

# Penerapan Metode Hierarchical Clustering Untuk Pengelompokan Kota/Kabupaten Di Indonesia Berdasarkan Indikator Kemiskinan

Brigitta Melati Kumarahadi<sup>1)</sup>; Hasih Pratiwi<sup>2)</sup>; Sri Subanti<sup>3)</sup>

<sup>1)</sup>Program Studi Statistika, Universitas Sebelas Maret

<sup>2)</sup>Program Studi Statistika, Universitas Sebelas Maret

<sup>3)</sup>Program Studi Statistika, Universitas Sebelas Maret

<sup>1)</sup>brigitmelati@student.uns.ac.id; <sup>2)</sup>hasihpratiwi@staff.uns.ac.id; <sup>3)</sup>sri\_subanti@yahoo.co.id<sup>3</sup>

## ABSTRACT

In 2024, the government sets a poverty target in Indonesia of 6-7%. Until September 2022, poverty still shows a figure of 9.57%. To achieve the target, it is necessary to determine priority areas so that government policies can be right on target. This study aims to group cities/regencies in Indonesia based on poverty indicators as a solution to obtain priority areas using the clustering method. This method is used to collect data into several groups based on the same criteria. Hierarchical clustering consists of several methods, including agglomerative nesting such as single linkage, complete linkage, average linkage, and Ward linkage, and divisive analysis. The results showed that the agglomerative nesting average linkage method is the better method because it has a greater cophenetic value and silhouette coefficient value, which is 0.90 and 0.71. The clustering results consist of two clusters, cluster 1 contains 493 areas with low poverty and cluster 2 contains 21 areas with high poverty.

**Keywords:** agglomerative nesting, average linkage, hierarchical clustering, poverty

## I. PENDAHULUAN

Sustainable Development Goals (SDGs) merupakan kesepakatan pembangunan global yang meliputi tujuan dan target untuk menjaga kesejahteraan masyarakat di bidang ekonomi, masyarakat atau sosial, dan lingkungan pada tahun 2030. Salah satu tujuan SDGs adalah mengakhiri kemiskinan dalam segala bentuk apapun. Kemiskinan merupakan suatu kondisi dimana masyarakat tidak dapat mewujudkan hak-hak dasarnya atas perlindungan dan pengembangan kehidupan yang bermartabat [1]. Peraturan Presiden Republik Indonesia (Perpres) Nomor 18 Tahun 2020 menyebutkan target kemiskinan Indonesia pada tahun 2024 adalah 6-7 persen. Kenyataannya, kemiskinan di Indonesia adalah 9,57% pada September 2022 [2]. Angka tersebut menunjukkan bahwa kemiskinan di Indonesia masih jauh dari target, sehingga diperlukan adanya kebijakan pemerintah, khususnya daerah prioritas, untuk mengurangi angka kemiskinan. Daerah prioritas merupakan kota/kabupaten dengan kemiskinan tinggi yang diperoleh dengan mengelompokkan kota/kabupaten di Indonesia berdasarkan indikator kemiskinan dengan menggunakan metode *clustering*. *Hierarchical clustering* adalah salah satu metode *clustering* yang bisa digunakan dalam penelitian ini. Metode tersebut meliputi metode *agglomerative nesting* dan *divisive analysis* [3]. Kedua metode tersebut dibandingkan dan didapatkan metode yang lebih baik dilakukan dalam penelitian ini.

Hasilnya akan diperoleh kelompok daerah beserta karakteristiknya.

## II. TINJAUAN PUSTAKA

### 2.1 Clustering

*Clustering* merupakan metode yang digunakan untuk mengumpulkan data menjadi beberapa kelompok berdasarkan kriteria yang sama [4]. Suatu data akan memiliki karakteristik yang sama dengan data lain yang berada dalam satu kelompok, sedangkan karakteristiknya akan berbeda dengan data lain yang berada di kelompok yang berbeda [5].

### 2.2 Hierarchical Clustering

*Hierarchical clustering* merupakan salah satu metode *clustering* yang didasarkan pada struktur seperti dendogram, yaitu membagi atau menggabungkan data seperti cabang-cabang pohon dengan mengelompokkan dua atau lebih data yang memiliki kesamaan terdekat [6]. *Hierarchical clustering* dapat dilakukan dengan menggunakan dua metode, yaitu *agglomerative nesting* dan *divisive analysis* [7].

### 2.3 Agglomerative Nesting (AGNES)

*Agglomerative nesting* merupakan metode *clustering* menggunakan strategi *bottom-up*, yaitu diawali dengan meletakkan setiap objek ke dalam sebuah *cluster* tersendiri kemudian menggabungkan dua *cluster* dengan jarak terdekat dan seterusnya hingga semua objek

menyatu ke dalam satu *cluster* [8]. *Agglomerative nesting* dapat dilakukan dengan beberapa metode.

### 2.3.1 Single Linkage

Pembentukan *cluster* dilakukan menggunakan jarak minimal atau jarak terdekat antara satu objek dengan objek yang lainnya [9]. Perhitungan dapat dilakukan dengan persamaan (1)

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \quad (1)$$

dengan  $d_{UW}$  adalah jarak antara *cluster* U dan W,  $d_{VW}$  adalah jarak antara *cluster* V dan W, dan  $d_{(UV)W}$  adalah jarak antara *cluster* UV dan W.

### 2.3.2 Complete Linkage

Pembentukan *cluster* dilakukan menggunakan jarak maksimal atau jarak terjauh antara satu objek dengan objek yang lainnya [10]. Perhitungan dapat dilakukan dengan persamaan (2)

$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\} \quad (2)$$

### 2.3.3 Average Linkage

Pembentukan *cluster* dilakukan menggunakan jarak rata-rata antar objek [11]. Perhitungan dapat dilakukan dengan persamaan (3)

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)} N_W} \quad (3)$$

### 2.3.4 Ward Linkage

Pembentukan *cluster* didasari oleh hilangnya informasi akibat dari penggabungan objek menjadi satu *cluster*. Pengukuran dilakukan menggunakan *Sum Squares of Error* (SSE) sebagai fungsi objektif, kemudian menggabungkan dua objek dengan fungsi objektif terkecil [12]. Perhitungan dapat dilakukan dengan persamaan (4)

$$SSE = \sum_{j=1}^p (\sum_{i=1}^n X_{ij}^2 - \frac{1}{n} (\sum_{i=1}^n X_{ij})^2) \quad (4)$$

dengan  $X_{ij}$  adalah nilai untuk objek ke- $i$  pada *cluster* ke- $j$ ,  $p$  adalah banyaknya variabel, dan  $n$  adalah banyaknya objek dalam *cluster* yang terbentuk.

## 2.4 Divisive Analysis (DIANA)

*Divisive analysis* merupakan metode *clustering* menggunakan strategi *top-down*, yaitu dimulai dengan menempatkan semua objek ke dalam satu *cluster* atau disebut dengan akar hierarki kemudian membagi akar *cluster* menjadi beberapa *cluster* yang lebih kecil [13].

## 2.5 Jarak Euclidean

Jarak *Euclidean* merupakan salah satu metode dalam perhitungan jarak dari dua atau

lebih objek [14]. Perhitungan dilakukan seperti yang tertera pada persamaan (5)

$$d_{i,j} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (5)$$

dengan  $d_{i,j}$  adalah jarak antara objek  $i$  dan  $j$ , serta  $x_{ik}$  dan  $x_{jk}$  berturut-turut adalah nilai objek  $i$  pada variabel ke- $k$  dan nilai objek  $j$  pada variabel ke- $k$ .

## 2.6 Korelasi Cophenetic

Korelasi *Cophenetic* merupakan pengujian validitas untuk *clustering* dengan metode *agglomerative nesting*. Persamaan (6) digunakan untuk menghitung koefisien korelasi *cophenetic* [15]

$$r_{Coph} = \frac{\sum_{i < j} (d_{i,j} - \bar{d})(d_{C_{ij}} - \bar{d}_C)}{\sqrt{[\sum_{i < j} (d_{i,j} - \bar{d})^2][\sum_{i < j} (d_{C_{ij}} - \bar{d}_C)^2]}} \quad (6)$$

dengan  $d_{i,j}$  dan  $d_{C_{ij}}$  berturut-turut adalah jarak *Euclidean* dan jarak *cophenetic* antara objek  $i$  dan objek  $j$ , serta  $\bar{d}$  dan  $\bar{d}_C$  berturut-turut adalah rata-rata dari  $d_{i,j}$  dan  $d_{C_{ij}}$ . Nilai  $r_{Coph}$  adalah koefisien korelasi *cophenetic* yang bernilai antara -1 sampai 1, dengan nilai yang mendekati 1 berarti menunjukkan *clustering* yang cukup baik [16].

## 2.7 Silhouette Coefficient

*Silhouette coefficient* merupakan pengujian validitas *cluster* yang dilakukan untuk mengetahui seberapa dekat relasi antar objek dalam satu *cluster* dan seberapa jauh jarak satu *cluster* dengan *cluster* lainnya [17].

$$a_i = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d_{i,j} \quad (7)$$

$$b_i = \frac{1}{|B|} \sum_{j \in B, j \neq i} d_{i,j} \quad (8)$$

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (9)$$

$$SC = \frac{1}{n} \sum_{i=1}^n s_i \quad (10)$$

Kriteria nilai *Silhouette Coefficient* ditampilkan pada Tabel 1.

**Tabel 1. Kriteria Nilai Silhouette Coefficient**

<i>Silhouette Coefficient</i>	Kategori
$0,7 < SC \leq 1$	<i>Strong Structure</i>
$0,5 < SC \leq 0,7$	<i>Medium Structure</i>
$0,25 < SC \leq 0,5$	<i>Weak Structure</i>
$SC \leq 0,25$	<i>No Structure</i>

## III. METODE PENELITIAN

Penelitian ini menggunakan data sekunder yang diperoleh dari *website* Badan Pusat Statistik. Objek penelitian yang dijadikan variabel yaitu kemiskinan di Indonesia pada

tahun 2021 dengan indikatornya ditampilkan pada Tabel 2.

**Tabel 2. Variabel Penelitian**

	Variabel
X1	Persentase perempuan pernah kawin umur 15-49 tahun yang proses melahirkan terakhirnya di fasilitas kesehatan
X2	Persentase rumah tangga yang memiliki akses terhadap layanan sumber air minum layak dan berkelanjutan
X3	Persentase rumah tangga yang memiliki akses terhadap layanan sanitasi layak dan berkelanjutan
X4	Persentase rumah tangga miskin dan rentan yang sumber penerangan utamanya listrik baik dari PLN dan bukan PLN

Tahapan analisis pada penelitian ini adalah melakukan *preprocessing* data, melakukan pengelompokan dengan keempat metode *agglomerative nesting* (*single linkage*, *complete linkage*, *average linkage*, dan *Ward linkage*), menentukan metode *agglomerative nesting* terbaik dengan membandingkan nilai koefisien korelasi *cophenetic*, menghitung nilai *silhouette coefficient* dari metode *agglomerative nesting* terbaik, melakukan pengelompokan dengan metode *divisive analysis* dan menghitung nilai *silhouette coefficient*, memilih metode *clustering* yang lebih baik dengan membandingkan nilai *silhouette coefficient* yang diperoleh kedua metode, mengidentifikasi *cluster* yang terbentuk, dan memvisualisasikan *cluster* yang terbentuk dengan membuat peta wilayah.

**IV. HASIL DAN PEMBAHASAN**

**4.1 Deskripsi Data**

Deskripsi data kemiskinan Indonesia ditampilkan pada Tabel 3 dengan detail sebagai berikut: nilai minimal X1, X2, dan X3 sebesar 0, sedangkan X4 sebesar 3,24; nilai maksimal X1, X2, dan X4 adalah 100, sedangkan X3 sebesar 99,97; nilai rata-rata X1 sebesar 92,58, X2 sebesar 85,32, X3 sebesar 77,35, dan X4 sebesar 97,15.

**Tabel 3. Deskripsi Data**

	X1	X2	X3	X4
Min	0	0	0	3,24
Max	100	100	99,97	100
Average	92,58	85,32	77,35	97,15

**4.2 Agglomerative Nesting (AGNES)**

Pengelompokan dengan keempat metode AGNES (*single linkage*, *complete linkage*, *average linkage*, dan *Ward linkage*) menghasilkan dua *cluster* untuk setiap metodenya. Dilakukan perbandingan nilai *cophenetic* untuk mendapatkan metode yang terbaik. Tabel 4 menunjukkan hasil nilai *cophenetic* yang didapatkan dari keempat metode, antara lain: *single linkage* sebesar 0,89; *complete linkage* sebesar 0,87; *average linkage* sebesar 0,90; dan *Ward linkage* sebesar 0,74. Berdasarkan hasil tersebut, bisa disimpulkan bahwa metode yang terbaik pada AGNES adalah *average linkage* dengan nilai *cophenetic* yang paling tinggi, yaitu sebesar 0,90.

**Tabel 4. Nilai Cophenetic AGNES**

Metode	Nilai Cophenetic
Single Linkage	0,89
Complete Linkage	0,87
Average Linkage	0,90
Ward Linkage	0,74

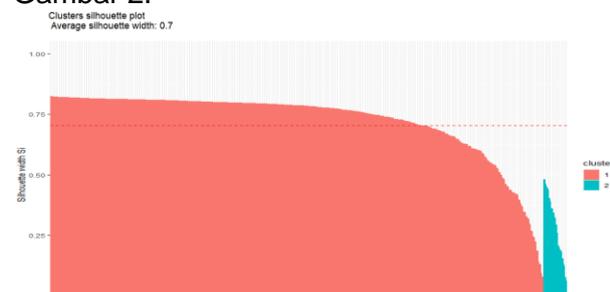
Selanjutnya, didapatkan nilai *Silhouette Coefficient* untuk metode AGNES terbaik, yaitu *Average Linkage* sebesar 0,71 yang ditunjukkan pada Gambar 1.



**Gambar 1. Silhouette Coefficient AGNES**

**4.3 Divisive Analysis (DIANA)**

Pengelompokan dengan metode DIANA menghasilkan dua *cluster* dan *Silhouette Coefficient* sebesar 0,70 yang ditunjukkan pada Gambar 2.



**Gambar 2. Silhouette Coefficient DIANA**

4.4 Perbandingan AGNES dan DIANA

Dilakukan perbandingan nilai *silhouette coefficient* untuk mengetahui metode yang lebih baik digunakan dalam penelitian ini. Tabel 5 menunjukkan metode AGNES mendapatkan nilai sebesar 0,71, sedangkan DIANA mendapatkan nilai sebesar 0,70. Dengan demikian, dapat disimpulkan bahwa metode yang terbaik adalah AGNES dengan selisih 0,01.

Tabel 5. *Silhouette Coefficient* AGNES dan DIANA

Metode	<i>Silhouette Coefficient</i>
<i>Agglomerative Nesting</i> (AGNES)	0,71
<i>Divisive Analysis</i> (DIANA)	0,70

4.5 Hasil Pengelompokan

Pengelompokan menggunakan metode AGNES *Average Linkage* menghasilkan dua cluster, dengan *cluster 1* berisi 493 kota/kabupaten dan *cluster 2* berisi 21 kota/kabupaten. Deskripsi data hasil *clustering* ditampilkan pada Tabel 6. *Cluster 1* memiliki nilai minimal *X1* sebesar 19,08, *X2* sebesar 19,78, *X3* sebesar 32,33, dan *X4* sebesar 66,75; nilai maksimal *X1*, *X2*, dan *X4* adalah 100, sedangkan *X3* sebesar 99,97; nilai rata-rata *X1* sebesar 94,13, *X2* sebesar 86,74, *X3* sebesar 86,74, dan *X4* sebesar 98,33. *Cluster 2* memiliki nilai minimal *X1*, *X2*, dan *X3* sebesar 0, sedangkan *X4* sebesar 3,24; nilai maksimal *X1* sebesar 96,67, *X2* sebesar 98,67, *X3* sebesar 65,37, dan *X4* sebesar 99,83; nilai rata-rata *X1* sebesar 56,15, *X2* sebesar 51,98, *X3* sebesar 17,84, dan *X4* sebesar 69,58. Terlihat bahwa *cluster 1* memiliki nilai yang lebih tinggi daripada *cluster 2*. Hal tersebut memiliki arti bahwa *cluster 2* merupakan kota/kabupaten yang memiliki kemiskinan lebih tinggi daripada *cluster 1*, sehingga dapat disimpulkan daerah prioritas adalah daerah yang berada di dalam *cluster 2*.

Tabel 5. Hasil Clustering

Cluster	Ket	X1	X2	X3	X4
1	Min	19,08	19,78	32,33	66,75
	Max	100	100	99,97	100
	Ave	94,13	86,74	79,89	98,33
2	Min	0	0	0	3,24
	Max	96,67	98,67	65,37	99,83

	Ave	56,15	51,98	17,84	69,58
--	-----	-------	-------	-------	-------

Gambar 3 menunjukkan peta tematik dari hasil analisis *cluster* yang dilakukan. Daerah yang berada dalam *cluster 1* ditandai dengan warna kuning, sedangkan daerah yang berada dalam *cluster 2* ditandai dengan warna merah. Daerah yang termasuk ke dalam *cluster 2* antara lain Kabupaten Pulau Taliabu, Kabupaten Jayawijaya, Kabupaten Paniai, Kabupaten Puncak Jaya, Kabupaten Mappi, Kabupaten Asmat, Kabupaten Yahukimo, Kabupaten Pegunungan Bintang, Kabupaten Tolikara, Kabupaten Nduga, Kabupaten Lanny Jaya, Kabupaten Mamberamo Tengah, Kabupaten Yalimo, Kabupaten Puncak, Kabupaten Dogiyai, Kabupaten Intan Jaya, Kabupaten Deiyai, Kabupaten Sorong Selatan, Kabupaten Nias, Kabupaten Nias Selatan, dan Kabupaten Kepulauan Anambas.



Gambar 3. Peta Hasil Pengelompokan

V. PENUTUP

Metode *agglomerative nesting average linkage* merupakan metode yang lebih baik digunakan untuk mengelompokkan kota/kabupaten di Indonesia berdasarkan indikator kemiskinan, dengan nilai *cophenetic* sebesar 0,90 dan *silhouette coefficient* sebesar 0,71. Hasil pengelompokan terdiri dari dua *cluster*, yaitu *cluster 1* yang berisi 493 daerah dengan kemiskinan rendah dan *cluster 2* yang berisi 21 daerah dengan kemiskinan tinggi.

DAFTAR PUSTAKA

[1] BPS, Penjelasan Data Kemiskinan, 2022, <https://www.bps.go.id/pressrelease/2011/01/27/884/penjelasan-data-kemiskinan.html>. (diakses pada tanggal 17 September 2022).

[2] BPS, Persentase Penduduk Miskin, 2022, <https://www.bps.go.id/indicator/23/192/1/persentase-penduduk-miskin-menurut-provinsi.html>. (diakses pada tanggal 17 September 2022).

[3] D. Exasanti, & A. Jananto, Analisa Hasil Pengelompokan Wilayah Kejadian Non-

- Kebakaran Menggunakan Agglomerative Hierarchical Clustering pada Dinas Pemadam Kebakaran Kota Semarang, *Jurnal Tekno Kompak*, 15(2), 63–75. <https://doi.org/10.33365/jtk.v15i2.1166>, 2021
- [4] J. V. de Oliveira, & W. Pedrycz, Advances in Fuzzy Clustering and its Applications, *In Advances in Fuzzy Clustering and its Applications*, <https://doi.org/10.1002/9780470061190>, 2007
- [5] H. Pallathadka, E. H. Ramirez-Asis, T. P. Loli-Poma, K. Kaliyaperumal, R. J. M. Ventayen, & M. Naved, Applications of Artificial Intelligence *In Business Management, E- Commerce, and Finance. Materials Today: Proceedings*. <https://doi.org/10.1016/j.matpr.2021.06.419>, 2021
- [6] S. Singh, & S. Srivastava, Review of Clustering Techniques in Control System. *Procedia Computer Science*, 173, 272–280. <https://doi.org/10.1016/j.procs.2020.06.032>, 2020.
- [7] H. Pratiwi, S. S. Handajani, & Respatiwan. Clustering dan Penerapannya dalam Bidang Seismologi, Deepublish Publisher, 2021.
- [8] M. Hajar, N.Y. Nasution, & S. Prangga, Penerapan Metode Agglomerative Nesting Kartanegara (Studi Kasus : Data Potensi Desa (PODES) Tahun 2018). *Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya*, 456–469, 2022.
- [9] Iis, Yahya, I., Wibawa, G. N. A., & L. Laome, Penggunaan Korelasi Cophenetic untuk Pemilihan Metode Cluster Berhierarki pada Mengelompokkan Kabupaten/Kota Berdasarkan Jenis Penyakit di Provinsi Sulawesi Tenggara Tahun 2020. *Prosiding Seminar Nasional Sains dan Terapan*, April, 1–16, 2022.
- [10] A. Septianingsih, Pemetaan Kabupaten Kota di Provinsi Jawa Timur Berdasarkan Tingkat Kasus Penyakit Menggunakan Pendekatan Agglomeratif Hierarchical Clustering. *Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, 3(2), 2022.
- [11] R. Novidianto, & A. T. Dani, Analisis Kluster Kasus Aktif Covid-19 Menurut Provinsi di Indonesia Berdasarkan Data Deret Waktu, *Jurnal Aplikasi Statistika & Komputasi Statistik*, 12(2), 15–24, 2020.
- [12] Z. F. Pusediktasari, W. G. Sasmita, W. R. Fitrilia, R. Fitriani, & S. Astutik, The Clustering of Provinces in Indonesia by The Economic Impact of Covid-19 using Cluster Analysis. *Indonesian Journal of Statistics and Its Applications*, 5(1), 117–129. <https://doi.org/10.29244/ijisa.v5i1p117-129>, 2021
- [13] J. Han, M. Kamber, & J. Pei, Data mining: Concepts and Techniques. In Elsevier Inc. (3 ed.), Morgan Kaufmann Publishers. <https://doi.org/10.1109/ICMIRA.2013.45>, 2012
- [14] N. N., Halim, & E. Widodo, Clustering Dampak Gempa Bumi di Indonesia Menggunakan Kohonen Self Organizing Maps. *Prosiding SI MaNIs (Seminar Nasional Integrasi Matematika dan Nilai Islami)*, 1(1), 188–194. <http://conferences.uin-malang.ac.id/index.php/SIMANIS/article/view/62>, 2017
- [15] A. D. Munthe, Penerapan Clustering Time Series Untuk Menggerombolkan Provinsi Di Indonesia Berdasarkan Nilai Produksi Padi. *Jurnal Litbang Sukowati: Media Penelitian dan Pengembangan*, 2(2), 11. <https://doi.org/10.32630/sukowati.v2i2.61>, 2019.
- [16] A.A Dzikrullah (2022), Pengelompokan Provinsi Berdasarkan Kualitas Jaringan Internet Dengan Metode Centroid Linkage. *Journal of Mathematics, Computations, and Statistics*, 5(1), 48–57, 2020
- [17] Y. K. Kumarahadi, B. M. Kumarahadi, & K. Sandradewi, Clustering Pelaksanaan Vaksinasi di Jawa Tengah Menggunakan Metode K-Means. *Jurnal Ilmiah SINUS*, 20(2), 77–86. <https://doi.org/10.30646/sinus.v20i2.620>, 2022