

Evaluasi *Agglomerative Hierarchical Clustering* pada Potensi Provinsi di Bidang Hiburan dan Olahraga Menggunakan Korelasi *Cophenetic*

Sylvia Syifa Rahayu *, Ilham Faishal Mahdy

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

ssr.sylviasr@gmail.com, ilham.faishal@unisba.ac.id

Abstract. The potential of provinces in Indonesia is highly diverse, reflecting the richness of resources and cultural diversity that can be leveraged to enhance development and societal welfare. Assessing provincial potential involves identifying strengths and opportunities to optimize resource utilization, encompassing economic, social, environmental, and infrastructural aspects. One significant potential lies in entertainment and sports facilities, which play a role in maintaining physical and mental health, strengthening social life, and delivering positive economic impacts through job creation and increased local economic activity. To determine the extent to which provincial potential has developed, the Agglomerative Hierarchical Clustering (AHC) method is employed. This method groups provinces based on characteristic similarities using the Mahalanobis distance matrix, forming a single large cluster, which is then evaluated for validity using the cophenetic correlation coefficient. This study utilizes secondary data from publications on the BPS website, covering information on the number of entertainment and sports facilities in each urban village/rural area across Indonesian provinces. Based on the clustering results, the study applied the Single Linkage, Complete Linkage, Average Linkage, Ward Method, and Centroid Method to classify provinces into two clusters: high and low potential. The validity results, measured using cophenetic correlation, show that the Average Linkage method achieves the highest correlation, making it the most optimal method. Conversely, the Ward Method exhibits the lowest correlation, indicating that its application requires further consideration.

Keywords: *Agglomerative Hierarchical Clustering, Cophenetic Correlation Coefficient, Mahalanobis Distance, Provincial Potential.*

Abstrak. Potensi provinsi di Indonesia sangat beragam, mencerminkan kekayaan sumber daya dan keanekaragaman budaya yang dapat dimanfaatkan untuk meningkatkan pembangunan dan kesejahteraan masyarakat. Penilaian potensi provinsi melibatkan identifikasi kekuatan dan peluang untuk mengoptimalkan penggunaan sumber daya, mencakup aspek ekonomi, sosial, lingkungan, dan infrastruktur. Salah satu potensi penting adalah fasilitas hiburan dan olahraga, yang berperan dalam menjaga kesehatan fisik dan mental, memperkuat kehidupan sosial, dan memberikan dampak ekonomi positif melalui penciptaan lapangan kerja dan peningkatan aktivitas ekonomi lokal. Untuk mengetahui sejauh mana potensi provinsi ini telah berkembang, digunakan metode *Agglomerative Hierarchical Clustering* (AHC) yang mengelompokkan berdasarkan kesamaan karakteristik menggunakan matrik jarak Mahalanobis, hingga membentuk satu *cluster* besar dan dievaluasi validitas menggunakan koefisien korelasi *cophenetic*. Penelitian ini menggunakan data sekunder dari publikasi di situs BPS, yang mencakup informasi tentang jumlah fasilitas hiburan dan olahraga perkelurahan/desa di setiap provinsi Indonesia. Berdasarkan hasil penggabungan *cluster* penelitian ini menggunakan metode *Single Linkage, Complete Linkage, Average Linkage, Ward Method, dan Centroid Method* menghasilkan dua *cluster* berdasarkan potensi provinsi tinggi dan rendah. Didapat hasil validitas dengan korelasi *cophenetic* menunjukkan bahwa *Average Linkage* memiliki korelasi tertinggi, menjadikannya metode yang paling optimal. Sebaliknya, *Ward Method* menunjukkan korelasi terendah, sehingga penggunaannya perlu dipertimbangkan lebih lanjut.

Kata Kunci: *Agglomerative Hierarchical Clustering, Jarak Mahalanobis, Koefisien Korelasi Cophenetic, Potensi Provinsi.*

A. Pendahuluan

Potensi provinsi di Indonesia mencerminkan kekayaan sumber daya dan budaya yang beragam, yang dapat dioptimalkan untuk pembangunan berkelanjutan dan kesejahteraan masyarakat. Salah satu aspek penting adalah hiburan dan olahraga, yang tidak hanya mendukung kesehatan fisik dan mental tetapi juga memperkuat ikatan sosial, mengembangkan keterampilan individu, dan memberikan dampak ekonomi positif melalui penciptaan lapangan kerja dan peningkatan aktivitas ekonomi lokal. Dengan memanfaatkan fasilitas ini secara efektif, provinsi dapat memperbaiki kualitas hidup, melestarikan budaya, dan mendorong pertumbuhan ekonomi yang inklusif. Untuk mencapai hal ini, diperlukan identifikasi dan analisis potensi setiap provinsi agar pengelolaannya lebih terarah dan berdampak luas.

Salah satu analisis yang dapat melihat potensi setiap provinsi adalah analisis *cluster*. Penelitian ini menggunakan analisis *cluster* hierarki dengan metode *Agglomerative Hierarchical Clustering* (AHC), yang dimulai dengan setiap objek berada dalam *cluster* terkecil sampai *cluster* terbesar, kemudian digabungkan dengan *cluster* lain yang memiliki karakteristik serupa (Sabara, 2022). Metode ini dimulai dengan menganggap setiap data sebagai *cluster* individu dan kemudian secara iteratif menggabungkan *cluster* – *cluster* tersebut berdasarkan matriks jarak atau kesamaan hingga semua data tergabung dalam satu *cluster* besar atau mencapai jumlah *cluster* yang diinginkan (Oktaviani & Rifai, 2024). Dalam penggabungan *cluster* penelitian ini menggunakan metode penggabungan yang digunakan meliputi *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Ward Method*, dan *Centroid Method*. Prosesnya divisualisasikan melalui dendrogram, yang menggambarkan hubungan hierarkis antar *cluster* di setiap tahap penggabungan. Matriks jarak yang digunakan pada penelitian ini adalah matriks mahalanobis. Dimana jarak Mahalanobis adalah ukuran jarak yang memperhitungkan pola korelasi antara variabel-variabel yang memiliki hubungan berbeda, yang dapat diidentifikasi dan dianalisis (Sari, D. P., & Kusdarwati, H, 2015). Dalam pengelompokan yang menggunakan jarak Mahalanobis, data asli tetap digunakan meskipun ada korelasi antar variabel, dan jika terdapat korelasi, analisis komponen utama tidak perlu dilakukan.

Untuk mengevaluasi seberapa baik dendrogram yang dihasilkan oleh *agglomerative hierarchical clustering* mencerminkan jarak atau kesamaan yang sebenarnya antara data, maka dilakukan pengukuran menggunakan koefisien korelasi *cophenetic*. Koefisien korelasi *cophenetic* merupakan pengukuran tingkat korelasi antara elemen – elemen dari matriks ketidakmiripan asli dengan elemen – elemen matriks *cophenetic* dari dendrogram (Silva & Dias, 2013). Nilai koefisien ini berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa dendrogram dengan baik mencerminkan struktur jarak yang asli, sedangkan nilai mendekati 0 atau negatif menunjukkan ketidaksesuaian.

Berdasarkan latar belakang yang telah diuraikan, penelitian ini bertujuan mengevaluasi metode *Agglomerative Hierarchical Clustering* (AHC) dalam menganalisis potensi provinsi di Indonesia berdasarkan jumlah fasilitas hiburan dan olahraga perkelurahan/desa tahun 2021, menggunakan validitas koefisien korelasi *cophenetic*. Oleh karena itu, penelitian ini diberi judul “Evaluasi *Agglomerative Hierarchical Clustering* pada Potensi Provinsi di Bidang Hiburan dan Olahraga Menggunakan Korelasi *Cophenetic*”. Selanjutnya, tujuan dalam penelitian ini sebagai berikut :

1. Bagaimana pengelompokan potensi provinsi di bidang hiburan dan olahraga di Indonesia pada tahun 2021 menggunakan metode *Agglomerative Hierarchical Clustering* (AHC)?
2. Bagaimana evaluasi pengelompokan dari hasil metode *Agglomerative Hierarchical Clustering* (AHC) pada potensi provinsi di bidang hiburan dan olahraga di Indonesia pada tahun 2021 menggunakan korelasi *cophenetic*?

B. Metode

Data yang digunakan adalah data sekunder yang bersumber dari buku publikasi di website Badan Pusat Statistik (BPS) yang berjudul Statistik Potensi Desa 2021. Dalam buku publikasi ini terdapat banyak data, dalam penelitian ini digunakan data potensi di bidang hiburan dan olahraga perkelurahan/desa pada setiap provinsi pada tahun 2021. Dalam melakukan pengujian analisis *cluster*, peneliti menggunakan bantuan *software Microsoft Excel* dan *Rstudio* dengan bahasa pemrograman *R*.

Langkah Analisis Data

1. Input data Potensi Provinsi.
2. Memeriksa ada tidaknya data *missing*. Jika ada maka dilakukan penanganan data *missing* dengan rumus sebagai berikut [13]:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Keterangan :

$\bar{x}$ = Rata – rata

x_i = Data ke- i

i = Data ke-1 sampai data ke- n

n = Banyaknya data

3. Hitung ukuran kesamaan antar objek menggunakan jarak Mahalanobis dengan rumus sebagai berikut [10]:

$$d(i, j) = \sqrt{(x_i - x_j)^T S^{-1} (x_i - x_j)}$$

Keterangan :

$d_{(i,j)}$ = Jarak pengamatan antara objek- i ke objek- j

x_i = Nilai vektor untuk objek ke- i

x_j = Nilai vektor untuk objek ke- j

S^{-1} = Invers matriks ragam peragam

4. Menentukan jumlah cluster menggunakan evaluasi internal *cluster* sebagai berikut [6]:
 - a. *Connectivity* yaitu mengukur seberapa baik titik-titik data yang berdekatan dikelompokkan ke dalam *cluster* yang sama. Dimana semakin kecil nilai konektivitas, semakin baik klusterisasi. Dengan rumus sebagai berikut:

$$Conn = \sum_{i=1}^N \sum_{j=1}^L x_{i,nn_{i(j)}}$$

Keterangan :

$Conn$ = Koefisien *Connectivity*

$x_{i,nn_{i(j)}}$ = Merupakan titik data ke- i dan tetangga terdekat ke- j dari data x_i

N = Jumlah data observasi

L = Paramater jumlah tetangga terdekat yang dipertimbangkan

- b. *Dunn Index* yaitu mengevaluasi kualitas klusterisasi dengan membandingkan jarak antar *cluster* dan ukuran dalam *cluster*. Dimana semakin tinggi nilai *Dunn Index*, semakin baik klusterisasi. Dengan persamaan berikut :

$$D = \frac{\min_{1 \leq i \neq j \leq k} \delta(C_i, C_j)}{\max_{1 \leq l \leq k} \Delta(C_l)}$$

Keterangan :

C_i dan C_j = Dua *cluster* yang berbeda

$\delta(C_i, C_j)$ = Jarak minimum antar-*cluster* C_i dan C_j (jarak centroid atau jarak - terdekat antar titik).

$\Delta(C_l)$ = Diameter maksimum dari *cluster* C_l (jarak maksimum antar titik dalam klaster)

- c. *Silhouette Coefficient* yaitu mengukur seberapa baik setiap titik data ditempatkan dalam *cluster* yang benar dengan membandingkan jarak rata-rata ke titik-titik dalam klaster yang sama dan jarak rata-rata ke *cluster* terdekat lainnya. Dimana nilai koefisien *silhouette* yang mendekati 1 menunjukkan klaster yang baik, sedangkan yang mendekati -1 dianggap kurang baik. Dengan rumus sebagai berikut::

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Keterangan :

$s(i)$ = Koefisien *Silhouette*

$a(i)$ = Jarak rata-rata antara objek ke- i dengan semua titik lain dalam *cluster* yang sama.

$b(i)$ = Jarak rata-rata antara objek ke- i dengan semua titik di *cluster* terdekat.

5. Lakukan pengelompokan dengan *Agglomerative Hierarchical Clustering* (AHC) dengan menggabungkan objek x dan y yang memiliki jarak terdekat, kemudian jadikan satu *cluster*. Dari *cluster* xy yang sudah terbentuk, hitung jaraknya dengan objek lain yang belum bergabung menggunakan rumus sebagai berikut [2]:
 - a. Single Linkage : $d_{(XY)Z} = \min\{d_{XZ}, d_{YZ}\}$
 - b. Complete Linkage : $d_{(XY)Z} = \max\{d_{XZ}, d_{YZ}\}$
 - c. Average Linkage : $d_{(XY)Z} = \text{avr}\{d_{XZ}, d_{YZ}\}$
 - d. Ward Method : $d_{(XY)Z} = SSE\{d_{XZ}, d_{YZ}\} = \frac{[(n_X+n_Z)d_{XZ} + (n_Y+n_Z)d_{YZ}] - n_Z d_{XY}}{n_X+n_Y+n_Z}$
 - e. Centroid Method : $d_{(XY)Z} = \frac{n_X}{n_X+n_Y} d_{XZ} + \frac{n_Y}{n_X+n_Y} d_{YZ} - \frac{n_X n_Y}{(n_X+n_Y)^2}$
6. Perbarui matriks ukuran jarak setiap kali ada penggabungan objek baru. Ulangi langkah tersebut hingga semua objek tidak dapat berpindah lagi dan bergabung menjadi satu *cluster* besar. Buat dendrogram dari masing-masing hasil pengelompokan
7. Interpretasikan hasil analisis *cluster* sesuai metode yang ditentukan.
8. Pilih metode terbaik dengan uji validitas menggunakan Koefisien Korelasi *Cophenetic* dengan rumus sebagai berikut [11]:

$$r_{coph} = \frac{\sum_{i < k} (d_{ik} - \bar{d})(d_{c_{ik}} - \bar{d}_c)}{\sqrt{[\sum_{i < k} (d_{ik} - \bar{d})^2][\sum_{i < k} (d_{c_{ik}} - \bar{d}_c)^2]}}$$

Keterangan :

r_{coph} = koefisien korelasi *cophenetic*

d_{ik} = jarak asli (jarak Mahalanobis) antara objek i dan k

$\bar{d}$ = rata – rata d_{ik}

$d_{c_{ik}}$ = jarak dendrogram (matriks *cophenetic*) yg terbentuk objek i dan k

$\bar{d}_c$ = rata – rata $d_{c_{ik}}$

Nilai r_{coph} dari rentang -1 sampai 1, dimana jika nilai r_{coph} mendekati 1 artinya proses klasterisasi yang dihasilkan baik.

9. Kesimpulan.

C. Hasil Penelitian dan Pembahasan

Jarak Mahalanobis

Langkah pertama adalah menentukan kedekatan antar objek menggunakan jarak Mahalanobis, didapat sebagai berikut.

$$D_{Mahalanobis} = \begin{pmatrix} 0 & 4.7503 & \cdots & 2.8602 & 2.9704 \\ 4.7503 & 0 & \cdots & 4.5659 & 5.1650 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 2.8602 & 4.5659 & \cdots & 0 & 1.6327 \\ 2.9704 & 5.1650 & \cdots & 1.6327 & 0 \end{pmatrix}$$

Setelah mendapatkan matriks jarak Mahalanobis, selanjutnya tentukan jarak data terdekat pada hasil matriks jarak Mahalanobis, didapat sebagai berikut.

$$\text{Min}(D_{Mahalanobis}) = d_{33,34} = 1.6327$$

Didapatkan jarak terdekat pada data ke – 33 (Papua Barat) dan data ke – 34 (Papua) sebesar 1.6327. Dari kedua data tersebut digabungkan menjadi satu kelompok. Ini merupakan langkah awal untuk setiap metode yang akan diterapkan.

Evaluasi Internal Cluster

Untuk menghindari subjektif ataupun interpretasi yang berbeda dalam menentukan jumlah *cluster* yang optimal dari dendrogram, maka diperlukan evaluasi internal *cluster* untuk menentukan banyak *cluster* hirarki yang optimal. Dalam penelitian ini, digunakan bantuan *software R-Studio* menggunakan *packages clValid*. Berikut hasil yang didapatkan.

Tabel 1. Validasi *Cluster* Internal

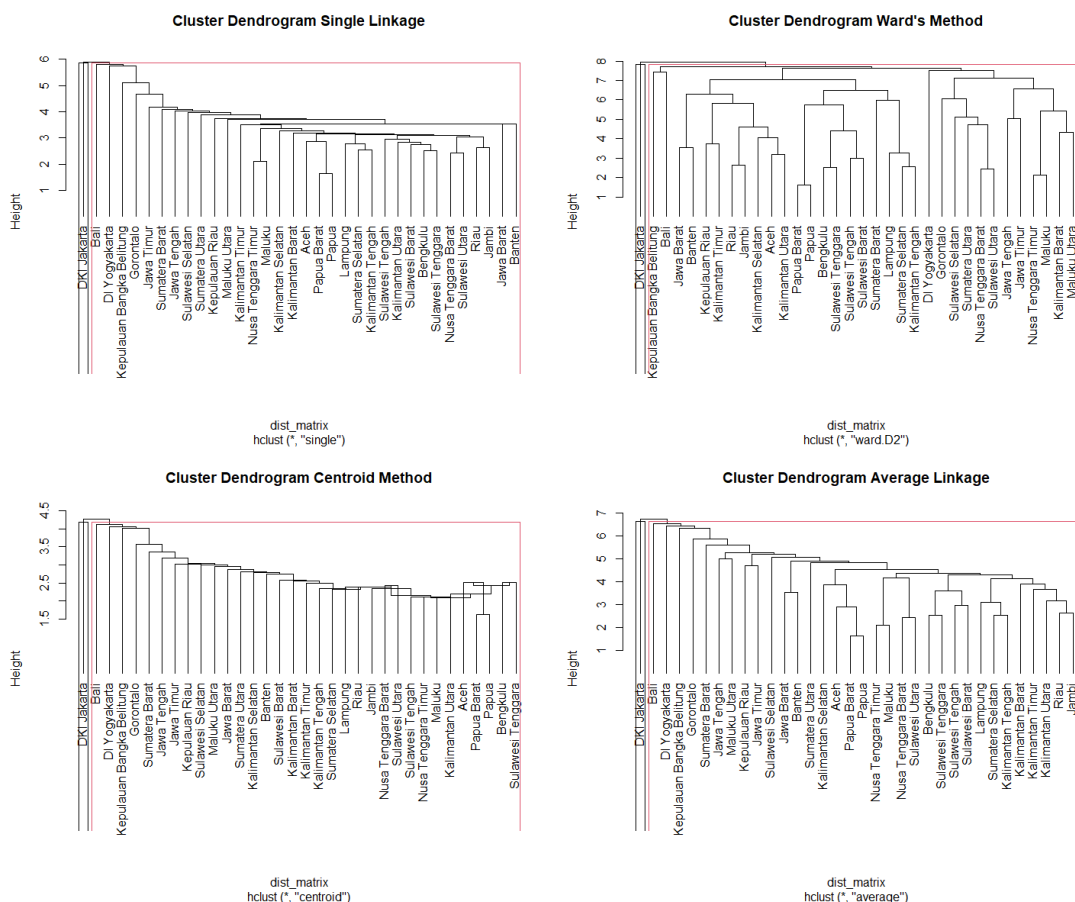
	Score	Clusters
Connectivity	2.9290	2
Dunn	0.6063	2
Silhouette	0.5277	2

Pada Tabel 1. Terlihat nilai *score* baik *connectivity*, *dunn index*, maupun *silhouette coefficient* menunjukkan hasil yang sama, yaitu 2 *cluster* yang memiliki nilai optimal. Oleh karena itu, pengelompokan Potensi Provinsi berdasarkan bidang olahraga dan hiburan di Indonesia dibagi menjadi 2 *cluster* yaitu tinggi dan rendah. Setelah itu, perhitungan *Agglomerative Hierarchical Clustering* dapat dilakukan pada setiap metode.

Agglomerative Hierarchical Clustering (AHC)

Hasil metode Single Linkage, Average Linkage, Ward Method dan Centroid Method

Pada tahap awal, telah ditemukan jarak terdekat, yaitu antara data ke – 33 (Papua Barat) dan data ke – 34 (Papua) sebesar 1.6327. Selanjutnya, langkah berikutnya adalah menghitung jarak antara kelompok yang terbentuk dari data ke – 33 dan 34 dengan kelompok – kelompok lain yang tersisa seperti kelompok 1, 2 , ... , 34 menggunakan rumus metode Dendrogram *Single Linkage*, *Average Linkage*, *Ward Method* dan *Centroid Method*. Sehingga, didapatkan hasil pengelompokan potensi provinsi di bidang hiburan dan olahraga sebagai berikut:



Gambar 1. Dendrogram Single Linkage, Average Linkage, Ward Method dan Centroid*Method*

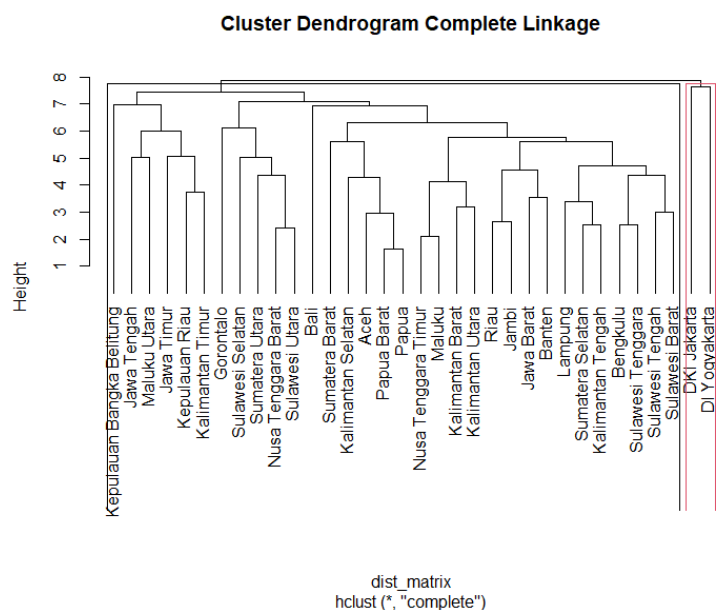
Didapat anggota tiap *cluster* sebagai berikut :

Tabel 2. Hasil klasterisasi Single Linkage, Average Linkage, Ward Method dan Centroid Method

Cluster	Potensi Provinsi	Jumlah
1	DKI Jakarta	1
2	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua	33

Hasil metode Complete Linkage

Pada tahap awal, telah ditemukan jarak terdekat, yaitu antara data ke – 33 (Papua Barat) dan data ke – 34 (Papua) sebesar 1.6327. Selanjutnya, langkah berikutnya adalah menghitung jarak antara kelompok yang terbentuk dari data ke – 33 dan 34 dengan kelompok – kelompok lain yang tersisa seperti kelompok 1, 2 , ... , 34 menggunakan rumus metode *Complete Linkage*. Sehingga, didapatkan hasil pengelompokan potensi provinsi di bidang hiburan dan olahraga sebagai berikut:

**Gambar 2.** Dendrogram *Complete Linkage*

Didapat anggota tiap *cluster* sebagai berikut :

Tabel 3. Hasil klasterisasi *Complete Linkage*

Cluster	Potensi Provinsi	Jumlah
1	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah,	32

Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua	
2	DKI Jakarta, DI Yogyakarta

Validitas Hasil Cluster

Untuk mengetahui metode *cluster* yang terbaik maka dicari validitas hasil *cluster* menggunakan koefisien korelasi *cophenetic*. Perhitungan nilai koefisien korelasi *cophenetic* (r_{coph}), yaitu koefisien korelasi antara objek pada matriks jarak (jarak Mahalanobis) dan objek pada matriks jarak *cophenetic*. Didapatkan hasil perhitungan koefisien korelasi *cophenetic* sebagai berikut.

Tabel 4. Hasil koefisien korelasi *Cophenetic*

Metode	Korelasi <i>Cophenetic</i>
<i>Single Linkage</i>	0.8054
<i>Complete Linkage</i>	0.6309
<i>Average Linkage</i>	0.8447
<i>Ward Method</i>	0.5304
<i>Centroid Method</i>	0.8148

Berdasarkan Tabel 4 dapat dilihat, baik itu metode *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Ward Method*, ataupun *Centroid Method* yang diuji menunjukkan korelasi yang sangat tinggi dengan nilai r_{coph} mendekati angka 1, yang menunjukkan bahwa masing-masing metode ini menghasilkan pengelompokan yang valid dan konsisten dengan struktur data asli

Di antara metode yang diuji, *Average Linkage* memiliki korelasi tertinggi sebesar 0.8447, yang berarti metode ini yang paling direkomendasikan untuk pengelompokan dalam analisis. *Ward Method* meskipun valid, memiliki korelasi yang sedikit lebih rendah sebesar 0.5304 dibandingkan dengan metode lainnya, sehingga perlu diperhatikan jika digunakan dalam analisis lebih lanjut. Secara keseluruhan, *Average Linkage* adalah metode paling konsisten dengan struktur data asli.

D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

1. Dari pengelompokan menggunakan metode *Agglomerative Hierarchical Clustering* (AHC) ditentukan 2 *cluster* potensi provinsi yang tinggi dan rendah. Dimana pada *Single Linkage*, *Average Linkage*, *Ward method* dan *Centroid Method*, *cluster* pertama terdapat 1 Provinsi dengan potensi provinsi tinggi dan *cluster* ke-2 terdapat 33 provinsi dengan potensi provinsi rendah. Sedangkan pada *Complete Linkage*, *cluster* pertama terdapat 32 Provinsi dengan potensi provinsi rendah dan *cluster* ke-2 terdapat 2 provinsi dengan potensi provinsi tinggi.
2. Hasil dari pengujian validitas menggunakan korelasi *cophenetic* pada metode *Agglomerative Hierarchical Clustering* (AHC). Menunjukkan bahwa *Average Linkage* memiliki korelasi tertinggi, menjadikannya metode yang paling terbaik. *Ward Method* memiliki korelasi lebih rendah, sehingga penggunaannya perlu dipertimbangkan lebih lanjut.

Ucapan Terimakasih

Peneliti mengucapkan terima kasih kepada Ilham Faishal Mahdy, S.Stat., M.Stat. selaku dosen pembimbing atas arahan dan saran yang diberikan, serta kepada keluarga, sahabat dan rekan mahasiswa yang telah memberikan dukungan moral.

Daftar Pustaka

- Oktaviani, N. K., & Rifai, N. A. K. (2024). Pemodelan ARFIMA dengan Estimasi Parameter Pembeda Menggunakan Metode Geweke Porter-Hudak. *Jurnal Riset Statistika*, 11–20. <https://doi.org/10.29313/jrs.v4i1.3835>

- Badan Pusat Statistik Indonesia. (2021). Statistik Potensi Desa Indonesia 2021. Bps.go.id; Badan Pusat Statistik Indonesia. <https://www.bps.go.id/id/publication/2022/03/24/ceab4ec9f942b1a4fdf4cd08/statistik-potensi-desa-indonesia-2021.html>.
- Dewi, A. F., & Ahadiyah, K. (2022). Agglomerative Hierarchy Clustering Pada Penentuan Kelompok Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pendidikan. *Zeta-Math Journal*, 7(2), 57-63.
- Fadliana, A. (2015). Penerapan metode Agglomerative Hierarchical Clustering untuk klasifikasi Kabupaten/Kota di Provinsi Jawa Timur berdasarkan kualitas pelayanan keluarga berencana (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. Seventh Ed. New Jersey: Prentice Hall International, Inc.
- Mardhiyah, F. A., & Herlina, M. (2024). Pengelompokan Data Jenis Kejahatan di Indonesia Menggunakan Metode Agglomerative Hierarchical Clustering (AHC) pada Tahun 2021. In *Bandung Conference Series: Statistics* (Vol. 4, No. 1, pp. 1-7).
- Prayoga, S. W., & Pramana, S. (2024). Pemetaan Kejadian Balita Stunting Melalui Integrasi Citra Satelit Multisumber dan Official Statistics di Provinsi Nusa Tenggara Timur Menggunakan Machine Learning. In *PROSIDING SEMINAR NASIONAL SAINS DATA* (Vol. 4, No. 1, pp. 434-444).
- Ryan, T. P. (1997). *Modern Regression Methods*. New York: John Wiley and Sons.
- Sabara, I. M. (2022). Analisis agglomerative hierarchical clustering berdasarkan pengurutan parsial graf hasse terhadap indikator kemiskinan di Jawa Timur (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).
- Sari, D. P., & Kusdarwati, H. (2015). Analisis Cluster Menggunakan Algoritma K-Means Untuk Mengelompokan Siswa Kelas Iv Sekolah Dasar Brawijaya Smart School Malang. *Malang: Jurnal Mahasiswa Statistik*, 86-89.
- Seber, G.A.F. (1983). *Multivariate Observations*. John Wiley and Sons, Inc. New York.
- Silva, A. R. D., & Dias, C. T. (2013). A cophenetic correlation coefficient for Tocher's method. *Pesquisa Agropecuaria Brasileira*, 48(6), 590-596.
- Surat Al-Anfal Ayat 72: Arab, Latin, Terjemah dan Tafsir Lengkap | Quran NU Online. (2023). Retrieved July 30, 2024, from Nu.or.id website: <https://quran.nu.or.id/al-anfal/72>.
- Syauqi, R. M., Sabrina, P. N., & Santikarama, I. (2023). K-Means Clustering with KNN and Mean Imputation on CPU Benchmark Compilation Data. *Journal of Applied Informatics and Computing*, 7(2), 231-239.