SMARTICS Journal, Vol.12 No.02, 2026. p106-116
ISSN online: 2476-9754, ISSN print: 2623-0429
DOI : https://doi.org/10.21067/smartics.v12i2.14712

Hierarchical Clustering Kabupaten/Kota Di Sumatera Berdasarkan Faktor-faktor yang Berhubungan dengan Indeks Pembangunan Manusia

Anwar Fitrianto[a*], Akbar Nugraha[b], I Made Sumertajaya[c]
[a,b,c] Sekolah Sains Data Matematika dan Informatika, IPB University, Bogor, Indonesia
[*]correspondence email : anwarstat@gmail.com

Abstract*—Components of the Human Development Index will be used to cluster regions with similar characteristics and provide information on development conditions in the region, especially in Sumatra. One of the clustering methods that will be used is the centralized hierarchical method using factor scores in principal component analysis to enable clustering. This study aims to examine the clustering of districts/cities in the Sumatra region based on factors related to human development. The results show that the centralized hierarchical clustering in Sumatra based on factors related to human development produces six optimal clusters, except for the centroid method which produces five optimal clusters. The optimal clustering results from the single linkage method, the average linkage method, and the centroid method have their respective similarities. Based on the clustering results for each method, the results of the six clusters of the Ward method provide better clustering performance results seen from the number of cluster members that are more evenly distributed and diverse compared to the other methods.

***Index Terms*—Analysis; Cluster; Hierarchical; Optimal; Ward**

Abstrak*—Komponen-komponen dalam Indeks Pembangunan Manusia (IPM) digunakan untuk melakukan penggerombolan daerah yang memiliki karakteristik yang sama dan memberikan informasi mengenai kondisi pembangunan di daerah tersebut khususnya di Sumatera. Salah satu metode penggerombolan yang akan digunakan yaitu dengan metode berhierarki terpusat dengan menggunakan skor faktor pada analisis komponen utama agar penggerombolan dapat dilakukan. Penelitian ini bertujuan untuk melihat penggerombolan kabupaten/kota di wilayah Sumatera berdasarkan faktor-faktor yang berhubungan dengan pembangunan manusia. Hasil penelitian menunjukkan penggerombolan berhierarki terpusat di Sumatera berdasarkan faktor yang berhubungan dengan pembangunan manusia menghasilkan enam gerombol optimal, kecuali pada metode centroid yang menghasilkan lima gerombol optimal. Hasil gerombol optimal dari metode pautan tunggal, pautan rataan, dan centroid memiliki kemiripan masing-masingnya. Berdasarkan hasil gerombol tiap metode, hasil enam gerombol metode Ward memberikan hasil kinerja gerombol yang lebih baik dilihat dari jumlah anggota gerombol yang lebih merata dan beragam dibanding metode yang lain.

***Kata Kunci*—Analisis; Gerombol; Hirarki; Optimal; Ward**

I. PENDAHULUAN

Pembangunan manusia merupakan suatu langkah dalam membuat suatu perubahan menjadi lebih baik dan meningkat untuk setiap individu maupun kelompok, [1]. Menurut UNDP, [2] pembangunan manusia sebagai suatu proses untuk memperluas pilihan-pilihan bagi penduduk. Pilihan-pilihan penduduk ini nantinya menggambarkan pencapaian kesejahteraan dari penduduk itu sendiri, [3]. Pembangunan manusia sendiri diukur menggunakan Indeks Pembangunan Manusia (IPM). Terdapat tiga faktor yang mempengaruhi Indeks Pembangunan Manusia (IPM) yaitu pendidikan, kesehatan, dan standar hidup layak, [4]. Beberapa faktor yang berhubungan dengan IPM dari suatu daerah adalah angka melek huruf, jumlah sekolah, angka harapan hidup, tenaga kesehatan, jumlah kelahiran, pendapatan per kapita, dan lain-lain. Pelaksanaan pembangunan manusia di Indonesia tidak dapat dilakukan secara bersamaan. Hal ini dikarenakan setiap daerah memiliki karakteristik yang berbeda-beda, [5].

Pengelompokan daerah-daerah yang memiliki karakteristik yang sama akan memudahkan untuk mengetahui bagaimana gambaran umum mengenai kondisi pembangunan manusia di Indonesia. Pengelompokan ini juga dapat menjadi bahan pertimbangan untuk memutuskan daerah yang perlu mendapat perhatian lebih terhadap kebijakan pembangunan khususnya untuk daerah-daerah di wilayah

Sumatera. Pengelompokan atau penggerombolan dapat dilakukan dengan metode analisis gerombol. Analisis gerombol merupakan salah satu teknik yang digunakan dalam mengelompokkan objek ke dalam beberapa kelompok yang relatif homogen, yang disebut dengan gerombol ([6], [7]). Metode ini telah banyak digunakan dalam penelitian terkini, misalnya untuk meninjau tren algoritma clustering dalam data mining, [8]. Selain itu, komparasi algoritma K-Means dan DBSCAN juga telah diterapka.n dalam pengelompokan indeks pembangunan gender di tingkat kabupaten/kota di Indonesia, [9].

Penggerombolan metode berhierarki menggabungkan objek satu dengan objek lain menjadi sebuah gerombol berdasarkan kemiripan yang dimilikinya. Metode yang akan digunakan pada penelitian ini yaitu dengan menggunakan metode berhierarki terpusat yang terdiri dari pautan tunggal, pautan lengkap, pautan rataan, metode Ward dan metode centroid. Perbedaan metode penggerombolan akan memberikan hasil gerombol yang berbeda-beda dan menarik untuk diketahui serta diharapkan dapat menghasilkan sebuah gerombol yang relevan sesuai dengan kondisi pembangunan manusia di Indonesia saat ini. Berdasarkan hal tersebut, penelitian ini bertujuan untuk menggerombolkan kabupaten/kota di wilayah Sumatera berdasarkan faktor-faktor yang berhubungan dengan Pembangunan Manusia dan membandingkan hasil gerombol berdasarkan beberapa metode penggerombolan ([10], [11], [12]).

II. METODE PENELITIAN

A. Data

Data yang digunakan untuk penelitian ini merupakan data yang berasal dari Badan Pusat Statistik (BPS) yang berupa data dari 10 provinsi di Sumatera. Peubah-peubah yang digunakan terdiri dari 3 aspek yang mempengaruhi IPM, [13], yaitu:

1. Kesehatan yaitu jumlah fasilitas kesehatan, jumlah tenaga kesehatan, dan kelahiran bayi.
2. Pendidikan yaitu jumlah sekolah (SD/SMP/SMA), rasio jumlah guru terhadap murid (SD/SMP/SMA).
3. Standar Hidup Layak yaitu kepadatan penduduk, Produk Domestik Regional Bruto (PDRB) harga berlaku, dan Pengeluaran per kapita.

Tabel 1 menyajikan penjelasan mengenai setiap indikator yang digunakan untuk melakukan penggerombolan:

Table 1 Indikator Pembangunan Manusia

Aspek	Peubah	Indikator
Pendidikan	X_1	Jumlah sekolah dasar (SD) tiap kabupaten/kota.
	X_2	Jumlah sekolah menengah pertama (SMP) tiap kabupaten/kota
	Peubah	Indikator
	X_3	Jumlah sekolah menengah atas (SMA) tiap kabupaten/kota
	X_4	Rasio guru terhadap murid SD
	X_5	Rasio guru terhadap murid SMP
	X_6	Rasio guru terhadap murid SMA
Kesehatan	X_7	Jumlah Dokter
	X_8	Jumlah Perawat
	X_9	Jumlah Bidan
	X_{10}	Jumlah fasilitas kesehatan tiap km^2
	X_{11}	Jumlah gizi buruk tiap kelahiran bayi
Standar Hidup Layak	X_{12}	Kepadatan penduduk
	X_{13}	PDRB (Triliun Rupiah)
	X_{14}	Pengeluaran per kapita (Juta Rupiah)

B. Prosedur Analisis Data

Prosedur analisis data yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Melakukan eksplorasi data dari tiap peubah yang digunakan seperti rataan, ragam, nilai maximum, dan nilai minimum,

2. pemilihan peubah-peubah yang akan dipergunakan untuk penggerombolan. Peubah yang dipilih berdasarkan penelitian sebelumnya, teori atau suatu pertimbangan berkenaan dengan hipotesis yang diuji,
3. menentukan pengukuran jarak pada penggerombolan. Pada penelitian ini digunakan jarak euclidean,
4. memulai penggerombolan dengan dua atau lebih objek yang mempunyai kesamaan paling dekat. Kemudian diteruskan pada objek lain dan seterusnya hingga gerombol membentuk pohon dimana terdapat tingkatan yang jelas antar objek,
5. melakukan penggerombol pada tiap-tiap metode yang akan digunakan,
6. menentukan jumlah gerombol terbaik menggunakan nilai keragaman dalam gerombol dan keragaman antar gerombol dari gerombol yang dihasilkan, *Compactness Index*, dan *Silhouette Coefficient*,
7. melihat perbandingan kemiripan antar gerombol optimal beberapa metode yang digunakan dengan Rand Index, dan
8. melakukan interpretasi terhadap gerombol yang telah terbentuk, yang pada intinya memberi karakteristik spesifik untuk menggambarkan isi gerombol tersebut.

C. Analisis Gerombol

Analisis gerombol merupakan suatu analisis statistika peubah ganda yang digunakan untuk mengelompokkan objek pengamatan berdasarkan kemiripan karakteristik yang dimiliki, sehingga objek-objek yang berada dalam gerombol yang sama memiliki kemiripan yang lebih besar dibandingkan objek dari gerombol yang berbeda, [14]. Kemiripan antar objek dapat diukur dengan menggunakan ukuran jarak, salah satunya jarak euclidean, [15]. Jarak euclidean merupakan tipe jarak yang paling umum dipilih karena perhitungannya yang sederhana. Jarak euclidean ditulis sebagai berikut:

$$d_{x,y} = \sqrt{\sum_i (x_i - y_i)^2}$$

dengan x_i dan y_i berurutan adalah unsur ke-i dari vektor x dan y , dan juga $d_{x,y}$ merupakan jarak antara vektor x dan y.

Menurut [14] , ada dua macam metode penggerombolan yaitu metode penggerombolan berhierarki dan metode penggerombolan tak berhierarki. Pada metode penggerombolan berhierarki banyaknya gerombol tidak ditentukan, sehingga jumlahnya tidak diketahui sampai penggerombolan selesai. Hasil penggerombolan ditampilkan dalam bentuk diagram dua dimensi yang disebut *dendrogram*, yang dapat mengilustrasikan gabungan objek dari tiap analisis gerombolnya, [16]. Sementara itu, pada metode penggerombolan tak berhierarki banyaknya gerombol sudah ditentukan sebelum penggerombolan. Pada penelitian ini pembahasan dibatasi pada metode penggerombolan berhierarki. Metode penggerombolan berhierarki yang biasa digunakan adalah metode pemusatan. Metode ini meliputi pautan tunggal, pautan lengkap, pautan rataan, metode ward, dan metode *centroid*.

1. Pautan tunggal

   Metode analisis gerombol pautan tunggal menggunakan prinsip penggerombolan didasarkan pada jarak terpendek antar anggotanya. Jarak antara gerombol gabungan *(i, j*) dengan gerombol lainnya (*k*) adalah $d_{(i,j)k}$= minimun (d_{ik},d_{jk}).
2. Pautan lengkap

   Metode analisis gerombol pautan lengkap merupakan kebalikan dari metode pautan tunggal, yaitu dengan menggunakan prinsip penggerombolan didasarkan pada jarak terjauh antar anggotanya. Jarak antara gerombol gabungan (*i, j*) dengan gerombol lainnya (*k*) adalah $d_{(i,j)k}$= maksimum (d_{ik},d_{jk}).
3. Pautan rataan

   Metode analisis gerombol pautan rataan didasarkan pada jarak rata-rata antar semua anggota dalam satu gerombol dengan semua anggota gerombol lain. Jarak antara gerombol gabungan (*i, j*) dengan gerombol lainnya (*k*) adalah

   $$d_{(i,j)k} = \frac{n_i}{n_i + n_j} d_{(i,k)} + \frac{n_j}{n_i + n_j} d_{(j,k)}$$
4. Metode ward

   Metode ward merupakan metode analisis gerombol yang didasarkan atas jumlah kuadrat antara dua gerombol untuk seluruh peubah. Jarak antara gerombol gabungan (*i, j*) dengan gerombol lainnya (*k*) adalah

$$d_{(i,j)k} = \frac{n_i+n_k}{\boldsymbol{n_i+n_j+n_k}}\, d_{(i,k)} + \frac{n_j+n_k}{n_i+n_j+n_k}\, d_{(j,k)} - \frac{n_k}{n_i+n_j+n_k}\, d_{(i,j)}$$

Metode ini berbeda dengan metode lainnya karena menggunakan pendekatan analisis ragam untuk menghitung jarak antar cluster atau metode ini meminimumkan jumlah kuadrat (ESS).

$$ESS = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$$

dengan x_i merupakan amatan ke-*i*

5. Metode *centroid*

Metode *centroid* didasarkan pada jarak antar *centroid*. *Centroid* merupakan jarak rata-rata semua objek dalam satu gerombol. Jarak antara gerombol gabungan (i, j) dengan gerombol lainnya (k) adalah

$$d_{(i,j)k} = \frac{n_i}{n_i+n_j}\, d_{(i,k)} + \frac{n_j}{n_i+n_j}\, d_{(j,k)} - \frac{n_i n_j}{n_i+n_j}\, d_{(i,j)}$$

Keterangan:

$d_{(i,j)k}$: jarak antara gerombol gabungan (i, j) dengan gerombol lainnya (*k*)
$d_{(i,k)}$: jarak antara gerombol ke-*i* dengan gerombol ke-*k*
$d_{(j,k)}$: jarak antara gerombol ke-*j* dengan gerombol ke-*k*
n_i : jumlah objek gerombol ke-*i*
n_j : jumlah objek gerombol ke-*j*
n_k : jumlah objek gerombol ke-*k*

Dalam melakukan penggerombolan nantinya akan dilakukan terlebih dahulu analisis Komponen Utama untuk mengatasi perbedaan skala pengukuran sebelum dilakukan penggerombolan. Komponen Utama merupakan kombinasi linear dari peubah yang diamati, informasi yang terkandung pada komponen utama merupakan gabungan dari semua peubah dengan bobot tertentu. Kombinasi linear yang dipilih merupakan kombinasi linear dengan ragam paling besar yang memuat informasi paling banyak. Antar Komponen Utama bersifat ortogonal, tidak berkorelasi dan informasinya tidak tumpang tindih, [15].

Misalkan x_1 , … , x_p merupakan *p* peubah yang menjadi perhatian, memiliki sebaran peubah ganda dengan vektor rataan dan matriks peragam. Komponen Utama seperti telah dijelaskan di atas merupakan kombinasi linier dari *p* peubah asal, atau dapat ditulis :

$$\mathbf{Y} = \mathbf{A}\,\mathbf{X}$$

dimana :

$$\mathbf{Y}=\begin{bmatrix} Y_1 \\ \dots \\ Y_p \end{bmatrix},\ \mathbf{A}=\begin{bmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pp} \end{bmatrix},\ \mathbf{X}=\begin{bmatrix} X_1 \\ \dots \\ X_p \end{bmatrix}$$

Sehingga komponen utama pertama dapat di tulis sebagai :

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

Keterangan:

a_{ij} : vektor ciri dari tiap peubah
X_i : peubah bebas
$\underline{x}$: vektor rataan
Y : matriks komponen utama
A : matriks dari vektor ciri
X : matriks vektor rataan

III. HASIL DAN PEMBAHASAN

A. Eksplorasi Data

Deskripsi pada data dilakukan untuk mengetahui sebaran nilai dari setiap indikator yang berhubungan dengan pembangunan manusia seperti nilai rata-rata, nilai minimum, nilai maksimum, korelasi, dan pencilan. Deskripsi data pada nilai rata-rata, nilai minimum dan nilai maksimum disajikan dalam bentuk statistik deskriptif untuk setiap aspek. Gambaran umum pembangunan manusia pada aspek pendidikan dapat dilihat pada Tabel 2 .

Beberapa data yang digunakan pada penelitian ini memiliki korelasi yang tinggi, untuk itu perlu dilakukan standardisasi pada indikator-indikator yang akan digunakan. Tabel 1 tersebut memperlihatkan bahwa nilai *minimum* dan *maximum* pada tiap peubah memperlihatkan masih ada kesenjangan yang cukup tinggi diantara kabupaten/kota di Sumatera. Nilai rataan pada setiap peubah cenderung lebih mendekati nilai *minimum* yang menandakan kabupaten/kota yang memiliki nilai yang tinggi merupakan kabupaten/kota yang menjadi pencilan atas dan juga secara umum kabupaten/kota di Sumatera masih berkisar dinilai yang kecil tiap peubahnya. Hal ini memberikan informasi masih belum meratanya jumlah sekolah di Sumatera. Kota Medan merupakan kabupaten/kota yang memiliki jumlah sekolah tertinggi, hal ini menandakan bahwa Kota Medan merupakan wilayah dengan jumlah fasilitas pendidikan yang terbaik untuk menampung siswa dalam menempuh pendidikan.

Table 2 Statistika Deskriptif

Peubah	Nilai			
	Minimum	Maksimum	Rataan	Simpangan baku
X_1	25.00	861.00	232.94	159.02
X_2	7.00	380.00	65.90	51.18
X_3	3.00	214.00	29.00	29.03
X_4	0.04	0.15	0.07	0.02
X_5	0.04	0.18	0.08	0.02
X_6	0.05	0.15	0.08	0.02
X_7	9.00	3143.00	146.00	300.73
X_8	65.00	5143.00	485.00	562.18
X_9	13.00	1521.00	416.00	262.54
X_{10}	0.00	1.03	0.11	0.20
X_{11}	0.00	0.04	0.00	0.01
X_{12}	15.00	8481.00	573.00	1228.30
X_{13}	1.00	204.30	20.09	28.52
X_{14}	5.59	25.92	10.14	2.41

Beberapa daerah seperti Kota Sabang dan Kota Solok memiliki jumlah sekolah yang paling sedikit, ini menandakan bahwa jumlah sekolah disana sudah cukup untuk menampung jumlah siswa yang sedikit di wilayah tersebut. Indikator yang berhubungan dengan rasio jumlah guru terhadap murid ada Kabupaten Pidie Jaya dan Kabupaten Bener Meriah memiliki rasio tertinggi yang berarti jumlah murid yang ditampung masih sedikit dibanding kabupaten/kota yang lain. Sedangkan, indikator jumlah fasilitas kesehatan tiap km^2 dan jumlah gizi buruk tiap kelahiran bayi, informasi yang bisa dilihat bahwa adanya perbedaan yang cukup jauh dari sisi jumlah fasilitas kesehatan tiap km^2 yang menggambarkan bahwa belum meratanya jumlah fasilitas kesehatan di Sumatera lalu gizi buruk tiap kelahiran bayi masih cukup kecil nilainya yaitu berkisar diantara 0 sampai 3.75%. Pada peubah jumlah fasilitas kesehatan tiap km^2 (X_{10}), rasio tertinggi terdapat di Kota Bukittinggi yang berarti bahwa Kota ini memiliki fasilitas kesehatan yang baik untuk mencakup masyarakat yang ada di wilayah Kota ini. Pada peubah jumlah gizi buruk tiap kelahiran bayi (X_{11}), rasio tertinggi terdapat di Kabupaten Nias Barat dengan nilainya itu 3.75% berarti untuk setiap kelahiran seratus bayi disana, tiga sampai empat bayinya mengalami gizi buruk.

Dari beberapa indikator yang ada, Kota Medan merupakan kabupaten/kota yang banyak memiliki nilai tertinggi yaitu pada indikator yang berhubungan dengan jumlah sekolah, tenaga kesehatan, kepadatan penduduk, dan PDRB. Khusus untuk Kota Banda Aceh merupakan kabupaten/kota yang memiliki pengeluaran per kapita paling tinggi di Sumatera. Diagram kotak garis untuk aspek kesehatan dan standar hidup layak dapat dilihat pada Lampiran 1. Pada diagram kotak garis, Kota Medan merupakan pencilan atas pada beberapa indikator yang berhubungan dengan pembangunan manusia. Pemeriksaan korelasi dapat

dilihat pada Lampiran 2. Terdapat lima pasang peubah yang memiliki korelasi diatas 0.8 atau -0.8. Koefisien korelasi antara indikator jumlah SD (X_1) dengan jumlah SMP (X_2) memiliki hubungan positif dengan nilai sebesar 0.867. Indikator jumlah SMP (X_2) dengan jumlah SMA (X_3) memiliki hubungan positif dengan nilai sebesar 0.882. Indikator proporsi guru terhadap murid SMP (X_5) dengan proporsi guru terhadap murid SMA (X_6) memiliki hubungan positif dengan nilai sebesar 0.857. Indikator jumlah dokter (X_7) dengan jumlah perawat (X_8) memiliki hubungan positif dengan nilai sebesar 0.896. Indikator jumlah gizi buruk tiap kelahiran bayi (X_{10}) dengan kepadatan penduduk (X_{12}).

B. Penentuan Jumlah Gerombol Optimal

Penggerombolan menggunakan metode pautan tunggal, pautan lengkap, pautan rataan, Ward, dan *centroid* menghasilkan gerombol yang berbeda-beda.Penggerombolan dilakukan menggunakan komponen utama, skor komponen tiap peubah terdapat pada Lampiran 3. Pemilihan gerombol yang memberikan hasil optimal menggunakan tiga metode pada penelitian ini yaitu nilai *Compactness Index*, rasio dari rata-rata simpangan baku dalam gerombol (Sw) dengan rata-rata simpangan baku antar gerombol (Sb), dan *Silhouette Coefficient*. Hasil nilai optimal untuk beberapa gerombol yang terbentuk disajikan dalam Tabel 3.

Nilai rasio yang kecil menunjukkan bahwa hasil gerombol yang terbentuk memiliki keragaman dalam gerombol yang rendah dan keragaman antar gerombol tinggi. Berdasarkan Tabel 3, hasil jumlah gerombol optimal untuk setiap metode berbeda. Hasil jumlah gerombol optimal pada metode pautan tunggal yang terbentuk yaitu enam gerombol dengan nilai rasio sebesar 0.097, metode pautan lengkap jumlah gerombol yang terbentuk sebanyak enam gerombol dengan nilai rasio sebesar 0.094, metode pautan rataan jumlah gerombol yang terbentuk sebanyak lima gerombol dengan nilai rasio sebesar 0.191, metode Ward jumlah gerombol yang terbentuk sebanyak enam gerombol dengan nilai rasio sebesar 1.946, dan terakhir metode *centroid* jumlah gerombol yang terbentuk sebanyak lima gerombol dengan nilai rasio sebesar 0.191.

Table 3 Kriteria Penentuan Jumlah Gerombol Optimal

Metode	Gerombol	Pautan Tunggal	Pautan Lengkap	Pautan Rataan	Ward	*Centroid*
Compactness Index	2	4.31	4.31	4.31	4.08	4.31
	3	4.22	3.94	4.22	3.72	4.22
	4	4.14	3.76	3.98	3.41	4.13
	5	4.06	3.65	3.92	3.25	4.04
	6	**3.98**	**3.30**	**3.82**	**3.10**	**3.94**
Rasio Sw/Sb	2	0.31	0.65	0.31	4.96	0.31
	3	0.29	0.30	0.39	2.28	0.39
	4	0.14	0.34	0.29	2.42	0.64
	5	0.11	0.40	**0.19**	2.30	**0.19**
	6	**0.10**	**0.09**	0.22	**1.95**	0.27
Silhouette Coefficient	2	0.42	0.42	0.42	0.81	0.42
	3	0.57	0.62	0.57	0.85	0.57
	4	0.58	0.78	0.62	0.87	0.74
	5	0.63	0.77	0.73	0.88	**0.74**
	6	**0.65**	**0.80**	**0.73**	**0.98**	0.72

Hasil jumlah gerombol optimal pada kriteria *Compactness Index* yaitu pada enam gerombol untuk setiap metode yang digunakan. Kesimpulannya bahwa untuk validasi gerombol menggunakan *Compactness Index* pada penelitian ini semakin tinggi jumlah gerombol maka nilai indeksnya semakin kecil atau semakin optimal jumlah gerombolnya. Hasil jumlah gerombol optimal kriteria *Silhoutte Coefficient* yaitu pada enam gerombol, kecuali pada metode *centroid* yaitu lima gerombol dengan nilai indeksnya itu 0.74. Semakin tinggi nilai indeksnya maka semakin optimal gerombol yang dihasilkan. Dari tiga metode yang digunakan untuk penentuan jumlah gerombol optimal dipilih minimal dua nilai yang terkecil untuk dilihat hasil perolehan gerombolnya. Berdasarkan hasil di setiap metode yang digunakan akan dilihat hasil dari enam gerombol, kecuali pada metode *centroid* akan dilihat hasil dari lima gerombol yang diperoleh.

C. Hasil Gerombol Optimal Berdasarkan Beberapa Metode Penggerombolan

Hasil jumlah gerombol optimal semuanya menghasilkan enam gerombol, kecuali pada metode *centroid* yang menghasilkan lima gerombol (Tabel 4). Sebagai ilustrasi, Tabel 5 memperlihatkan hasil penggerombolan kabupaten/kota di Sumatera tiap metode yang digunakan. Pada metode pautan tunggal gerombol satu, tiga, empat, lima, dan enam terbentuk masing-masing dari satu kabupaten/kota. Gerombol dua terbentuk dari 149 kabupaten/kota. Hal ini menggambarkan bahwa anggota gerombol selain dari

gerombol dua merupakan pencilan yang membentuk masing-masing gerombol tersendiri. Gerombol satu yang merupakan Kota Medan memiliki rata-rata jumlah sekolah yang tinggi di setiap jenjangnya, jumlah tenaga kesehatan, kepadatan penduduk, dan PDRB yang tinggi. Kota Batam yang merupakan anggota gerombol tiga menggambarkan jumlah gizi buruk yang tinggi dibanding gerombol lain serta pengeluaran per kapita yang besar dikarenakan Kota Batam merupakan salah satu kota yang menjadi pusat bisnis dan industri di wilayah Sumatera. Kota Pekanbaru dan Palembang masing-masing berada pada gerombol empat dan lima yang memiliki karakteristik yang baik di aspek kesehatan yaitu rendahnya gizi buruk dan jumlah fasilitas kesehatan yang cukup tinggi. Gerombol enam yang merupakan Kabupaten Bener Meriah memiliki rata-rata tinggi pada rasio jumlah guru terhadap murid di setiap jenjangnya yaitu SD, SMP, dan SMA.

Table 4 Hasil Gerombol Optimal Berdasarkan Metode Penggerombolan

Metode	Gerombol	Jumlah Anggota	Simpangan Baku Dalam Gerombol	Simpangan Baku Antar Gerombol
Pautan Tunggal	1	1	0.00	8807.06
	2	149	23631.81	
	3	1	0.00	
	4	1	0.00	
	5	1	0.00	
	6	1	0.00	
Pautan Lengkap	1	1	0.00	16305.91
	2	115	22391.41	
	3	8	48377.45	
	4	12	15012.43	
	5	15	8647.9	
	6	3	2030.24	
Pautan Rataan	1	1	0.00	19282.16
	2	142	22195.12	
	3	1	0.00	
	4	6	51980.27	
	5	1	0.00	
	6	3	2030.24	
Ward	1	7	68835.89	21483.77
	2	23	32113.83	
	3	16	13839.81	
	4	31	8217.59	
	5	62	13726.67	
	6	15	8647.9	
Centroid	1	1	0.00	10520.31
	2	150	26300.79	
	3	1	0.00	
	4	1	0.00	
	5	1	0.00	

Penggerombolan dengan metode pautan lengkap menjadikan Kota Medan membentuk gerombol sendiri yaitu pada gerombol satu dikarenakan banyaknya perbedaan dengan kabupaten/kota yang lain dan menyebabkan kemiripan dengan kabupaten/kota yang lain sangat kecil. Gerombol satu memiliki rataan yang tinggi di semua indikator kecuali rasio jumlah guru terhadap murid dan jumlah gizi buruk tiap kelahiran. Gerombol dua yang terdiri dari 115 kabupaten/kota memiliki rataan yang tinggi pada jumlah gizi buruk tiap kelahiran yang menandakan anggota gerombol ini memiliki aspek kesehatan yang rendah dibanding gerombol lain. Kabupaten Pidie Jaya, Kota Sabang, dan Kabupaten Kerinci merupakan kabupaten/kota dengan rasio jumlah guru terhadap murid tinggi yang menggambarkan selisih jumlah guru dengan muridnya cukup rendah. Pada metode pautan rataan Kota Medan, Kabupaten Deli Serdang, dan Kabupaten Bener Meriah membentuk gerombol sendiri yaitu pada gerombol satu, tiga, dan lima dikarenakan banyaknya perbedaan karakteristik dengan kabupaten/kota yang lain dan menyebabkan kemiripan dengan kabupaten/kota yang lain sangat kecil.

Table 5 Karakteristik Rataan dan Anggota Gerombol Metode Ward

Peubah	Gerombol					
	1	2	3	4	5	6
X_1	395.000	**484.000**	85.000	131.000	217.000	207.000
X_2	**166.000**	132.000	26.000	35.000	57.000	59.000
X_3	**97.000**	62.000	17.000	13.000	21.000	25.000
X_4	0.049	0.060	0.056	0.083	0.064	**0.099**
X_5	0.063	0.066	0.070	0.095	0.072	**0.109**
X_6	0.066	0.064	0.073	0.090	0.070	**0.097**
X_7	**1060.000**	171.000	193.000	54.000	80.000	95.000
X_8	**2450.000**	567.000	557.000	253.000	337.000	453.000
X_9	**792.000**	692.000	273.000	231.000	362.000	578.000
X_{10}	0.391	0.025	**0.549**	0.048	0.035	0.026
X_{11}	0.002	0.003	0.003	**0.004**	0.004	0.002
X_{12}	**3379.000**	231.000	2639.000	218.000	140.000	102.000
X_{13}	**97414.000**	44772.000	10413.000	6985.000	13810.000	9489.000
X_{14}	**14343.000**	9937.000	12962.000	9431.000	9725.000	8630.000

Gerombol dua memiliki karakteristik yang menonjol pada tingginya jumlah gizi buruk terhadap kelahiran bayi . Gerombol tiga yaitu Kabupaten Deli Serdang memiliki karakteristik jumlah sekolah SD yang tinggi, gerombol empat memiliki rataan yang tinggi namun masih dibawah gerombol satu jadi gerombol empat ini memiliki kesamaan karakteristik dengan gerombol satu dan juga merupakan kota besar di wilayah Sumatera. Gerombol lima yaitu Kabupaten Bener Meriah memiliki karakteristik rasio jumlah guru terhadap murid SMP dan SMA yang tinggi. Sedangkan, gerombol enam memiliki karakteristik rasio jumlah guru terhadap murid SD yang tinggi. Secara umum, metode Ward menyajikan jumlah gerombol yang lebih merata, yang memperlihatkan bahwa metode ini menghasilkan jumlah gerombol yang lebih baik dibandingkan metode yang lainnya. Berikut disajikan pada Tabel 5 karakteristik rataan dan anggota tiap gerombol pada metode Ward.

D. Perbandingan Hasil Gerombol Optimal

Perbandingan hasil gerombol optimal pada masing-masing metode untuk melihat kemiripan tiap gerombolnya menunjukkan suatu indikasi bahwa pada tiap meode yang digunakan ada kemiripan dari hasil penggerombolnya. Hal ini akan diuji menggunakan validasi gerombol *Rand Index*. Tabel 6 memperlihatkan matriks nilai indeksnya.

Table 6 Matriks Hasil Rand Index Berdasarkan Metode Linkage yang Digunakan

	Tunggal	Lengkap	Rataan	*Centroid*	Ward
Tunggal	1				
Lengkap	0.633	1			
Rataan	0.913	0.718	1		
Centroid	0.962	0.621	0.902	1	
Ward	0.301	0.638	0.369	0.286	1

Dari nilai matriks diatas terlihat bahwa hasil gerombol metode pautan tunggal dan pautan rataan memiliki tingkat kemiripan yang tinggi dibandingkan gerombol lain. Namun beberapa metode memiliki nilai indeks yang cukup tinggi yaitu antara metode pautan tunggal dengan pautan rataan, lalu ada metode pautan rataan dengan pautan centroid, dan juga metode pautan lengkap dengan pautan rataan. Dari empat perbandingan tersebut memiliki nilai indeks > 0.7.

E. Hasil Gerombol Terbaik

Secara umum, kelima metode telah menunjukkan bahwa Kota Medan membentuk gerombol tersendiri karena memiliki karakteristik yang tinggi hampir di setiap aspek pembangunan manusia sehingga kemiripan dengan kabupaten/kota lain sangat kecil. Sementara itu, hasil enam gerombol pada metode Ward (Tabel 7) memperlihatkan hasil yang cukup representatif, karena dapat memisahkan kabupaten/kota yang memiliki karakteristik tertentu dan juga dikuatkan dengan nilai *Silhouette Coefficient* dan *Compactness Index* yang kecil dibanding metode yang lain yaitu 3.102 dan 0.02.

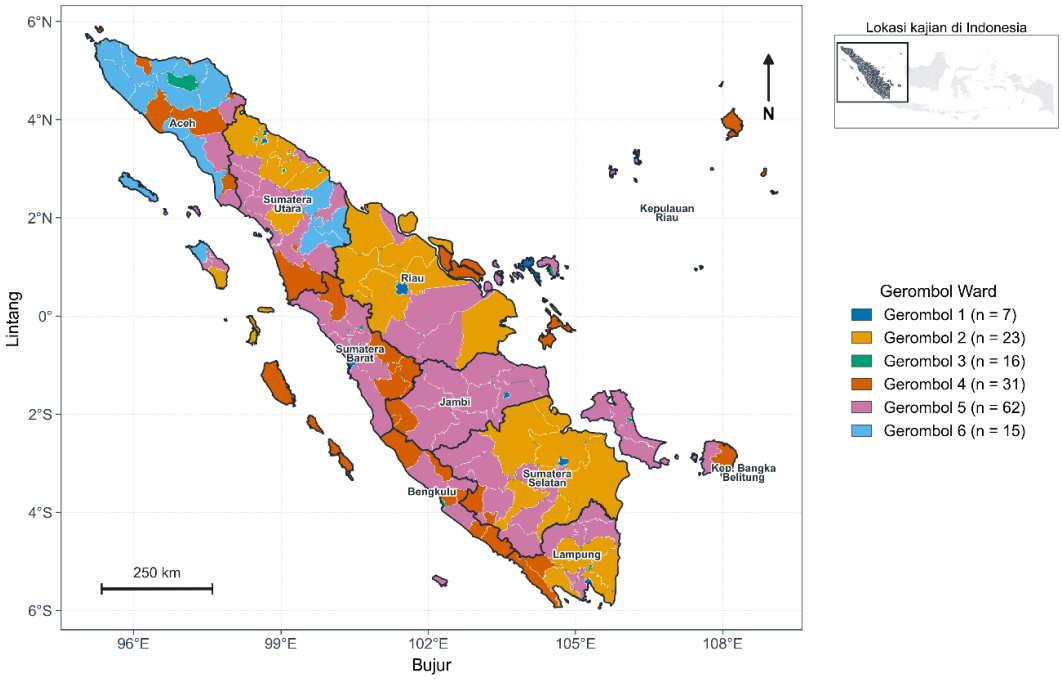
Table 7 Hasil Penggerombolan dengan Metode Ward

Gerombol	Jumlah Anggota	Anggota Gerombol
1	7	Medan, Jambi, Bandar Lampung, Padang, Batam, Pekanbaru, dan Palembang.
2	23	Deli Serdang, Siak, Musi Banyuasin, Lampung Tengah, dan 19 kabupaten/k lainnya.
3	16	Bengkulu, Payakumbuh, Sibolga, Binjai, Banda Aceh, Pangkal Pinang, Tanj Pinang, Metro, Kota Solok, Padang Panjang, Bukittinggi, Pariaman, Tanjungba Tebing Tinggi, Pematangsiantar, dan Bener Meriah.
4	31	Padangsidimpuan, Langsa, Lebong, Pasaman, dan 27 kabupaten/kota lainnya.
5	62	Pringsewu, Belitung, Bangka, Dumai, dan 58 kabupaten/kota lainnya.
6	15	Padang Lawas Utara, Labuhanbatu Utara, Labuhanbatu Selatan, Nias Ut Simeulue, Aceh Selatan, Aceh Timur, Aceh Tengah, Aceh Barat, Aceh Besar, Pi Bireuen, Aceh Utara, Aceh Barat Daya, dan Aceh Jaya.

Tabel 7 menunjukkan bahwa gerombol satu terbentuk dari tujuh kabupaten/kota, gerombol dua terbentuk dari 23 kabupaten/kota, gerombol tiga terbentuk dari enam belas kabupaten/kota, gerombol empat terbentuk dari 31 kabupaten/kota, gerombol lima terbentuk dari 62 kabupaten/kota, dan gerombol enam terbentuk dari lima belas kabupaten/kota. Pada metode Ward anggota tiap gerombolnya lebih beragam dan antar gerombolnya tidak jauh beda jumlahnya dibanding metode sebelumnya.

Terlihat pada gerombol satu anggotanya merupakan kota-kota besar di wilayah Sumatera yang memiliki karakterisitik rataan tingginya jumlah SMP (X_2) dan SMA (X_3), gerombol dua karakteristiknya yaitu tingginya jumlah SD (X_1), pada gerombol tiga secara umum anggota gerombolnya merupakan kota kecil, kecuali Kabupaten Bener Meriah yang menggambarkan jumlah fasilitas kesehatan cukup baik yang ditandai dengan tingginya rataan jumlah fasilitas kesehatan tiap km^2 (X_{10}). Gerombol empat memiliki karakteristik rataan yang tinggi pada jumlah gizi buruk tiap kelahiran bayi (X_{11}), lalu untuk gerombol lima tidak memiliki rataanya yang tinggi pada tiap peubah tetapi informasi yang bisa didapatkan memiliki karakteristik yang mirip dengan gerombol empat dan gerombol enam anggota gerombolnya berada di wilayah Provinsi Nanggroe Aceh Darussalam dan Sumatera Utara dengan karakteristiknya tinggi pada rasio jumlah guru terhadap murid SD (X_4), SMP (X_5), dan SMA (X_6). Terakhir, pada metode *centroid* memiliki kemiripan jumlah anggota gerombol dengan metode pautan tunggal tetapi memiliki perbedaan karakteristik di beberapa gerombol, pada indikator jumlah sekolah dasar (SD) yang tinggi memiliki gerombol tersendiri yaitu pada gerombol tiga. Gerombol satu yang merupakan Kota Medan memiliki karakteristik jumlah sekolah yang tinggi dijenjang SMP dan SMA, jumlah tenaga kesehatan yang tinggi, kepadatan penduduk yang tinggi, dan tingginya PDRB. Gerombol dua memiliki karakteristik jumlah gizi buruk tiap kelahiran bayi yang tinggi. Gerombol empat yang merupakan Kota Palembang memiliki karakteristik jumlah fasilitas kesehatan tiap km^2 yang tinggi dan pengeluaran per kapita yang tinggi juga. Gerombol lima yaitu Kabupaten Bener Meriah memiliki karakteristik rasio jumlah guru terhadap murid SD, SMP,dan SMA yang tinggi.

Gambar 1 menyajikan peta tematik dari gerombol yang terbentuk. Berdasarkan kategorisasi dan lokasi geografisnya pada Gambar 1, Gerombol 1 murni beranggotakan 7 kota besar utama di Pulau Sumatera (Medan, Jambi, Bandar Lampung, Padang, Batam, Pekanbaru, dan Palembang), Gerombol 2 didominasi oleh kabupaten berwilayah luas seperti Deli Serdang, Siak, Musi Banyuasin, dan Lampung Tengah, sedangkan Gerombol 3 hampir seluruhnya merupakan kategori kota kecil (seperti Banda Aceh, Bukittinggi, Pangkal Pinang, dan Kota Solok) dengan pengecualian Kabupaten Bener Meriah. Sementara itu, Gerombol 4 dan Gerombol 5 terdiri dari kombinasi kabupaten dan kota berkembang yang tersebar luas di berbagai provinsi di Sumatera, serta Gerombol 6 memiliki lokasi yang sangat spesifik karena seluruh anggotanya terkonsentrasi di wilayah utara Pulau Sumatera, yaitu di Provinsi Aceh dan Sumatera Utara



Gambar 1. Peta tematik hasil penggerombolan kabupaten/kota di pulau Sumatera

IV. KESIMPULAN

Penggerombolan berhierarki terpusat di Sumatera berdasarkan faktor yang berhubungan dengan pembangunan manusia menghasilkan enam gerombol optimal, kecuali pada metode centroid yang menghasilkan lima gerombol optimal. Secara umum, disetiap metode Kota Medan menjadi gerombol tersendiri. Namun, pada enam gerombol metode Ward kota besar di Sumatera membentuk gerombol tersendiri yang beranggotakan Kota Medan, Kota Jambi, Kota Pekanbaru, Kota Padang, Kota Batam, Kota Palembang, dan Kota Bandar Lampung. Berdasarkan hasil gerombol tiap metode, hasil enam gerombol metode Ward memberikan gerombol dengan karakteristik yang lebih baik dibanding metode lainnya. Hal ini dapat dilihat dari nilai *Silhouette Coefficient* yang tinggi dan *Compactness Index* yang terkecil. Perbandingan beberapa metode menggunakan Rand Index memperlihatkan bahwa hasil gerombol optimal pada metode pautan tunggal dengan pautan rataan, pautan tunggal dengan centroid, dan pautan rataan dengan centroid memiliki kemiripan hasil gerombol.

REFERENCES

[1] Sen, A. 2017. *Development as freedom revisited.* Journal of *Human* Development and Capabilities, 18(3), 1–15. https://doi.org/10.1080/19452829.2017.1342363

[2] UNDP United Nations Development Program. 1996. *Human Development Report 1996, Economic and Human Development*. New York (US): Oxford University Press.

[3] Ranieri, R., & Ramos, R. A. 2019. *Human Development and Inequality: A Multidimensional Approach.* World Development, 122, 1–13. https://doi.org/10.1016/j.worlddev.2019.05.019

[4] Badan Pusat Statistik, 2018. *Indeks Pembangunan Manusia*. Jakarta (ID): BPS- Indonesia.

[5] Permana, R., Sari, D., & Nugroho, A. 2021. *Regional Disparities in Human Development Index Across Indonesia: A Clustering Analysis.* Social Indicators Research, 156(2), 567–589. https://doi.org/10.1007/s11205-021-02645-7

[6] Hanniva, A., Kurnia, A., Rahardiantoro, S., & Mattjik, A. A. 2022. *Penggerombolan Kabupaten/Kota di Indonesia Berdasarkan Indikator Indeks Pembangunan Manusia Menggunakan Metode K-Means dan Fuzzy C-Means.* Jurnal Xplore, 11(1). DOI: 10.29244/xplore.v11i1.855

[7] Tapsirudin, T., & Sari, R. 2026. *Regional Human Development in Indonesia: Education, Economy, and Digitalization Perspectives.* Jurnal Ekonomi dan Pembangunan Indonesia, 26(2).

[8] Sihombing, J. K., & Wijaya, B. A. 2025. *Implementasi Algoritma Clustering dan Classification dalam Data Mining: Systematic Literature Review terhadap Tren dan Tantangan Terkini.* Jurnal Aplikasi Sains, Informasi, Elektronika dan Komputer, 4(3), 59–72. DOI: 10.55606/jupsim.v4i3.5240

[9] Sofwani, S. F., Fachrodhiya, M. R., Wati, M., & Widians, J. A. 2026. *Komparasi Algoritma K-Means dan DBSCAN dalam Klasterisasi Indeks Pembangunan Gender Tingkat Kabupaten/Kota di Indonesia.* Jurnal Aplikasi Sains, Informasi, Elektronika dan Komputer, 6(1), 33–48. DOI: 10.57152/ijirse.v6i1.2838

[10] Meilani, D., Masjkur, M., & Afendi, F. M. 2023. *Grouping Provinces in Indonesia Based on the Causes of Stunting Variables using Hierarchical Clustering Analysis. Indonesian Journal of Statistics and Its Applications, 7*(1), 32–43. https://doi.org/10.29244/ijsa.v7i1p32-43

[11] Purwanto, D., Pratama, R. A., Lein, R. B., Prastyo, I., & Al Haris, M. 2025. *Clustering of Districts in Central Java According to People's Welfare Indicators Using Ward's Method. VARIANCE: Journal of Statistics and Its Applications, 7*(1), 73–82. https://doi.org/10.30598/variancevol7iss1page73-82

[12] Lubis, S. K., & Tarigan, E. D. 2026. *Pengelompokan Provinsi di Indonesia Berdasarkan Prevalensi Balita Stunting Tahun 2024 Menggunakan Metode Agglomerative Hierarchical Clustering dengan Ward Linkage. USU Institutional Repository.*

[13] Melliana A, Zain I. 2013. Analisis Statistika Faktor *yang* Mempengaruhi Indeks Pembangunan Manusia di Kabupaten/Kota Provinsi Jawa Timur dengan Menggunakan Regresi Panel. *Jurnal Sains dan Seni Pomits*. 2 (2), D237–D242.

[14] Johnson RA, Wichern DW. 2007. *Applied Multivariate Statistical Analysis*. New Jersey (US): Prentice Hall.

[15] Mattjik AA, Sumertajaya IM. 2011. Sidik Peubah *Ganda* dengan Menggunakan SAS. Bogor (ID): Institut Pertanian Bogor.

[16] Everitt BS, Dunn G. 2001. *Applied Multivariate Data Analysis Second Edition*. London (UK): Wiley.