

Penentuan Cluster Hierarki Optimum dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan dengan Indikator Kemiskinan

Hasmira

Program Studi Matematika, Universitas Islam Negeri Alauddin Makassar, 60600118029@uin-alauddin.ac.id

Wahidah Alwi

Program Studi Matematika, Universitas Islam Negeri Alauddin Makassar, wahidah.alwi@uin-alauddin.ac.id

Khalilah Nurfadilah

Program Studi Matematika, Universitas Islam Negeri Alauddin Makassar, khalilah.nurfadilah@uin-alauddin.ac.id

ABSTRAK. Kemiskinan merupakan salah satu permasalahan sosial yang dihadapi oleh setiap negara, salah satunya di Indonesia. Dalam artikel ini bertujuan untuk mengelompokkan Kabupaten/Kota di Sulawesi Selatan berdasarkan nilai indikator kemiskinan, dengan menggunakan metode analisis *cluster* hierarki. Dalam analisis *cluster* hierarki terdapat beberapa metode yang dapat digunakan untuk proses pengelompokan seperti; *single linkage*, *complete linkage*, *average linkage*, dan beberapa metode yang lain. Banyaknya metode dalam analisis *cluster* hierarki, tentunya terdapat salah satu metode yang optimum (terbaik) dalam melakukan proses pengelompokan Kabupaten/Kota Provinsi Sulawesi Selatan. Dari hasil analisis, diperoleh bahwa metode yang optimum berdasarkan nilai korelasi *cophenetic* dalam proses pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan adalah metode *average linkage*. Dengan tiga *cluster*, dengan masing-masing anggota: *Cluster 1* dengan jumlah anggota 1 Kabupaten, *Cluster 2* dengan jumlah anggota 21 Kabupaten/Kota, *Cluster 3* dengan jumlah anggota 1 Kota.

Kata Kunci: Indikator Kemiskinan, Analisis Cluster Hierarki, Metode Optimum, Provinsi Sulawesi Selatan

1. PENDAHULUAN

Kemiskinan merupakan salah satu permasalahan sosial yang dihadapi oleh setiap negara, salah satunya di Indonesia. Berdasarkan informasi dari Badan Pusat Statistik (BPS), jumlah kemiskinan di Indonesia mencapai 27,5 juta jiwa per Maret 2021. Berdasarkan jumlah tersebut, tingkat kemiskinan di Indonesia mencapai 10,14% dari total populasi nasional. Masalah kemiskinan ini merupakan permasalahan yang sangat pokok bagi bangsa Indonesia. Sehingga, menjadi salah satu prioritas pemerintahan dalam melakukan perencanaan pembangunan nasional.

Sulawesi Selatan merupakan salah satu provinsi yang ada di Indonesia. Berdasarkan hasil Survei Sosial Ekonomi Nasional (Susenas)

Maret 2021, jumlah penduduk miskin di Sulawesi Selatan tercatat sebanyak 784,98 ribu jiwa atau sekitar 8,78% dari total penduduk Sulawesi Selatan. Jumlah tersebut menunjukkan tingkat kemiskinan mengalami peningkatan sebesar 1,05% dibandingkan dari tahun sebelumnya, dimana jumlah kemiskinan tahun sebelumnya yaitu pada tahun 2020 menunjukkan sebesar 776,83 ribu jiwa.

Dalam mengukur tingkat kemiskinan, ada beberapa indikator yang dapat digunakan yaitu persediaan tempat tinggal, persediaan mutu pangan, persediaan air, layanan kesehatan, sanitasi, pendapatan serta persediaan pekerjaan. Permasalahan kemiskinan sendiri juga dipengaruhi oleh banyak faktor. Akan tetapi secara umum, faktor yang mempengaruhi secara signifikan terhadap masalah kemiskinan dalam suatu negara adalah tingkat pengangguran, pendapatan, serta pendidikan.

Pemerintah memiliki tugas tertinggi dalam penyelenggaraan urusan pemerintahan negara, terutama dalam pengentasan masalah kemiskinan. Oleh karena itu, pemerintah harus melakukan upaya-upaya dalam mengatasi kemiskinan secara merata dan mencakup segala aspek-aspek kehidupan dalam masyarakat. Hal tersebut merupakan hak-hak yang seharusnya di dapatkan bagi masyarakat.

Dalam mengatasi kemiskinan, pemerintah tentunya harus mengetahui kondisi atau keadaan pada setiap wilayah. Sebagaimana dengan Kabupaten/Kota di Sulawesi Selatan tentunya memiliki perbedaan karakteristik masing-masing, banyaknya keragaman karakteristik tersebut membuat sulit untuk memahami kondisi kemiskinan setiap daerah secara detail. Untuk mengetahuinya, perlu dilakukan pengelompokan untuk memahami kondisi setiap daerah tersebut.

Salah satu metode yang dapat digunakan adalah analisis *cluster*. Dengan melakukan pengelompokan berdasarkan indikator kemiskinan, tentunya mendapatkan hasil yang dapat digunakan dalam membantu dan memudahkan pemerintah dalam mengambil suatu keputusan di masa mendatang.

Salah satu bagian dari analisis *cluster* adalah analisis *cluster* hierarki. Dalam analisis *cluster* hierarki, terdapat beberapa metode seperti *single linkage*, *average linkage*, *complete linkage*, *ward*, dan metode yang lainnya. Analisis *cluster* hierarki merupakan suatu analisis yang tepat digunakan untuk melakukan proses pengelompokan berdasarkan karakteristik suatu objek. Banyaknya metode dalam analisis *cluster* hierarki, tentunya memiliki salah satu metode yang menunjukkan hasil terbaik dalam proses pengelompokan. Oleh karena itu, peneliti mengangkat permasalahan untuk menentukan metode yang optimum pada analisis *cluster* hierarki dalam proses pengelompokan Kabupaten /Kota di Provinsi Sulawesi Selatan berdasarkan indikator kemiskinan.

2. TINJAUAN PUSTAKA

Analisis Cluster

Analisis *cluster* (*Cluster analysis*) merupakan suatu teknik analisis multivariat yang mempunyai tujuan utama untuk melakukan proses pengelompokan objek-objek pengamatan yang dibagi menjadi beberapa kelompok berdasarkan karakteristik yang dimilikinya. Dalam analisis *cluster*, tidak ada asumsi yang dibuat mengenai jumlah kelompok atau struktur *cluster*. Pengelompokan dilakukan berdasarkan kesamaan atau jarak dari setiap objek-objeknya. Analisis *cluster* terbagi menjadi dua, yaitu: metode *cluster* hierarki dan *non-hierarki*.

Prinsip dasar dalam analisis *cluster* adalah melakukan proses pengelompokan objek (observasi) pada suatu *cluster* yang memiliki kemiripan yang besar dengan objek yang lainnya dalam *cluster* yang sama, tetapi sangat tidak mirip dengan anggota objek yang lain pada *cluster* yang berbeda. Hal tersebut, menunjukkan bahwa *cluster* yang baik akan mempunyai sifat:

1. Sifat *Homogenitas*: yaitu sifat yang memiliki kesamaan yang tinggi antar anggota dalam suatu *cluster* (*within cluster*).
2. Sifat *Heterogenitas*: yaitu sifat yang memiliki perbedaan/ketidaksamaan antar *cluster* yang satu dengan yang lainnya (*between cluster*).

Prosedur Analisis Cluster

Dalam menggunakan analisis *cluster* sebagai alat analisis data, terdapat beberapa prosedur yaitu:

1. Merumuskan Masalah
 Tujuan dari perumusan masalah pada tahapan ini adalah pemilihan variabel yang nantinya digunakan dalam melakukan proses pengelompokan atau pembentukan sebuah *cluster*. Variabel-variabel yang dipilih, sebaiknya harus memiliki kemiripan (*similarity*) antar objek dan harus relevan dengan masalah yang diteliti. Variabel yang dipilih juga harus berdasarkan pertimbangan-pertimbangan yang bersesuaian dengan dugaan yang diteliti nantinya.
2. Standarisasi Data

Tujuan analisis *cluster* adalah mengelompokan objek yang memiliki kemiripan sama dalam suatu *cluster*. Objek dan jarak lebih dekat, akan memiliki kemiripan satu sama lain dibanding dengan jarak yang jauh. Jika nilai pengamatan antar objek memiliki skala pengukuran yang berbeda dan cukup besar, maka data diperlukan untuk dilakukan standarisasi. Dengan standarisasi data, dapat menghilangkan pengaruh dari skala pengukuran serta memperkecil perbedaan tiap *cluster*. Standarisasi data dapat dilakukan dalam bentuk Z_{score} dengan rumus:

$$Z_{score} = \frac{x - \bar{x}}{s} \quad (2.1)$$

Dengan,
 x = Nilai amatan
 \bar{x} = Rata – rata amatan
 s = Standar Deviasi

3. Mendeteksi *Outliers*

Outliers merupakan data yang digunakan dengan nilai esktrim yang muncul dikarenakan memiliki karakteristik tersendiri (unik) atau berbeda jauh dengan observasi-observasi lainnya. Dalam mendeteksi data yang *outliers* dapat dilakukan dengan standarisasi data terlebih dahulu dan setelah itu mengecek data tersebut, dimana data dikatakan *outliers* jika nilai tidak terletak dari rentang ($\pm 2,5$).

4. Menguji Kecukupan Sampel

Untuk melihat apakah sampel telah mewakili populasi atau melihat syarat kecukupan sampel, dapat diuji dengan Uji KMO (*Kaiser Meyer Olkin*). Jika nilai $KMO < 0,5$, maka sampel yang digunakan tidak mewakili populasi. Sehingga dianjurkan untuk menambah sampel penelitian.

5. Multikolinearitas

Dalam sebuah sampel, sebaiknya data tidak saling berhubungan, dikarenakan akan menyebabkan hasil yang bias pada variabel yang memiliki hubungan. Jika data saling berhubungan, maka data yang saling berhubungan disebut multikolinearitas. Salah satu cara identifikasi multikolinearitas adalah dengan menghitung nilai VIF:

$$VIF = \frac{1}{1-R^2} \quad (2.2)$$

R^2 merupakan nilai koefisien determinasi variabel dependen dan independen. Multikolinearitas terjadi apabila nilai $VIF > 10$. Selain menghitung nilai VIF, multikolinearitas dapat dilihat dengan menghitung koefisien korelasi sederhana (*korelasi pearson*) antar variabel bebas. Jika nilai korelasi mencapai 0,8 atau melebihi, maka terjadi multikolinearitas. Jika suatu data terjadi multikolinearitas, maka direduksi dengan metode analisis komponen utama.

6. Menentukan Ukuran Kemiripan Objek

Analisis *cluster* memiliki tujuan untuk mengelompokkan objek yang memiliki kesamaan (mirip) yang sama. Sehingga, diperlukan ukuran untuk mengetahui seberapa mirip atau berbedanya objek tersebut. Mengukur kemiripan objek dapat

menggunakan pendekatan yang dinyatakan dalam jarak (*distance*) antar setiap pasangan objek. Beberapa ukuran jarak yang dapat digunakan, yaitu: ukuran jarak euclid (*Euclidean distance*), ukuran jarak *squared euclidean*, ukuran jarak manhattan, ukuran jarak *chebyshev*, ukuran jarak mahalanobis, dan ukuran jarak lainnya.

7. Metode Peng-Cluster-an

Analisis *cluster* merupakan proses mengelompokkan objek yang terbagi menjadi dua, yaitu: analisis *cluster* hierarki dan non-hierarki. Analisis *cluster* hierarki terdiri dari beberapa metode yaitu *single linkage*, *complete linkage*, *average linkage*, dan metode lainnya. Sedangkan analisis *cluster* non-hierarki, salah satu metodenya adalah k-means. Dalam pemilihan metode peng-*cluster*-an, dipilih sesuai dengan kebutuhan dari peneliti.

8. Menginterpretasi *Cluster*

Tahapan ini merupakan tahapan dalam mencari karakteristik setiap *cluster*. Tahapan ini dapat dilakukan dengan membandingkan nilai mean (rata-rata) pada masing-masing kelompok yang didasarkan dengan tujuan pemberian label pada masing-masing kelompok tersebut. Adapun cara menghitungnya dengan menggunakan rumus (*centroid*):

$$C = \frac{\sum_{j=1}^n \sum_{k=1}^p X_{jk}}{N} \quad (2.3)$$

Dimana:

C = Nilai Rata-Rata

X_{jk} = Nilai/data objek ke-j, variabel ke-k

N = Jumlah objek

Analisis Cluster Hierarki

Analisis *cluster* hierarki merupakan metode yang memenuhi pengelompokannya dua ataupun lebih objek yang memiliki kesamaan yang paling dekat, kemudian proses dilanjutkan ke objek yang memiliki kesamaan dengan jarak yang terdekat kedua. Pada umumnya, *cluster* hierarki dibagi menjadi dua jenis yaitu *Agglomerative* (Pemusatan/Penggabungan) dan *Divisive* (Pemecahan).

1. Metode Penggabungan (*Agglomerative*)

Metode penggabungan atau disebut *agglomerative* merupakan suatu metode

yang masing-masing objek dianggap satu *cluster* tersendiri, kemudian objek yang memiliki kemiripan yang paling mirip berdasarkan ukuran jarak terdekat akan bergabung menjadi satu *cluster*. Dalam metode *agglomerative* ini, dikenal beberapa metode untuk membentuk sebuah *cluster*, yaitu: *Single linkage*, *complete linkage*, *average linkage*, dan *Ward*.

Beberapa metode dalam analisis *cluster* hierarki *agglomerative* adalah sebagai berikut:

a. *Single Linkage* (Pautan Tunggal)

Metode *single linkage* (pautan tunggal) merupakan proses *peng-cluster-an* yang didasarkan pada jarak terdekat antar objek. Jika dua objek terpisah dari jarak yang pendek, maka akan digabung menjadi satu *cluster* dari kedua objek tersebut. Metode ini dimulai dengan menentukan objek-objek yang memiliki jarak terkecil dalam matriks *proximity* $D = \{d_{jk}\}$. Objek-objek yang memiliki jarak terkecil digabung, misalkan: U dan V merupakan objek yang akan di *clusterkan*, sehingga diperoleh *cluster* (UV). Untuk mencari ukuran jarak dari *cluster* (UV) dan *cluster* W, maka diperoleh:

$$d_{(UV)W} = \min \{d_{UW}, d_{VW}\} \quad (2.4)$$

Dengan,

d_{UW} = jarak terdekat dari *cluster* U dan W
 d_{VW} = jarak terdekat dari *cluster* V dan W

b. *Complete Linkage* (Pautan Lengkap)

Complete linkage (pautan lengkap) merupakan kebalikan dari *single linkage*, yaitu pengelompokan dengan prinsip jarak terjauh antar objek.

$$d_{(UV)W} = \max \{d_{UW}, d_{VW}\} \quad (2.5)$$

dengan,

d_{UW} = jarak terjauh dari *cluster* U dan W
 d_{VW} = jarak terjauh dari *cluster* V dan W

c. *Average Linkage* (Pautan Rata-Rata)

Average linkage (pautan rata-rata) merupakan proses pengelompokan dengan prinsip jarak rata-rata dari seluruh objek

dalam suatu *cluster* terhadap seluruh objek pada *cluster* lain. Rumus:

$$d_{(UV)W} = \frac{\sum_i \sum_j d_{ij}}{n_{UV} n_W} \quad (2.6)$$

Dengan,

d_{ij} = jarak antara objek ke-i pada *cluster* UV dan objek ke-j pada *cluster* W

$n_{(UV)W}, n_W$ = jumlah objek pada *cluster* (UV) dan W.

d. *Ward*

Metode *ward* merupakan metode *pengclusteran* yang bersifat *agglomerative* (penggabungan) untuk memperoleh kelompok yang memiliki varian internal sekecil mungkin. Metode ini menggunakan perhitungan yang lengkap dan memaksimalkan homogenitas dalam suatu kelompok. Untuk *pengclusteran* metode *ward*, jumlah kelompok juga ditentukan terlebih dahulu berdasarkan dendogram. Dalam metode *ward*, ukuran yang digunakan adalah *Sum of Square Error* (SSE) variabel. SSE hanya dapat dihitung jika *cluster* memiliki elemen lebih dari satu objek. SSE *cluster* yang hanya memiliki satu objek adalah nol. Adapun bentuk persamaan metode *ward* adalah:

$$SSE = \sum_{i=1}^n (X_i - \bar{X})' - (X_i - \bar{X}) \quad (2.7)$$

Dengan,

X_i = Vektor kolom yang entrinya nilai objek ke-i, $i = 1, 2, 3, \dots, n$

\bar{X} = Vektor kolom yang entrinya rata-rata nilai objek dalam *cluster*

n = Banyaknya objek dalam *cluster* yang terbentuk

Jarak Euclide (*Euclidean Distance*)

Jarak Euclid merupakan ukuran jarak dalam analisis *cluster* yang paling umum digunakan untuk mengukur jarak dari objek ke pusat. Jarak ini digunakan apabila variabel yang digunakan tidak terjadi korelasi satu sama lain atau saling orthogonal yang memiliki satuan dan skala pengukuran yang sama. *Euclidean distance* disimbolkan dengan d_{ij} dengan variabel ke-k =

1,2, ..., p. Nilai d_{ij} diperoleh perhitungan jarak Euclid:

$$d_{ij} = \sqrt{\sum_{k=1}^n (X_{ik} - X_{jk})^2} \quad (2.8)$$

Dengan,

d_{ij} = Jarak antara objek ke-i dan objek ke-j

n = Jumlah variabel yang diamati

X_{ik} = Data dari objek ke-i pada variabel ke-k

X_{jk} = Data dari objek ke-j pada variabel ke-k

Beberapa syarat yang harus dipenuhi dalam penggunaan ukuran jarak Euclidean, sebagai berikut:

1. Data tidak saling berkorelasi
2. Mempunyai skala pengukuran yang sama
3. Pengukuran pembakuan mempunyai rata-rata nol dan standar deviasi satu.

Pemilihan Metode Optimum

Untuk pemilihan metode optimum dapat membandingkan nilai korelasi cophenetic. Korelasi cophenetic adalah koefisien korelasi antar elemen asli matriks (matriks jarak Euclidean) dengan elemen yang dihasilkan matriks cophenetic. Koefisien korelasi ini dapat digunakan untuk melihat keeratan hubungan antar variabel. Dalam proses peng-*cluster*-an, untuk melihat kebaikan hasil analisis *cluster* dapat menggunakan koefisien korelasi cophenetic. Saracli, dkk (2013), menuliskan formula untuk menghitung korelasi cophenetic sebagai berikut:

$$r_{coph} = \frac{\sum_{i < k} (d_{ik} - \bar{d})^2 (d_{c_{ik}} - \bar{d}_c)^2}{\sqrt{[\sum_{i < k} (d_{ik} - \bar{d})^2] - [\sum_{i < k} (d_{c_{ik}} - \bar{d}_c)^2]}} \quad (2.9)$$

Dengan,

r_{coph} = Koefisien Korelasi Cophenetic

d_{ik} = Jarak Euclid antara objek i dan k

\bar{d} = Rata-Rata d_{ik}

$d_{c_{ik}}$ = Jarak Cophenetic objek i dan k

\bar{d}_c = Rata-Rata $d_{c_{ik}}$

Kemiskinan

Kemiskinan merupakan masalah global yang dihadapi dan menjadi perhatian orang-orang di penjuru dunia. Negara miskin di hadapkan antara masalah pertumbuhan dan distribusi pendapatan yang tidak merata. Di lain sisi, banyak negara berkembang yang mengalami

pertumbuhan ekonomi yang tinggi namun kurang memberi manfaat bagi penduduk miskinnya.

Dalam tujuan pembangunan berkelanjutan, penurunan angka kemiskinan menjadi isu yang mendapatkan perhatian serius. Masalah kemiskinan memang telah lama menjadi problem sejak dahulu. Pada masa lalu umumnya masyarakat menjadi miskin bukan karena pangan, tetapi miskin dalam bentuk minimnya kebutuhan, kemudahan atau materi. Dari ukuran kehidupan modern pada masa kini banyak dari warga negara tidak menikmati fasilitas Pendidikan, pelayanan Kesehatan, dan fasilitas-fasilitas lainnya.

Indikator-indikator kemiskinan yang dikeluarkan oleh BAPPENAS (Badan Perancangan Pembangunan Nasional), yaitu:

1. Terbatasnya kecukupan mutu pangan.
2. Terbatasnya akses dan rendahnya mutu layanan Kesehatan.
3. Terbatasnya kesempatan kerja.
4. Lemahnya perlindungan terhadap aset usaha serta perbedaan upah.
5. Terbatasnya akses terhadap air bersih.

3. METODOLOGI

Jenis penelitian yang digunakan adalah penelitian terapan dengan Jenis data dalam penelitian ini adalah data sekunder dan sumber data yaitu data yang diambil dari situs resmi Badan pusat Statistik (BPS) Provinsi Sulawesi Selatan untuk periode 2022 yang berisi data tahun 2021. Penelitian ini dilakukan dari Januari-Juni 2022. Variabel yang digunakan dalam penelitian terdiri dari 11 variabel, yaitu:

Tabel 3. 1 Variabel Penelitian

Symbol	Variabel
X_1	Persentase Jumlah Penduduk (%)
X_2	Persentase Jumlah Keluarga Penerima Manfa (%)
X_3	Rata-Rata Pengeluaran per Kapita (%)
X_4	Kepadatan Penduduk (%)
X_5	Tingkat Pengangguran Terbuka (%)
X_6	Penerima Bantuan Iuran Kesehatan (%)
X_7	Layanan Sanitasi Tidak Layak (%)
X_8	Status Kepemilikan Rumah (%)
X_9	Persentase Sumber Air Minum (%)
X_{10}	Persentase Sumber Penerangan (%)
X_{11}	Persentase Bahan Baku Utama Memasak(%)

Prosedur Analisis

Analisis data yang digunakan pada penelitian ini menggunakan aplikasi R dengan prosedur sebagai berikut:

1. Mengumpulkan Data
2. Deskriptif Data
3. Standarisasi Data
4. Mendeteksi Outlier
5. Menguji Kecukupan Sampel
6. Menguji Multikolinearitas
7. Mengukur Kemiripan Objek
8. Melakukan Proses Peng-Cluster-an
9. Menentukan Metode Optimum

4. PEMBAHASAN

Hasil Penelitian

Data Penelitian

Data diperoleh dari buku Badan Pusat Statistik yang berjudul “Provinsi Sulawesi Selatan dalam Angka 2022”, buku tersebut berisi tentang data-data Sulawesi Selatan pada tahun 2021 dan buku tersebut didapatkan dari situs resmi Badan Pusat Statistik (BPS) Provinsi Sulawesi Selatan. Data penelitian terdiri dari 11 variabel yaitu indikator-indikator kemiskinan yang telah dijelaskan pada metodologi penelitian.

Statistika Deskriptif

Statistika deskriptif dilakukan dengan tujuan untuk mengetahui secara umum karakteristik masing-masing variabel atau indikator yang digunakan dalam penelitian ini. Berikut statistika deksriptif dari indikator atau variabel dalam penelitian ini menggunakan tabel. Tabel 4.1 menunjukkan hasil statistika deskriptif untuk variabel indikator kemiskinan. Nilai tersebut menunjukkan bahwa setiap variabel tentunya memiliki perbedaan karakteristik. Sebagai contoh pada variabel X_1 yaitu jumlah penduduk terbanyak adalah Kota Makassar dan jumlah penduduk paling sedikit adalah Kabupaten Kepulauan Selayar.

Standarisasi Data

Standarisasi data dilakukan jika terdapat perbedaan satuan yang signifikan di antara variabel-variabel yang sedang diteliti. Karena pada beberapa variabel yang digunakan

Tabel 4. 1 Statistika Deskriptif Variabel

	Minimum	Median	Mean	Maksimum
X_1	138.000	312.950	380.817	1.427.600
X_2	5.690	19.375	23.296	61.639
X_3	Rp824.737,03	Rp993.683,00	Rp1.048.965,00	Rp1.622.136,85
X_4	43	212,5	603,6	7161
X_5	2,34	4,11	4,748	13,18
X_6	62,15	76,19	77,47	96,74
X_7	1,48	7,915	9,833	24,43
X_8	3,69	10,525	13,259	39,86
X_9	0	5,83	9,073	28,08
X_{10}	0	0,5	0,6146	2,61
X_{11}	0	6,355	9,414	39,25

Sumber: Hasil Analisis *software-R* dalam penelitian ini berbeda, sehingga dilakukan standarisasi dengan menggunakan persamaan (2.1). Hasil standarisasi adalah sebagai berikut:

Tabel 4.2. Hasil Standarisasi Data

No	Kabupaten /Kota	ZX ₁	ZX ₂	...	ZX ₁₁
1	Kep. Selayar	-0,87824226	-0,87869579	...	-0,1455196
2	Bulukumba	0,214421562	0,08222981	...	-0,2653685
3	Bantaeng	-0,66159028	-0,95141644	...	0,47900526
4	Jeneponto	0,089277012	1,26356048	...	-0,4751039
5	Takalar	-0,28253974	0,35697638	...	-0,7466365
⋮	⋮	⋮	⋮	⋮	⋮
23	Pare-Pare	-0,82435054	-1,27391986	...	-0,8814664
24	Palopo	-0,69992937	-1,15257406	...	-0,8440137

Sumber: Hasil Olah data

Mendeteksi Outlier

Dalam analisis *cluster* terdapat tahapan mendeteksi outlier. Data yang outlier dapat dilihat berdasarkan hasil standarisasi data, dikatakan outlier jika tidak berada diantara nilai $\pm 2,5$. Berikut ini akan ditunjukkan data yang outlier:

Tabel 4.3 Hasil Standarisasi Data

Kab/Kota	Standarisasi	Outlier
Kep.Selayar	ZX ₁₀	3,1011616
Bone	ZX ₂	2,7744857
Tana Toraja	ZX ₁₁	2,7935862
Makassar	ZX ₁	3,7861049
	ZX ₃	2,5594961
	ZX ₄	4,5797457

	ZX ₅	3,5100941
	ZX ₈	2,6689668
Pare-Pare	ZX ₃	2,7214732

Sumber: Hasil Olah data

Berdasarkan tabel 4.3 di atas bahwa terdapat 5 Kabupaten/Kota yang outlier dari 24 Kabupaten/Kota. Pada penelitian ini, data yang outlier tetap dipertahankan karena termasuk representasi dari populasi yang sedang diteliti. Data outlier tersebut menggambarkan bahwa memiliki variabel yang sangat berbeda karakteristiknya dari Kabupaten/Kota yang lain.

Menguji Kecukupan Sampel

Hipotesis dalam uji kecukupan sampel:

H₀ : Sampel belum mencukupi untuk dianalisis

H₁ : Sampel sudah mencukupi untuk dianalisis

Adapun kriteria nilai dari KMO yaitu jika nilai $KMO > 0,5$, maka Hipotesis Nol (H₀) gagal diterima artinya sampel layak dianalisis. Berdasarkan hasil komputasi dengan R-Program, diperoleh output uji kecukupan yaitu sebesar 0,525637. Berdasarkan hasil tersebut, dapat ditarik kesimpulan bahwa sampel layak untuk dianalisis lebih lanjut.

Menguji Multikolinearitas

Adanya multikolinearitas dalam analisis *cluster* merupakan suatu pelanggaran karena dapat mempengaruhi hasil *cluster*. Untuk mengetahui apakah dalam variabel yang digunakan terjadi multikolinearitas, dapat melihat nilai korelasi dari matriks korelasi. Multikolinearitas terjadi apabila nilai korelasinya menunjukkan $> 0,80$.

Pada penelitian ini, tidak terjadi multikolinearitas antar variable, sehingga tidak ada langkah penanganan yang dilakukan.

Mengukur Kemiripan Objek

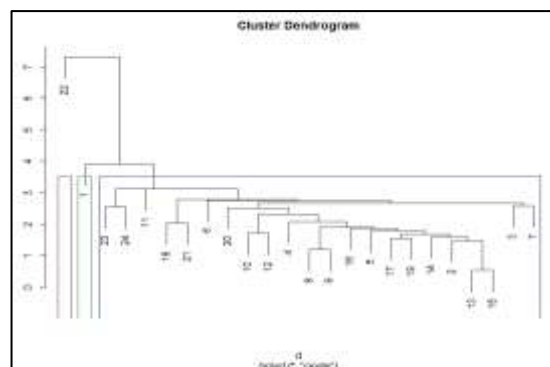
Berdasarkan data indikator kemiskinan, berikut merupakan contoh perhitungan manual antara Objek 1 dan 2, Objek 1 dan 3, Objek 2 dan 3, Objek 2 dan 4, serta Objek 3 dan 4. Keempat Objek tersebut dibandingkan menggunakan 10 variabel dengan tujuan untuk mencari dua objek yang paling mirip. Adapun persamaan yang digunakan, yaitu persamaan 2.8.

Dengan menggunakan persamaan 2.8 didapat hasil perhitungan kemiripan antara objek 1 dan 2 yaitu Kabupaten Kep. Selayar dan Bulukumba sebesar 5,431070. Untuk perhitungan kemiripan objek lainnya juga dapat menggunakan persamaan 2.8.

Melakukan Proses Peng-Cluster-an

1. *Single Linkage*

Dalam proses *cluster* hierarki dengan *single linkage*, didapatkan hasil sebagai berikut:



Gambar 1. Output Cluster Hirarki Single Linkage

Tabel 4.3 Anggota-Anggota Cluster Single Linkage

Cluster	Anggota Cluster
1 (1 Anggota)	Kep.Selayar
2 (22 Anggota)	Bulukumba, Bantaeng, Jeneponto, Takalar, Gowa, Sinjai, Maros, Pangkep, Barru, Bone, Soppeng, Wajo, Sidrap, Pinrang, Enrekang, Tana Toraja, Toraja Utara, Luwu, Luwu Utara, Luwu Timur, Pare-Pare, Palopo
3 (1 Anggota)	Makassar

Sumber: Hasil Olah data

Berdasarkan nilai centroid tiap *cluster*, interpretasi *single linkage* adalah pada *cluster* 1 memiliki nilai rata-rata tertinggi pada variabel X₆, X₇, dan X₁₀. Pada *cluster* ini, memiliki 5 variabel dengan rata-rata terendah yaitu X₁, X₂, X₃, X₄, dan X₅.

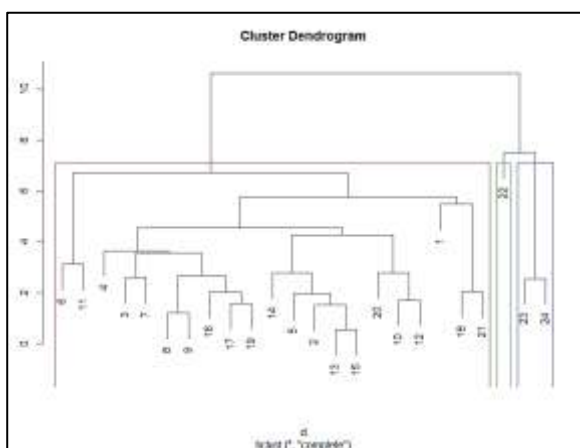
Untuk *cluster* 2 memiliki dua variabel dengan rata-rata tertinggi, yaitu: X₉ dan X₁₁. *Cluster* 2 hanya memiliki variabel dengan rata-rata terendah yaitu X₈. Sedangkan variabel lainnya, memiliki rata-rata yang

sedang atau berada diantara tertinggi dan terendah.

Untuk *cluster* 3 memiliki nilai rata-rata tertinggi terbanyak dibandingkan *cluster* 1 dan 2. Adapun rata-rata variabel tertinggi pada *cluster* ini adalah X_1, X_2, X_3, X_4, X_5 dan X_8 . Dalam *cluster* 3 ini juga memiliki lima variabel dengan rata-rata terendah yaitu X_6, X_7, X_9, X_{10} dan X_{11} .

2. *Complete Linkage*

Dalam proses *cluster* hierarki dengan *complete linkage*, didapatkan hasil sebagai berikut:



Gambar 2. Output Cluster Hierarki *Complete Linkage*

Berdasarkan nilai centroid tiap *cluster*, interpretasi *complete linkage* adalah pada *cluster* 1 memiliki nilai rata-rata tertinggi pada variabel X_7, X_9, X_{10} dan X_{11} . Pada *cluster* ini, memiliki 5 variabel dengan rata-rata terendah yaitu X_3, X_4, X_5 , dan X_6 .

Tabel 4.4. Anggota-Anggota Cluster *Complete Linkage*

Cluster	Anggota Cluster
1 (21 Anggota)	Kep.Selayar, Bulukumba, Bantaeng, Jeneponto, Takalar, Gowa, Sinjai, Maros, Pangkep, Barru, Bone, Soppeng, Wajo, Sidrap, Pinrang, Enrekang, Tana Toraja, Toraja Utara, Luwu, Luwu Utara, Luwu Timur
2 (1 Anggota)	Makassar

Sumber: Hasil Olah Data

Untuk *cluster* 2 memiliki dua variabel dengan rata-rata tertinggi, yaitu X_1, X_2, X_3, X_4, X_5 , dan X_8 . *Cluster* 2 hanya memiliki variabel dengan rata-rata terendah yaitu X_9 dan X_{11} . Sedangkan variabel lainnya, memiliki rata-rata yang sedang atau berada diantara tertinggi dan terendah.

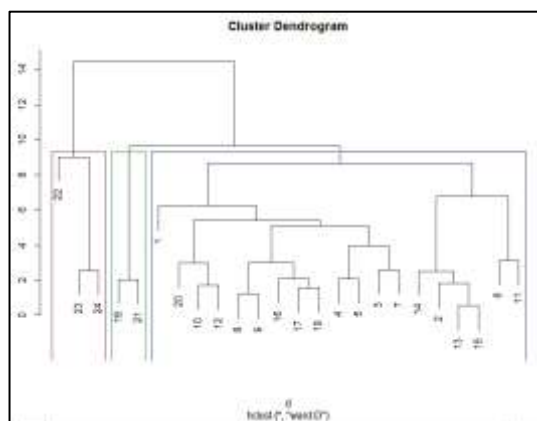
Untuk *cluster* 3 memiliki nilai rata-rata tertinggi pada variabel X_6 . Dalam *cluster* 3 ini juga memiliki dua variabel dengan rata-rata terendah yaitu X_2 dan X_{10} .

3. *Average Linkage*

Berdasarkan hasil komputasi dengan software R, didapatkan hasil pengelompokan *average linkage* sama dengan *single linkage*. Sehingga hasil pengelompokan dapat dilihat pada penjelasan *single linkage*.

4. *Ward*

Dalam proses *cluster* hierarki dengan *ward*, didapatkan hasil:



Gambar 3. Output Cluster Hierarki *Ward*

Tabel 4.5 Anggota-Anggota Cluster *Complete Linkage*

Cluster	Anggota Cluster
1 (21 Anggota)	Kep.Selayar, Bulukumba, Bantaeng, Jeneponto, Takalar, Gowa, Sinjai, Maros, Pangkep, Barru, Bone, Soppeng, Wajo, Sidrap, Pinrang, Enrekang,

Tana Toraja, Toraja Utara, Luwu, Luwu Utara, Luwu Timur

2 (1 Anggota)	Makassar
3 (2 Anggota)	Pare-Pare dan Palopo

Sumber: Hasil Olah data

Berdasarkan nilai centroid tiap *cluster*, hasil interpretasi metode *ward* adalah pada *cluster* 1 memiliki nilai rata-rata tertinggi pada variabel X_2 . Pada *cluster* ini, memiliki 5 variabel dengan rata-rata terendah yaitu X_3 , dan X_8 .

Untuk *cluster* 2 memiliki dua variabel dengan rata-rata tertinggi, yaitu : X_7, X_9, X_{10} dan X_{11} . *Cluster* 2 hanya memiliki variabel dengan rata-rata terendah yaitu X_1, X_4, X_5 , dan X_6 . Sedangkan variabel lainnya, memiliki rata-rata yang sedang atau berada diantara tertinggi dan terendah.

Untuk *cluster* 3 memiliki nilai rata-rata tertinggi terbanyak dibandingkan *cluster* 1 dan 2. Adapun variabel tertinggi yaitu X_1, X_3, X_4, X_5 dan X_8 . Dalam *cluster* 3 ini juga memiliki lima variabel dengan rata-rata terendah yaitu X_7, X_9, X_{10} dan X_{11} .

Metode Optimum

Dalam menentukan metode yang optimum dalam proses klasterisasi berdasarkan indikator kemiskinan di Provinsi Sulawesi Selatan dapat dilakukan dengan membandingkan nilai korelasi cophenetic. Berikut tabel nilai korelasi cophenetic yang diperoleh dari aplikasi R:

Tabel 4.6 Korelasi Cophenetic

Metode Cluster	Korelasi Cophenetic
Single Linkage	0,887
Complete Linkage	0,752
Average Linkage	0,925
Ward	0,704

Sumber: Hasil olah data

Berdasarkan Tabel 4.6 dapat disimpulkan bahwa metode yang terbaik atau optimum berdasarkan indikator kemiskinan di Provinsi Sulawesi Selatan adalah metode *Average*

Linkage dengan nilai korelasi Cophenetic tertinggi sebesar 0.925.

5. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan metode analisis *cluster* hierarki dalam proses pengelompokan Kabupaten/ Kota di Provinsi Sulawesi Selatan dengan indikator kemiskinan dapat diambil kesimpulan bahwa:

1. Metode analisis *cluster* hierarki yang optimum dalam proses pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan berdasarkan indikator kemiskinan adalah metode Average Linkage dengan nilai korelasi Cophenetic sebesar 0,925.
2. Hasil pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan dengan metode analisis *cluster* hierarki aglomerative didapatkan nilai korelasi Cophenetic yang terbesar yaitu metode average linkage dengan tiga cluster.

Cluster	Anggota Cluster
1 (1 Anggota)	Kep.Selayar
2 (22 Anggota)	Bulukumba, Bantaeng, Jeneponto, Takalar, Gowa, Sinjai, Maros, Pangkep, Barru, Bone, Soppeng, Wajo, Sidrap, Pinrang, Enrekang, Tana Toraja, Toraja Utara, Luwu, Luwu Utara, Luwu Timur, Pare-Pare, Palopo
3 (1 Anggota)	Makassar

6. DAFTAR PUSTAKA

[1] Alwi, Wahidah dan Muh. Hasrul. “Analisis Klaster untuk Pengelompokan Kabupaten/ Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat”. Jurnal Matematika Statistika dan Aplikasinya 6, no.1 (2018) : h. 35-42.

[2] Badan Pusat Statistik. 2022. “Provinsi Sulawesi Selatan dalam Angka 2022”

[3] Dani, Andreas Tri Rian, dkk. “Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu”.

Jambura Journal of Mathematics 1, no.2 (2019): h. 64-78.

- [4] Fathia, Annisa Nur, dkk. 2016. “Analisis Kluster Kecamatan di Kabupaten Semarang Berdasarkan Potensi Desa Menggunakan Metode Ward dan Single Linkage”. Jurnal Gaussian 5, no. 4: h.801-810.
- [5] Johnson, Richardso A dan Dean W. Wichern. 2007 “Applied Multivariate Statistical Analysis”. New York: Prentice Hall International.
- [6] Pangkey, Rani A, dkk. 2018. “Aplikasi Analisis Komponen Utama dan Analisis Gerombol pada Varietas Tanaman Hias Krisan di Kota Tomohon”. Jurnal Matematika dan Aplikasi deCartesian 7, no. 2: h. 73-77.
- [7] Prawoto, Nano. 2019. “Memahami Kemiskinan dan Strategi Penanggulangannya”. Jurnal Ekonomi dan Studi Pembangunan.
- [8] Rachmatin, Devi dan Kania Sawitri. 2019. “Perbandingan antara Metode Agglomerative, Divisif dan K-Means dalam Analisis Cluster”. Institut Teknologi Nasional Bandung: h.9-17.
- [9] Rini, Ayu Setyo, dkk. 2016. “Faktor-Faktor Penentu Kemiskinan di Indonesia”. Jurnal Ekonomi Terapan 01, no.2: h. 17-33.
- [10] Simamora, Hamelia dan Bagus Sumargo. 2019. “Pengklasifikasian Penggunaan Internet Lingkungan Pedesaan Menurut Jenjang Pendidikan di Indonesia Menggunakan Cluster Average Linkage”. Jurnal Statistika dan Aplikasinya 3, no.1: h. 22-29.
- [11] Srinadi, I Gusti Ayu Made dan I Made Agus Gelgel Wirasuta. 2014. “Penerapan Analisis Gerombol dalam Pengelompokan Lengkuas Berdasarkan Kandungan Bioaktif”. Seminar Nasional Matematika Universitas Udayana: h. 359-366.
- [12] Wahyuni, Rezki, dkk. 2015. “Analisis Kluster dengan Menggunakan Single Linkage dan Metode K-Means”. Jurnal Statistika UNIB Bengkulu: h.1-9.
- [13] Widodo, Edy, dkk. 2018. “Analsis Cluster Penderita Disabilitas Mental di Provinsi Daerah Istimewa Yogyakarta Tahun 2016”. Jurnal Program Studi Pendidikan Matematika FKIP UMS: h.577-586.