



Analisis Cluster untuk Pengelompokan Provinsi di Indonesia berdasarkan Tingkat Kemiskinan menggunakan Metode Average Linkage

Dede Saputra^a, Azrianti Ardania^b, Syaftirridho Putri^c, Adis Tia Juli Agil Asri^d, Lisa Harsyah^{e}*

- a. Program Studi Matematika, Universitas Mataram, Indonesia
- b. Program Studi Matematika, Universitas Mataram, Indonesia
- c. Program Studi Matematika, Universitas Mataram, Indonesia
- d. Program Studi Statistika, Universitas Mataram, Indonesia
- e. Program Studi Statistika, Universitas Mataram, Indonesia

*Corresponding author (email): lisa_harsyiah@unram.ac.id

ABSTRACT

Poverty is a major economic and social issue in Indonesia because it is a serious problem that can affect social welfare. Poverty is influenced by many factors including school enrollment rate, life expectancy, gross regional domestic product, human development index and open unemployment rate. Cluster analysis is a technique in multivariate statistics where objects are grouped based on proximity or similarity of properties so that objects that have close proximity (similar properties) will be in the same group (cluster). The purpose of this study is to cluster provinces in Indonesia based on poverty levels using the average linkage method. The results of this study obtained 5 clusters, where cluster 1 consists of Nanggroe Aceh Darussalam, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Bangka Belitung Islands, Central Java, East Java, Bali, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, Central Kalimantan, South Kalimantan, North Kalimantan, Central Sulawesi, South Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku and West Papua. Cluster 2 consists of Riau Islands, West Java, Banten and North Sulawesi. Cluster 3 consists of DKI Jakarta and East Kalimantan. Cluster 4 consists of DI Yogyakarta and the last cluster consists of Papua.

Keywords: Poverty, Cluster, Hierarchy, Average Linkage

ABSTRAK

Kemiskinan menjadi masalah utama dalam bidang ekonomi dan sosial di Indonesia karena merupakan masalah yang serius yang dapat memengaruhi kesejahteraan sosial. Kemiskinan dipengaruhi oleh banyak faktor diantaranya angka partisipasi sekolah, angka harapan hidup, produk domestik regional bruto,



indeks pembangunan manusia dan tingkat pengangguran terbuka. Analisis cluster merupakan teknik dalam statistika multivariat dimana objek-objek dikelompokkan berdasarkan kedekatan jarak atau kemiripan sifat yang dimilikinya sehingga objek-objek yang memiliki kedekatan jarak (sifat yang mirip) akan berada dalam kelompok (cluster) yang sama. Tujuan dari penelitian ini yaitu untuk mengelompokkan provinsi di Indonesia berdasarkan tingkat kemiskinan menggunakan metode average linkage sehingga mampu menciptakan strategi Pembangunan yang lebih terukur, berfokus pada pengurangan kemiskinan dan mencapai pemerataan kesejahteraan di seluruh wilayah. Hasil dari penelitian ini yaitu diperoleh 5 cluster, dimana cluster 1 terdiri dari Nangroe Aceh Darussalam, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Jawa Tengah, Jawa timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara dan Papua Barat. Cluster 2 terdiri dari, Kepulauan Riau, Jawa Barat, Banten dan Sulawesi Utara. Cluster 3 terdiri dari, DKI Jakarta dan Kalimantan Timur. Cluster 4 terdiri dari DI Yogyakarta dan cluster yang terakhir terdiri dari Papua. Setiap cluster yang terbentuk selanjutnya akan dianalisis untuk memahami karakteristik tiap kelompok sehingga dapat diambil kebijakan terkait penurunan angka kemiskinan di Indonesia.

Kata kunci: Kemiskinan, Cluster, Hierarki, Average Linkage

Diterima: 10-10-2024; Disetujui: 29-11-2024;

1. Pendahuluan

Kondisi dimana seseorang atau sekelompok orang tidak dapat memenuhi hak-hak dasar yang diperlukan untuk hidup dan menjalani hidup yang bermartabat disebut kemiskinan (Arifuzzaman et al., 2021). Kemiskinan menjadi masalah utama dalam bidang ekonomi dan sosial di Indonesia karena merupakan masalah yang serius yang dapat memengaruhi kesejahteraan sosial. Tingkat kemiskinan di Indonesia pada Maret 2023 mencapai 29,50 juta, turun sebesar 0,18% dari Maret 2022, menurut Badan Pusat Statistik. Kemiskinan ini dapat dipengaruhi oleh banyak faktor, faktor-faktor ini dapat berbeda di antara provinsi di Indonesia.

Kemiskinan sudah menjadi masalah yang sangat kompleks dan harus segera diatasi. Kemiskinan merupakan suatu kondisi dimana kebutuhan makanan, tempat tinggal, pendidikan, dan kesehatan tidak dapat terpenuhi. Kondisi tersebut dapat terjadi akibat dari sulitnya akses pekerjaan ataupun akibat dari kelangkaan alat untuk memenuhi kebutuhan dasar. Jika kemiskinan dibiarkan terus menerus atau tidak diatasi dengan tepat dan cepat, maka akan menimbulkan berbagai dampak buruk seperti angka kematian yang meningkat dan angka kriminalitas yang tinggi (Rahmawati, 2023).

Kemiskinan dipengaruhi oleh banyak faktor diantaranya angka partisipasi sekolah, angka harapan hidup, produk domestik regional bruto, indeks pembangunan manusia dan tingkat pengangguran terbuka (Bhinadi, 2017). Penelitian yang dilakukan oleh Indah Purboningtyas et al. (2020), dengan studi kasus tingkat kemiskinan provinsi Jawa Tengah menunjukkan bahwa variabel yang signifikan dalam mempengaruhi tingkat kemiskinan adalah variabel indeks pembangunan manusia dan tingkat pengangguran terbuka. Tingkat kemiskinan adalah salah satu indikator penting dalam mengukur kesejahteraan masyarakat dan kinerja pembangunan suatu daerah. Oleh karena itu, memahami distribusi kemiskinan di berbagai provinsi di Indonesia menjadi sangat penting untuk merumuskan kebijakan yang efektif dan tepat sasaran. Analisis yang mendalam yang terkait dengan pengelompokan provinsi berdasarkan kesamaan karakteristik merupakan analisis yang tepat untuk ini agar pemerintah dapat mengeluarkan kebijakan yang sama dalam penanganan kemiskinan di provinsi tersebut. Harapannya adalah analisis statistika tersebut dapat menekan jumlah kemiskinan dengan cara meminimalkan faktor-faktor tersebut.

Analisis cluster merupakan metode statistika yang fungsinya memasukkan kumpulan objek menjadi 2 kelompok, bahkan lebih menggunakan analisis cluster, yang merupakan metode analisis statistik yang didasarkan pada bagaimana masing-masing objek memiliki karakteristik yang mirip. Analisis kelompok adalah metode analisis yang bertujuan untuk memilih objek dalam sejumlah kelompok yang memiliki karakteristik yang berbeda satu sama lain. Tujuan dari analisis cluster adalah untuk memastikan bahwa variasi objek dalam setiap kelompok sekecil mungkin atau bahwa semua anggota kelompok adalah homogen. Penelitian yang dilakukan oleh Nurissaidah et al. (2022) yang melakukan pengelompokan berdasarkan tingkat penyakit menular pada provinsi di Indonesia dengan metode pautan berjenis lengkap, rata-rata dan ward memberikan hasil bahwa terbentuk enam cluster dengan ukuran statistik yang diperoleh adalah sebagai berikut nilai S_w sebesar 0,18405, S_b sebesar 2,12284, dan rasio S_w terhadap S_b sebesar 0,08670. Penelitian yang dilakukan oleh Sulthan dkk. (2019) yang mengelompokkan kecamatan-kecamatan yang ada di Kabupaten Sidoarjo berdasarkan faktor hewan ternak. Berdasarkan tiga jenis metode yang digunakan dalam pengelompokan tersebut menunjukkan bahwa metode pautan lengkap adalah metode yang paling baik karena memiliki rasio simpangan baku terkecil yaitu sebesar 0,222.

Berdasarkan masalah yang sudah dipaparkan, peneliti ingin mengelompokkan provinsi di Indonesia berdasarkan tingkat kemiskinan di Indonesia menggunakan metode average linkage dengan tujuan mengidentifikasi wilayah yang membutuhkan perhatian khusus. Provinsi dengan tingkat kemiskinan tinggi mungkin memerlukan program bantuan sosial, pembangunan infrastruktur, atau investasi ekonomi yang lebih besar.

2. Metode

2.1 Alat dan Data Penelitian

Alat yang digunakan dalam penelitian ini adalah software SPSS versi 24. Data yang digunakan dalam penelitian ini yaitu data terkait faktor-faktor yang mempengaruhi kemiskinan di Indonesia tahun 2023 yang bersumber dari Badan Pusat Statistik (BPS).

2.2 Variabel Penelitian

Beberapa variabel yang digunakan dalam penelitian ini yaitu.

Tabel 1 Variabel Penelitian

No.	Simbol	Variabel	Satuan
1	Y	Jumlah penduduk miskin di Indonesia	Juta jiwa
2	X_1	Angka partisipasi sekolah	Persen
3	X_2	Angka harapan hidup	Tahun
4	X_3	Produk domestik regional bruto	Rupiah
5	X_4	Indeks pembangunan manusia	Persen
6	X_5	Tingkat pengangguran terbuka	Persen

2.3 Prosedur Penelitian

Secara garis besar ada dua metode yang digunakan dalam analisis cluster yaitu metode hirarki dan metode non-hirarki (Komalasari, 2016). Proses analisis cluster dalam penelitian ini menggunakan metode hirarki. Metode hierarki adalah metode yang terdiri dari dua tipe dasar yaitu aglomeratif dan deisif. Proses clustering dalam metode aglomeratif yaitu tiap objek ditempatkan pada cluster tersendiri sehingga banyak cluster yang terbentuk sama dengan banyak objek yang diobservasi. Selanjutnya adalah untuk setiap jarak yang terdekat yang dihasilkan dari dua cluster

maka keduanya dapat dihimpun membentuk cluster yang baru. Sedangkan metode devisif merupakan kebalikan dari metode algromeratif dimana proses clustering dimulai dari satu cluster yang mencakup semua observasi kemudian akan dilakukan proses pemisahan dimana observasi yang tidak memiliki kesamaan akan dihimpun menjadi kelompok-kelompok kecil untuk membentuk cluster. Terdapat bermacam-macam pembagian metode hirarki. Penelitian ini menggunakan metode pautan rata-rata (*average linkage*) dalam penentuan clusternya (Maiyanti, dkk., 2024).

Metode pautan rata-rata (*average linkage*) adalah metode dalam analisis cluster hirarki aglomeratif yang mengelompokkan suatu observasi berdasarkan jarak rata-rata antar observasi. Pengelompokkan pada metode ini dimulai dari tengah atau pasangan observasi dengan jarak paling mendekati jarak rata-rata. Oleh karena itu, dua observasi dapat dimasukkan ke dalam cluster yang sama berdasarkan jarak yang terkecil dan observasi-observasi yang lain akan bergabung berdasarkan perhitungan jarak rata-rata. Penelitian ini menggunakan jarak kuadrat euclid digunakan untuk melihat kemiripan antar dua objek pada proses analisis cluster menggunakan metode *average linkage*.

Langkah-langkah analisis data adalah sebagai berikut.

1. Menentukan deskripsi data yaitu rata-rata dan simpangan baku dari faktor-faktor yang mempengaruhi tingkat kemiskinan menggunakan persamaan berikut:

$$S_{X_j} = \sqrt{\frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}{n - 1}} \quad (2.1)$$

Keterangan:

\bar{X}_j = rata-rata variabel independen

n = banyaknya data

dengan:

$$\bar{X}_j = \frac{\sum_{i=1}^n x_{ij}}{n} \quad (2.2)$$

2. Melakukan standarisasi data penelitian dengan dua tahapan berupa pemusatan dan penskalaan data menggunakan persamaan:

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{S_{X_j}} \quad (2.3)$$

Keterangan:

Z_{ij} = variabel independen hasil transformasi

X_{ij} = nilai data pada objek ke- i dan variabel ke- j dimana $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, p$

\bar{X}_j = nilai rata-rata pada variabel ke- j

S_{X_j} = simpangan baku variabel independen

Menurut Kurtner, *et al.* (2005) dalam Rahmawati (2023) standarisasi data adalah tahapan dalam analisa statistika yang meliputi dua metode yaitu pemusatan dan penskalaan. Pemusatan data perlu dilakukan untuk menyederhanakan model regresi yaitu dengan menghilangkan intersep regresi maka dilakukan melalui pemusatan data sedangkan untuk mentransformasi data dari variabel yang dikaji dilakukan melalui penskalaan data. Penyebab proses perhitungan suatu analisis menjadi tidak valid adalah apabila terdapat perbedaan satuan pada data yang digunakan. Proses standarisasi dilakukan apabila terdapat perbedaan ukuran satuan yang besar pada variabel-variabel yang diteliti (Hasrul, 2018).

3. Melakukan uji multikolinearitas berdasarkan nilai VIF yang diperoleh dari persamaan

$$VIF = \frac{1}{(1 - R_j^2)} \quad (2.4)$$

Multikolinearitas adalah suatu keadaan dimana ada hubungan atau korelasi yang kuat diantara variabel-variabel independen yang menyebabkan keakuratan model berkurang. Variabel yang mempunyai multikolinearitas umumnya membawa informasi yang sama sehingga membuat hasil *clustering* menjadi kurang akurat (Komalasari, 2016). Kriteria yang dilakukan dalam mendeteksi adanya multikolinearitas adalah menggunakan faktor inflasi ragam (*Variance Inflation Factor/VIF*). Gejala multikolinearitas ditunjukkan dengan nilai VIF yang lebih besar dari 10 (Hidayatullah, dkk., 2024).

Jika terdapat multikolinearitas, maka akan dilanjutkan dengan metode PCA. Jika tidak terjadi multikolinearitas, maka akan dilanjutkan dengan uji KMO yang dihitung menggunakan persamaan (2.5).

$$KMO = \frac{\sum_i^n \sum_{j \neq i}^n r_{ij}^2}{\sum_i^n \sum_{j \neq i}^n r_{ij}^2 + \sum_i^n \sum_{j \neq i}^n a_{ij}^2} \quad (2.5)$$

dimana r_{ij} merupakan koefisien korelasi dan a_{ij} merupakan koefisien korelasi parsial yang dirumuskan sebagai berikut (Telussa, dkk., 2013):

$$r_{ij} = \frac{n \sum_{i=1}^k \sum_{j=1}^k Z_i Z_j - \sum_{j=1}^k Z_i \sum_{j=1}^k Z_j}{\sqrt{n \sum_{j=1}^k Z_i^2 - (\sum_{i=1}^k Z_i)^2} \sqrt{n \sum_{j=1}^k Z_j^2 - (\sum_{j=1}^k Z_j)^2}} \quad (2.6)$$

$$a_{ij} = \frac{r_{ij} - r_{ij} \cdot r_{jk}}{\sqrt{(1 - r_{ik}^2)(1 - r_{jk}^2)}} \quad (2.7)$$

Angka KMO lebih besar dari 0.5 menunjukkan bahwa sampel pengamatan yang digunakan telah memenuhi kecukupan sehingga layak untuk dianalisis lebih lanjut dalam analisis cluster. Rumus perhitungan KMO adalah sebagai berikut (Nugroho, 2008):

4. Menentukan jarak kuadrat euclidean menggunakan persamaan:

$$d_{ij} = \sum_{k=1}^p (x_{ik} - x_{jk})^2 \quad (2.8)$$

dimana d_{ij} adalah jarak antara objek ke- i dan ke- j , x_{ik} adalah data dari objek ke- i pada variabel ke- k , dan x_{jk} adalah data dari objek ke- j pada variabel ke- k

5. Menentukan analisis cluster metode *average linkage* dengan persamaan:

$$d_{(UV)W} = \frac{\sum_{i=1}^n \sum_{k=1}^k d_{ik}}{n_{UV}n_W} \quad (2.9)$$

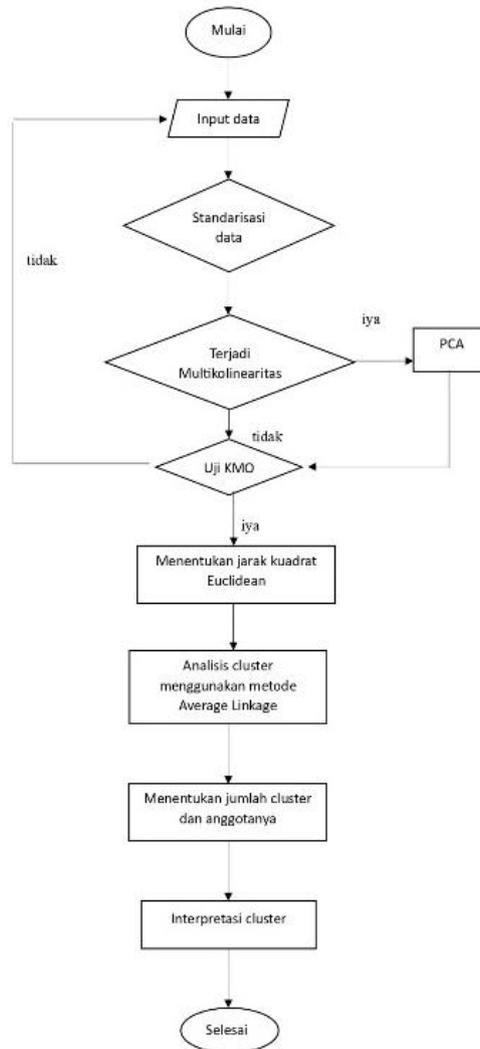
Pembentukan cluster pada metode *average linkage* dimulai dengan membentuk matriks jarak berdasarkan jarak kuadrat euclid untuk setiap observasi yang dinotasikan dengan $D = \{d_{ij}\}$ untuk melihat kedekatan (kemiripan) pada setiap observasi. Observasi yang memiliki jarak paling dekat (memiliki kemiripan) akan dimasukkan dalam cluster yang

sama. Kedekatan antar cluster yang terbentuk dapat dilihat berdasarkan jarak yang dirumuskan sebagai berikut (Ramadani & Salma, 2022).

Adapun $d_{(UV)W}$ adalah jarak antara cluster (UV) dan cluster W, d_{ik} adalah jarak antara objek i pada cluster (UV) dan objek k pada cluster W, n_{UV} adalah jumlah objek yang ada di cluster (UV) dan n_W adalah jumlah objek yang ada di cluster W.

6. Menentukan jumlah cluster dan kelompoknya berdasarkan output yang didapatkan.
7. Interpretasi hasil.

Adapun diagram alir (*flowchart*) dari prosedur penelitian sebagai berikut.



Gambar 1. *Prosedur Penelitian*

3. Hasil dan Pembahasan

3.1 Deskripsi Data

Berdasarkan persamaan (2.1) dan (2.2) diperoleh hasil statistika deskriptif data berupa rata-rata dan standar deviasi yang disajikan pada Tabel 2.

Tabel 2. Deskripsi Data

Variabel	Rata-rata	Simpangan baku
X_1	28.2671	6.35088
X_2	72.7197	2.40800
X_3	76822.9118	60913.99059
X_4	72.6221	3.85971
X_5	4.4635	1.30908

3.2 Standarisasi Data

Berdasarkan data yang diperoleh pada Tabel 2 dapat diketahui bahwa nilai rata-rata dan standar deviasi yang pada setiap variabel memiliki selisih yang cukup jauh satu sama lain. Hal tersebut disebabkan karena setiap variabel yang digunakan dalam penelitian ini memiliki satuan yang berbeda. Oleh karena itu perlu dilakukan standarisasi data menggunakan persamaan (2.3).

3.3 Uji Multikolinieritas

Multikolinieritas adalah satu keadaan dimana ada hubungan atau korelasi yang kuat diantara variabel-variabel independen yang menyebabkan keakuratan model berkurang. Variabel yang mempunyai multikolinieritas umumnya membawa informasi yang sama sehingga membuat hasil *clustering* menjadi kurang akurat (Komalasari, 2016). Pendeteksian multikolinieritas dilakukan dengan melihat nilai (*Variance Inflation Factor/VIF*) untuk setiap variabel yang diteliti. Gejala multikolinieritas ditunjukkan dengan nilai VIF yang lebih besar dari 10. Berdasarkan persamaan (2.4) diperoleh nilai VIF variabel independen terstandarisasi (Z) yang disajikan pada tabel berikut.

Tabel 3. Nilai VIF pada Deteksi Multikolinieritas

Variabel	VIF	Keputusan
Z_1	1.276	Tidak ada multikolinieritas
Z_2	2.780	Tidak ada multikolinieritas
Z_3	1.551	Tidak ada multikolinieritas
Z_4	1.116	Tidak ada multikolinieritas
Z_5	3.663	Tidak ada multikolinieritas

Berdasarkan Tabel 3 diperoleh bahwa nilai VIF untuk setiap variabel yang diteliti kurang dari 10 sehingga tidak terjadi multikolinieritas. Oleh karena itu, analisis dapat dilanjutkan dengan uji KMO.

Berdasarkan persamaan (2.6) dan (2.7) secara berturut-turut diperoleh nilai koefisien korelasi dan korelasi parsial pada data yang telah terstandarisasi yang disajikan pada tabel berikut:

Tabel 4. Nilai Koefisien Korelasi

Variabel	Z_1	Z_2	Z_3	Z_4	Z_5
Z_1	1.000	0.092	-0.091	0.276	-0.150
Z_2	0.092	1.000	0.417	0.788	0.129
Z_3	-0.091	0.417	1.000	0.525	0.275
Z_4	0.276	0.788	0.525	1.000	0.179

Variabel	Z_1	Z_2	Z_3	Z_4	Z_5
Z_5	-0.150	0.129	0.275	0.179	1.000

Tabel 5. Nilai Koefisien Korelasi Parsial

Variabel	Z_1	Z_2	Z_3	Z_4	Z_5
Z_1	1.000	0.259	-0.026	0.545	-0.135
Z_2	0.259	1.000	0.253	0.642	0.088
Z_3	-0.026	0.253	1.000	0.394	0.258
Z_4	0.545	0.642	0.394	1.000	0.153
Z_5	-0.135	0.088	0.258	0.153	1.000

Berdasarkan perolehan nilai koefisien korelasi pada Tabel 4 dan nilai koefisien korelasi parsial pada Tabel 5, selanjutnya dilakukan substitusi setiap komponen tersebut kedalam persamaan (2.5) sehingga diperoleh nilai KMO sebesar $0.548 > 0.500$. Hasil tersebut menunjukkan bahwa sampel yang diamati adalah sampel yang layak karena memenuhi kecukupan untuk dianalisis lebih lanjut dalam analisis cluster.

3.4 Analisis Cluster

3.4.1. Matriks Jarak Kuadrat Euclidian

Komponen matriks jarak kuadrat euclidian dengan menghitung jarak dari setiap observasi dengan persamaan (2.8). Berikut adalah perolehan matriks jarak kuadrat euclid $\{D = \{d_{ij}\}\}$.

$$D_{34 \times 34} = \begin{bmatrix} 0.000 & 0.800 & 0.692 & \dots & 13.308 \\ 0.800 & 0.000 & 2.562 & \dots & 11.523 \\ 0.692 & 2.562 & 0.000 & \dots & 16.361 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 13.308 & 11.523 & 16.361 & \dots & 0.000 \end{bmatrix}$$

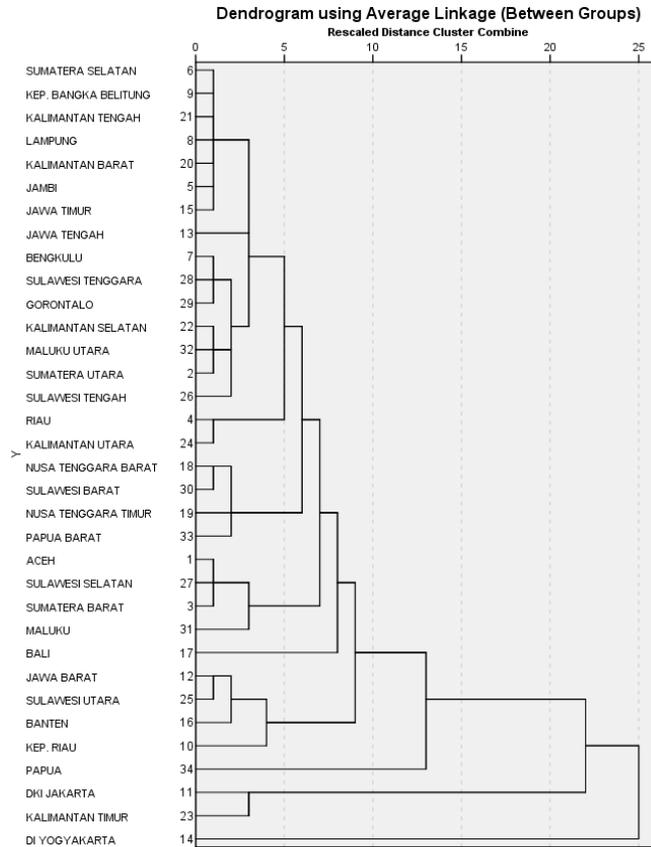
3.4.2. Analisis Cluster Metode Average Linkage

Proses analisis cluster dilakukan dengan menggabungkan dua observasi yang memiliki jarak terdekat. Berdasarkan matriks jarak kuadrat euclid komponen matriks dengan observasi provinsi Sumatera Selatan (6) dan provinsi Kepulauan Bangka Belitung (9) merupakan komponen yang memiliki jarak kuadrat euclid terkecil (d_{69}) sebesar 0.293. Oleh karena itu, kedua provinsi tersebut akan bergabung kedalam cluster yang sama. Selanjutnya adalah akan dihitung jarak observasi lainnya terhadap cluster (69) berdasarkan persamaan (2.9). Proses perhitungan dilakukan dengan cara yang serupa hingga semua observasi tergabung dalam cluster tunggal seperti yang terlihat pada Gambar 2.

Selanjutnya dibentuk matriks jarak yang baru dengan jarak ditentukan berdasarkan persamaan (2.9) yang disajikan dalam bentuk tabel sebagai berikut:

Tabel 6. Matriks Jarak Baru

Cluster	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14	14
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14	0.000	28.332
14	28.332	0.000



Gambar 2. Dendrogram Pengelompokan Metode Average Linkage

Berdasarkan dendrogram yang dihasilkan, peneliti membentuk lima cluster dalam mengelompokkan provinsi di Indonesia berdasarkan tingkat kemiskinan. Tidak ada alasan khusus dalam penentuan jumlah cluster yang terbentuk. Namun peneliti ingin terfokus pada lima cluster tersebut agar lebih terpusat dan mudah untuk dipahami. Selanjutnya akan dilakukan pemotongan pada dendrogram tersebut agar terbentuk lima cluster. Pemotongan ini dilakukan pada jarak dengan 5 sampai dengan 10 satuan dan saat penggabungan objek yang memiliki perbedaan jarak terbesar. Hasilnya, diperoleh lima cluster dengan jumlah anggota masing-masing klaster adalah 1, 7, 2, dan 24, dengan anggota cluster disajikan pada tabel berikut:

Tabel 7. Hasil Pembentukan Cluster

Cluster	Anggota Cluster
1	Nanggroe Aceh Darussalam, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Jawa Tengah, Jawa timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara dan Papua Barat.
2	Kepulauan Riau, Jawa Barat, Banten dan Sulawesi Utara
3	DKI Jakarta dan Kalimantan Timur
4	DI Yogyakarta
5	Papua

Berdasarkan hasil analisis, terbentuk 5 cluster dimana provinsi yang tergabung dalam satu cluster berarti memiliki kemiripan karakteristik kemiskinan misalnya variabel atau faktor yang mempengaruhi kemiskinan di wilayah tersebut sama.

4. Kesimpulan

Berdasarkan analisis yang telah dilakukan pada bagian hasil dan pembahasan didapat kesimpulan bahwa analisis cluster menggunakan metode average linkage pada pengelompokan tingkat kemiskinan provinsi di Indonesia berdasarkan lima cluster yang terbentuk adalah sebagai berikut:

- a. Cluster 1 meliputi provinsi Nanggroe Aceh Darussalam, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Jawa Tengah, Jawa timur, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara dan Papua Barat.
- b. Cluster 2 meliputi provinsi Kepulauan Riau, Jawa Barat, Banten dan Sulawesi Utara.
- c. Cluster 3 meliputi provinsi DKI Jakarta dan Kalimantan Timur.
- d. Cluster 4 meliputi Provinsi DI Yogyakarta.
- e. Cluster 5 meliputi provinsi Papua.

Provinsi yang tergabung dalam satu cluster berarti memiliki kemiripan karakteristik kemiskinan seperti mempunyai kesamaan pada variabel atau faktor yang mempengaruhi kemiskinan di wilayah tersebut. Untuk hasil yang lebih mendetail bisa menggunakan analisis lanjutan untuk menguji ketepatan clustering serta mengetahui variabel apa saja yang mempengaruhi kemiskinan di cluster pertama, ke-dua sampai cluster ke-lima.

DAFTAR PUSTAKA

- Arifuzzaman, M., Rafee, F. E. R., & Islam, M. J. (2021). Exploration of Poverty and Human Rights Violation: A Legal Analysis. *Asian Journal of Social Sciences and Legal Studies*, 3(1), 10–22. <https://doi.org/10.34104/ajssls.021.010022>
- Alwi, W., & Hasrul, M. (2018). Analisis Klaster Untuk Pengelompokan Kabupaten/Kota Di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal Matematika dan Statistika serta Aplikasinya*, 6(1), 35-35.
- Ardito Bhinadi, A. (2017). *Penanggulangan Kemiskinan & Pemberdayaan Masyarakat (Studi Kasus Daerah Istimewa Yogyakarta)*.
- Komalasari, D. (2016). *Buku Ajar Statistika Multivariat*. FMIPA-Universitas Mataram.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linier Statistical Models (Fifth Edition)*. McGraw-Hill.
- Maiyanti, S. I., Dwipurwani, O., & Aisyah, N. (2024). Pengelompokan Provinsi Di Indonesia Berdasarkan Konsumsi Kalori Per Kapita Sehari Menurut Kelompok Komoditas/Makanan Menggunakan Average Linkage Dan Ward's Method. *MULTIPLE: Journal of Global and Multidisciplinary*, 2(5), 1698-1713.
- Nugroho, S., (2008). *Statistika Multivariat Terapan Edisi Pertama*. Bengkulu: UNIB Press.
- Nurissaidah Ulinnuah dan Rafika Veriani, (2020), Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage dan Ward, *InforTekjar: jurnal Nasional Informatika dan Teknologi Jaringan*, 5 (1).
- Rahmawati, Raehani. (2023). Perbandingan Regresi Ridge dan Principal Component Regression dalam Mengatasi Multikolinieritas pada Faktor-Faktor yang Memengaruhi Kemiskinan di Indonesia. (Skripsi, Universitas Mataram).
- Ramadani, R., & Salma, A. (2022). Perbandingan Kinerja Metode Average Linkage dan Ward dalam Pengelompokan Tingkat Kesejahteraan Masyarakat Provinsi Sumatera Barat Menurut Kabupaten/Kota Tahun 2021. *Journal of Mathematics UNP*, 7(3), 11-24.

- Saribulan, Nur, dkk.,(2023). Perkembangan penelitian penanggulangan kemiskinan di Indonesia: Analisis Bibliometrik dan Analisis Konten. *Jurnal Ilmu Sosial dan Humaniora*. 12(2). 309-321.
- Sulthan Fikri Mu'afa, dkk., (2019), Perbandingan Metode Single Linkage, Complete Linkage Dan Average Linkage dalam Pengelompokan Kecamatan Berdasarkan Variabel Jenis Ternak Kabupaten Sidoarjo, *Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*.4(2).
- Telussa, A. M., Persulesy, E. R., & Leleury, Z. A. (2013). Penerapan analisis korelasi parsial untuk menentukan hubungan pelaksanaan fungsi manajemen kepegawaian dengan efektivitas kerja pegawai. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 7(1), 15-18.