Perbandingan Metode *K-Means*, *K-Medoids*, dan *AHC Ward* dalam Klasterisasi Kabupaten/Kota di Indonesia berdasarkan Indikator IPM

Salzi Rais Putra Kuswara *, Fauziah Roshafara

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

salzirais13@gmail.com, fauziahroshafara@unisba.ac.id

Abstract. This study aims to compare three clustering methods *K-Means*, *K-Medoids*, and *Agglomerative Hierarchical Clustering* (AHC) using the *Ward Method* in grouping 514 regencies/cities in Indonesia based on the Human Development Index (HDI) indicators, namely Life Expectancy, Expected Years of Schooling, Mean Years of Schooling, and Adjusted Per Capita Expenditure. The data were obtained from Statistics Indonesia (BPS) for the year 2024. Prior to clustering, data normalization was performed using the Min-Max Scaling method. The quality of clustering results was evaluated using the *Calinski-Harabasz Index*, with the number of clusters varied from k = 2 to k = 20. The results showed that *K-Means* and *K-Medoids* achieved optimal clustering at k = 2, with CHI values of 407.26 and 424.01, respectively. Meanwhile, AHC *Ward* yielded the optimal result at k = 3, with a CHI value of 301.17. Among the three methods, *K-Medoids* produced the highest index value, indicating the best clustering structure. These findings suggest that choosing the appropriate clustering method plays a crucial role in forming groups that optimally represent the regional characteristics.

Keywords: *AHC Ward Method, HDI, K-Means.*

Abstrak. Penelitian ini bertujuan untuk membandingkan tiga metode klasterisasi *K-Means*, *K-Medoids*, dan *Agglomerative Hierarchical Clustering* (AHC) dengan pendekatan *Ward* dalam mengelompokkan 514 kabupaten/kota di Indonesia berdasarkan indikator pembentuk Indeks Pembangunan Manusia (IPM), yaitu Umur Harapan Hidup, Harapan Lama Sekolah, Rata-rata Lama Sekolah, dan Pengeluaran per Kapita Disesuaikan. Data yang digunakan bersumber dari Badan Pusat Statistik tahun 2024. Sebelum dilakukan klasterisasi, data dinormalisasi menggunakan metode *Min-Max Scaling*. Evaluasi kualitas hasil klasterisasi dilakukan menggunakan *Calinski-Harabasz Index* dengan variasi jumlah klaster dari k = 2 hingga k = 20. Hasil analisis menunjukkan bahwa metode *K-Means* dan *K-Medoids* menghasilkan jumlah klaster optimal pada k = 2, dengan nilai CHI masing-masing sebesar 407,26 dan 424,01. Sementara itu, AHC *Ward* menunjukkan hasil optimal pada k = 3 dengan nilai CHI sebesar 301,17. Dari ketiga metode, *K-Medoids* memberikan nilai indeks tertinggi, yang mengindikasikan struktur klaster terbaik. Temuan ini menunjukkan bahwa pemilihan metode klasterisasi yang tepat berperan penting dalam menghasilkan kelompok yang merepresentasikan karakteristik daerah secara optimal.

Kata Kunci: *AHC Ward Method, IPM, K-Means.*

*salzirais13@gmail.com

A. Pendahuluan

Indeks Pembangunan Manusia (IPM) merupakan salah satu indikator penting dalam mengukur kemajuan pembangunan suatu daerah. IPM mencakup tiga dimensi utama, yaitu kesehatan, pendidikan, dan standar hidup layak, yang secara keseluruhan mencerminkan kualitas hidup masyarakat (BPS Kabupaten Bantaeng, 2018). Di Indonesia, disparitas dalam IPM antar kabupaten/kota sangat signifikan, mencerminkan ketidakmerataan pembangunan yang terjadi di berbagai wilayah (BPS, 2024). Oleh karena itu, penting untuk melakukan analisis yang mendalam mengenai pengelompokan kabupaten/kota berdasarkan indikator-indikator pembangunan manusia, sehingga mendukung perencanaan kebijakan yang lebih terfokus dan efektif. Salah satu pendekatan yang digunakan dalam menganalisis struktur data seperti ini adalah Analisis Klaster.

Analisis klaster merupakan salah satu pendekatan dalam analisis data yang bertujuan untuk mengelompokkan observasi ke dalam beberapa klaster berdasarkan tingkat kemiripan antar data (Jafar & Sivakumar, 2012). Teknik ini termasuk dalam cabang *data mining* dan menggunakan pendekatan *unsupervised learning*, di mana proses pengelompokan dilakukan secara otomatis melalui algoritma tertentu. Dalam praktiknya, telah dikembangkan berbagai metode klasterisasi yang memiliki keunggulan serta keterbatasan masing-masing. Perbedaan metode tersebut seringkali menghasilkan hasil pengelompokan yang tidak sama, meskipun diterapkan pada dataset yang identik. Sebagai ilustrasi, penelitian oleh Suraya & Wijatanto (2022) mengenai pengelompokan provinsi di Indonesia berdasarkan indeks khusus penanganan stunting menunjukkan variasi dalam jumlah dan komposisi klaster yang terbentuk. Metode *Agglomerative Hierarchical Clustering* (AHC) menghasilkan empat klaster, sementara *K-Means* dan *K-Medoids* menghasilkan dua klaster, dan *Fuzzy C-Means* membentuk tiga klaster.

Setiap metode klasterisasi memiliki kelebihan dan keterbatasan tersendiri, sehingga menentukan metode yang paling sesuai sering menjadi tantangan tersendiri dalam proses analisis. Tiga algoritma klasterisasi yang banyak digunakan dan cukup populer adalah *K-Means*, *K-Medoids*, dan *Agglomerative Hierarchical Clustering* (AHC) dengan pendekatan *Ward*. Ketiganya termasuk dalam kelompok *hard clustering*, di mana setiap objek hanya dapat dimasukkan ke dalam satu klaster secara eksklusif. Pemilihan ketiga metode tersebut dilakukan berdasarkan keunggulan dan karakteristik unik dari masing-masing algoritma. *K-Means* dipilih karena kemampuannya dalam mengolah data berukuran besar dengan efisiensi tinggi dan kompleksitas komputasi yang rendah (Amanah et al., 2024). Metode ini juga dikenal karena kemudahannya dalam implementasi dan interpretasi, sehingga menjadi salah satu teknik klasterisasi yang paling luas diterapkan dalam berbagai disiplin ilmu. Sementara itu, *K-Medoids* digunakan karena ketahanannya terhadap nilai pencilan dan kemampuannya dalam menangani bentuk klaster yang tidak beraturan (Lopes & Gosling, 2021). Di sisi lain, *Agglomerative Hierarchical Clustering* (AHC) dipilih karena dapat menangkap struktur hierarkis dalam data, memungkinkan pemahaman yang lebih dalam terhadap pola hubungan antar kelompok. Kelebihan lainnya adalah tidak perlunya menentukan jumlah klaster di awal analisis, memberikan fleksibilitas dalam eksplorasi jumlah klaster yang optimal. AHC juga lebih mampu mengakomodasi data dengan distribusi yang kompleks dibandingkan *K-Means* dan *K-Medoids*.

Oleh sebab itu, pemilihan metode klasterisasi yang paling optimal perlu dilakukan secara sistematis untuk menghasilkan klasterisasi yang representatif terhadap struktur data sebenarnya. Salah satu pendekatan untuk mencapainya adalah dengan membandingkan performa berbagai metode klasterisasi. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membandingkan metode *K-Means*, *K-Medoids*, dan AHC *Ward* dalam mengelompokkan kabupaten/kota di Indonesia berdasarkan indikator-indikator pembentuk IPM, guna memperoleh hasil klasterisasi yang akurat dan mendukung pengambilan kebijakan pembangunan yang lebih tepat sasaran.

B. Metode

Pada penelitian kali ini data yang digunakan merupakan data sekunder. Data didapatkan dari laman resmi Badan Pusat Statistik (BPS) yaitu data indikator dan nilai dari indeks pembangunan manusia dari 514 Kabupaten/Kota di Indonesia pada tahun 2024 yang terdiri dari variabel Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah, dan Pengeluaran per Kapita yang Disesuaikan.

Normalisasi Min-Max

Metode *Min-Max Normalization* merupakan teknik normalisasi yang mengubah skala data secara linier dari rentang awal ke rentang baru. Proses ini bertujuan untuk menjaga keseimbangan nilai perbandingan antara data sebelum dan setelah transformasi. Normalisasi Min-Max dilakukan dengan menyusun ulang data dalam skala antara 0 hingga 1, di mana nilai terkecil dipetakan menjadi 0, sementara nilai terbesar menjadi 1. Proses perhitungan normalisasi ini dapat dirumuskan sebagai berikut:

$$v_i' = \frac{v_i - min_A}{max_A - min_A}(new\_max_A - new\_min_A) + new\_min_A \quad (1)$$

Dimana:

v_i' = normalisasi nilai ke-i
v_i = data ke-i
min_A = nilai minimum pada data
max_A = nilai maksimum pada data
new_min_A = nilai minimum baru yang diinginkan pada data
new_max_A = nilai maksimum baru yang diinginkan pada data

K-Means

K-Means merupakan metode inisialisasi yang sederhana, di mana objek dibagi secara acak ke dalam *K* klaster, kemudian dihitung rata-rata nilai dalam setiap klaster sebagai pusat klaster. Selanjutnya, data dikelompokkan berdasarkan kedekatannya dengan pusat klaster (centroid), yang diperbarui secara berulang di setiap iterasi. Objek dikelompokkan ke dalam klaster yang memiliki karakteristik paling mirip, berdasarkan jarak euclidean antara objek dan rata-rata klaster (pusat klaster) (Han *et al*., 2012). Berikut adalah tahapan dalam algoritma *K-Means* (Han et al., 2012):

1. Tentukan jumlah klaster yang diinginkan sebanyak *K*.
2. Pilih secara acak *K* centroid awal sebagai pusat klaster.
3. Hitung jarak setiap objek ke *centroid* awal menggunakan jarak *Euclidean* dengan rumus berikut:

$$d(x_{il}, c_{jl}) = \sqrt{\sum_{l=1}^{n}(x_{il} - c_{jl})^2} \quad (2)$$

Dimana:

$d(x_{ij}, c_{ij})$ = Jarak *euclidean* antara objek ke-i dengan *centroid* ke-j.
x_{il} = Nilai objek ke-i pada variabel ke-l.
c_{jl} = Nilai *centroid* klaster ke-j pada variabel ke-l.

4. Kelompokkan setiap objek ke klaster dengan *centroid* terdekat.
5. Perbarui *centroid* setiap klaster menggunakan rumus berikut:

$$c_{jl} = \frac{1}{n_j}\sum_{l=1}^{n_j} x_{il} \quad (3)$$

Dimana:

c_{jl} = Nilai *centroid* klaster ke-j pada variabel ke-l.
x_{il} = Nilai objek ke-i pada variabel ke-l.
n_j = Jumlah objek yang menjadi anggota klaster ke-j.

6. Kelompokkan Kembali setiap objek ke klaster dengan *centroid* baru terdekat.
7. Ulangi Langkah 3 dan 6 hingga tidak ada perubahan pada *centroid* dan keanggotaan

klaster.

K-Medoids

K-Medoids adalah salah satu metode dalam analisis klaster yang mirip dengan *K-Means*, tetapi menggunakan *medoid* sebagai pusat klaster. *Medoid* adalah objek yang memiliki nilai *dissimilarity* (ketidaksamaan) minimum dengan semua objek lain dalam klaster. *K-Medoids* lebih tahan terhadap *outlier* dibandingkan dengan *K-Means*, yang menggunakan rata-rata (*centroid*) sebagai pusat klaster (Wierzchoń & Kłopotek, 2018). Berikut adalah tahapan dalam algoritma *K-Medoids* (Han et al., 2012):

1. Tentukan jumlah klaster sebanyak *K*.
2. Tentukan nilai *medoid* setiap klaster secara acak sebagai *medoid* awal.
3. Lakukan perhitungan jarak setiap objek dengan *medoid* menggunakan jarak *Euclidean* dengan rumus berikut:

$$d(x_{il}, m_{jl}) = \sqrt{\sum_{l=1}^{n} (x_{il} - m_{jl})^2} \quad (4)$$

Dimana :

$d(x_{il}, m_{jl})$ = Jarak *euclidean* antara objek ke-*i* dengan *medoid* ke-*j*.
x_{il} = Nilai objek ke-*i* pada variabel ke-*l.*
m_{jl} = Nilai *medoid* klaster ke-*j* pada variabel ke-*l.*

4. Kelompokkan setiap objek ke klaster dengan *Medoid* terdekat.
5. Hitung total jarak dengan menggunakan persamaan berikut:

$$E_b = \sum_{i=1}^{n} D_i \, , i = 1,2, \dots , n \quad (5)$$

Dimana D_i merupakan minimum jarak antara objek ke-i dengan *medoid* awal tiap klaster.

6. Pilih secara acak objek pada masing-masing klaster sebagai *medoid* baru.
7. Menghitung jarak setiap objek yang berada pada masing-masing klaster dengan *medoid* baru dengan persamaan (2.7).
8. Hitung total jarak dengan menggunakan persamaan (2.8).
9. Menghitung selisih total jarak S dengan persamaan berikut:

$$S = E_b - E_{b-1} \quad (6)$$

Keterangan :

S = Perubahan total jarak antara iterasi b dan iterasi sebelumnya (b−1).
E_b = Total jarak pada iterasi ke-b
E_{b-1} = Total jarak pada iterasi ke-b-1

10. Jika selisih total jarak $S < 0$, maka tukar objek dengan data klaster lainnya untuk membentuk sekupulan *K* objek baru sebagai *medoid*.
11. Lakukan pengulangan langkah 6 sampai 10 hingga nilai $S > 0$.

Agglomerative Hierarchical Clustering Ward Method

Metode *Agglomerative Hierarchical Clustering Ward Method* mengajukan pendekatan pengelompokan yang didasarkan pada minimisasi kehilangan informasi dengan cara menggabungkan objek ke dalam kelompok. Proses ini dievaluasi melalui jumlah deviasi kuadrat rata-rata dalam setiap klaster untuk masing-masing observasi. *Sum of Square Error* (SSE) digunakan sebagai kriteria utama dalam menentukan pengelompokan. Dua objek atau kelompok akan digabungkan apabila menghasilkan nilai fungsi tujuan yang paling kecil dibandingkan dengan kemungkinan kombinasi lainnya. Metode *Ward* juga dikenal sebagai salah satu metode yang baik dalam menangani pencilan, karena cenderung membentuk klaster yang seimbang dan mengurangi dampak objek-objek ekstrem terhadap struktur pengelompokan (Ward, 1963). Berikut adalah langkah penerapan *Ward Method* (Johnshon & Wichern, 2007):

a. Menghitung jarak setiap objek dengan persamaan (2.10).

$$d(x_{il}, y_{jl}) = \sqrt{\sum_{l=1}^{n} (x_{il} - y_{jl})^2} \quad (7)$$

Dimana :
$d(x_{il}, y_{jl})$= Jarak *euclidean* antara objek ke-*i* dengan objek ke-*j*.
x_{il} = Nilai objek ke-*i* pada variabel ke-*l.*
y_{jl} = Nilai objek ke-*j* pada variabel ke-*l.*
.

b. Mengidentifikasi pasangan objek dengan jarak paling kecil dalam $\boldsymbol{D} = \{d_{xy}\}$.
c. Menggabungkan objek tersebut, misalnya *A* dan *B*, untuk membentuk klaster baru (*AB*).
d. Menghitung jarak antara klaster (*AB*) dan klaster lainnya, misalnya *C*, menggunakan rumus berikut:

$$d_{(xy)z} = SS\,E\{d_{xz}, d_{yz}\} = \frac{[(n_x + n_z)d_{xz} + (n_y + n_z)d_{yz}] - n_z d_{xy}}{n_x + n_y + n_z} \quad (8)$$

Dimana:
$d_{(xy)z}$ = Jarak klaster *xy* dengan klaster *z*
d_{xy} = Jarak antara *x* dan *y*
d_{xz} = Jarak antara *x* dan *z*
d_{yz} = Jarak antara *y* dan *z*
n_x = Jumlah individu pada klaster *x*
n_y = Jumlah individu pada klaster *y*
n_z = Jumlah individu pada klaster *z*

e. Lakukan pengulangan langkah b sampai d hingga semua objek menjadi 1 klaster.

Calinski-Harabasz Index

Calinski-Harabasz Index (CHI), yang juga dikenal dengan sebutan *Variance Ratio Criterion*, merupakan salah satu indikator yang digunakan untuk menilai kualitas hasil klasterisasi dalam analisis klaster. Indeks ini menghitung perbandingan antara variansi antar klaster dengan variansi di dalam klaster. Semakin besar nilai CHI yang diperoleh, maka semakin baik tingkat pemisahan antar klaster yang terbentuk. Adapun rumus perhitungan *Calinski-Harabasz Index* disajikan sebagai berikut:

$$CH = \frac{n-k}{k-1} \times \frac{\boldsymbol{tr}(\boldsymbol{B})}{\boldsymbol{tr}(\boldsymbol{W})} \quad (9)$$

Keterangan :
$\boldsymbol{tr}(\boldsymbol{B})$ = varians antar klaster
$\boldsymbol{tr}(\boldsymbol{W})$ = varians dalam klister
k = jumlah klaster
n = jumlah banyak data.

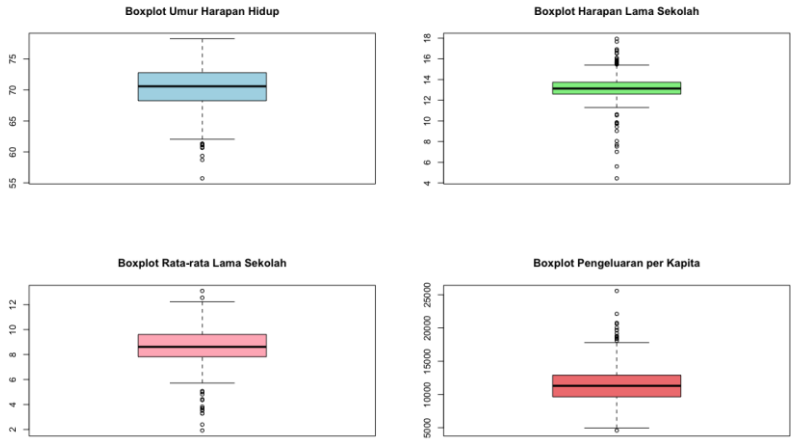
C. Hasil Penelitian dan Pembahasan

Bagian Analisis deskriptif digunakan untuk mendeskripsikan atau menjelaskan data yang akan digunakan dalam analisis. Dalam konteks ini, variabel yang dianalisis merupakan variabel dengan tipe data numerik seperti, UHH, HLS, RLS, dan Pengeluaran per Kapita Disesuaikan. Tabel 4.1 berikut menyajikan statistik deskriptif dari data yang dianalisis.

Tabel 1. Deskriptif Data

	UHH	HLS	RLS	Pengeluaran per Kapita Disesuaikan
Minimum	55.74	4.45	1.92	4597
Median	70.59	13.13	8.62	11305
Rata-rata	70.41	13.20	8.74	11435
Maksimum	78.26	17.94	13.1	25573
Simpangan Baku	3.39	1.30	1.61	2820.51

Berdasarkan tabel nilai minimum dan maksimum menunjukkan adanya variasi yang cukup besar antar kabupaten/kota, terutama pada indikator pengeluaran. Rata-rata dan median pada indikator UHH, HLS, dan RLS memiliki nilai yang cukup dekat, yang menandakan bahwa data ketiga indikator tersebut tersebar secara relatif merata dan tidak terlalu condong ke salah satu sisi. Sementara itu, simpangan baku terbesar terdapat pada indikator pengeluaran, menunjukkan bahwa nilai pengeluaran per kapita antar daerah sangat bervariasi dibandingkan indikator lainnya. Untuk melihat penyebaran dan potensi adanya pencilan (outlier) pada masing-masing indikator IPM secara visual, dapat diamati melalui gambar boxplot berikut:

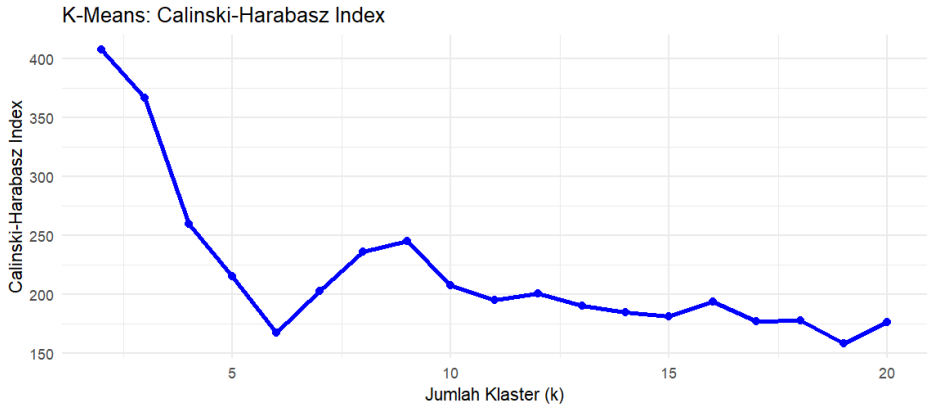


Gambar 1. Boxplot tiap variabel

Berdasarkan boxplot yang ditampilkan, dapat diindikasikan adanya nilai pencilan (outlier) pada keempat indikator IPM, yaitu Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran per Kapita Disesuaikan. Hal ini terlihat dari keberadaan titik-titik data yang berada di luar batas whisker pada masing-masing boxplot. Berdasarkan hal tersebut *K-Means* mungkin tidak terlalu cocok digunakan. Namun, *K-Means* akan tetap digunakan pada penelitian kali ini karena memiliki keunggulan dalam menangani data dengan jumlah besar.

Hasil Analisis *K-Means*

Setelah dilakukan proses normalisasi data menggunakan metode Min-Max Scaling, analisis klasterisasi selanjutnya dilakukan menggunakan algoritma *K-Means* dengan jumlah klaster (k) yang divariasikan dari 2 hingga 20. Hasil evaluasi kualitas klaster berdasarkan nilai *Calinski-Harabasz Index* untuk setiap k disajikan pada diagram berikut:

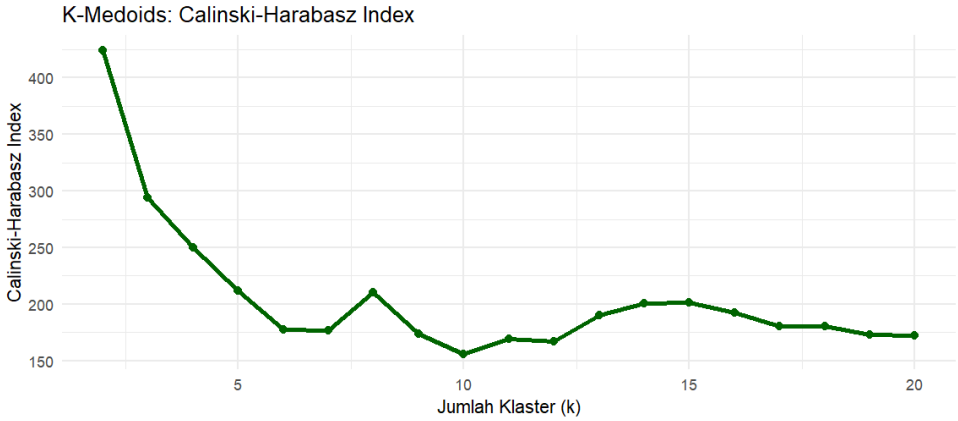


Gambar 2. Diagram niali CHI metode *K-Means*

Berdasarkan diagram tersebut, dapat dilihat bahwa nilai *Calinski-Harabasz Index,* tertinggi diperoleh pada saat jumlah klaster (k) = 2, yaitu sebesar 407,2613864, yang mengindikasikan bahwa struktur klaster paling optimal pada metode *K-Means* terjadi ketika data dibagi menjadi dua kelompok.

Hasil Analisis *K-Medoids*

Selanjutnya, analisis klasterisasi dilakukan menggunakan metode *K-Medoids* dengan jumlah klaster (k) yang divariasikan dari 2 hingga 20, seperti halnya pada metode *K-Means*. *K-Medoids* merupakan algoritma hard clustering yang mirip dengan *K-Means*, namun menggunakan medoid sebagai pusat klaster sehingga lebih tahan terhadap nilai pencilan atau data ekstrem. Evaluasi hasil klasterisasi untuk setiap nilai k dilakukan menggunakan *Calinski-Harabasz Index*, dan nilai-nilainya disajikan dalam bentuk diagram untuk mempermudah interpretasi.



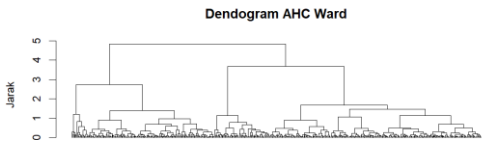
Gambar 3. Diagram niali CHI metode *K-Medoids*

Berdasarkan diagram *Calinski-Harabasz Index* pada metode *K-Medoids*, dapat dilihat bahwa nilai tertinggi diperoleh saat jumlah klaster (k) = 2, dengan nilai sebesar 424,0075833. Hal ini menunjukkan bahwa pengelompokan paling optimal menurut metode *K-Medoids* terjadi ketika data dibagi menjadi dua klaster. Setelah k = 2, nilai indeks cenderung menurun dan tidak menunjukkan peningkatan signifikan, yang mengindikasikan bahwa pembentukan lebih banyak klaster justru menghasilkan struktur klaster yang kurang baik dibandingkan dengan pembagian dua klaster.

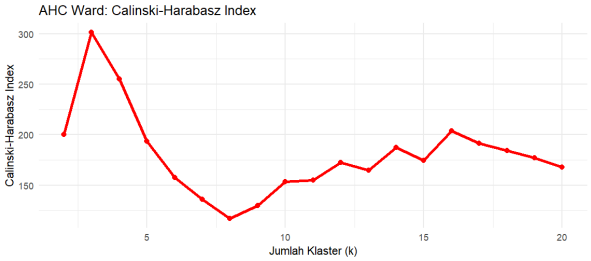
Hasil Analisis Agglomerative Hierarchical Clustering Ward Method

Selanjutnya, analisis klasterisasi dilakukan menggunakan metode *Agglomerative Hierarchical Clustering* (AHC) dengan pendekatan *Ward*. Berbeda dengan metode sebelumnya, AHC membentuk klaster secara bertahap melalui proses penggabungan antar objek atau kelompok yang paling mirip berdasarkan matriks jarak. Pendekatan *Ward* digunakan karena mampu meminimalkan variasi dalam klaster yang terbentuk. Hasil dari proses ini divisualisasikan dalam bentuk dendogram untuk

menggambarkan struktur hierarki pengelompokan, serta diagram *Calinski-Harabasz Index* untuk mengevaluasi kualitas klaster yang terbentuk dari jumlah klaster 2 hingga 20. Berikut disajikan dendogram hasil AHC dan grafik nilai *Calinski-Harabasz Index* pada berbagai jumlah klaster.



Gambar 4. Dendogram AHC *Ward*



Gambar 5. Diagram niali CHI metode AHC *Ward*

Berdasarkan dendogram yang dihasilkan oleh metode AHC *Ward*, terlihat bahwa struktur hierarki pengelompokan terbentuk secara bertahap melalui proses penggabungan antar objek yang memiliki kemiripan tertinggi. Dendogram ini memberikan gambaran visual mengenai hubungan antar data dan membantu dalam menentukan jumlah klaster yang optimal. Selain itu, berdasarkan diagram garis *Calinski-Harabasz Index*, diketahui bahwa nilai tertinggi diperoleh saat jumlah klaster **k = 3**, yaitu sebesar 301,1736490. Hal ini menunjukkan bahwa pada metode AHC *Ward*, struktur klaster yang paling baik terbentuk ketika data dibagi menjadi tiga kelompok.

Perbandingan Metode

Berdasarkan hasil evaluasi menggunakan *Calinski-Harabasz Index* , diperoleh jumlah klaster optimal yang berbeda untuk setiap metode klasterisasi. Metode *K-Means* menunjukkan nilai indeks tertinggi sebesar 407,2614 pada saat jumlah klaster k = 2, yang mengindikasikan bahwa struktur klaster terbaik terbentuk ketika data dibagi menjadi dua kelompok. Hasil serupa juga ditunjukkan oleh metode *K-Medoids*, dengan nilai indeks tertinggi mencapai 424,0076 pada k = 2, yang sekaligus menjadi nilai CHI tertinggi di antara ketiga metode. Sementara itu, metode *Agglomerative Hierarchical Clustering* (AHC) dengan pendekatan *Ward* menghasilkan nilai Calinski-Harabasz optimal sebesar 301,1736 pada k = 3, sehingga struktur klaster terbaik pada metode ini terbentuk saat data dibagi menjadi tiga kelompok. Secara umum, ketiga metode menghasilkan jumlah klaster optimal yang relatif kecil, namun nilai indeks tertinggi diperoleh oleh *K-Medoids*, yang menunjukkan bahwa metode ini memberikan pengelompokan dengan kualitas pemisahan antar klaster yang paling baik.

D. Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan, diperoleh bahwa ketiga metode klasterisasi, yaitu *K-Means*, *K-Medoids*, dan *Agglomerative Hierarchical Clustering* (AHC) dengan pendekatan *Ward*, menghasilkan jumlah klaster optimal yang berbeda. *K-Means* dan *K-Medoids* menunjukkan hasil pengelompokan terbaik saat data dibagi menjadi dua klaster, dengan nilai *Calinski-Harabasz Index* masing-masing sebesar 407,2614 dan 424,0076. Sementara itu, metode AHC *Ward* menunjukkan hasil optimal pada tiga klaster dengan nilai indeks sebesar 301,1736. Dari ketiga metode yang digunakan, *K-Medoids* menghasilkan nilai indeks tertinggi, yang mengindikasikan bahwa metode ini mampu memberikan struktur klaster dengan kualitas pemisahan yang paling baik.

Ucapan Terimakasih

Penulis menyadari bahwa menyelesaikan artikel ini bukanlah hal yang mudah tanpa adanya bantuan dan dukungan dari berbagai pihak. Oleh karena itu, penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada Ibu Fauziah Roshafara, S.Stat., M.Stat. selaku dosen pembimbing yang telah meluangkan waktu, memberikan tenaga, ilmu, serta bimbingan yang sangat berarti. Ucapan terima kasih juga disampaikan kepada seluruh dosen Program Studi Statistika atas ilmu dan wawasan yang diberikan selama masa perkuliahan, serta kepada teman-teman sejurusan yang selalu memberikan semangat dan motivasi sepanjang proses studi.

Daftar Pustaka

Amanah, F., Roshafara, F., Lestari, P. I., Salsabila, S., & Maharani, R. (2024). Utilizing K-Means Clustering for Constructing Black-Litterman Portfolio Models. *Jurnal Matematika, Statistika Dan Komputasi*, *20*(3), 670–679. https://doi.org/10.20956/j.v20i3.34165

Athira, N., & Herlina, M. (2022). Identifikasi Faktor yang Mempengaruhi Data Driven Decision pada Pemerintah Desa Menggunakan SEM GSCA. *Jurnal Riset Statistika*, 145–152. https://doi.org/10.29313/jrs.v2i2.1458

BPS. (2024). *Indeks Pembangunan Manusia 2023* (Vol. 18).

BPS Kabupaten Bantaeng. (2018). Indeks Pembangunan Manusia Kabupaten Bantaeng 2017. In *BPS Kabupaten Bantaeng* (Issue 73).

Budiman, A. S., & Hajarisman, N. (2024). Analisis Mediasi Multipel Paralel Kausal Step pada Data Stunting menurut Kabupaten/Kota. *Jurnal Riset Statistika*, 31–40. https://doi.org/10.29313/jrs.v4i1.3860

Camartya, D., & Achmad, A. I. (2022). Analisis Korespondensi pada Jumlah Pengangguran Terbuka Menurut Kabupaten/Kota Berdasarkan Pendidikan Tertinggi. *Jurnal Riset Statistika*, 119–128. https://doi.org/10.29313/jrs.v2i2.1424

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. In *IIE Transactions*.

Jafar, M., & Sivakumar, R. (2012). A study on possibilistic and fuzzy possibilistic C-means clustering algorithms for data clustering. *IEEE Proceedings of the INternational Conference On Emerging Trends in Science Engineering and Technology: Recent Advancements on Science and Engineering Innovation, INCOSET 2012*. https://doi.org/10.1109/incoset.2012.6513887

Johnshon, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis. In *Technometrics* (6th Editio). Pearson.

Lopes, H. E. G., & Gosling, M. de S. (2021). Cluster Analysis in Practice: Dealing with Outliers in Managerial Research. *Revista de Administração Contemporânea*, *25*(1), 1–19.

Putri, H. J., & Ramdani, Y. (2024). Prediksi Hujan Rencana Tahunan Di Stasiun Palolo, Sulawesi Tengah menggunakan Metode Gumbel, Log Normal, Dan Log Pearson III. *DataMath: Journal of Statistics and Mathematics*, *2*(1), 17–24.

Suraya, G. R., & Wijayanto, A. W. (2022). Comparison of Hierarchical Clustering, K-Means, K-Medoids, and Fuzzy C-Means Methods in Grouping Provinces in Indonesia according to the Special Index for Handling Stunting. *Indonesian Journal of Statistics and Its Applications*, *6*(2), 180–201. https://doi.org/10.29244/ijsa.v6i2p180-201

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. In *Journal of the American Statistical Association* (Vol. 58, Issue 301, pp. 236–244). https://doi.org/10.1080/01621459.1963.10500845

Wierzchoń, S. T., & Kłopotek, M. A. (2018). *Modern Algorithms of Cluster Analysis*.