

ANALISIS *HIERARCHICAL CLUSTERING MULTISCALE BOOTSTRAP* (KASUS: INDIKATOR KEMISKINAN DI PROVINSI SULAWESI SELATAN TAHUN 2020)

Musdalifah M. Ramly¹, Sudarmin², Bobby Poerwanto³

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Makassar, Indonesia

Keywords: hierarchical clustering analysis, poverty indicators, multiscale bootstrap

Abstract:

Hierarchical cluster analysis is a statistical analysis used to group data based on their similarities. The single linkage, complete linkage and average linkage methods can be used to group data using distance techniques. There is a large difference in the number of poor people in urban and rural areas in South Sulawesi Province, so an analysis is needed to classify areas that have the same characteristics based on poverty indicators. For this reason, these three methods are used. However, the results of this analysis are only based on the similarity measure based on the distance technique used. Thus, the multiscale bootstrap method is used to obtain the validity of the resulting clusters. The results of the research using these three methods are four clusters with different characteristics. By using multiscale bootstrap, it is found that in single linkage there are four valid clusters, for complete linkage there is only one valid cluster and on average linkage there are three valid clusters. So it is found that single linkage is the best method in classifying these cases.

1. Pendahuluan

Kemiskinan merupakan awal dari permasalahan sosial yang secara tidak langsung dapat mengganggu suatu ketahanan dan stabilitas. Di Indonesia, masih banyak terdapat penduduk miskin yang tersebar di berbagai provinsi, salah satunya di Sulawesi Selatan yang persentase penduduk miskinnya mengalami kenaikan hampir dua kali lipat pada tahun 2020. Berdasarkan hasil Susenas Maret 2020, di tahun 2019 persentase penduduk miskin sebesar 8,69% dan jumlah penduduk miskin ini mengalami peningkatan di tahun 2020 yaitu sebesar 0,003% menjadi 8,72%, jumlah penduduk miskin ini terus mengalami fluktuasi setiap tahunnya. Perbedaan jumlah penduduk miskin di Sulawesi Selatan yang berada di perkotaan ataupun pedesaan juga memiliki perbedaan jumlah yang sangat jauh (Badan Pusat Statistik, 2021). Untuk mengetahui kondisi tiap daerah perlu dilakukan pengelompokan berdasarkan tingkat kemiskinan yang terjadi pada suatu daerah. Pengelompokan ini dilakukan dengan analisis *clustering*.

Analisis *clustering* mengelompokkan objek penelitian yang memiliki kemiripan atau kesamaan dengan objek lainnya berada dalam satu *cluster* yang sama. Objek yang terdapat pada masing-masing *cluster* cenderung memiliki kemiripan satu sama lain dan jauh berbeda dengan objek yang berada di *cluster* lainnya (Supranto, 2004). Analisis *cluster* terbagi atas dua yaitu metode *clustering* hirarki dan *clustering* nonhirarki.

Analisis *cluster* memiliki kelemahan yaitu tidak dapat menunjukkan hasil validasi dari hasil pengelompokan sehingga digunakan metode lainnya untuk mengetahui hasil validasi dari analisis *cluster* yang dilakukan. Metode *multiscale bootstrap* adalah metode yang dapat menunjukkan validitas dari masing-masing *cluster*

* Corresponding author.

E-mail address: musdharamly@gmail.com



yang terbentuk pada analisis *cluster* yang dilakukan sebelumnya. Metode ini bekerja dengan menghitung nilai *Approximately Unbiased* (AU) melalui pendekatan *bootstrap* resampling untuk setiap *cluster* yang terbentuk. Nilai *Approximately Unbiased* (AU) *p-values* menghasilkan taksiran yang lebih baik dalam mengatasi bias (Safarina et al., 2019).

2. Tinjauan Pustaka

2.1 Analisis Multivariat

Analisis multivariate adalah teknik statistika yang digunakan untuk menganalisis banyak variabel secara simultan. Menurut Laeli (2014), analisis multivariate terdiri dari beberapa jenis metode yaitu metode dependensi dan independensi.

2.2 Analisis Cluster

Analisis *cluster* merupakan salah satu teknik *multivariate* yang digunakan untuk mengelompokkan suatu objek-objek berdasarkan kesamaan yang dimiliki tiap objek tersebut. Objek-objek ini dikelompokkan ke dalam dua kelompok atau lebih sehingga objek yang berada dalam satu *cluster* akan memiliki kemiripan satu sama lain. *Cluster* yang baik adalah *cluster* yang terbentuk memiliki:

1. Homogenitas (kesamaan) yang tinggi antar anggota dalam satu *cluster* (*within-cluster*).
2. Heterogenitas (perbedaan) yang tinggi antar tiap kelompok (*between-cluster*).

Asumsi-asumsi yang harus terpenuhi dalam analisis *cluster* (Abidin, 2017):

1. Sampel yang diambil dapat benar-benar bisa mewakili populasi yang ada.
2. Tidak terdapat multikolinearitas. Multikolinearitas adalah adanya hubungan linear sempurna (*perfect*) atau pasti (*exact*) di antara beberapa atau semua variabel (Gujarati, 2009). Untuk melihat adanya multikolinearitas digunakan faktor inflasi ragam (*Variance Inflation Factor/VIF*). Nilai VIF yang lebih besar dari 10 menandakan adanya masalah multikolinearitas. Nilai VIF dapat dihitung dengan rumus (Sari et al., 2016):

$$VIF = \frac{1}{1 - R_j^2} \quad (2.1)$$

Dimana nilai R_j^2 adalah koefisien determinasi antara X_j dengan variabel independen lainnya dengan $j = 1, 2, \dots, k$ yang dirumuskan dengan:

$$R^2 = \left[\frac{n \sum_{i=1}^n X_1 X_2 - \sum_{i=1}^n X_1 \sum_{i=1}^n X_2}{\sqrt{[n \sum_{i=1}^n (X_1)^2 - (\sum_{i=1}^n X_1)^2][n \sum_{i=1}^n (X_2)^2 - (\sum_{i=1}^n X_2)^2]}} \right]^2 \quad (2.2)$$

Jika terdapat multikolinearitas maka yang perlu dilakukan adalah mengeluarkan variabel yang tidak berhubungan atau saling terkait. Hal ini perlu diperhatikan sehingga variabel-variabel yang bersifat multikolinearitas dapat dipertimbangkan dengan teliti (Veriani, 2020). Jika asumsi multikolinearitas telah terpenuhi maka jarak *Euclidean* dapat digunakan dalam menentukan ukuran kemiripan (Sholihah, 2021).

Tahapan-tahapan dalam analisis *cluster* adalah sebagai berikut:

- a. Merumuskan masalah. Dalam analisis *cluster*, pemilihan variabel-variabel yang digunakan sangat penting.
- b. Melakukan standarisasi data apabila variabel dari data yang digunakan dalam penelitian memiliki satuan yang saling berbeda. Untuk penelitian ini menggunakan nilai skor standar (*Z-Score*). Nilai *z-score* dapat dihitung dengan menggunakan rumus:

$$Z = \frac{X - \bar{X}_i}{s} \quad (2.3)$$

Keterangan:

Z = nilai *z-score*

X = objek penelitian

\bar{X} = rata-rata objek penelitian

s = standar deviasi

Di mana untuk mencari nilai standar deviasi menggunakan rumus, yaitu sebagai berikut:

$$s = \sqrt{\frac{\sum_{j=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (2.4)$$

Keterangan:

s = standar deviasi

x_i = nilai objek penelitian ke- i

\bar{x} = rata-rata objek penelitian

- c. Menghitung kemiripan dan ketidakmiripan antar tiap objek. Analisis *cluster* mengelompokkan objek-objek berdasarkan kemiripan yang dimiliki tiap objek. Jarak yang digunakan adalah jarak *Euclidean* yang menghitung nilai jumlah kuadrat perbedaan masing-masing variabel. Jarak antar objek ke- i dan objek ke- j disimbolkan dengan d_{ij} dan untuk variabel ke- k dengan $k = 1, \dots, p$. Nilai d_{ij} didapatkan dengan rumus (Veriani, 2020):

$$d_{ij} = \sqrt{\left(\sum_{k=1}^p x_{ik} - x_{jk}\right)^2} \quad (2.5)$$

Keterangan:

d_{ij} = jarak *Euclidean* dari objek ke- i dan objek ke- j

p = jumlah variabel *cluster*

x_{ik} = nilai objek ke- i pada variabel ke- k

x_{jk} = nilai objek ke- j pada variabel ke- k

- d. Memilih metode analisis
- e. Menentukan banyak *cluster*. Menurut Mulyaningrum (2020) ada beberapa pendekatan yang bisa dijadikan acuan dalam menentukan banyak *cluster* yaitu:
1. Pertimbangan teoritis, konseptual, praktis dapat digunakan untuk menentukan berapa banyak *cluster* yang sebenarnya.
 2. Dalam analisis *cluster* dengan metode hirarki, jarak *cluster* digabung dapat dipergunakan sebagai kriteria.
 3. Banyaknya *cluster* seharusnya bermanfaat/berguna.
- f. Menginterpretasikan hasil *cluster*. Interpretasi terhadap *cluster-cluster* yang terbentuk dilakukan untuk memberi nama yang dapat menggambarkan isi *cluster* dan selanjutnya dilakukan proses *profiling* untuk menjelaskan masing-masing *cluster* yang terbentuk berdasarkan profil tertentu (Santoso, 2017).
- g. Melakukan validasi. Proses validasi dalam analisis *cluster* bertujuan untuk mengetahui apakah hasil *clustering* bisa mewakili populasi dan dapat digeneralisasikan ke objek yang lain (Afandi, 2020).

2.3 Metode Analisis Cluster

Metode *clustering* dibagi menjadi dua yaitu metode *clustering* hirarki dan metode *clustering* non-hirarki. Metode *clustering* hirarki dimulai dengan mengelompokkan dua atau lebih objek yang mempunyai kesamaan paling dekat. Kemudian proses dilanjutkan ke objek lain yang mempunyai kedekatan kedua. Proses ini terus berlanjut hingga *cluster* membentuk seperti ‘pohon’ yaitu terdapat hirarki (tingkatan) yang jelas antar objek dari yang paling mirip hingga paling tidak mirip (Goreti et al., 2017).

a. Single Linkage

Pengelompokkan pada metode *Single Linkage* ini didasari pada jarak minimum atau jarak terdekat antar objek. Proses dimulai dengan mengelompokkan dua objek dengan jarak terdekat menjadi *cluster* pertama, proses ini terus berulang hingga seluruh objek membentuk satu *cluster* (Fadliana, 2015).

Menurut Johnson dan Wichern (2007), proses pengelompokkan metode ini dimulai dengan memilih jarak terkecil dalam $D = \{d_{ij}\}$ dan menggabungkan objek-objek yang bersesuaian misalnya U dan V untuk mendapatkan *cluster* (UV). Selanjutnya, jarak di antara (UV) dan *cluster* lainnya, misalnya W .

$$d_{(UV)W} = \min(d_{UW}, d_{VW}) \quad (2.6)$$

Keterangan:

d_{UW} = jarak antara tetangga terdekat dari *cluster* U dan W

d_{VW} = jarak antara tetangga terdekat dari *cluster* V dan W

b. Complete Linkage

Metode pautan lengkap (*Complete Linkage*) juga disebut dengan *Maksimum Linkage*, yaitu jarak antar dua *cluster* dengan menggunakan jarak yang terjauh antara anggota *cluster* yang satu dengan *cluster* yang kedua. Hal pertama yang dilakukan adalah mencari nilai minimum dalam matriks D_{ij} dan mengelompokkan objek-objek yang

bersesuaian, misalnya I dan J , untuk mendapatkan *cluster* (IJ). Lalu, jarak antara (IJ) dan *cluster* K lainnya dihitung dengan cara:

$$d_{(IJ)K} = \max(d_{IK}, d_{JK}) \quad (2.7)$$

Keterangan:

d_{IK} = jarak antara tetangga terjauh dari *cluster* I dan K

d_{JK} = jarak antara tetangga terjauh dari *cluster* J dan K

c. Average Linkage

Metode pautan rata-rata (*Average Linkage*) merupakan variasi dari *Single Linkage* dan *Complete Linkage* yaitu menghitung jarak antara dua *cluster* dengan menggunakan jarak rata-rata dimana jarak tersebut dihitung pada tiap *cluster* dengan meminimumkan jarak rata-rata objek yang digabung (Ningsih et al., 2016).

$$d_{(IJ)K} = \frac{\sum a \sum b d_{ab}}{N_{(IJ)}N_K} \quad (2.8)$$

Keterangan:

d_{ab} = jarak antara objek a dalam *cluster* (IJ) dan objek b dalam *cluster* K

$N_{(IJ)}$ = jumlah objek pada *cluster* (IJ)

N_K = jumlah objek pada *cluster* K

2.4 Validasi Analisis Cluster

Menurut Kapita dan Abdullah (2020), validasi *cluster* merupakan suatu tahap dalam analisis *cluster* yang digunakan untuk memberikan informasi penilaian hasil dari analisis *cluster* dengan pendekatan jumlah atau kuantitatif dan sesuai dengan kondisi yang sebenarnya atau objektif. Menurut Susilowati, Sugiarto dan Mardianto (2020), validasi *cluster* merupakan salah satu metode yang dapat digunakan untuk mengevaluasi hasil dari solusi *clustering* yang telah dilakukan untuk memperoleh jumlah *cluster* terbaik. Dikatakan suatu *clustering* yang terbentuk dengan jumlah *cluster* terbaik apabila solusi *cluster* yang dihasilkan tidak berbeda jauh dengan pendekatan yang digunakan.

2.5 Metode Bootstrap

Bootstrap merupakan salah satu metode pengambilan sampel ulang dari data yang diteliti (*resampling*) yang dikembangkan melalui menggabungkan metode-metode *resampling* untuk membuat inferensia tertentu yang lebih sederhana menjadi suatu metode *resampling* baru yaitu *bootstrap* nonparametrik sederhana. Pengambilan sampel ulang (*resampling*) data ini dilakukan dengan pengembalian dari sampel hasil observasi dengan replikasi B .

Berdasarkan penelitian Efron dan Tibshirani (1998), terdapat dua poin penting dalam menentukan besaran B :

1. Menggunakan replikasi *bootstrap* yang sedikit, misalnya $B = 25$, biasanya sudah cukup informative, $B = 50$ sudah cukup memberikan estimasi simpangan baku dengan baik. Dalam analisis *cluster* sendiri, jumlah replikasi yang dibutuhkan $B = 100$ sampai dengan 500.
2. Penggunaan replikasi $B = 200$ ke atas jarang digunakan untuk mengukur simpangan baku. Penggunaan B yang lebih besar diperlukan untuk membuat interval kepercayaan *bootstrap*.

2.6 Multiscale Bootstrap

Menurut Efron dan Tibshirani (1993), metode *multiscale bootstrap* dalam analisis *cluster* digunakan sebagai salah satu metode yang digunakan untuk mengetahui validitas dari hasil analisis *cluster* sebelumnya. Pada analisis *cluster*, *multiscale bootstrap* bekerja dengan cara mengambil sampel yang telah *bootstrap* yang dianggap telah mewakili distribusi data asli. Lalu, sampel *bootstrap* ini dikelompokkan dengan menggunakan analisis *cluster*. Setelah diperoleh hasil pengelompokkan dari sampel *bootstrap* maka dapat diketahui hasil analisis *cluster* yang dianalisis.

Multiscale bootstrap resampling digunakan untuk menghitung nilai *probability bootstrap* (BP) *value* dan *Approximately unbiased* (AU) *p-values* (Anuraga, 2015). Nilai *probability bootstrap* (BP) adalah suatu nilai peluang sampel data asli yang diperoleh dari proses *resampling bootstrap*. Nilai *approximately unbiased* (AU) adalah nilai yang menunjukkan validitas suatu *cluster*. Dimana suatu *cluster* dianggap telah valid apabila nilai $AU \geq 0,95$ (Shimodaira, 2004).

Berikut ini merupakan ilustrasi penerapan *multiscale bootstrap* pada *hierarchical clustering*:

1. Sampel data \vec{D} didefinisikan sebagai data sampel berukuran n yang terdiri dari $\vec{D}_i = \vec{D}_1, \vec{D}_2, \dots, \vec{D}_n$ dimana \vec{D}_i merupakan vector data pengamatan.
2. Sampel data diambil secara acak dan diperoleh \vec{D}^* sebagai sampel baru.

- Langkah ke-2 dilakukan secara berulang sebanyak B kali sehingga didapatkan himpunan data *bootstrap* $(\overrightarrow{D}^{*1}, \overrightarrow{D}^{*2}, \dots, \overrightarrow{D}^{*B})$ dimana masing-masing sampel *bootstrap* merupakan sampel acak yang saling bebas.
- Nilai v dan c diperoleh dari hasil perhitungan $z(N_i^*)$. Nilai $z(N_i^*)$ dapat dihitung dengan rumus sebagai berikut:

$$z(N_i^*) = v \sqrt{\frac{N_i'}{N}} + c \sqrt{\frac{N_i}{N'}} \quad (2.9)$$

Keterangan:

$z(N_i^*)$ = nilai z

N = jumlah data pengamatan

N' = jumlah data pengamatan *bootstrap*

v dan c = parameter untuk AU

- Nilai *Approximately Unbiased* (AU) menggunakan rumus sebagai berikut:

$$AU = z(-v + c) \quad (2.10)$$

2.7 Kemiskinan

Kemiskinan adalah suatu keadaan yang sering dikaitkan dengan kesulitan dan kekurangan dalam memenuhi kebutuhan atas konsumsi dasar untuk meningkatkan kualitas hidup seseorang (Rahmawati et al., 2021). Sehingga dapat disimpulkan bahwa kemiskinan merupakan suatu keadaan dimana seseorang atau kelompok tertentu memiliki kemampuan sosial ekonomi yang lemah yaitu kesulitan dan kekurangan dalam memenuhi kebutuhan untuk meningkatkan kualitas hidupnya.

2.8 Indikator Kemiskinan

Menurut BPS terdapat beberapa indikator kemiskinan yang berhubungan dengan kondisi di setiap kabupaten/kota, yaitu (Mustafidah, 2017):

- Luas lantai bangunan tempat tinggal <20 m² per orang.
- Lantai bangunan tempat tinggal yang terbuat dari tanah/bamboo/kayu/ berkualitas rendah.
- Dinding tempat tinggal terbuat dari bamboo/rumbia/tembok tanpa plester/kayu berkualitas rendah.
- Tidak memiliki fasilitas buang air besar/bersama-sama dengan rumah tangga lain.
- Sumber penerangan rumah tangga tidak menggunakan listrik.
- Sumber air minum berasal dari sumur/mata air tidak terlindungi/sungai/air hujan.
- Bahan bakar untuk memasak sehari-hari adalah kayu bakar/arang/minyak tanah.
- Hanya mengkonsumsi daging/susu/ayam satu kali dalam seminggu.
- Hanya membeli satu sel pakaian baru dalam setahun.
- Hanya sanggup makan sebanyak dua kali dalam sehari.
- Tidak sanggup membayar biaya pengobatan di puskesmas poliklinik.
- Sumber penghasilan kepala rumah tangga adalah: petani dengan luas lahan 0,5ha, buruh tani, nelayan, buruh perkebunan atau pekerjaan lainnya dengan pendapatan dibawah Rp. 600.000 (enam ratus ribu rupiah) perbulan.
- Pendidikan tertinggi kepala rumah tangga: tidak sekolah/tidak tamat SD/hanya SD.
- Tidak memiliki tabungan/barang yang mudah dijual dengan nilai Rp. 500.000 (lima ratus ribu rupiah), seperti: sepeda motor, (kredit/non kredit) emas, kapal motor, atau barang modal lainnya.

3. Metode Penelitian

3.1 Jenis Penelitian

Jenis penelitian yang digunakan yaitu pendekatan deksriptif kuantitatif.

3.2 Sumber Data

Data yang digunakan dalam penelitian ini adalah indikator kemiskinan di Provinsi Sulawesi Selatan tahun 2020 yang merupakan data sekunder bersumber dari Badan Pusat Statistik Provinsi Sulawesi Selatan yang dipublikasikan pada tahun 2021.

3.3 Definisi Operasional Variabel

X_1 : Persentase Penduduk Miskin Usia 15 Tahun Ke Atas Tidak Bekerja

X_2 : Persentase Penduduk Miskin Usia 15 Tahun Ke Atas Yang Bekerja di Sektor Pertanian

X_3 : Persentase Rumah Tangga Yang Pernah Membeli Beras Raskin

X_4 : Persentase Penduduk Miskin Usia 15 Tahun Yang Tidak Tamat SD

X_5 : Angka Melek Huruf Penduduk Miskin Usia 15-55 Tahun

- X₆ : Angka Partisipasi Sekolah Usia 13-15 Tahun
- X₇ : Rata-Rata Lama Sekolah
- X₈ : Persentase Rumah Tangga Miskin Yang Menggunakan Air Bersih
- X₉ : Persentase Rumah Tangga Yang Menggunakan Jamban Sendiri/Bersama

3.4 Prosedur Penelitian

Prosedur yang dilakukan dalam penelitian ini adalah:

1. Mengumpulkan referensi atau sumber-sumber informasi yang dibutuhkan dalam penelitian.
2. Mengumpulkan data indikator kemiskinan di Provinsi Sulawesi Selatan tahun 2020 berdasarkan variabel-variabel yang digunakan di penelitian ini.
3. Melakukan analisis deskriptif untuk melihat data kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan indikator kemiskinan.
4. Melakukan pengolahan data dan pengelompokkan dengan menggunakan metode *single linkage*, *complete linkage* dan *average linkage*. Lalu, untuk melihat validitas dari hasil pengelompokkan menggunakan *multiscale bootstrap*.
5. Membuat kesimpulan berdasarkan masalah yang telah dibahas.
6. Menyusun laporan penelitian.

3.5 Teknik Analisis Data

Adapun langkah-langkah yang dilakukan dalam penelitian ini adalah:

1. Melakukakn analisis deskriptif terhadap masing-masing variabel.
2. Melakukan standarisasi data.
3. Melakukakn uji multikolinearitas.
4. Membuat matriks jarak antar objek.
5. Melakukan pengelompokkan dengan metode *single linkage*, *complete linkage*, dan *average linkage*
6. Melakukan validasi dari hasil pengelompokkan dengan metode *multiscale bootstrap* dengan resampling sebanyak B=500.
7. Interpretasi hasil analisis

4. Hasil dan Pembahasan

4.1 Hasil Penelitian

4.1.1 Standarisasi Data

Standarisasi data dilakukan apabila data yang digunakan memiliki satuan yang berbeda. Berdasarkan data penelitian pada Lampiran 1, terdapat variabel yang memiliki satuan berbeda yaitu variabel rata-rata lama sekolah (X₇) dengan satuan tahun. Untuk itu, sebelum data dianalisis lebih lanjut maka perlu distandarisasi terlebih dahulu. Standarisasi data dilakukan dengan menggunakan nilai *z-score* dengan rumus pada persamaan (2.3).

Tabel 4.1 Standarisasi Data Pada Data Penelitian

X1	X2	X3	...	X7	X8	X9
-0,03	0,29	-0,48	...	-0,23	0,75	-244,49
117,58	0,24	-168,94	...	-0,41	0,62	0,23
⋮	⋮	⋮	⋮	⋮	⋮	⋮
-0,93	-208,21	0,80	...	195,67	0,92	0,99
0,59	-162,92	0,19	...	222,027	0,56	0,74

4.1.2 Asumsi Non- Multikolinearitas

Uji multikolinearitas digunakan untuk melihat ada atau tidaknya korelasi yang tinggi antar variabel-variabel independen dalam suatu model regresi linear berganda. Adanya korelasi yang tinggi di antara variabel-variabel independen akan menyebabkan hubungan variabel independen dengan variabel dependen terganggu. Pada penelitian ini, uji multikolinearitas dilihat dari nilai *Tolerance* dan VIF (*Variance Inflation Factor*) serta besaran korelasi antar variabel independen sebagai tolak ukur. Tidak terjadi multikolinearitas apabila nilai VIF ≤ 10. Diperoleh bahwa semua variabel dependen pada penelitian ini tidak mengalami gejala multikolinearitas.

4.1.3 Jarak Euclidean

Sebelumnya diketahui bahwa tidak terjadi multikolinearitas sehingga dapat disimpulkan bahwa asumsi analisis *cluster* telah terpenuhi. Selanjutnya adalah menghitung matriks jarak antar data. Dalam penelitian ini digunakan jarak *Euclidean*. Menurut Tiskadewi (2017) kelebihan jarak *Euclidean* adalah tingkat identifikasi

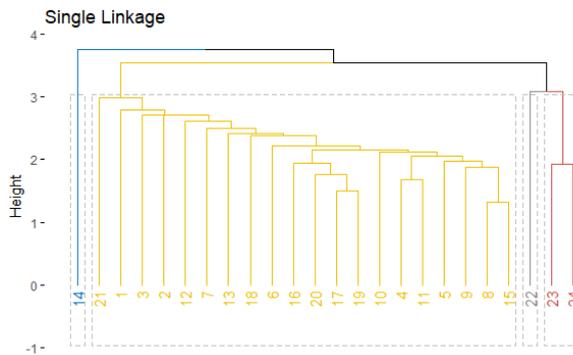
kemiripan (*similarity*) lebih tinggi. Jarak *Euclidean* dapat dihitung menggunakan persamaan (2.5) lalu membuat matriks jaraknya.

Tabel 4.2 Matriks Jarak *Euclidean*

	1	2	3	...	22	23	24
1	0,00	4,76	2,91	...	5,63	5,27	5,09
2	4,76	0,00	4,26	...	5,89	6,48	6,13
3	2,91	4,26	0,00		7,38	6,91	7,14
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
22	5,63	5,89	7,38	...	0,00	3,66	3,08
23	5,27	6,48	6,91	...	3,66	0,00	1,93
24	5,09	6,13	7,14	...	3,08	1,93	0,00

4.1.4 Metode *Single Linkage*

Pengelompokkan pada metode *Single Linkage* ini didasari pada jarak minimum atau jarak terdekat antar objek. Hasil pengelompokkan dengan menggunakan metode *single linkage* ini dapat dilihat pada gambar berikut.



Gambar 4.1 Dendrogram *Hierarchical Clustering Single Linkage*

Cluster 1: Pinrang, Maros, Pangkajene dan Kepulauan, Takalar, Bone, Jenepono, Barru, Luwu Utara, Luwu, Luwu Timur, Enrekang, Gowa, Tana Toraja, Wajo, Sinjai, Soppeng, Bulukumba, Bantaeng, Kepulauan Selayar, dan Toraja Utara.

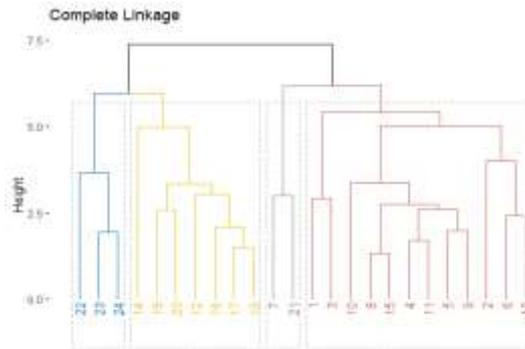
Cluster 2 : Sidenreng Rappang

Cluster 3 : Kota Makassar

Cluster 4 : Kota Pare-Pare dan Kota Palopo

4.1.5 Metode *Complete Linkage*

Metode *Complete Linkage* dilakukan dengan menggunakan jarak yang terjauh antara anggota *cluster* yang satu dengan *cluster* yang kedua. Hasil pengelompokkan dengan menggunakan *complete linkage* ini ditampilkan pada gambar berikut.



Gambar 4.2 Dendrogram *Hierarchical Clustering Complete Linkage*

Cluster 1: Wajo, Gowa, Bulukumba, Pangkajene dan Kepulauan, Takalar, Bone, Jeneponto, Pinrang, Maros, Barru, Bantaeng, Kepulauan Selayar,

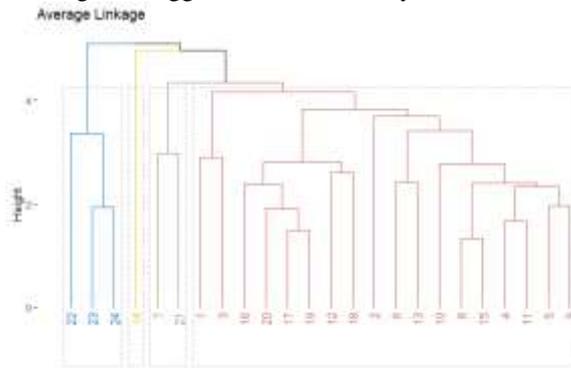
Cluster 2 : Toraja Utara dan Sinjai

Cluster 3 : Luwu Utara, Luwu, Enrekang, Luwu Timur, Tana Toraja Sidenreng Rappang

Cluster 4 : Kota Palopo, Kota Pare-Pare dan Kota Makassar.

4.1.6 Metode Average Linkage

Metode *Average Linkage* bekerja dengan cara menghitung jarak antara dua *cluster* dengan menggunakan jarak rata-rata dimana jarak tersebut dihitung pada tiap *cluster* dengan meminimumkan jarak rata-rata objek yang digabung. Hasil pengelompokkan dengan menggunakan metode ini yaitu:



Gambar 4.3 Dendrogram *Hierarchical Clustering Average Linkage*

Cluster 1: Pangkajene dan Kepulauan, Takalar, Bone, Jeneponto, Pinrang, Maros, Barru, Wajo, Gowa, Bulukumba, Tana Toraja, Soppeng, Luwu Utara, Luwu, Luwu Timur, Enrekang, Bantaeng, dan Kepulauan Selayar

Cluster 2 : Toraja Utara dan Sinjai.

Cluster 3 : Sidenreng Rappang

Cluster 4 : Kota Pare-Pare, Kota Palopo dan Kota Makassar

4.1.7 Metode Multiscale Bootstrap

Metode ini digunakan untuk menghitung nilai *Approximately Unbiased (AU) p-values* yang berfungsi memberikan taksiran lebih baik dalam mengatasi bias. *Cluster* dengan $AU \geq (0,95)$ maka dapat diinterpretasikan sebagai *cluster* yang valid. Apabila suatu *cluster* dinyatakan valid maka artinya *cluster* tersebut telah akurat dan dapat menggambarkan populasi sebenarnya.

Hasil validasi dari hasil pengelompokkan dengan metode *single linkage*, *complete linkage* dan *average linkage* menggunakan metode *multiscale bootstrap* ditampilkan dalam Tabel 4.3.

Tabel 4.3 Validasi Hasil Pengelompokkan

Metode	Cluster	Approximately Unbiased (AU)	Keterangan
<i>Single Linkage</i>	1	0,96	Valid
	2	1,00	Valid
	3	1,00	Valid
	4	0,96	Valid
<i>Complete Linkage</i>	1	0,93	Tidak Valid
	2	0,93	Tidak Valid
	3	0,93	Tidak Valid
	4	0,95	Valid
<i>Average Linkage</i>	1	0,94	Tidak Valid
	2	0,96	Valid
	3	1,00	Valid
	4	0,97	Valid

Tabel 4.3, menunjukkan bahwa hasil validasi dengan menggunakan metode *single linkage* terdapat empat *cluster* yang valid, pada *complete linkage* terdapat hanya satu *cluster* yang valid dan untuk *average linkage* terdapat tiga *cluster* yang valid. Sehingga diperoleh bahwa *single linkage* merupakan metode terbaik dalam mengelompokkan kabupaten/kota di Sulawesi Selatan berdasarkan indikator kemiskinan.

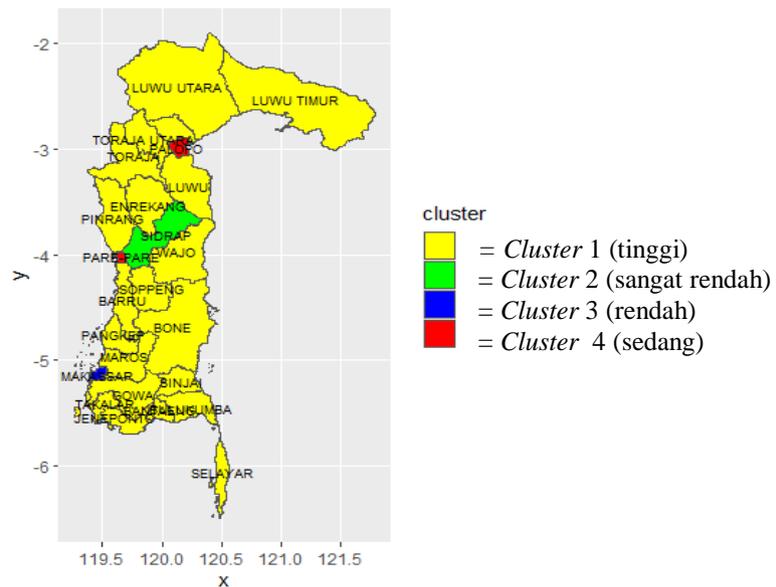
4.2 Pembahasan

Dari hasil pengelompokkan yang dilakukan dengan menggunakan metode *single linkage*, *complete linkage* dan *average linkage* diperoleh bahwa *single linkage* merupakan metode terbaik dalam mengelompokkan kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan indikator kemiskinan. Lalu, dilakukan pemberian ciri spesifik untuk menggambarkan keanggotaan dari tiap *cluster* yang terbentuk dengan metode *single linkage*. Hal ini dilakukan dengan melihat rata-rata variabel indikator kemiskinan di Provinsi Sulawesi Selatan.

Tabel 4.4 Rata-Rata *Cluster* yang Terbentuk

	Variabel									Rata-Rata
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	
Cluster 1	46,76	35,82	34,19	34,72	91,55	89,25	7,76	64	82,44	54,06
Cluster 2	55,57	16,49	6,01	22	97,57	100	7,84	55,75	96,18	50,82
Cluster 3	49,65	35,04	30,89	18,99	96,98	65,78	11,21	95,42	94,38	55,37
Cluster 4	45,99	5,76	37,825	18,64	98,44	88,63	10,61	80,92	95,69	53,61

Dimana berdasarkan Tabel 4.4, diketahui bahwa *cluster 1 cluster* dengan wilayah indikator kemiskinan yang sedang dengan rata-rata 54,06%, Pada *cluster 2* yang terdiri dari satu kabupaten yaitu Sidenreng Rappang merupakan *cluster* dengan tingkat indikator kemiskinan paling rendah dengan rata-rata 50,82%, *Cluster 3* yang terdiri dari satu daerah yaitu Kota Makassar merupakan *cluster* dengan indikator kemiskinan tertinggi dengan rata-rata 55,37% dan *Cluster 4* yang terdiri dari Kota Pare-Pare dan Palopo termasuk *cluster* dengan tingkat indikator kemiskinan yang rendah dengan rata-rata 53,61%. Hasil pengelompokkan ini dapat dilihat pada gambar di bawah ini.



Gambar 4.4 Peta Ilustrasi Hasil Pengelompokkan

Gambar 4.4 menunjukkan Kabupaten Sidenreng Rappang dan Makassar masing-masing berada pada tingkat indikator kemiskinan yang sangat rendah dan tinggi. Dilansir pada laman Satu Data Kota Pare-Pare (2020), data angka kemiskinan pada kabupaten/kota di Sulawesi Selatan tahun 2020, Kota Makassar merupakan daerah dengan kemiskinan terendah yaitu dengan persentase sebesar 4,54%. Lalu, kemiskinan terendah kedua yaitu Sidenreng Rappang sedangkan daerah yang berada di *cluster 1* memiliki persentase yang besar.

Hasil pengelompokkan ini juga sejalan dengan penelitian sebelumnya yang dilakukan oleh Fajriani (2019) mengenai persebaran tingkat kemiskinan di Sulawesi Selatan dengan hasil Kabupaten Sidenreng Rappang tergolong kabupaten/kota dengan tingkat kemiskinan yang rendah. Namun, pada hasil pengelompokkan di penelitian ini adalah

Kota Makassar tergolong ke tingkat indikator kemiskinan yang tinggi dikarenakan oleh variabel-variabel lainnya di luar variabel penelitian.

5. Kesimpulan

Berdasarkan hasil penelitian diperoleh:

1. Hasil pengelompokkan dengan menggunakan metode *single linkage* diperoleh empat *cluster* dengan *cluster* 1 terdiri dari 20 kabupaten/kota, *cluster* 2 terdiri satu kabupaten/kota, *cluster* 3 terdiri dari satu kabupaten/kota dan *cluster* 4 terdiri dua kabupaten/kota. Pada metode *complete linkage* dengan *cluster* terdiri dari 12 kabupaten/kota, *cluster* 2 terdiri dari dua kabupaten/kota, *cluster* 3 terdiri dari tujuh kabupaten/kota dan *cluster* 4 terdiri dari tiga kabupaten/kota. Dengan metode *average linkage* diperoleh *cluster* 1 yang terdiri dari 18 kabupaten/kota, *cluster* 2 terdiri dari dua kabupaten/kota, *cluster* 3 terdiri dari satu kabupaten/kota dan *cluster* 4 terdiri dari tiga kabupaten/kota.
2. Hasil validasi dari hasil pengelompokkan dengan metode *single linkage*, *complete linkage* dan *average linkage* menggunakan metode *multiscale bootstrap*, diperoleh *single linkage* terdapat empat *cluster* yang valid, *complete linkage* hanya satu *cluster* yang valid dan *average linkage* terdapat tiga *cluster* yang valid. Untuk itu, *single linkage* merupakan metode terbaik dalam mengelompokkan kabupaten/kota di Sulawesi Selatan berdasarkan indikator kemiskinan.

References

- Abidin, Z. (2017). *Pengelompokkan Kabupaten/Kota Di Jawa Timur Berdasarkan Indikator Kemiskinan Dengan Menggunakan Analisis Cluster Hierarki*. Skripsi. Surabaya: Institut Teknologi Sepuluh Noverber.
- Afandi, M. A. (2020). *Analisis Cluster Hierarki Dengan Metode Complete Linkage Pada Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Indikator Kemiskinan*. Skripsi. Malang: Universitas Islam Negeri Maulana Malik Ibrahim.
- Anuraga, G. (2015). Hierarchical Clustering Multiscale Bootstrap untuk Pengelompokkan Kemiskinan di Jawa Timur. *Statistika*, 1, 27–33.
- Badan Pusat Statistik. (2020). *Data dan Informasi Kemiskinan Provinsi Sulawesi Selatan*. Makassar: Badan Pusat Statistik Provinsi Sulawesi Selatan. diakses dari : <https://sulsel.bps.go.id>
- Efron, B. dan Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman Hall.
- Fajriani, F. (2019). Persebaran Tingkat Kemiskinan di Sulawesi Selatan Menggunakan *K-Means Clustering Analysis*. *Jurnal Semantik* 2019, 276-286. diakses dari <http://journal.uncp.ac.id/index.php/semantik/article/view/1526>
- Gujarati, D. N. (2009). *Basic Econometrics*. New Delhi: Tata McGraw-Hill Education.
- Johnson, R. A. dan Wichern, D. W. (2007). *Applied Multivariat Statistical Analysis, 6th Edition*. New Jersey: Person Prentice Hall.
- Kapita, S. N., & Abdullah, S. Do. (2020). Pengelompokkan Data Mutu Sekolah Menggunakan Jaringan Syaraf Tiruan Kohonen-SOM. *JIKO (Jurnal Informatika Dan Komputer)*, 3(1), 56–61. <https://doi.org/10.33387/jiko>
- Laeli, S. (2014). *Analisis Cluster dengan Average Linkage Method dan Ward ' s Method untuk Data Responden Nasabah Asuransi Jiwa Unit Link*. Skripsi. Yogyakarta: Universitas Negeri Yogyakarta.
- Mulyaningrum, D., Nusrang, M., & Sudarmin. (2020). *Analisis Cluster Pendekatan Metode Hierarhical Clustering Terhadap Pertumbuhan Ekonomi di Provinsi Sulawesi Selatan*. Makassar: Universitas Negeri Makassar. diakses dari <http://eprints.unm.ac.id/19073/>
- Mustafidah, R. (2017). *Pengelompokkan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Kemiskinan Dengan C-Means dan Fuzzy C-Means Clustering*. Skripsi. Surabaya: Institut Teknologi Sepuluh Noverber.
- Ningsih, S., Wahyuningsih, S., & Nasution, Y. N. (2016). PERBANDINGAN KINERJA METODE COMPLETE LINKAGE DAN AVERAGE LINKAGE DALAM MENENTUKAN HASIL ANALISIS CLUSTER (Studi Kasus : Produksi Palawija Provinsi Kalimantan Timur 2014 / 2015). *Prosiding Seminar Sains Dan Teknologi FMIPA Unmul*, 1(1), 46–50.
- Rahmawati, A., Lutfiani, L., Yunia, Z. R., Rofiqoh, I., Zahrok, F. F., & Wahyuningtyas, D. (2021). Dampak Pandemic Covid- 19 Terhadap Indeks Pembangunan Ekonomi Inklusif Jawa Timur Indikator Tingkat Kemiskinan Dan Ketimpangan. *Efektor*, 8(1), 79–88. <https://doi.org/10.29407/e.v8i1.15708>
- Safarina, Y., Hayati, M. N., & Nasution, Y. N. (2019). Penerapan Metode Hierarchical Clustering Multiscale Bootstrap (Studi Kasus: Indeks Pembangunan Manusia (Ipm) Di Kalimantan Timur Tahun 2017. *Prosiding*

- Seminat Nasional Matematika, Statistika Dan Aplikasinya*, 29–35.
<http://jurnal.fmipa.unmul.ac.id/index.php/SNMSA/article/view/523>
- Santoso, S. (2017). *Statistik Multivariat dengan SPSS*. Jakarta: PT Elex Media Komputindo.
- Sari, N., Yasin, H., & Prahutama, A. (2016). Geographically Weighted Regression Principal Component Analysis (Gwrpca) Pada Pemodelan Pendapatan Asli Daerah Di Jawa Tengah. *Jurnal Gaussian*, 5(4), 717–726.
<http://ejournal-s1.undip.ac.id/index.php/gaussian>.
- Satu Data Kota Pare-Pare. (2020). Angka Kemiskinan Kab/Kota di Sulawesi Selatan Tahun 2020 (%). diakses dari <https://satudata.pareparekota.go.id/>
- Supranto, J. (2004). *Analisis Multivariat Arti & Interpretasi*. Jakarta: Penerbit Rineka Cipta.
- Susilowati, T., Sugiarto, D., & Mardianto, I. (2020). *Jurnal Resti*. *Resti*, 4(6), 1171–1178.
- Tiskadewi, N. (2017). *Identifikasi Citra Iris Mata Dengan Metode KNN (K-Nearest Neighbor)*. Skripsi Thesis. Yogyakarta: STMIK AKAKOM. diakses dari <https://eprints.utdi.ac.id/4974/>
- Veriani, R. (2020). *Analisis Cluster Dalam Pengelompokkan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage Dan Ward*. Skripsi. Surabaya: Universitas Islam Negeri Sunan Ampel.