode klaster K-Means dan AHC. Percobaan penyusunan klaster dilakukan menggunakan tiga macam jumlah data yang berbeda dengan jumlah klaster 2, 3, dan 4. Melalui data percobaan yang sama, hasil klaster kedua metode tersebut dibandingkan. Hasil penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik pada proses penyusunan klaster data dan dapat dijadikan dasar untuk memilih metode klaster yang lebih sesuai.

Berikut ini adalah beberapa penelitian mengenai penerapan metode klaster K-Means dan AHC. Beberapa penelitian yang telah dilakukan belum menjelaskan secara detail proses penyusunan klaster. Kedua metode ini digunakan untuk mengelompokkan data pada berbagai macam penerapan.

a. Penelitian yang menerapkan metode klaster K-Means.

Dewi dan kawan-kawan [2] menentukan kelompok pendalaman materi ujian nasional, yaitu kelas umum dan kelas khusus. Sari dan kawan-kawan [3] mengelompokkan data kasus Covid-19 di Indonesia, jumlah kasus tinggi, kasus sedang, dan kasus rendah. Suraya dan kawan-kawan [4] mengelompokkan indeks prestasi (IP) akademik mahasiswa, yaitu IP tinggi, IP sedang, dan IP rendah. Nurahman dan Susanto [5] menentukan kelompok penerima bantuan langsung tunai (BLT). Syahra dan Kawan-kawan [6] meneliti tingkat penanganan dan pelayanan kesehatan masyarakat, dengan tiga kategori penilaian oleh pasien yaitu: sangat puas, puas, dan tidak puas. Aranski dan Kawan-kawan [7] mengelompokkan data penerima bantuan subsidi rumah yaitu layak dan tidak layak mendapatkan bantuan. Muzaqqi dan kawan-kawan [8] mengidentifikasi kelompok masyarakat yang berhak menerima program bantuan langsung dari pemerintah.

b. Penelitian yang menggunakan metode AHC.

Widyawati dan kawan-kawan [9] membahas penerapan AHC untuk mengelompokkan pelanggan berdasarkan kesamaan karakteristiknya. Setyawan dan Fadilla [10] meneliti klusterisasi media pembelajaran daring yang efektif pada era pandemi Covid-19. Nellie dan kawan-kawan [11] menyusun sistem rekomendasi film dengan mengelompokkan film berdasarkan kesamaan kategori. Lubis dan Kawan-kawan [12] mengelompokkan produk terlaris di swalayan menjadi tiga cluster. Anggraeni dan Kawan-kawan [13] melalui penelitiannya terbentuk tiga tingkat prevalensi stunting balita atau jumlah kasus stunting balita dalam suatu populasi pada periode tertentu, yaitu: prevalensi rendah, sedang, dan tinggi.

c. Penelitian yang menggunakan secara bersama metode kluster K-Means dan AHC.

Lubis dan Kawan-kawan [14] memberikan informasi mengenai pengelompokan provinsi di Indonesia berdasarkan tingkat kejahatan. Rozaq [15] mengelompokkan jenis rumah berdasarkan total biaya yang dikeluarkan untuk kebutuhan semen, luas bangunan, dan komposisi campuran material.

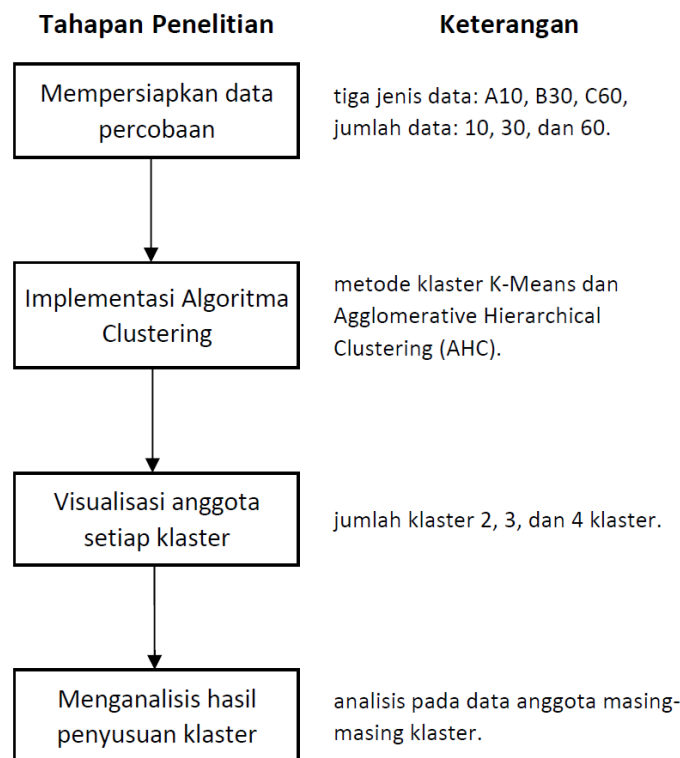
Metode AHC menyusun klusterisasi data secara hirarki yaitu penggabungan data yang mirip secara bertahap sampai dengan membentuk kluster tunggal yang secara visual mirip dengan pohon kluster data. Pohon ini dibangun dengan menggabungkan kluster yang mirip [16]. Metode kluster K-Means adalah proses klusterisasi yang bekerja dengan menetapkan nilai pusat kluster (*centroid*) awal secara acak. Data dialokasikan ke dalam kluster terdekat berdasarkan *centroid* awal. Selanjutnya *centroid* dihitung ulang berdasarkan anggota kluster [1]. Data dialokasikan kembali ke dalam kluster terdekat menggunakan nilai *centroid* baru. Proses ini diulang sehingga tercapai konvergensi.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian diawali dengan mempersiapkan tiga jenis data percobaan yaitu data A10 dengan jumlah data 10 buah, B30 jumlah data 30, dan C60 dengan 60 buah data. Seluruh data mencatat dua nilai fitur yaitu fitur x dan y. Data percobaan memiliki nilai acak dan unik. Dalam penerapannya, fitur ini dapat mewakili pendapatan, biaya produksi, nilai hasil suatu pengamatan, dan lain sebagainya. Dengan dua nilai fitur maka data dapat divisualkan pada grafik dua dimensi.

Tahap penelitian selanjutnya adalah implementasi algoritma kluster K-Means dan Agglomerative Hierarchical Clustering (AHC) menggunakan bahasa Python [17] [18]. Hasil implementasi adalah data anggota setiap kluster dan visualisasi kluster. Tahapan penelitian dapat dilihat pada Gambar 1. Analisis hasil penyusunan kluster dilakukan dengan membandingkan hasil kluster kedua metode ini.



Gambar 1. Tahap Penelitian

2.2 Kluster K-Means

Algoritma kluster K-Means bekerja dengan cara membagi data menjadi sejumlah kluster, dengan setiap kluster memiliki titik pusatnya masing masing [1]. Empat langkah utama dalam algoritma ini adalah:

- Inisialisasi pemilihan titik pusat suatu kluster atau nilai *centroid*, *centroid* awal setiap kluster dipilih secara acak. *Centroid* awal dapat dipilih dari salah satu data atau nilai yang berdekatan dengan data.

- b. Setiap data menjadi anggota kluster yang terdekat.
Setiap kluster diwakili oleh masing-masing nilai centroid-nya. Jarak antara data dengan centroid digunakan untuk menentukan kluster yang terdekat [19] [20]. Jarak antara data dan centroid dapat dihitung menggunakan beberapa macam metrik, antara lain jarak Euclidean (*Euclidean distance*), jarak Manhattan (*Manhattan distance*), atau jarak Minkowski (*Minkowski distance*).
- c. Nilai centroid diperbarui, diperoleh dari data yang menjadi anggota kluster.
Centroid baru dari setiap kluster diperoleh dari menghitung titik pusat seluruh data anggota kluster tersebut.
- d. Apabila nilai centroid baru berbeda dengan nilai centroid lama, maka kembali ke langkah b.
Apabila sebaliknya, tidak ada perubahan nilai centroid atau telah sampai kepada batas iterasi yang ditentukan maka data anggota setiap kluster telah diperoleh.

2.3 Agglomerative Hierarchical Clustering

Algoritma Hierarchical Clustering (AHC) bekerja dengan cara membangun pohon kluster dari data. Pohon ini dibangun menggabungkan kluster yang mirip satu sama lain [21] [22] [23]. Algoritma ini bekerja dalam dua pendekatan, yaitu:

- a. Agglomerative, membangun pohon kluster dimulai dari setiap data sebagai kluster individu. Pada tahap selanjutnya, dua kluster yang paling mirip digabung menjadi satu kluster baru.
- b. Divisive, dimulai dengan seluruh data menjadi anggota satu kluster. Pada setiap tahap, kluster tersebut dipisah menjadi dua kluster yang lebih kecil.

Pendekatan agglomerative, dimulai dengan setiap data sebagai kluster tersendiri. Pada setiap tahap, dua kluster yang paling mirip digabungkan menjadi satu kluster baru. Jarak antara dua kluster dapat dihitung menggunakan berbagai metrik, antara lain: jarak Euclidean, jarak Manhattan, atau jarak Minkowski. Langkah ini diulangi sampai dengan terbentuk sebuah kluster tunggal.

Pendekatan divisive, dimulai dengan semua data menjadi anggota satu kluster. Pada setiap tahap, kluster tersebut dibagi menjadi dua kluster yang lebih kecil. Pemisahan kluster ditentukan berdasarkan jarak antar data. Data yang paling berjauhan akan dipisahkan menjadi dua kluster. Langkah ini diulangi sampai dengan setiap data menjadi anggota tunggal dalam klasternya sendiri.

3. HASIL DAN PEMBAHASAN

3.1 Data Percobaan

Data percobaan terdiri dari tiga jenis yaitu data A10, B30, dan C60, dengan jumlah kluster bervariasi seperti dicatat pada Tabel 1. Seluruh data disini adalah acak dan unik untuk mencatat dua nilai fitur yaitu fitur x dan y. Nilai yang dicatat adalah bilangan bulat 0 sampai dengan 30. Dengan dua nilai fitur maka data dan anggota kluster dapat divisualkan dalam grafik dua dimensi.

Tabel 1. Keterangan data percobaan

Nama Data	Jumlah Data	Jumlah Kluster
A10	10	2, 3
B30	30	2, 3, 4
C60	60	2, 3, 4

Pada Tabel 2 mencatat fitur data percobaan. Fitur data (nomor data, fitur x, fitur y) A10 dengan nomor data 0 (0, 4, 21) sampai dengan 9 (9, 12, 21). Data B30 dengan nomor data 0 (0, 7, 18) sampai dengan 29 (29, 17, 10). Data C60 dengan nomor data 0 (0, 3, 23) sampai dengan 59 (59, 25, 18). Data A10 berjumlah 10 data dikelompokkan menjadi 2 dan 3 kluster. Data B30 berjumlah 30 data dan C60 berjumlah 60 data dikelompokkan menjadi 2, 3, dan 4 kluster.

Grafik data percobaan dapat dilihat pada Tabel 2 kolom Plot Data. Pengamatan visual pada grafik plot data, data A10 terbagi menjadi dua kelompok yaitu data berada di kiri bawah dan kanan atas yang masing-masing berjumlah lima data. Data B30 menempati posisi dari kiri atas sampai ke kanan bawah. Tidak terlihat dengan jelas batas data yang memiliki nilai fitur x kecil dan fitur y besar atau menempati area grafik kiri atas, sampai data yang memiliki nilai fitur x besar dan fitur y kecil, atau menempati area grafik kanan bawah. Data C60 menempati area grafik yang tersebar secara merata. Data B30 dan C60 tidak mudah untuk menemukan batas setiap kluster melalui pengamatan grafik secara langsung. Pada percobaan selanjutnya, kluster data dikerjakan menggunakan algoritma kluster K-Means dan AHC.

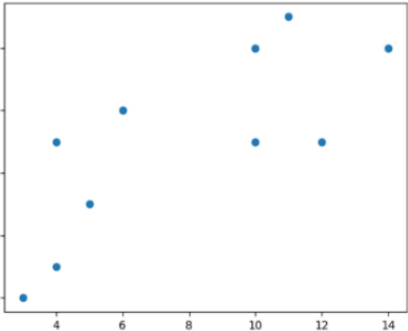
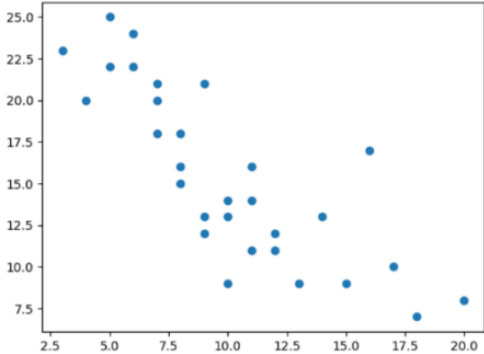
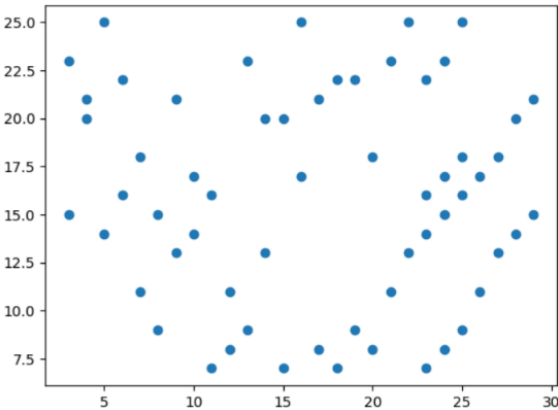
3.2 Kluster Data

Algoritma atau metode kluster K-Means bekerja dengan cara membagi data menjadi sejumlah kluster, dimana setiap kluster memiliki titik pusat. Sedangkan AHC, bekerja dengan cara membangun pohon kluster dari data. Pohon ini dibangun dengan menggabungkan kluster yang mirip satu sama lain. Visualisasi hasil kluster K-Means dan AHC dapat dilihat pada Gambar 2. Pada kluster K-Means anggota setiap kluster dibedakan oleh warna plot data.

Data hasil kluster K-Means dan AHC dicatat pada Tabel 3. Data A10 dengan metode kluster K-Means pada pengelompokan dua kluster, anggota kluster 1 (K1) adalah data nomor 0, 1, 3, 4, 7, dan anggota kluster 2 (K2) adalah data nomor 2, 5, 6, 8, 9, masing-masing beranggotakan 5 data. Pada pengelompokan tiga kluster, data yang berada di daerah

kiri bawah dibagi menjadi dua kluster yaitu K1 data nomor 3, 4 dan K2 data nomor 0, 1, 7, serta satu kluster di daerah kanan atas yaitu K3 data nomor 2, 5, 6, 8, 9.

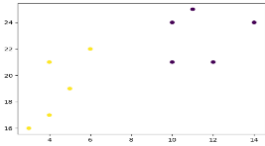
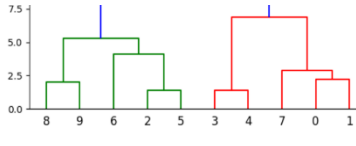
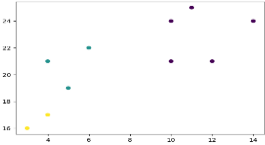
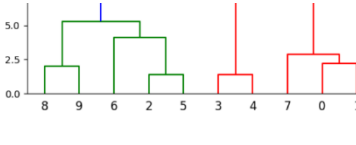
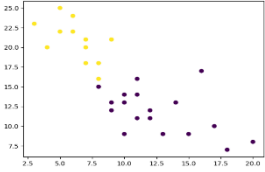
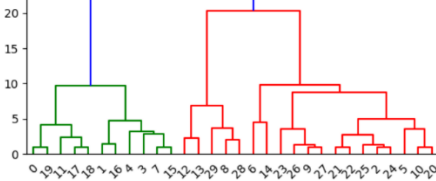
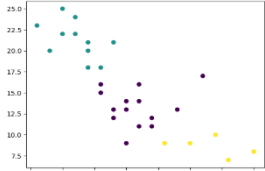
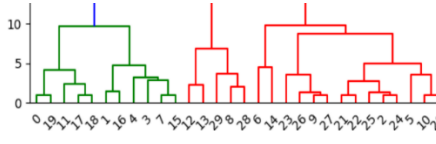
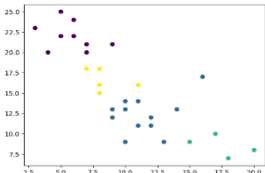
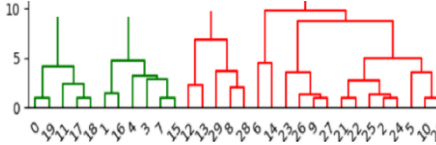
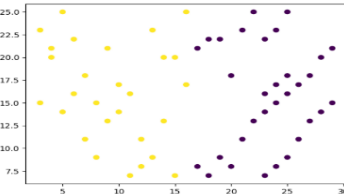
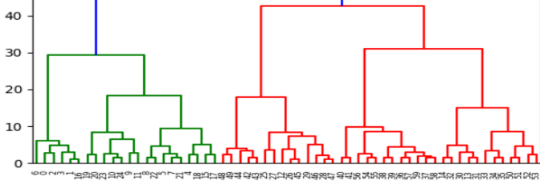
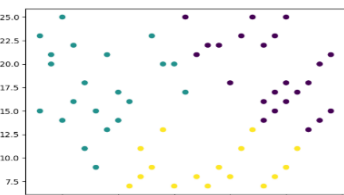
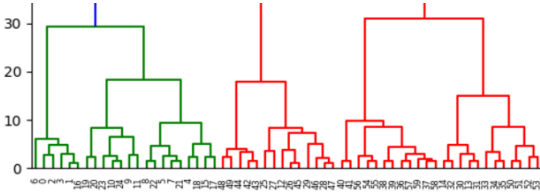
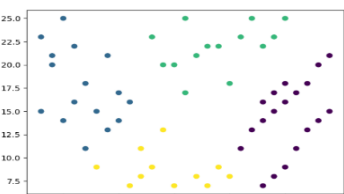
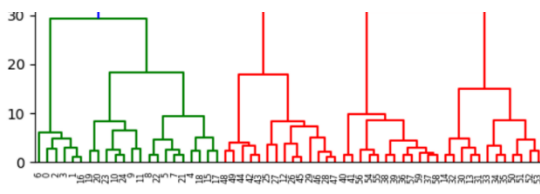
Tabel 2. Nilai fitur dan grafik data percobaan

Nama Data	Jumlah Data	Fitur Data (nomor data, fitur x, fitur y)	Plot Data
A10	10	[(0, 4, 21), (1, 5, 19), (2, 10, 24), (3, 4, 17), (4, 3, 16), (5, 11, 25), (6, 14, 24), (7, 6, 22), (8, 10, 21), (9, 12, 21)]	
B30	30	[(0, 7, 18), (1, 5, 25), (2, 10, 14), (3, 4, 20), (4, 3, 23), (5, 11, 16), (6, 14, 13), (7, 6, 22), (8, 13, 9), (9, 12, 11), (10, 8, 15), (11, 9, 21), (12, 20, 8), (13, 18, 7), (14, 16, 17), (15, 5, 22), (16, 6, 24), (17, 7, 21), (18, 7, 20), (19, 8, 18), (20, 8, 16), (21, 9, 13), (22, 9, 12), (23, 10, 9), (24, 10, 13), (25, 11, 14), (26, 11, 11), (27, 12, 12), (28, 15, 9), (29, 17, 10)]	
C60	60	[(0, 3, 23), (1, 4, 20), (2, 5, 25), (3, 6, 22), (4, 7, 18), (5, 8, 15), (6, 9, 21), (7, 10, 14), (8, 11, 16), (9, 12, 11), (10, 13, 9), (11, 14, 13), (12, 16, 17), (13, 18, 7), (14, 20, 8), (15, 3, 15), (16, 4, 21), (17, 5, 14), (18, 6, 16), (19, 7, 11), (20, 8, 9), (21, 9, 13), (22, 10, 17), (23, 11, 7), (24, 12, 8), (25, 13, 23), (26, 14, 20), (27, 16, 25), (28, 18, 22), (29, 20, 18), (30, 15, 7), (31, 17, 8), (32, 19, 9), (33, 21, 11), (34, 22, 13), (35, 23, 14), (36, 24, 15), (37, 25, 16), (38, 26, 17), (39, 27, 18), (40, 28, 20), (41, 29, 21), (42, 23, 22), (43, 24, 23), (44, 25, 25), (45, 15, 20), (46, 17, 21), (47, 19, 22), (48, 21, 23), (49, 22, 25), (50, 23, 7), (51, 24, 8), (52, 25, 9), (53, 26, 11), (54, 27, 13), (55, 28, 14), (56, 29, 15), (57, 23, 16), (58, 24, 17), (59, 25, 18)]	

Data A10 dengan metode AHC, pada pengelompokan dua kluster, K1 dan K2 masing-masing beranggotakan 5 data, yaitu K1 data nomor 2, 5, 6, 8, 9 dan K2 data nomor 0, 1, 3, 4, 7. Pada pengelompokan tiga kluster, maka anggota K1 data nomor 2, 5, 6, 8, 9, anggota K2 data nomor 3, 4, dan anggota K3 data nomor 0, 1, 7. Disini anggota K2 dan K3 berasal dari pemisahan salah satu kluster hasil pengelompokan data dua kluster.

Data B30 dengan metode kluster K-Means pada pengelompokan dua kluster, anggota K1 adalah 12 data yaitu data nomor 0, 1, 3, 4, 7, 11, 15, 16, 17, 18, 19, 20, dan anggota K2 adalah 18 data yaitu data nomor 2, 5, 6, 8, 9, 10, 12, 13, 14, 21, 22, 23, 24, 25, 26, 27, 28, 29. Pada pengelompokan tiga kluster, anggota K1 adalah 11 data yaitu data nomor 0, 1, 3, 4, 7, 11, 15, 16, 17, 18, 19, anggota K2 adalah 14 data yaitu data nomor 2, 5, 6, 9, 10, 14, 20, 21, 22, 23, 24, 25, 26, 27, dan anggota K3 adalah 5 data yaitu data nomor 8, 12, 13, 28, 29. Pada pengelompokan empat kluster, anggota K1 adalah 9 data yaitu data nomor 1, 3, 4, 7, 11, 15, 16, 17, 18, anggota K2 adalah 5 data yaitu data nomor 0, 5, 10, 19, 20, anggota K3 adalah 12 data yaitu data nomor 2, 6, 8, 9, 14, 21, 22, 23, 24, 25, 26, 27, dan anggota K4 adalah 4 data yaitu data nomor 12, 13, 28, 29.

Tabel 3. Visualisasi hasil kluster K-Means dan AHC

No.	Nama Data	Jumlah Kluster	Metode Kluster K-Means	Metode Agglomerative Hierarchical Clustering (AHC)
1	A10	2		
2	A10	3		
3	B30	2		
4	B30	3		
5	B30	4		
6	C60	2		
7	C60	3		
8	C60	4		

Data B30 dengan metode AHC pada pengelompokan dua kluster, anggota K1 adalah 11 data yaitu data nomor 0, 1, 3, 4, 7, 11, 15, 16, 17, 18, 19, dan anggota K2 adalah 19 data yaitu data nomor 2, 5, 6, 8, 9, 10, 12, 13, 14, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29. Pada pengelompokan tiga kluster, anggota K1 adalah 11 data yaitu data nomor 0, 1, 3, 4, 7, 11, 15, 16, 17, 18, 19, anggota K2 adalah 5 data yaitu data nomor 8, 12, 13, 28, 29, dan anggota K3 adalah 14 data yaitu data nomor 2, 5, 6, 9, 10, 14, 20, 21, 22, 23, 24, 25, 26, 27. Anggota K2 dan K3 berasal dari pemisahan salah satu kluster hasil pengelompokan data dua kluster. Pada pengelompokan empat kluster, anggota K1 adalah 5 data yaitu data nomor 0, 11,

17, 18, 19, anggota K2 adalah 6 data yaitu data nomor 1, 3, 4, 7, 15, 16. Anggota K1 dan K2 berasal dari pemisahan salah satu kluster hasil pengelompokan data tiga kluster. Anggota K3 adalah 5 data yaitu data nomor 8, 12, 13, 28, 29, dan anggota K4 adalah 14 data yaitu data nomor 2, 5, 6, 9, 10, 14, 20, 21, 22, 23, 24, 25, 26, 27.

Tabel 3. Data hasil kluster K-Means dan AHC

No.	Nama Data	Jumlah Kluster	Jumlah Data dan Nomor Data Anggota Kluster	
			Metode K-Means	Metode AHC
1	A10	2	K1(5) = 0, 1, 3, 4, 7 K2(5) = 2, 5, 6, 8, 9	K1(5) = 2, 5, 6, 8, 9 K2(5) = 0, 1, 3, 4, 7
2	A10	3	K1(2) = 3, 4 K2(3) = 0, 1, 7 K3(5) = 2, 5, 6, 8, 9	K1(5) = 2, 5, 6, 8, 9 K2(2) = 3, 4 K3(3) = 0, 1, 7
3	B30	2	K1(12) = 0, 1, 3, 4, 7, 11, 15, 16, 17, 18, 19, 20 K2(18) = 2, 5, 6, 8, 9, 10, 12, 13, 14, 21, 22, 23, 24, 25, 26, 27, 28, 29	K1(11) = 0, 1, 3, 4, 7, 11, 15, 16, 17, 18, 19 K2(19) = 2, 5, 6, 8, 9, 10, 12, 13, 14, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29
4	B30	3	K1(11) = 0, 1, 3, 4, 7, 11, 15, 16, 17, 18, 19 K2(14) = 2, 5, 6, 9, 10, 14, 20, 21, 22, 23, 24, 25, 26, 27 K3(5) = 8, 12, 13, 28, 29	K1(11) = 0, 1, 3, 4, 7, 11, 15, 16, 17, 18, 19 K2(5) = 8, 12, 13, 28, 29 K3(14) = 2, 5, 6, 9, 10, 14, 20, 21, 22, 23, 24, 25, 26, 27
5	B30	4	K1(9) = 1, 3, 4, 7, 11, 15, 16, 17, 18 K2(5) = 0, 5, 10, 19, 20 K3(12) = 2, 6, 8, 9, 14, 21, 22, 23, 24, 25, 26, 27 K4(4) = 12, 13, 28, 29	K1(5) = 0, 11, 17, 18, 19 K2(6) = 1, 3, 4, 7, 15, 16 K3(5) = 8, 12, 13, 28, 29 K4(14) = 2, 5, 6, 9, 10, 14, 20, 21, 22, 23, 24, 25, 26, 27
6	C60	2	K1(28) = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 30, 45 K2(32) = 13, 14, 28, 29, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59	K1(22) = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 K2(38) = 12, 13, 14, 25, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59
7	C60	3	K1(21) = 0, 1, 2, 3, 4, 5, 6, 7, 8, 12, 15, 16, 17, 18, 19, 20, 21, 22, 25, 26, 45 K2(23) = 27, 28, 29, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 46, 47, 48, 49, 54, 55, 56, 57, 58, 59 K3(16) = 9, 10, 11, 13, 14, 23, 24, 30, 31, 32, 33, 34, 50, 51, 52, 53	K1(22) = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 K2(14) = 12, 25, 26, 27, 28, 29, 42, 43, 44, 45, 46, 47, 48, 49 K3(24) = 13, 14, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59
8	C60	4	K1(16) = 0, 1, 2, 3, 4, 5, 6, 7, 8, 15, 16, 17, 18, 19, 21, 22 K2(14) = 12, 25, 26, 27, 28, 29, 42, 43, 44, 45, 46, 47, 48, 49 K3(11) = 9, 10, 11, 13, 14, 20, 23, 24, 30, 31, 32 K4(19) = 33, 34, 35, 36, 37, 38, 39, 40, 41, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59	K1(22) = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 K2(14) = 12, 25, 26, 27, 28, 29, 42, 43, 44, 45, 46, 47, 48, 49 K3(12) = 36, 37, 38, 39, 40, 41, 54, 55, 56, 57, 58, 59 K4(12) = 13, 14, 30, 31, 32, 33, 34, 35, 50, 51, 52, 53

3.3 Hasil Kluster

Label atau nama kluster K1, K2, K3, dan K4 akan menyesuaikan atau dapat diganti dengan konteks penerapannya. Contoh dua kluster dapat diberi nama “lulus” atau “gagal”. Tiga kluster dapat diberi nama “mentah”, “matang”, atau “busuk”. Empat kluster dapat diberi nama “istimewa”, “baik”, “cukup”, atau “kurang”.

Pada Tabel 4 merangkum perbandingan hasil kluster K-Means dan AHC. Data A10 dengan dua kluster, data anggota setiap kluster adalah sama yaitu 5 data. Data A10 dengan tiga kluster, data anggota setiap kluster adalah 2, 3, dan 5 data. Jumlah anggota masing-masing kluster untuk kedua metode adalah sama.

Data B30 dengan dua kluster metode K-Means diperoleh 12 dan 18 anggota, untuk metode AHC diperoleh 11 dan 19 anggota. Terdapat perbedaan jumlah anggota kluster untuk kedua metode tersebut. Data B30 dengan tiga kluster metode K-Means dan AHC diperoleh 11, 14, dan 5 anggota. Jumlah anggota masing-masing kluster untuk kedua metode adalah sama. Data B30 dengan empat kluster metode K-Means diperoleh 9, 5, 12, dan 4 anggota, untuk metode AHC diperoleh 5, 6, 5, dan 14 anggota. Terdapat perbedaan jumlah anggota kluster untuk kedua metode tersebut.

Data C60 dengan dua kluster metode K-Means diperoleh 28 dan 32 anggota, untuk metode AHC diperoleh 22 dan 38 anggota. Data C60 dengan tiga kluster metode K-Means diperoleh 21, 23, dan 16 anggota, untuk metode AHC diperoleh 22, 14 dan 24 anggota. Data C60 dengan empat kluster metode K-Means diperoleh 16, 14, 11, dan 19 anggota, untuk metode AHC diperoleh 22, 14, 12, dan 12 anggota. Terdapat perbedaan jumlah anggota kluster untuk kedua metode tersebut.

Tabel 4. Perbandingan hasil kluster K-Means dan AHC

Nama Data	Jumlah Data	Jumlah Kluster	Hasil Kluster K-Means dan AHC
A10	10	2	Sama
		3	Sama
B30	30	2	Berbeda
		3	Sama
		4	Berbeda
C60	60	2	Berbeda
		3	Berbeda
		4	Berbeda

3.4 Implementasi

Kode program kluster K-Means dapat dilihat pada Gambar 3, dengan menyesuaikan data percobaan A10, B30, atau C60 serta pilihan jumlah kluster [17]. Pada kode program ini menggunakan data percobaan C60 yaitu $x=[3, 4, 5, 6, \dots, 24, 25]$, $y=[23, 20, 25, 22, \dots, 17, 18]$. Pilihan jumlah kluster dapat disesuaikan melalui perintah `kmeans = KMeans(n_clusters=4)`, disini adalah contoh empat kluster. Kode program AHC dapat dilihat pada Gambar 4, dengan menyesuaikan data percobaan A10, B30, atau C60 [18]. Pada kode program ini menggunakan data percobaan C60.

```
#Three lines to make our compiler able to draw:
import sys
import matplotlib
matplotlib.use('Agg')

import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

x = [ 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 18, 20, 3, 4, 5, 6, 7, 8, 9, 10,
11, 12, 13, 14, 16, 18, 20, 15, 17, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 23, 24,
25, 15, 17, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 23, 24, 25]

y = [ 23, 20, 25, 22, 18, 15, 21, 14, 16, 11, 9, 13, 17, 7, 8, 15, 21, 14, 16, 11,
9, 13, 17, 7, 8, 23, 20, 25, 22, 18, 7, 8, 9, 11, 13, 14, 15, 16, 17, 18, 20, 21,
22, 23, 25, 20, 21, 22, 23, 25, 7, 8, 9, 11, 13, 14, 15, 16, 17, 18]

data = list(zip(x, y))

kmeans = KMeans(n_clusters=4)
kmeans.fit(data)

plt.scatter(x, y, c=kmeans.labels_)
plt.show()

#Two lines to make our compiler able to draw:
plt.savefig(sys.stdout.buffer)
sys.stdout.flush()
```

Gambar 2. Kode program kluster K-Means

```
#Three lines to make our compiler able to draw:
import sys
import matplotlib
matplotlib.use('Agg')

import numpy as np
import matplotlib.pyplot as plt
from scipy.cluster.hierarchy import dendrogram, linkage

x = [ 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 18, 20, 3, 4, 5, 6, 7, 8,
9, 10, 11, 12, 13, 14, 16, 18, 20, 15, 17, 19, 21, 22, 23, 24, 25, 26, 27,
28, 29, 23, 24, 25, 15, 17, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 23, 24,
25]

y = [ 23, 20, 25, 22, 18, 15, 21, 14, 16, 11, 9, 13, 17, 7, 8, 15, 21, 14,
16, 11, 9, 13, 17, 7, 8, 23, 20, 25, 22, 18, 7, 8, 9, 11, 13, 14, 15, 16, 17,
18, 20, 21, 22, 23, 25, 20, 21, 22, 23, 25, 7, 8, 9, 11, 13, 14, 15, 16, 17,
18]

data = list(zip(x, y))

linkage_data = linkage(data, method='ward', metric='euclidean')
dendrogram(linkage_data)

plt.show()

#Two lines to make our compiler able to draw:
plt.savefig(sys.stdout.buffer)
sys.stdout.flush()
```

Gambar 3. Kode program Agglomerative Hierarchical Clustering

4. KESIMPULAN

Hasil kluster data menggunakan K-Means clustering dan Agglomerative Hierarchical Clustering (AHC) pada 10 data (A10) adalah sama, namun pada 60 data (C60) adalah berbeda. Hasil kluster sama yaitu dalam jumlah anggota kluster dan nomor datanya. Hasil kluster berbeda apabila terdapat perbedaan jumlah anggota kluster. Jumlah data yang akan dikluster dan nilai fitur data berpengaruh kepada perbedaan hasil kluster. Pada 30 data (B30) hasil tiga kluster adalah sama, sedangkan pada dua dan empat kluster adalah berbeda. Metode kluster K-Means menjalankan algoritma untuk setiap n-kluster dengan data anggota dapat berpindah antar kluster. Metode AHC mengelompokkan atau memecah anggota kluster mengikuti hirarki yang telah disusun.

REFERENCES

- [1] J. Han, J. Pei dan H. Tong, *Data Mining Concepts and Techniques*, 4th Edition, Cambridge: Morgan Kaufmann, 2023.
- [2] Y. N. Dewi, H. Rianto, C. Budihartanti dan F. W. Fibriany, "Penerapan Metode K-Means Dalam Menentukan Kelompok Pendalaman Materi Ujian Nasional," *JISAMAR (Journal of Information System, Applied, Management, Accounting and Research)*, vol. 6, no. 1, pp. 26-31, 2022.
- [3] N. Sari, H. H. Handayani dan A. M. Siregar, "Implementasi Clustering Data Kasus Covid 19 Di Indonesia Menggunakan Algoritma K-Means," *Jurnal Bianglala Informatika*, vol. 11, no. 1, pp. 7-12, 2023.
- [4] S. Suraya, M. Sholeh dan D. Andayati, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Indeks Prestasi Akademik Mahasiswa," *Jurnal SKANIKA*, vol. 6, no. 1, pp. 51-60, 2023.
- [5] Nurahman dan J. Susanto, "Klasterisasi Data Penerima Bantuan Langsung Tunai Menggunakan Algoritma K-Means," *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 2, p. 461-470, 2023.
- [6] Y. Syahra, D. R. Habibie, M. Nasution, H. N. Nasution dan A. H. Nasyuha, "Klasterisasi Data Penanganan dan Pelayanan Kesehatan Masyarakat dengan Algoritma K-Means," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 5, p. 1423-1433, 2022.
- [7] A. W. Aranski, S. Astiti, R. A. Putra dan D. Darmansah, "Pengaplikasian Data Mining Dalam Mengelompokkan Data Penerima Bantuan Subsidi Rumah dengan Menggunakan Metode K-Means Clustering," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 1, pp. 480-489, 2024.
- [8] F. H. Muzaqqi, D. Rakhmawati dan A. B. Wijaya, "K-Means Clustering Algorithm for Grouping Eligible Community Recipients of Government Assistance," *Jurnal Media Informatika Budidarma*, vol. 8, no. 3, pp. 1494-1502, 2024.
- [9] W. Widyawati, W. L. Y. Saptomo dan Y. R. W. Utami, "Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan," *Jurnal Sinus*, vol. 18, no. 1, pp. 75-87, 2020.
- [10] R. A. Setyawan dan R. M. Fadilla, "Klasterisasi Media Pembelajaran Daring di Era Pandemi Covid-19 Menggunakan Metode Agglomerative," *Jurnal Informasi Interaktif*, vol. 5, no. 3, pp. 100-105, 2020.
- [11] V. Nellie, V. C. Mawardi dan N. J. Perdana, "Implementasi Metode Agglomerative Hierarchical Clustering untuk Sistem Rekomendasi Film," *Jurnal Jiksi*, vol. 11, no. 1, pp. 1-6, 2023.
- [12] R. M. F. Lubis, J.-P. Huang, P.-C. Wang, K. Khoifin, Y. Elvina dan D. A. Kusumaningtyas, "Agglomerative Hierarchical Clustering (AHC) Method for Data Mining Sales Product Clustering," *Building of Informatics, Technology and Science (BITS)*, vol. 5, no. 1, p. 285-294, 2023.
- [13] M. R. Anggraeni, U. Yudatama dan M. Maimunah, "Clustering Prevalensi Stunting Balita Menggunakan Agglomerative Hierarchical Clustering," *Jurnal Media Informatika Budidarma*, vol. 7, no. 1, pp. 351-359, 2023.
- [14] R. M. F. Lubis, J.-P. Huang, P.-C. Wang, N. Damanik, A. C. Sitepu dan C. D. Simanullang, "K-Means and AHC Methods for Classifying Crime Victims by Indonesian Provinces: A Comparative Analysis," *Building of Informatics, Technology and Science (BITS)*, vol. 5, no. 1, p. 295-307, 2023.
- [15] A. Rozaq, "Implementation of K-Means and Agglomerative Hierarchical Methods to House Clusterization," *Jurnal Media Informatika Budidarma*, vol. 6, no. 2, pp. 933-942, 2022.
- [16] L. M. C. Cabezas, R. Izbicki dan R. B. Stern, "Hierarchical clustering: Visualization, feature importance and model selection," *Applied Soft Computing*, vol. 141, no. July 2023, p. Article 110303, 2023.
- [17] w3schools, "python_ml_k-means," w3schools, 1 8 2024. [Online]. Available: https://www.w3schools.com/python/python_ml_k-means.asp. [Diakses 1 9 2024].
- [18] w3schools, "Machine Learning - Hierarchical Clustering," w3schools, 1 8 2024. [Online]. Available: https://www.w3schools.com/python/python_ml_hierarchical_clustering.asp. [Diakses 1 9 2024].
- [19] B. Hartono, S. Eniyati dan K. Hadiono, "Perbandingan Metode Perhitungan Jarak pada Nilai Centroid dan Pengelompokan Data Menggunakan K-Means Clustering," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 3, p. 503-509, 2023.
- [20] S. I. Hassan, A. Samad, O. Ahmad dan A. Alam, "Partitioning and hierarchical based clustering: a comparative empirical assessment on internal and external indices, accuracy, and time," *International Journal of Information Technology*, vol. 12, no. 4, p. 1377-1384, 2020.
- [21] B. S. Everitt, S. Landau, M. Leese dan D. Stahl, *Cluster analysis*, 5th Edition, London: John Wiley & Sons, 2011.
- [22] S. E. Saqila, I. P. Ferina dan A. Iskandar, "Analisis Perbandingan Kinerja Clustering Data Mining Untuk Normalisasi Dataset," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 5, no. 2, p. 356-365, 2023.
- [23] J. Faran dan R. T. Aldisa, "Analisis Data Mining dalam Komparasi Average Linkage AHC dan K-Means Clustering untuk Dataset Facebook Live Sellers," *Jurnal Media Informatika Budidarma*, vol. 7, no. 4, pp. 2041-2050, 2023.