
Perbandingan Metode Gerombol Pautan Lengkap dan Pautan Rataan untuk Pengelompokan Kemiskinan Kabupaten/Kota di Indonesia

Akhya Wijaya¹, Nurhasanah^{2*}, Fitriana AR³, Asep Rusyana⁴

^{1,2,3,4}Jurusan Statistika, FMIPA, Universitas Syiah Kuala, Banda Aceh
E-mail: akhyar17wijaya@gmail.com, nurhasanah@unsyiah.ac.id*, fithri_a_r@yahoo.com, asepr.rusyana@unsyiah.ac.id

* = corresponding author

Abstrak

Masalah kemiskinan merupakan masalah yang kompleks dan bersifat multidimensional karena sangat erat kaitannya dengan berbagai aspek kehidupan baik sosial, ekonomi, budaya, dan aspek lainnya. Penelitian ini bertujuan untuk membandingkan dan mengelompokkan kabupaten/kota di Indonesia berdasarkan indikator tingkat kemiskinan, serta membandingkan kedua metode gerombol berhirarki dengan nilai-nilai validitas internal. Kabupaten/kota pada penelitian ini dibagi menjadi tiga wilayah waktu yaitu WIB, WITA, dan WIT. Metode yang digunakan adalah metode pautan lengkap dan metode pautan rata-rata. Data yang digunakan adalah tingkat kemiskinan kabupaten/kota yang bersumber dari Badan Pusat Statistik Indonesia tahun 2018. Hasil penelitian menunjukkan bahwa pengelompokan dengan kedua metode tersebut menghasilkan 4 gerombol. Gerombol 1 dari masing-masing zona memiliki nilai rata-rata tertinggi yang menggambarkan karakteristik tingkat kemiskinan pada gerombol 1 lebih baik dari gerombol lainnya. Metode terbaik antara metode pautan lengkap dan metode pautan rata-rata untuk pengelompokan kabupaten/kota pada masing-masing wilayah waktu adalah metode pautan rata-rata, sehingga metode pautan rata-rata memiliki kinerja yang lebih baik dari metode pautan lengkap.

Abstract

The problem of poverty is a complex and multidimensional problem because it is closely related to various aspects of life, such as social, economic, cultural, and other aspects. This study aims to compare and classify districts or cities in Indonesia based on poverty level indicators, as well as to compare the two hierarchical cluster methods with internal validity values. Districts or cities in this study are divided into three time zones, namely WIB, WITA, and WIT. The methods used are the complete linkage method and the mean linkage method. The data used is the district or city poverty level sourced from the Indonesian Central Bureau of Statistics in 2018. The results show that grouping with the two methods produces 4 clusters. Cluster 1 from each foreign zone has the highest average value which describes the characteristics of the poverty level in cluster 1 better than the other groups. The best method between the complete linkage method and the mean linkage method for grouping

Informasi Artikel

Sejarah Artikel:

Diajukan 28 Juni 2020
Diterima 5 Agt, 2020

Kata Kunci:

Indikator kemiskinan,
Pautan lengkap,
Pautan rata-rata.

Keyword:

Poverty indicators
Complete linkage
Mean linkage.

districts or cities in each time zone is the mean linkage method, so the mean linkage method has better performance than the complete linkage method.

1. Pendahuluan

Masalah kemiskinan merupakan masalah yang kompleks dan bersifat multidimensional karena sangat erat kaitannya dengan berbagai aspek kehidupan baik sosial, ekonomi, budaya, dan aspek lainnya [1]. Multidimensional tidak dapat hanya dinilai dari sisi ekonomi oleh sebab itu United Nation Development Program (UNDP) mengembangkan indeks kemiskinan multidimensi menjadi tiga dimensi yaitu kesehatan, pendidikan, dan standar kehidupan. Kemiskinan dipengaruhi oleh berbagai faktor yang saling berkaitan, antara lain: tingkat pendapatan, akses terhadap barang dan jasa, lokasi, geografis, gender, dan kondisi lingkungan [2].

Indonesia memiliki 34 provinsi dan 514 kabupaten/kota, dari 514 kabupaten/kota tidak seluruh kabupaten/kota berkategori miskin dan setiap provinsi di Indonesia memiliki tingkat kemiskinan kabupaten/kota yang berbeda-beda. *Tuberkulosis* diklasifikasikan menjadi dua yaitu *Tuberkulosis* paru dan *Tuberkulosis* Pada penelitian ini kabupaten/kota di Indonesia terlalu banyak sehingga sulit untuk mengelompokkan kabupaten/kota dengan metode pautan lengkap dan metode pautan rata-rata sehingga kabupaten/kota di Indonesia dibagi menjadi tiga zona waktu yaitu: Indonesia bagian barat, Indonesia bagian tengah dan, Indonesia bagian timur. Pembagian zona waktu dilakukan untuk mempermudah mengelompokkan kabupaten/kota miskin dan untuk mengetahui karakteristik kemiskinan kabupaten/kota pada masing-masing zona. Metode yang digunakan untuk mengelompokkan kabupaten/kota di Indonesia adalah dengan menggunakan analisis cluster atau analisis gerombol.

Analisis gerombol merupakan teknik peubah ganda yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan kemiripan karakteristik yang dimilikinya [3] [4]. Analisis gerombol bertujuan untuk mengelompokkan objek-objek berdasarkan karakteristik sehingga dapat diketahui ciri khas dari tiap kelompok.

Metode pengelompokan yang digunakan pada penelitian ini adalah metode pautan lengkap (*complete linkage*) dan pautan rata-rata (*average linkage*). Pautan lengkap mengelompokkan objek pengamatan berdasarkan jarak terjauh sedangkan pautan rata-rata mengumpulkan objek pengamatan berdasarkan jarak rata-rata dengan menggunakan cara pengukuran jarak kedekatan *euclidean*.

2. Tinjauan Pustaka

2.1. Analisis Gerombol

Analisis gerombol (cluster) merupakan teknik peubah ganda yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan kemiripan karakteristik yang dimilikinya. Karakteristik objek-objek dalam suatu gerombol memiliki tingkat kemiripan yang tinggi, sedangkan karakteristik antar objek pada suatu gerombol dengan gerombol lain memiliki tingkat kemiripan yang rendah. Dengan kata lain, keragaman dalam suatu gerombol minimum sedangkan keragaman antar gerombol maksimum [3] [4]. Penggerombolan digunakan untuk mencari kelompok dari data yang tidak memiliki kelompok secara alami juga atau sering disebut dengan *unsupervised learning* [5].

Pengelompokan ini kemudian dijadikan dasar sejumlah teknik pengolahan data seperti peringkasan (summarization) dan compression. Menurut [6] metode yang digunakan untuk menghitung jarak euclidian adalah sebagai berikut:

2.2. Jarak Euclidian

Jarak euclidian adalah jarak yang sering digunakan sebagai ukuran kemiripan atau ketidakmiripan antar objek. Jarak euclidian didefinisikan sebagai berikut:

$$D(x_1, x_2) = \|x_1, x_2\|^2 = \sqrt{\sum_{k=1}^n (x_{1k} - x_{2k})^2} \quad (1)$$

di mana :

x_{1k} : koordinat objek ke-1 dan variabel ke-k

x_{2k} : koordinat objek ke-2 dan variabel ke-k

N : banyaknya data

Menurut [7] terdapat beberapa metode gerombol berhirarki agglomerative, yakni: single linkage atau nearest neighbor method, complete linkage atau farthest neighbor method, average linkage, centroid method dan ward's method. Sedangkan k-means clustering adalah metode tak-berhirarki. Banyak metode-metode gerombol yang sudah diperkenalkan. Namun pada penelitian ini hanya fokus pada dua metode yaitu pautan lengkap (*complete linkage clustering*) dan pautan rata-rata (*average linkage clustering*).

2.3. Pautan Lengkap

Pautan lengkap atau complete linkage adalah suatu metode yang menggunakan prinsip jarak minimum yang diawali dengan mencari jarak terjauh antar dua buah gerombol dan keduanya membentuk gerombol baru [8]. Metode complete linkage atau disebut juga metode pautan lengkap, ditentukan dari jarak terjauh antara dua objek pada gerombol yang berbeda (farthest neighbor). Metode ini dapat digunakan dengan baik untuk kasus di mana objek-objek yang ada berasal dari kelompok yang benar-benar berbeda [3].

Algoritma aglomeratif umum dimulai dengan menemukan entri minimum pada $D = \{d_{ik}\}$ dan menggabungkan objek yang sesuai seperti U dan V , untuk mendapatkan gerombol (UV). Jarak antara (UV) dan setiap gerombol W dihitung dengan menggunakan persamaan berikut [9]:

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

di mana D_{UW} dan D_{VW} adalah jarak antara anggota yang paling jauh dari kelompok U dan W dan kelompok V dan W secara berurutan.

2.4. Pautan Rataan

Metode Pautan rata-rata atau disebut juga metode *average linkage*, ditentukan dari rata-rata jarak seluruh objek suatu gerombol terhadap seluruh objek pada gerombol lainnya [3]. Pautan rata-rata memperlakukan jarak antara dua kelompok sebagai jarak rata-rata antara semua pasangan *item* di mana satu anggota pasangan menjadi milik masing-masing kelompok [9]. Pada metode pautan lengkap, penentuan gerombol dicari melalui rata – rata dari semua pasangan data [10].

Algoritma berhirarki secara umum memiliki tahapan sebagai berikut [11]:

1. Menentukan objek bersesuaian yang memiliki jarak terdekat dalam matriks jarak $D = \{d_{ik}\}$,
2. Menggabungkan objek yang bersesuaian tersebut, katakanlah objek U dan objek V , yang kemudian didapatkan gerombol (UV).
3. Mengitung jarak antara gerombol (UV) dengan objek lain katakanlah W yang belum bergabung, dengan rumus :

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

di mana :

d_{ik} = Jarak antara objek i pada gerombol UV dan objek k pada gerombol

$N_{(UV)}$ = Jumlah masing-masing objek dalam gerombol UV

N_W = Jumlah masing-masing objek dalam gerombol W

4. Mengulangi Langkah 2 sampai bergabung menjadi satu gerombol.

2.5. Metode Elbow

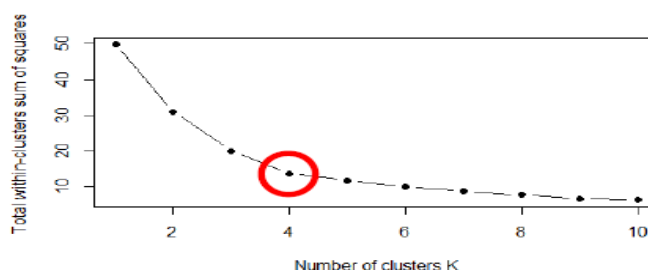
Menurut [12] disebutkan bahwa metode *Elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah gerombol terbaik dengan cara melihat persentase hasil perbandingan antara jumlah gerombol yang akan membentuk siku pada suatu titik. Metode ini digunakan dalam analisis gerombol untuk interpretasi dan uji performa tingkat konsistensi jumlah gerombol yang tepat dengan melihat nilai SSE. Untuk mendapatkan perbandingannya adalah dengan menghitung SSE (*Sum of Square Error*) dari masing-masing nilai gerombol. Karena semakin besar jumlah gerombol k maka nilai SSE akan semakin kecil [13]. Rumus untuk menghitung nilai SSE adalah sebagai berikut :

$$SSE = \sum_{k=1}^k \sum_{x_j \in S_k} \|X_i - C_k\|^2 \quad (4)$$

di mana :

- k = jumlah gerombol
- X_i = data ke- i
- C_k = rata-rata dari variabel suatu gerombol

Nilai SSE tersebut kemudian digunakan pada grafik *elbow*. Grafik pada metode *elbow* terdiri dari sumbu x dan y . Sumbu x merupakan jumlah gerombol (k) dan sumbu y merupakan nilai SSE. Penentuan jumlah gerombol (k) menggunakan kriteria siku, di mana pada titik tertentu akan terjadi penurunan secara drastis namun tidak landai dengan sebuah lekukan yang disebut dengan kriteria siku. Nilai itu kemudian menjadi jumlah gerombol (k) yang terbaik [12]. Gambar 1 menunjukkan grafik dari metode *elbow*.



Gambar 1 Grafik metode *elbow* (Taqwim et al, 2019)

2.6. Validitas Gerombol

Validitas gerombol dalam analisis gerombol sudah menjadi bagian penting dalam proses penggerombolan untuk mengevaluasi hasil penggerombolan yang telah dilakukan. Pada penelitian ini, hasil gerombol diuji tingkat validitasnya menggunakan validitas internal. Validitas internal digunakan untuk memilih algoritma gerombol terbaik serta jumlah gerombol (k) optimal tanpa informasi tambahan apapun [14]. Pengukuran validitas internal dalam analisis gerombol terdiri dari *connectivity*, *dunn index*, dan *silhouette index* [5].

2.7. Connectivity

Validitas internal menggambarkan tiga ukuran yaitu *compactness* (kekompakan), *connectedness* (keterkaitan), dan *separation* (pemisahan) [5]. *Connectivity* merupakan pengukuran berdasarkan aspek *connectedness* (keterkaitan). *Connectivity* diukur untuk melihat seberapa jauh konektivitas antara observasi dengan tetangga terdekatnya yang ditempatkan dalam kelompok yang sama dalam ruang data. Nilai *connectivity* berada diantara 0 sampai ∞ , di mana jika nilai *connectivity* semakin mendekati nol maka hasil penggerombolan yang dihasilkan semakin baik. Jika $nn_{(i)(j)}$ adalah tetangga terdekat dari observasi i dan x_i , $nn_{(i)(j)}$ akan bernilai nol jika i dan j berada pada gerombol yang sama. *Connectivity* untuk k gerombol didefinisikan sebagai berikut:

$$Conn(C) = \sum_{i=1}^N \sum_{j=1}^L x_i, nn_{i(j)} \quad (5)$$

untuk: $j = 1, 2, \dots, p$; $i = 1, 2, \dots, N$; $C = \{C_1, C_2, \dots, C_k\}$

di mana:

$nn_{i(j)}$ = Tetangga terdekat pengamatan ke- i

k = jumlah gerombol

N = jumlah pengamatan

x_i = sampel ke- i

L = parameter yang menentukan jumlah tetangga konduktivitas

2.8. Dunn Index

Dunn index dapat mengidentifikasi set gerombol yang terpisah dengan baik dan kompak (Saitta et al., 2007). Menurut [15] *dunn index* menghitung validitas gerombol menggunakan diameter gerombol (kohesi) dan jarak antara dua gerombol (separasi). Untuk mendapatkan diameter sebuah gerombol ke- i dilakukan dengan menghitung jarak pasangan dua data dalam sebuah gerombol, kemudiandiambil yang terbesar, seperti yang dinyatakan oleh persamaan berikut:

$$\Delta i = \max_{x, y \in C_i} \{d(x, y)\} \quad (6)$$

Dunn index untuk k gerombol didefinisikan sebagai berikut (Brock et al., 2008):

$$D(C) = \frac{\min_{C_k, C_l \in C, C_k \neq C_l} \left(\min_{i \in C_k, j \in C_l} \right)}{\max_{C_m \in C} \text{diam}(C_m)} \quad (7)$$

di mana, $D(C)$ adalah *dunn index* dan k adalah jumlah gerombol. Nilai *dunn index* berada diantara 0 sampai ∞ . $\text{diam}(C_m)$ adalah jarak maksimum antara pengamatan di kelompok C_m Jika nilai *dunn index* semakin besar maka hasil penggerombolan yang dihasilkan semakin baik.

2.9. Silhouette Index

Silhouette Index adalah rata-rata dari setiap nilai *silhouette* (SI) untuk setiap observasi. Nilai *silhouette index* berada diantara -1 sampai 1, di mana jika nilai *silhouette index* semakin mendekati 1 maka penggerombolan yang dilakukan semakin baik.

Nilai *silhouette index* (SI) merupakan nilai yang dapat digunakan untuk memvalidasi gerombol yang mengabungkan nilai kohesi dan separasi. Untuk menghitung nilai SI dari sebuah data ke- i , terdapat 2 komponen yaitu a_i dan b_i . a_i adalah rata-rata jarak data ke- i terhadap semua data lainnya dalam satu gerombol, sedangkan b_i didapatkan dengan menghitung rata-rata jarak data ke- i terhadap semua data dari gerombol lain yang tidak dalam satu gerombol dengan data ke- i , kemudian diambil yang terkecil. Berikut adalah formula untuk menghitung a_i^j [15]:

$$a_i^j = \frac{1}{m_j - 1} \sum_{\substack{r=1 \\ r \neq i}}^{m_j} d(x_i^j, x_r^j) \quad (8)$$

untuk : $i = 1, 2, \dots, m_j$

di mana:

a_i^j = rata-rata jarak data ke- i terhadap semua data lainnya dalam gerombol j

m_j = jumlah data dalam gerombol ke- j

k = jumlah gerombol

x_i^j = jarak data ke- i dalam satu gerombol j

x_r^j = jarak data ke- r dalam satu gerombol j
 berikut adalah formula untuk menghitung nilai *silhouette index* (SI) data ke- i menggunakan persamaan berikut:

$$SI_i^j = \frac{b_i^j - a_i^j}{\max\{a_i^j, b_i^j\}} \quad (9)$$

Nilai a_i mengukur seberapa tidak mirip sebuah data dengan gerombol yang diikutinya, nilai semakin kecil menandakan semakin tepatnya data tersebut berada dalam gerombol tersebut. Nilai b_i yang besar menandakan seberapa jeleknya data terhadap gerombolan lain.

2.10. Kemiskinan

Kemiskinan merupakan salah satu penyakit dalam ekonomi yang dianggap sebagai sumber berbagai kejahatan dan kegiatan sumbang, sehingga harus disembuhkan atau paling tidak dikurangi. Permasalahan kemiskinan merupakan permasalahan yang kompleks dan bersifat multidimensional [16]. Penyebab kemiskinan dipandang dari segi ekonomi adalah akibat dari rendahnya kualitas sumber daya manusia. Rendahnya kualitas sumber daya manusia ini disebabkan oleh rendahnya pendidikan, Kualitas sumber daya manusia yang rendah berarti produktivitasnya juga rendah, yang pada gilirannya upahnya juga rendah. Kemiskinan dapat menyebabkan kurang gizi bagi masyarakat. Kurang gizi ini dapat ditunjukkan dengan skor Indeks Masa Tubuh (IMT) [17] dan [18].

Terdapat dua jenis kemiskinan, yaitu kemiskinan alamiah dan kemiskinan buatan. Kemiskinan alamiah adalah kemiskinan yang terbentuk sebagai akibat adanya kelangkaan sumber daya alam dan minimnya atau ketiadaan pra sarana umum (jalan raya, listrik, dan air bersih), dan keadaan tanah yang kurang subur. Kemiskinan alamiah juga bisa terjadi karena bencana alam. Bencana alam telah terjadi sebanyak 2.700 kali di Indonesia dalam rentang 2016 s.d. 2018 [19]. Kemiskinan buatan adalah kemiskinan yang diakibatkan oleh sistem modernisasi atau pembangunan yang menyebabkan masyarakat tidak memiliki banyak kesempatan untuk menguasai sumber daya, sarana, dan fasilitas ekonomi secara merata.

2.11. Data dan Variabel Penelitian

Penelitian ini menggunakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS) Indonesia. Data yang digunakan adalah data kemiskinan di Indonesia tahun 2018 sebanyak 514 kabupaten/kota yang ada di Indonesia. Terdapat 23 variabel indikator kemiskinan multidimensi pada penelitian ini, namun berdasarkan dari BPS hanya 9 variabel yang digunakan pada penelitian ini, hal ini karena pada 9 variabel tersebut terdapat data lengkap di setiap kabupaten/kota. Tabel 1. adalah daftar variabel yang digunakan pada penelitian ini.

Tabel 1 Variabel penelitian

Variabel	Keterangan
X ₁	Persentase penduduk miskin usia 15 tahun ke atas menurut kabupaten/kota dan pendidikan yang ditamatkan, tahun 2018
X ₂	Persentase angka melek huruf penduduk miskin menurut kabupaten/kota dan golongan umur 15-55 tahun, tahun 2018
X ₃	Persentase angka partisipasi sekolah penduduk miskin menurut kabupaten/kota dan golongan umur 7-12 tahun, tahun 2018
X ₄	Persentase penduduk miskin usia 15 tahun ke atas menurut kabupaten/kota status tidak bekerja, tahun 2018
X ₅	Persentase pengeluaran per kapita untuk makanan menurut kabupaten/kota status miskin, tahun 2018
X ₆	Persentase perempuan berstatus miskin usia 15-49 tahun yang menggunakan alat KB menurut kabupaten/kota, tahun 2018

Variabel	Keterangan
X ₇	Persentase rumah tangga miskin yang menggunakan air layak menurut kabupaten/kota, tahun 2018
X ₈	Persentase rumah tangga miskin yang menggunakan jamban sendiri/bersama menurut kabupaten/kota, tahun 2018
X ₉	Persentase rumah tangga miskin yang menerima beras miskin (raskin)/beras sejahtera (rastra) menurut kabupaten/kota, tahun 2018

2.12. Prosedur Analisis Data

Prosedur analisis data yang digunakan untuk menyelesaikan penelitian ini digunakan langkah-langkah sebagai berikut:

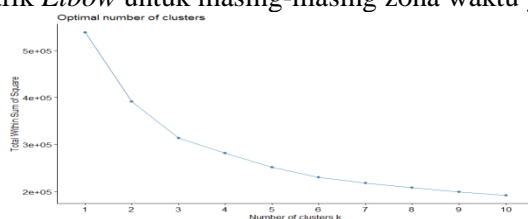
Adapun langkah-langkah yang dilakukan untuk mencapai tujuan penelitian adalah sebagai berikut:

- Memberikan kode untuk penamaan kabupaten/kota pada dendrogram berdasarkan masing-masing zona waktu.
- Melakukan analisis gerombol berdasarkan metode pautan lengkap dan pautan rata-rata untuk masing-masing zona waktu.
- Metode pautan lengkap dan pautan rata-rata adalah bagian dari gerombol berhirarki agglomerative. Berikut ini merupakan langkah-langkah dalam algoritma pengelompokan berhirarki agglomerative:
 - a) Menggabungkan objek yang berdekatan satu gerombol.
 - b) Menghitung jarak objek yang bergabung menjadi satu gerombol dengan objek yang lain.
 - c) Menghitung jumlah gerombol metode pautan lengkap dari jarak terjauh dan menghitung jumlah gerombol metode pautan rata-rata dari jarak rata-rata.
 - d) Mencari dua gerombol yang mempunyai jarak paling dekat dengan menggunakan metode pautan lengkap.
 - e) Ulangi langkah dua, tiga dan empat sampai tersisa satu kelompok dan berhenti ketika jumlah gerombol yang terbentuk sama dengan 1.
- Menentukan jumlah gerombol menggunakan metode *elbow*
- Membandingkan hasil gerombol antara metode pautan lengkap dan pautan rata-rata dengan menggunakan validitas gerombol yaitu : *Connectivity*, *Dunn index*, *Silhouette Index*.
- Kesimpulan.

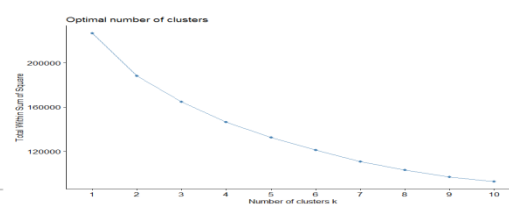
3. Hasil dan Pembahasan

3.1. Penentuan Perkiraan Jumlah Gerombol 3 Zona Waktu

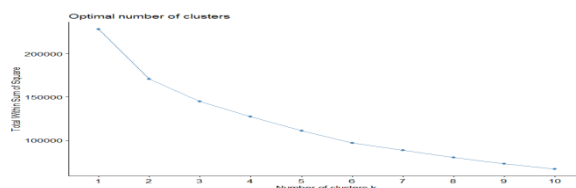
Penentuan perkiraan jumlah gerombol (k) pada penelitian ini menggunakan metode *Elbow*. Metode gerombol yang digunakan pada penelitian ini ada 2 yaitu, pautan lengkap dan pautan rata-rata. Indonesia memiliki tiga zona waktu Indonesia yaitu, WIB, WITA, dan WIT. Karena terdapat dua metode dan tiga zona waktu dengan jarak yang digunakan oleh kedua metode adalah sama, maka SSE (*Sum of Square Error*) untuk kedua metode akan sama sehingga, terdapat 3 grafik *Elbow* untuk masing-masing zona waktu yang ditunjukkan pada Gambar 2 – Gambar 4.



Gambar 2 Grafik *Elbow* zona WIB



Gambar 3 Grafik *Elbow* zona WITA



Gambar 4 Grafik *Elbow* zona WIT

Gambar 2 menjelaskan penurunan nilai Within Sum of Square saat jumlah gerombol (K) semakin besar. Pada zona WIB lekukan siku pada grafik *Elbow* yang diperkirakan menjadi jumlah gerombol terbaik pada saat jumlah gerombol 4 ($K=4$), 5 ($K=5$), 6 ($K=6$), 7 ($K=7$), dan 8 ($K=8$). Pada zona WITA (Gambar 3) perkiraan jumlah gerombol terbaik pada saat jumlah gerombol 4 ($K=4$), 5 ($K=5$), 6 ($K=6$), dan 7 ($K=7$), begitu juga pada zona WIT (Gambar 4) jumlah gerombol terbaik sama dengan zona WITA. Setelah mendapatkan perkiraan jumlah gerombol terbaik pada masing-masing zona, selanjutnya dilakukan penentuan jumlah gerombol terbaik dengan menggunakan validitas internal.

3.2. Penentuan Jumlah Gerombol Terbaik

Penentuan jumlah gerombol (K) terbaik dilakukan untuk memilih k terbaik dari perkiraan jumlah gerombol (K) yang telah ditentukan. Penentuan gerombol terbaik ditentukan berdasarkan masing-masing metode pada setiap zona waktu. Hasil penentuan gerombol terbaik masing-masing metode pada setiap zona waktu dapat dilihat pada Tabel 2.

Tabel 2 Nilai validitas internal untuk metode pautan lengkap dan pautan rata-rata serta zona waktu Indonesia

Zona waktu	Metode	Validitas Internal	Jumlah Gerombol			
			4	5	6	7
WIB	Pautan Lengkap	<i>Connectivity</i>	107.57	134.52	138.29	145.81
		<i>Dunn Index</i>	0.1338	0.1427	0.1431	15.38
		<i>Silhouette Index</i>	0.2077	0.2007	0.2021	0.1987
	Pautan Rataan	<i>Connectivity</i>	70.55	76.81	79.73	99.78
		<i>Dunn Index</i>	0.137	0.137	0.137	0.158
		<i>Silhouette Index</i>	0.2004	0.1603	0.1484	0.1607
WITA	Pautan Lengkap	<i>Connectivity</i>	76.24	99.98	102.13	117.69
		<i>Dunn Index</i>	0.1704	0.1561	0.1593	0.1803
		<i>Silhouette Index</i>	0.1193	0.1292	0.1398	0.1243
	Pautan Rataan	<i>Connectivity</i>	9.053	20.71	32.93	60.96
		<i>Dunn Index</i>	0.2532	0.21	0.21	0.1955
		<i>Silhouette Index</i>	0.128	0.0519	0.0351	0.0922
WIT	Pautan Lengkap	<i>Connectivity</i>	36.4	43.95	50.3	51.36
		<i>Dunn Index</i>	0.2467	0.2584	0.2515	0.2625
		<i>Silhouette Index</i>	0.1982	0.1582	0.1827	0.1897
	Pautan Rataan	<i>Connectivity</i>	15.07	19.03	21.2	24.13
		<i>Dunn Index</i>	0.3278	0.3278	0.3278	0.3225
		<i>Silhouette Index</i>	0.2925	0.2564	0.235	0.1222

Jumlah gerombol terbaik yang diperoleh untuk metode pautan lengkap pada semua zona waktu Indonesia adalah sama yaitu $K=4$ (Tabel 2). Sedangkan pautan rata-rata juga diperoleh hasil yang sama seperti pautan lengkap. Hal tersebut dikarenakan gerombol $k=4$ yang memenuhi kriteria teori validitas internal di mana nilai *Connectivity* terkecil, sedangkan untuk *Dunn Index* dan *Silhouette Index* bernilai tertinggi. Kriteria validitas internal akan terus digunakan untuk tiga

zona waktu Indonesia dan hanya satu metode gerombol yang digunakan untuk menganalisis ketiga zona waktu Indonesia.

3.3. Penentuan Metode Terbaik

Penentuan metode terbaik pada penelitian ini menggunakan tiga pengukuran validitas internal yaitu *Connectivity*, *Dunn Index*, dan *Silhouette Index*. Penentuan metode terbaik dilakukan dengan membandingkan kedua metode yaitu pautan lengkap dan pautan rata-rata dengan jumlah gerombol (K) yang digunakan adalah 4. Nilai validitas internal untuk penentuan metode terbaik ditunjukkan pada Tabel 3.

Tabel 3 Nilai validitas internal untuk penentuan metode terbaik

Zona Waktu	Validasi Internal	Jumlah Gerombol (4)	
		Pautan Lengkap	Pautan Rataan
WIB	<i>Connectivity</i>	107.57	70.55
	<i>Dunn Index</i>	0.1338	0.137
	<i>Silhouette Index</i>	0.2077	0.2004
WITA	<i>Connectivity</i>	76.24	9.053
	<i>Dunn Index</i>	0.1704	0.2532
	<i>Silhouette Index</i>	0.1193	0.128
WITA	<i>Connectivity</i>	36.4	15.07
	<i>Dunn Index</i>	0.2467	0.3278
	<i>Silhouette Index</i>	0.1982	0.2925

Tabel 3 menjelaskan nilai validitas internal untuk setiap metode pada masing-masing zona waktu Indonesia. Nilai validitas internal yang ditebalkan pada Tabel 3. menunjukkan hasil terbaik dari validitas internal dan berdasarkan tabel tersebut, metode terbaik yang terpilih yaitu metode pautan rata-rata, di mana pada metode pautan rata-rata mempunyai nilai *Connectivity* lebih kecil dari pada metode pautan lengkap, sedangkan untuk nilai *Dunn Index*, dan *Silhouette Index* lebih tinggi dari metode pautan lengkap. Berdasarkan validitas internal metode pautan rata-rata terpilih sebagai metode terbaik untuk menganalisis zona waktu WIB, WITA, dan WIT pada data kemiskinan tahun 2018.

3.4. Metode Pautan Rataan

Hasil penentuan validitas internal diperoleh metode pautan rata-rata sebagai metode terbaik yaitu pada jumlah gerombol 4 ($k=4$). Analisis gerombol dengan menggunakan metode pautan rata-rata akan dilakukan pada masing-masing zona waktu Indonesia (WIB, WITA, dan WIT). Gerombol yang sudah terbentuk pada masing-masing zona waktu Indonesia akan dilakukan pemeringkatan berdasarkan jumlah rata-rata dari setiap variabel tingkat kemiskinan.

a) Zona Waktu Indonesia Bagian Timur (WIT)

Gerombol yang sudah terbentuk selanjutnya dihitung jumlah rata-rata pada setiap variabel dalam masing-masing gerombolnya di mana jumlah rata-rata terendah dari setiap variabel kemiskinan merupakan peringkat untuk mengetahui karakteristik dari suatu gerombol. Jumlah rata-rata dari setiap gerombol dan variabeln zona WIT dapat dilihat pada Tabel 4.

Tabel 4 Rata-rata indikator kemiskinan setiap gerombol untuk zona WIT

Gerombol	Indikator Kemiskinan								
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
1	49.86	92.24	93.74	38.87	64.91	45.93	61.47	59.09	59.56
2	15.9	30.44	63.80	8.32	64.30	16.78	6.03	97.25	99.22
3	18.28	36.13	53.74	14.93	66.23	47.27	19.74	51.57	6.70
4	22.85	54.22	72.55	6.86	71.49	37.06	32.94	13.43	76.43

Penentuan peringkat gerombol mendapatkan karakteristik dari setiap gerombol yang sudah terbentuk, sehingga gerombol yang sudah terbentuk memiliki ciri spesifik untuk menggambarkan karakteristik-karakteristik dari gerombol yang terbentuk. Berikut adalah karakteristik-karakteristik dari keempat gerombol yang terbentuk:

- 1) Gerombol 1 memiliki nilai rata-rata tertinggi yaitu hanya pada variabel (X_4), di mana pada variabel (X_4) menyatakan semakin tinggi persentasenya maka semakin miskin di suatu daerah.
- 2) Gerombol 2 memiliki nilai rata-rata terendah yaitu pada variabel (X_1), (X_2), (X_5), (X_6), (X_7). Sedangkan pada variabel (X_9) semakin tinggi persentasenya maka semakin miskin.
- 3) Gerombol 3 hanya memiliki nilai rata-rata terendah pada variabel (X_3).
- 4) Gerombol 4 memiliki nilai rata-rata terendah yaitu hanya pada variabel (X_8).

b) *Zona Waktu Indonesia Bagian Tengah (WITA)*

Tabel 5 Rata-rata indikator kemiskinan setiap gerombol untuk zona WITA

Gerombol	Indikator Kemiskinan								
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	20.92	100	100	22.13	66.20	70.72	97.71	97.71	2.29
2	17.64	95.47	89.71	43.91	64.37	81.94	93.57	85.35	91.33
3	42.65	100	100	38.96	59.86	44.89	81.57	34.82	100
4	49.86	95.60	98.29	41.77	64.25	70.59	65.21	69.57	55.34

Penentuan peringkat gerombol dilakukan supaya mendapatkan karakteristik dari setiap gerombol yang terbentuk, sehingga gerombol yang sudah terbentuk memiliki ciri spesifik untuk menggambarkan karakteristik-karakteristik dari gerombol yang terbentuk. Berikut adalah karakteristik-karakteristik dari keempat gerombol yang terbentuk:

- 1) Gerombol 1 tidak memiliki nilai rata-rata yang menyatakan miskin di setiap variabel kemiskinan. Hal ini menunjukkan kabupaten/kota di gerombol 1 tidak miskin.
- 2) Gerombol 2 memiliki nilai rata-rata terendah yaitu pada variabel (X_2), (X_3), (X_4).
- 3) Gerombol 3 memiliki nilai rata-rata terendah yaitu pada variabel (X_1), (X_5), (X_6), dan (X_8). Sedangkan pada variabel (X_9) semakin tinggi persentasenya maka semakin miskin.
- 4) Gerombol 4 memiliki karakteristik tersendiri di mana nilai rata-rata terendah yaitu pada variabel (X_7) yang menyatakan semakin tinggi persentasenya maka semakin miskin di suatu daerah.

c) *Zona Waktu Indonesia Bagian Barat (WIB)*

Tabel 6 Rata-rata indikator kemiskinan setiap gerombol untuk zona WIB

Gerombol	Indikator Kemiskinan								
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	52.97	98.79	99.21	43.78	63.33	72.31	75.08	88.48	29.19
2	52.89	97.92	97.87	35.80	67.09	66.96	46.21	42.88	38.27
3	52.83	98.70	99.25	14.48	67.53	43.14	47.42	68.94	75.66
4	55.66	97.93	99.12	40.19	66.97	75.10	55.33	69.87	69.19

Penentuan peringkat gerombol dilakukan supaya mendapatkan karakteristik dari setiap gerombol yang terbentuk, sehingga gerombol yang sudah terbentuk memiliki ciri spesifik untuk menggambarkan karakteristik-karakteristik dari gerombol yang terbentuk. Berikut adalah karakteristik-karakteristik dari keempat gerombol yang terbentuk:

- 1) Gerombol 1 memiliki nilai rata-rata terendah yaitu pada (X_5) dan (X_4) yang menyatakan semakin tinggi persentasenya maka semakin miskin di suatu daerah.
- 2) Gerombol 2 memiliki nilai rata-rata terendah yaitu pada variabel (X_1), (X_2), (X_3), (X_7) dan (X_8).

- 3) Gerombol 3 memiliki nilai rata-rata terendah yaitu pada variabel (X_1), Sedangkan pada (X_9) semakin tinggi persentasenya maka semakin.
- 4) Gerombol 4 tidak terdapat nilai rata-rata terendah di setiap indikator tingkat kemiskinan. Hal ini menunjukkan kabupaten/kota di gerombol 4 tidak miskin.

Jumlah Kabupaten/Kota dan Persentase Tingkat Kemiskinan setiap Zona Berdasarkan Gerombol

Tabel 7 Persentase Tingkat Kemiskinan berdasarkan gerombol pada zona WIT

Gerombol	Jumlah kabupaten/kota	Persentase
Gerombol 1	54	85.7 %
Gerombol 2	1	1.59 %
Gerombol 3	2	3.18 %
Gerombol 4	6	9.53 %
Total	63	100 %

Tabel 8 Persentase Tingkat Kemiskinan berdasarkan gerombol pada zona WITA

Gerombol	Jumlah kabupaten/kota	Persentase
Gerombol 1	1	0.67 %
Gerombol 2	1	0.67 %
Gerombol 3	1	0.66 %
Gerombol 4	147	98.0 %
Total	150	100 %

Tabel 9 Persentase Tingkat Kemiskinan berdasarkan gerombol pada zona WITA

Gerombol	Jumlah Kabupaten/Kota	Persentase
Gerombol 1	102	33.88 %
Gerombol 2	32	10.63 %
Gerombol 3	5	1.67 %
Gerombol 4	162	53.82 %
Total	301	100 %

4. Kesimpulan dan Saran

4.1. Kesimpulan

Kesimpulan yang diperoleh untuk penggerombolan kabupaten/kota di Indonesia pada masing-masing zona waktu adalah sebagai berikut:

- 1) Metode terbaik antara metode pautan lengkap dan metode pautan rataan untuk penggerombolan kabupaten/kota pada masing-masing zona adalah metode pautan rataan. Hal ini diketahui dari perbandingan nilai-nilai validitas internal dari 3 zona antara metode pautan lengkap dan pautan rataan.
- 2) Penggerombolan kabupaten/kota di Indonesia pada masing-masing zona waktu menghasilkan 4 gerombol untuk metode pautan lengkap dan pautan rataan, didapat dari kriteria validitas internal (*Connectivity*, *Dunn Index*, dan *Silhouette Index*).
- 3) Jumlah kabupaten/kota pada tiap gerombol untuk zona WIB secara berturut-turut adalah 102, 32, 5, dan 162 kabupaten/kota. Jumlah kabupaten/kota pada tiap gerombol zona WITA secara berturut-turut adalah 1, 1, 1, dan 147 kabupaten/kota. Sedangkan jumlah kabupaten/kota pada tiap gerombol zona WIT secara berturut-turut adalah 54, 1, 2, dan 6 kabupaten/kota.
- 4) Nilai rata-rata indikator kemiskinan yang relatif paling miskin untuk zona WIT yaitu pada gerombol 2 sebesar 1,59% kabupaten/kota. Kabupaten/kota yang relatif paling miskin pada zona WITA didominasi oleh gerombol 2 dan gerombol 3 yaitu sebesar 1,33%. Sedangkan

pada zona WIB, kabupaten/kota yang relatif paling miskin berada pada gerombol 2 yaitu sebesar 10,63%.

4.2. *Saran*

Penelitian ini hanya mengkaji tentang dua metode gerombol berhirarki yaitu metode pautan lengkap dan metode pautan rata-rata. Untuk penelitian selanjutnya diharapkan dapat mengkaji tentang metode-metode gerombol yang lain dalam berbagai bidang, seperti bidang kesehatan, sosiologi, kriminologi, dan lain sebagainya. Serta mencoba untuk membentuk gerombol dengan menggunakan zona-zona lain seperti zona 5 pulau besar Indonesia.

Daftar Kepustakaan

- [1] Badan Pusat Statistika. 2016. *Perhitungan dan Analisis Kemiskinan Makro Indonesia 2016*.
- [2] Syakti, F. 2013. "Sistem Informasi Data Kemiskinan Kabupaten Banyuasin Sumatera Selatan. Seminar Nasional Informatika 2013. ISSN: 1979-2328.,"
- [3] Mattjik, A.A. and Sumertajaya, I.M. 2011. *Sidik Peubah Ganda dengan Menggunakan SAS*. Bogor: IPB Press.
- [4] Husna, I. Rusyana, A. Muslem, Idroes, G.M. Suhendra R. and Idroes, R. 2020. "Grouping of Retention Index on Gas Chromatography using Cluster Analysis," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 79, no. 1, p. 012064., 2020.
- [5] Brock, G. Pihur, V. Datta, S. and Datta, S. 2008. "clValid: An R Package for Cluster Validation," *J. Stat. Softw.*, vol. 25, no. 4, 2.
- [6] Agusta, Y. 2007. "K-Means – Penerapan, Permasalahan dan Metode Terkait," *J. Sist. dan Inform.*, vol. 3, pp. 1–14.
- [7] Rivani, E. 2010. "Aplikasi K-Means Cluster Untuk Pengelompokan Provinsi Berdasarkan Produksi Padi, Jagung, Kedelai, Dan Kacang Hijau Tahun 2009," *J. Mat Stat*, vol. 10, no. 2, pp. 122–134.
- [8] Pradnyana, A.G. and Ngurah, A.S. 2012. "Perancangan Dan Implementasi Automated Document Integration Dengan Menggunakan Algoritma Complete Linkage Agglomerative Hierarchical Clustering," *J. Ilmu Komput.*, vol. 5, no. 2, pp. 1–10.
- [9] Johnson, R.A. and Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice-Hall.
- [10] Subekti, R. Kusumawati, R. Sari, E.R. J. 2017. "K-Means Clustering dan Average Linkage dalam Pembentukan Portfolio Saham," *Semin. Mat. Dan Pendidik. Mat. Uny*, pp. 219–224.
- [11] Alwi, W. and Hasrul, M. 2018. "Analisis Klaster Untuk Pengelompokan Kabupaten/Kota Di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat," *J. MSA (Mat. dan Stat. serta Apl.)*, vol. 6, no. 1, p. 35.
- [12] Madhulatha, T.S. 2012 "An Overview On Clustering Methods," *IOSR J. Eng.*, vol. 2, no. 4, p. 723.
- [13] Muningsih, E. and Kiswati, S. 2018. "Sistem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan Clustering Pelanggan," *Joutica*, vol. 3, no. 1, p. 117.
- [14] Liu, Y. Li, Z. Xiong, H. Gao, X. and Wu, J. 2010. "Understanding of Internal Clustering Validation Measures," *IEEE Int. Conf. Data Min.*, vol. 911.
- [15] Prasetyo, E. 2014. *Data Mining : Mengolah Data Menjadi Informasi Menggunakan*

Matlab. Yogyakarta: Andi Publisher.

- [16] Benazir and Azharsyah 2017. "Analisis Faktor-Faktor Yang Mempengaruhi Kemiskinan Di Kabupaten Pidie Jaya", pp. 1–80.
- [17] Kesuma, Z.M. Rusdiana, S. Rusyana, A. Rahayu, L. dan Rosadi, R. 2019. Aplikasi Analisis Korespondensi Berganda Terhadap Status Gizi Remaja di Kota Banda Aceh. *Buletin Penelitian Kesehatan*, vol 7, no 1, pp 47-54.
- [18] Kesuma, Z.M. Rusyana, A. dan Rahayu, L. Factors affecting adolescent nutritional status in Banda Aceh, Indonesia. *Journal of Physics: Conference Series*, vol 1490, no 1, 012049.
- [19] Yana, M.S. Setiawan, L. Ulfa, E. M. dan Rusyana, A. 2018. "Penerapan Metode K-Means dalam Pengelompokan Wilayah Menurut Intensitas Kejadian Bencana Alam di Indonesia Tahun 2013-2018". *Journal of Data Analysis*, vol 1, no 2, pp. 93-102.