

STUDI KOMPARATIF METODE HIERARKI DAN NON-HIERARKI DALAM KLAUSTERISASI WILAYAH BERDASARKAN INDIKATOR SOSIAL EKONOMI DI SUMATERA UTARA 2024 UNTUK PEMBANGUNAN BERKELANJUTAN

Rindamia Silvana Fincantiery^{1*}, Kholidyah Tosa Pertiwi², Raafi Cahya Andrino³, Rieska Damai
Yanti⁴, Gamareza Arhyan Prasetya⁵, Muhammad Riefky⁶

^{1,2,3,4,5,6}Program Studi Demografi dan Pencatatan Sipil, Universitas Sebelas Maret, Surakarta,
Indonesia

*Penulis korespondensi: rindamia.fincantiery@student.uns.ac.id

ABSTRAK

Tingkat kesejahteraan sosial ekonomi di Sumatera Utara masih menunjukkan ketimpangan antar wilayah, di mana daerah perkotaan memiliki akses ekonomi dan pendidikan lebih baik dibandingkan pedesaan atau terpencil. Kemiskinan dan pengangguran masih terkonsentrasi di wilayah tertentu akibat keterbatasan infrastruktur serta kualitas sumber daya manusia yang rendah. Timpangnya distribusi pendapatan dan perbedaan kepadatan penduduk menyebabkan kesejahteraan masyarakat belum merata secara menyeluruh. Tujuan penelitian ini untuk mengelompokkan wilayah di Sumatera Utara tahun 2024 berdasarkan indikator sosial ekonomi menggunakan pendekatan komparatif antara metode hierarki (*Average Linkage*, *Complete Linkage*, *Single Linkage*) dan non-hierarki (*K-Means*). Penelitian ini memakai jenis kuantitatif dengan data diperoleh merupakan data sekunder berasal dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara tahun 2024. Variabel penelitian yang digunakan yaitu persentase kemiskinan, persentase Tingkat Pengangguran Terbuka (TPT), kepadatan penduduk, dan *Gini Ratio*. Hasil penelitian menunjukkan bahwa *Single Linkage* dan *Complete Linkage* sebagai metode hierarki merupakan metode yang paling optimum dengan nilai *Silhouette* 0,4821 dan jumlah kluster terbentuk sebanyak 2 kluster yaitu kluster 1 sebanyak 31 kabupaten/kota sebagai kluster dengan nilai indikator sosial ekonomi rendah, kluster 2 sebanyak 2 kabupaten/kota yaitu Kota Medan dan Kota Sibolga sebagai kluster dengan nilai indikator sosial ekonomi tinggi. Temuan ini diharapkan dapat memberikan gambaran bagi pemerintah Provinsi Sumatera Utara mengenai pola ketimpangan kesejahteraan antar wilayah serta menjadi dasar perumusan kebijakan pembangunan untuk mendukung pencapaian *Sustainable Development Goals* (SDGs), khususnya SDG 1 (Tanpa Kemiskinan), SDG 8 (Pekerjaan Layak dan Pertumbuhan Ekonomi), SDG 10 (Berkurangnya Kesenjangan), serta SDG 11 (Kota dan Permukiman yang Berkelanjutan).

Kata kunci: Hierarki, Non-Hierarki, SDGs, Sosial-Ekonomi, Sumatera Utara

1 PENDAHULUAN

Masalah kemiskinan merupakan fenomena yang siklikal dan persisten. Hal tersebut menunjukkan bahwa kemiskinan cenderung terjadi secara berulang pada setiap generasi dan berlangsung jangka panjang (Arifin, 2020). Kemiskinan juga bersifat dimensional di mana suatu kondisi disebut miskin bukan hanya karena keterbatasan pendapatan, tetapi juga ketidakmampuan dalam mengakses berbagai kebutuhan yang dapat menunjang kesejahteraan hidup (Wahyuni, 2024). Keterbatasan dalam mengakses berbagai kebutuhan tersebut juga dapat dipengaruhi oleh faktor demografi dan ketidakmerataan pembangunan infrastruktur. Kondisi tersebut menyebabkan adanya kesenjangan dalam akses layanan dan fasilitas, yang berimplikasi pada terjadinya ketimpangan kualitas hidup antarwilayah. Pemahaman yang mendalam mengenai kemiskinan diperlukan untuk mengetahui determinan utama yang

memengaruhi kondisi tersebut, sehingga dapat digunakan sebagai dasar perumusan kebijakan pembangunan kependudukan oleh pemerintah.

Kondisi ketimpangan cenderung terjadi pada wilayah dengan keberagaman karakteristik baik dari aspek geografis, sosial, maupun ekonomi. Salah satunya adalah Provinsi Sumatera Utara dengan penduduk terpadat keempat di Indonesia terdiri dari 33 kabupaten dan kota (Sinaga et al., 2022). Perbedaan tingkat pembangunan antar kabupaten/kota menyebabkan ketimpangan kesejahteraan masyarakat wilayah tersebut yang tercermin pada perbedaan tingkat kemiskinan di daerah kepulauan, seperti Kepulauan Nias yang cenderung berada pada kondisi relatif lebih tertinggal. Di sisi lain, Kota Medan yang merupakan pusat ekonomi dengan penduduk terpadat di Sumatera Utara justru memiliki tingkat kesejahteraan yang lebih rendah, seperti tingginya pengangguran dan ketimpangan pendapatan. Dalam menganalisis ketimpangan wilayah tersebut, digunakan pendekatan *clustering* untuk mengelompokkan wilayah berdasarkan karakteristik yang serupa (Kristanti et al., 2024). Penelitian ini membandingkan antara metode hierarki (*Average Linkage*, *Complete Linkage*, *Single Linkage*) dengan non-hierarki (*K-Means*) untuk menentukan metode klasifikasi yang paling efektif dalam pengelompokan wilayah di Sumatera Utara tahun 2024. Metode ini dipilih untuk mengklasifikasikan kabupaten dan kota di Sumatera Utara menurut indikator utama sosial ekonomi, yaitu persentase kemiskinan, kepadatan penduduk, Tingkat Pengangguran Terbuka (TPT), dan *gini ratio*.

Teori kesejahteraan sosial yang dikemukakan oleh John Rawls (1971) relevan dalam konteks kemiskinan dan pengangguran karena menekankan bahwa suatu kebijakan publik perlu disusun dengan mempertimbangkan keadilan distribusi sehingga dapat secara signifikan memberikan keuntungan bagi kelompok masyarakat yang paling rentan (Meytriana et al., 2025). Ragnar Nurkse dalam teori lingkaran setan kemiskinan (*Vicious Circle of Poverty*) (1953) menjelaskan bahwa terdapat beberapa faktor pemicu kemiskinan, seperti ketidakmerataan kepemilikan sumber daya sehingga distribusi pendapatan tidak seimbang, perbedaan kualitas sumber daya manusia, serta keterbatasan modal (Lindrianti, 2024). Beberapa penelitian terdahulu seperti pada penelitian Wahyuni dan Jatmiko (2019) membahas mengenai pengelompokan wilayah berdasarkan faktor-faktor kemiskinan menggunakan metode *clustering* hierarki *Average Linkage* dan menemukan adanya perbedaan karakteristik kemiskinan antar kelompok wilayah. Di sisi lain, penelitian Amalia et al. (2025) memanfaatkan metode *clustering* non-hierarki *K-Means* yang dikombinasikan dengan PCA untuk mengidentifikasi pola ketimpangan kesejahteraan dengan mengelompokkan kondisi sosial ekonomi wilayah di Indonesia. Ramadhan (2024) turut membahas terkait *clustering K-Means* di Jawa Timur, menghasilkan pengelompokan wilayah berdasarkan tingkat pembangunan manusia dan ketimpangan pendapatan. Arfindi dan Rahmawati (2025) membandingkan metode hierarki dan *K-means* serta mengevaluasi kualitas klaster berdasarkan beberapa indeks validitas, yang menunjukkan bahwa keberagaman pengelompokan bergantung pada metode yang digunakan. Penelitian lain oleh Ibrahim dan Marlina (2025) mengidentifikasi perbedaan hasil pengelompokan pada *K-Means* dan *Fuzzy C-Means* dengan indikator pengangguran menghasilkan variasi dalam struktur klaster. Sementara itu, Hanifah dan Primandari (2023) menggunakan *K-Means* dalam pengelompokan wilayah Nusa Tenggara Barat menurut indikator pendidikan dengan hasil pengelompokan wilayah sesuai variasi capaian pendidikan antardaerah.

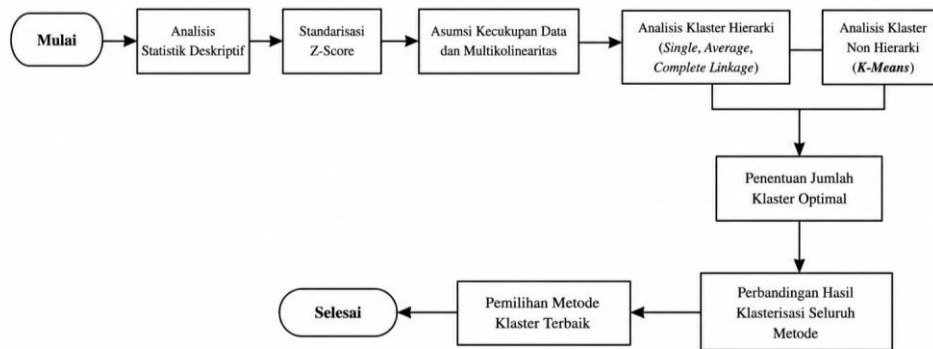
Berbeda dengan penelitian sebelumnya yang menggunakan sebagian indikator sosial ekonomi dan metode analisis tunggal, penelitian ini menggabungkan beberapa variabel sekaligus, yaitu persentase kemiskinan, kepadatan penduduk, tingkat pengangguran, serta ketimpangan pendapatan. Kebaruan lain terletak pada perbandingan penggunaan metode klusterisasi yang paling efektif dalam pengelompokan suatu wilayah. Penelitian ini bertujuan untuk mengelompokkan wilayah di Sumatera Utara menurut indikator sosial ekonomi

menggunakan komparatif antara kluster hierarki (*Average Linkage*, *Complete Linkage*, *Single Linkage*) dan non-hierarki (*K-Means*). Sehingga, penelitian ini diharapkan mampu memberikan informasi mengenai pengelompokan wilayah menurut kondisi nyata sosial ekonomi masyarakat sehingga perumusan kebijakan pembangunan dilakukan tepat sasaran dan mendukung pencapaian *Sustainable Development Goals* (SDGs), khususnya SDG 1 (Tanpa Kemiskinan), SDG 8 (Pekerjaan Layak dan Pertumbuhan Ekonomi), SDG 10 (Berkurangnya Kesenjangan), serta SDG 11 (Kota dan Permukiman yang Berkelanjutan) sebagai upaya peningkatan kesejahteraan masyarakat dan pengurangan ketimpangan antar wilayah.

2 METODE

2.1 Tahapan Pengolahan Data

Penelitian ini menggunakan pendekatan kuantitatif dengan data sekunder bersumber dari Badan Pusat Statistik Provinsi Sumatera Utara tahun 2024 berupa indikator kesejahteraan sosial-ekonomi pada 33 kabupaten/kota di Sumatera Utara, yang meliputi persentase kemiskinan, TPT, kepadatan penduduk, dan *gini ratio*. Analisis kluster merupakan salah satu teknik analisis multivariat bertujuan untuk mengelompokkan objek-objek ke dalam beberapa kelompok berdasarkan tingkat kemiripannya, sehingga objek dalam kelompok yang sama memiliki karakteristik yang relatif serupa dan berbeda dengan objek pada kelompok lainnya. Tahapan pengolahan data dapat disajikan pada Gambar 1.



Gambar 1. Diagram Alir Tahapan Pengolahan Data

2.1.1 Analisis Deskriptif

Analisis deskriptif merupakan teknik analisis yang digunakan untuk mendeskripsikan karakteristik data tanpa melakukan generalisasi. Analisis tersebut dilakukan dengan menyajikan ukuran pemusatan dan penyebaran data, yang meliputi nilai minimum, maksimum, rata-rata (*mean*), dan standar deviasi dari setiap variabel penelitian (Darmawan, 2023; Malakar, 2023).. Penelitian ini menggunakan 4 variabel yaitu persentase kemiskinan (X1), untuk mengukur persentase penduduk yang berada di bawah ambang batas kemiskinan, persentase pengangguran terbuka (X2) yang menggambarkan jumlah angkatan kerja yang belum mendapatkan pekerjaan, kepadatan penduduk (X3) yang menggambarkan rata-rata pertambahan jumlah penduduk pertahun dalam jangka waktu tertentu, dan *gini ratio* (X4) untuk menggambarkan tingkat ketimpangan distribusi pendapatan/pengeluaran suatu wilayah.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Satuan
X1	Persentase Kemiskinan	Persen (%)
X2	Persentase Pengangguran Terbuka	Persen (%)
X3	Kepadatan Penduduk	Jiwa/km ²
X4	<i>Gini Ratio</i>	-

(Sumber: BPS Sumatera Utara 2024)

2.2 Asumsi Persyaratan Analisis (*Pre-processing*)

2.2.1 Asumsi Multikolinearitas

Asumsi multikolinearitas dilakukan untuk mengidentifikasi suatu model regresi ditemukan adanya korelasi antar variabel independen (Ghozali, 2016). Asum multikolinearitas dapat dilihat dari nilai korelasi antar variabel bebas $> 0,8$, maka terjadi multikolinearitas. Namun, jika sebaliknya, maka tidak terjadi multikolinearitas.

2.2.2 Asumsi Kecukupan Data

Asumsi kecukupan data digunakan untuk mengevaluasi tingkat kecukupan dan representativitas sampel dalam penelitian. Pendekatan *Kaiser-Meyer-Olkin* (KMO) merupakan suatu indeks yang digunakan untuk mengukur kecukupan sampel keseluruhan berdasarkan hubungan antar indikator yang digunakan dalam penelitian. Asumsi sampel telah mewakili populasi atau bersifat representatif dinyatakan terpenuhi apabila nilai koefisien KMO berada pada rentang 0,5 sampai dengan 1 sebagaimana Tabel 2 berikut (Mahmudan, 2020).

Tabel 2. Indikator KMO

Nilai KMO	Asumsi
KMO $> 0,90$	Sangat Baik
0,80 – 0,89	Baik
0,70 – 0,79	Cukup
0,60 – 0,69	Sedang
0,50 – 0,59	Rendah
KMO $< 0,50$	Tidak Layak

Sumber: (Mahmudan, 2020)

Tabel 3. Kriteria Pengujian *Barlett's Test of Sphericity*

Hasil Barlett's Test	Asumsi
Nilai Signifikansi (Sig.) $< 0,05$	Data layak dianalisis
Nilai Signifikansi (Sig.) $> 0,05$	Data tidak layak dianalisis

Sumber: (Mahmudan, 2020)

2.3 Standarisasi (*Z-Scoring*)

Variabel yang digunakan memiliki satuan yang berbeda (persentase, jiwa/km², dan *ratio*), maka dilakukan proses standarisasi menggunakan *Z-score*. Tujuan standarisasi adalah untuk menyetarakan skala data agar tidak menimbulkan dominannya variabel tertentu dalam mengkalkulasikan jarak pada proses klasterisasi sebagaimana ditulis pada Persamaan (1).

$$Z = \frac{(X - \mu)}{\sigma} \quad (1)$$

Keterangan:

- X = Nilai asli variabel
- μ = Rata-rata variabel
- σ = Standar deviasi variabel

2.4 Klaster Hierarki

Klaster hierarki (*hierarchical clustering*) merupakan teknik pengelompokan yang mengelompokkan objek berdasarkan tingkat kemiripan atau jarak antarobjek secara bertahap sehingga membentuk struktur bertingkat (*hierarchy*) yang dapat divisualisasikan dalam bentuk dendrogram. Proses pengelompokan dilakukan secara berurutan melalui penggabungan atau pemisahan objek hingga terbentuk kelompok-kelompok yang memiliki karakteristik yang

relatif homogen di dalam kluster dan heterogen antar kluster (Vichi et al., 2022; Rompas et al., 2023). Metode ini umumnya digunakan ketika jumlah kluster terbentuk belum diketahui sebelumnya. Selain itu, metode hierarki cocok diterapkan pada data yang memiliki hubungan bertingkat atau tersusun dalam beberapa sub kelompok saling berkaitan, serta sering dipakai pada penelitian dengan jumlah sampel terbatas. Hasil pengelompokan pada metode ini disajikan dalam bentuk diagram pohon yang dikenal sebagai dendrogram.

2.4.1 Single Linkage

Single Linkage yaitu metode pengelompokan hierarki yang menggabungkan dua kluster menurut jarak terdekat antar anggota dari kedua kluster tersebut sebagaimana ditulis pada Persamaan (2) berikut.

$$D_{(AB),C} = \min(d_{AC}, d_{BC}) \quad (2)$$

Keterangan:

- $D_{(AB),C}$ = jarak kluster baru (AB) dengan kluster C
- d_{AC} = jarak antara anggota kluster A dan C
- d_{BC} = jarak antara anggota kluster B dan C

2.4.2 Average Linkage

Metode *Average Linkage* menggabungkan kluster berdasarkan rata-rata jarak seluruh anggota antar dua kluster sebagaimana ditulis pada Persamaan (3) berikut.

$$D_{(AB),C} = \frac{\sum_{i \in (AB)} \sum_{j \in C} d_{ij}}{n_{(AB)}n_c} \quad (3)$$

Keterangan:

- $D_{(AB),C}$ = rata-rata jarak antara kluster (AB) dan kluster C
- d_{ij} = jarak objek ke-i dengan objek ke-j
- $n_{(AB)}$ = jumlah anggota kluster (AB)
- n_c = jumlah anggota kluster C

2.4.3 Complete Linkage

Complete Linkage menggabungkan kluster berdasarkan jarak terjauh antar anggota dari kedua kluster sebagaimana ditulis pada Persamaan (4) berikut.

$$D_{(AB),C} = \max(d_{AC}, d_{BC}) \quad (4)$$

Keterangan:

- $D_{(AB),C}$ = jarak kluster baru (AB) dengan kluster C
- d_{AC} = jarak antara anggota kluster A dan C
- d_{BC} = jarak antara anggota kluster B dan C

2.5 Kluster Non Hierarki (K-Means)

Penelitian ini menggunakan metode *K-Means* Klastering untuk mengelompokkan 33 kabupaten/kota menjadi dua kluster berdasarkan kemiripan karakteristik sosial ekonomi. Pengelompokan dilakukan dengan menggunakan ukuran jarak *Euclidean* sebagaimana ditulis pada Persamaan (5) berikut.

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (5)$$

Keterangan:

- d_{ij} = Jarak *Euclidean*
- x_{ik} = Objek 1
- x_{jk} = Objek 2

3 HASIL DAN PEMBAHASAN

3.1 Karakteristik Data

Karakteristik data digunakan untuk menggambarkan kondisi suatu wilayah. Tabel 1 menggambarkan kondisi kemiskinan di Provinsi Sumatera Utara tahun 2024 ditinjau dari aspek persentase penduduk miskin, tingkat pengangguran terbuka, kepadatan penduduk, dan *gini ratio*.

Tabel 4. Hasil Statistik Deskriptif

Variabel	Mean	Minimum	Maximum	Std. Deviation
Persentase Kemiskinan	9,73	3,44	22,68	4,22
Tingkat Pengangguran Terbuka	4,36	0,89	8,13	2,37
Kepadatan Penduduk	1159,93	41,16	8902,16	2221,83
<i>Gini Ratio</i>	0,26	0,21	0,36	0,04

Sumber: (Rstudio, diolah)

Tabel 4 dijelaskan bahwa kondisi sosial ekonomi dan demografi di Provinsi Sumatera Utara menunjukkan tingkat variasi antarwilayah yang beragam. Kepadatan penduduk menjadi variabel dengan ketimpangan spasial paling ekstrem yang dibuktikan oleh rentang nilai sangat lebar antara daerah paling lengang (41,16 jiwa/km²) hingga kawasan terpadat (8.902,16 jiwa/km²), serta nilai standar deviasi yang tinggi mencapai 2.221,83 jiwa/km² yang menegaskan bahwa konsentrasi penduduk sangat menumpuk di pusat perkotaan dan tidak merata. Menurut aspek kesejahteraan, persentase kemiskinan (rata-rata 9,73%) juga menunjukkan kesenjangan yang cukup signifikan dengan standar deviasi 4,22% yang merentang dari angka 3,44% hingga 22,68%. Sementara itu, tingkat pengangguran terbuka yang berada pada rata-rata 4,36% memiliki sebaran data yang lebih terkendali dengan rentang 0,89% hingga 8,13% dan standar deviasi 2,37%. Di sisi lain, variabel *gini ratio* yang mengukur ketimpangan pendapatan justru mencatatkan kondisi yang paling homogen di seluruh wilayah pengamatan dengan rata-rata 0,26, rentang data di antara 0,21 hingga 0,36, serta standar deviasi sebesar 0,04. Hal tersebut menunjukkan bahwa tingkat ketimpangan pendapatan relatif seragam antarwilayah tanpa ada perbedaan ekstrem yang mencolok.

3.2 Asumsi Kecukupan Data Kaiser-Meyer-Olkin (KMO)

Kaiser-Meyer-Olkin (KMO) merupakan ukuran statistik digunakan untuk menilai kecukupan sampel dan kekuatan korelasi antar variabel dengan rentang nilai antara 0 hingga 1. Dalam klasterisasi, uji KMO digunakan sebagai langkah awal untuk memastikan variabel-variabel yang digunakan memiliki keterkaitan memadai sehingga menghasilkan klaster yang representatif. Nilai KMO di atas 0,50 dianggap layak, sedangkan nilai di atas 0,80 dikategorikan sangat baik.

Tabel 5. Hasil Asumsi KMO

<i>Kaiser-Meyer-Olkin factor Adequacy</i>	
<i>Overall MSA</i>	0,54

Sumber: (Rstudio, diolah)

Tabel 5 diperoleh nilai *Overall MSA* yaitu 0,54 lebih besar dari nilai minimum KMO yaitu 0,5 dapat diartikan data sampel berstatus cukup dan dapat dilanjutkan untuk analisis klaster.

3.3 Asumsi Multikolinearitas

Multikolinearitas merupakan kondisi korelasi yang sangat tinggi antar variabel yang digunakan dalam analisis. Pengujian asumsi tersebut penting dilakukan karena variabel yang berkorelasi terlalu tinggi akan memberikan redundansi informasi, sehingga bobot pengaruhnya

terhadap pembentukan kluster menjadi tidak proporsional dan hasil pengelompokan menjadi kurang valid. Nilai korelasi antarvariabel disyaratkan berada di bawah 0,80 sebagai batas toleransi yang dapat diterima.

Tabel 6. Korelasi *Pearson*

	Persentase Kemiskinan	Tingkat Pengangguran Terbuka	Kepadatan Penduduk	<i>Gini Ratio</i>
Persentase Kemiskinan	1	-0,349	-0,098	-0,222
Tingkat Pengangguran Terbuka	-0,349	1	0,478	0,317
Kepadatan Penduduk	-0,098	0,478	1	0,669
<i>Gini ratio</i>	-0,222	0,317	0,669	1

Sumber: (Rstudio, diolah)

Berdasarkan Tabel 6 dengan menggunakan ambang batas nilai $< 0,8$, hubungan antar variabel sosial ekonomi di Sumatera Utara pada tahun 2024 menunjukkan pola yang bervariasi. Persentase kemiskinan secara konsisten memiliki hubungan berbanding terbalik (negatif) dan lemah terhadap seluruh variabel lainnya, yakni tingkat pengangguran terbuka, kepadatan penduduk, dan *gini ratio*. Sebaliknya, hubungan berbanding lurus (positif) yang tergolong kuat hanya terjadi antara kepadatan penduduk dengan *gini ratio*. Sementara itu, hubungan berbanding lurus dengan kategori lemah ditemukan pada interaksi antara tingkat pengangguran terbuka dengan *gini ratio*, serta antara kepadatan penduduk dengan tingkat pengangguran terbuka.

3.4 Klusterisasi

Berikut merupakan hasil dari pendekatan klusterisasi hierarki dan non-hierarki, dengan menggunakan metrik pengukuran *Silhouette Score*. Metrik *Silhouette* ini berfungsi untuk menilai seberapa mirip suatu objek dengan objek lain di dalam kluster sendiri dibandingkan dengan objek di kluster lain. Semakin tinggi nilai evaluasinya (mendekati angka 1), maka semakin baik dan tepat pembagian kluster tersebut.

Tabel 7. Hasil *Silhouette Score*

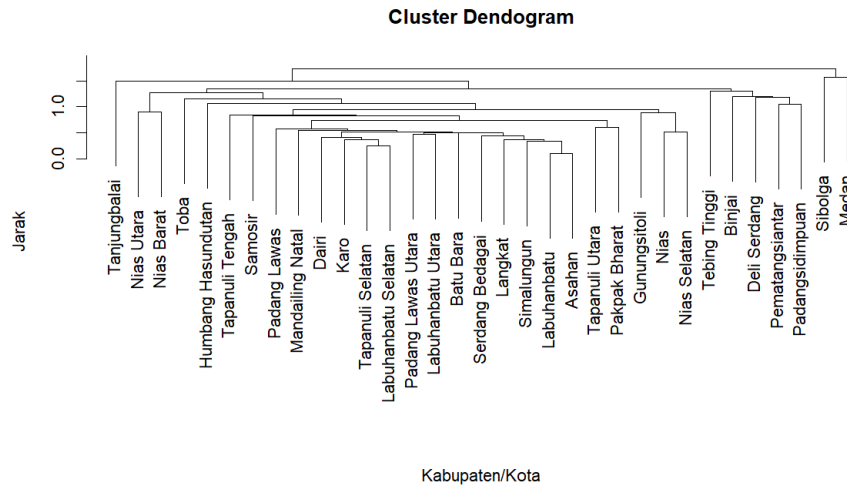
Metode	Cluster	Score	Metode	Cluster	Score	Metode	Cluster	Score	Metode	Cluster	Score
Single Linkage	2	0,4821	Average Linkage	2	0,4183	Complete Linkage	2	0,4821	K-Means	2	0,4058
	3	0,3900		3	0,3949		3	0,3767		3	0,3498
	4	0,0617		4	0,3778		4	0,3402		4	0,3663
	5	0,0588		5	0,3457		5	0,3014		5	0,3573

Sumber: (Rstudio, diolah)

Pada kelompok klusterisasi hierarki, terdapat tiga metode *linkage* (penggabungan) yang diuji. Pengujian pertama menggunakan metode *Single Linkage* dan menghasilkan skor *silhouette* sebesar 0,4821. Selanjutnya, pada pengujian kedua dengan metode *Average Linkage*, kualitas kluster terlihat sedikit menurun dengan perolehan skor sebesar 0,4183. Pengujian ketiga menggunakan metode *Complete Linkage* menghasilkan skor yang sama persis dengan *Single Linkage*, yaitu 0,4821. Hal tersebut dijelaskan bawah di dalam batasan pendekatan

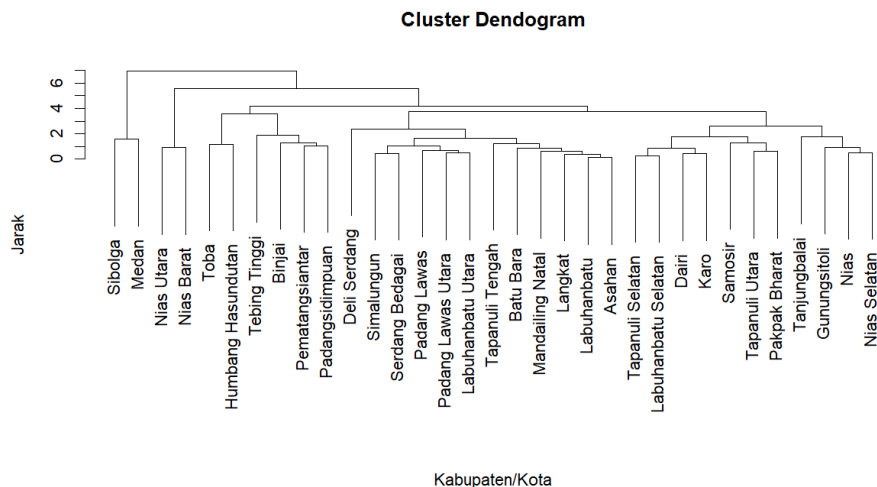
hierarki, metode *Single* dan *Complete Linkage* memberikan performa pemisahan data yang paling baik.

Sementara itu, kelompok klasterisasi non-hierarki yang dievaluasi dalam Tabel 7 menggunakan algoritma *K-Means*. Meskipun algoritma ini juga menyetujui bahwa dua klaster adalah jumlah pembagian yang ideal, skor *silhouette* yang dihasilkannya hanya sebesar 0,4058. Jika seluruh hasil dari kedua pendekatan tersebut disintesis, dapat disimpulkan bahwa metode *Single Linkage* dan *Complete Linkage* adalah algoritma yang paling optimal untuk mengelompokkan data pada kasus ini karena memiliki skor *silhouette* tertinggi. Sebaliknya, algoritma *K-Means* menunjukkan performa yang paling rendah karena menghasilkan skor evaluasi yang paling kecil dibandingkan semua metode klaster yang telah digunakan.



Gambar 2. Dendrogram *Single Linkage* (RStudio, diolah)

Visualisasi Gambar 2 menunjukkan hasil pengelompokan hierarki menggunakan metode *Single Linkage*. Metode pautan tunggal ini secara khas menghasilkan pola berantai (*chaining effect*) dengan rentang jarak antar klaster yang sempit. Terlihat bahwa sebagian besar wilayah memiliki tingkat kesamaan yang tinggi dan langsung mengelompok secara berurutan pada jarak terdekat. Sebaliknya, wilayah dengan karakteristik yang sangat berbeda, seperti Sibolga dan Medan, bergabung ke dalam kelompok pada tingkat jarak tertinggi di akhir proses pengelompokan.



Gambar 3. Dendrogram *Complete Linkage* (RStudio, diolah)

Visualisasi Gambar 3 menampilkan hasil pengelompokan pada data yang sama namun menggunakan metode *Complete Linkage*. Penggunaan pautan lengkap ini menghasilkan rentang skala jarak yang lebih besar dan membentuk struktur pohon hierarki yang jauh lebih seimbang tanpa adanya pola berantai. Visualisasi ini mempermudah pembagian data menjadi beberapa klaster yang lebih solid dan terdefinisi dengan baik, di mana kelompok wilayah tertentu, seperti Sibolga dan Medan, dapat membentuk sub-klasternya sendiri secara terpisah sebelum akhirnya bergabung dengan klaster wilayah lain pada jarak yang lebih tinggi.

Tabel 8. Anggota Setiap Klaster

Klaster	Jumlah Anggota	Kabupaten/Kota
1	31	Nias, Mandailing Natal, Tapanuli Selatan, Tapanuli Tengah, Tapanuli Utara, Toba, Labuhanbatu, Asahan, Simalungun, Dairi, Karo, Deli Serdang, Langkat, Nias Selatan, Humbang Hasundutan, Pakpak Bharat, Samosir, Serdang Bedagai, Batu Bara, Padang Lawas Utara, Padang Lawas, Labuhanbatu Selatan, Labuhanbatu Utara, Nias Utara, Nias Barat, Tanjungbalai, Pematangsiantar, Tebing Tinggi, Binjai, Padangsidempuan, Gunungsitoli
2	2	Sibolga dan Medan

Sumber: (Rstudio, diolah)



Gambar 4. Visualisasi *Clustering* (QGIS, diolah)

Berdasarkan pemetaan *clustering* pada Gambar 4, ditampilkan visualisasi dua klaster yang terbentuk di Sumatera Utara berdasarkan indikator sosial ekonomi. Klaster satu terdiri dari sebagian besar wilayah di Sumatera Utara yaitu sebanyak 31 kabupaten/kota, menunjukkan karakteristik wilayah dengan tingkat kesejahteraan rendah. Hal tersebut tercermin dari persentase kemiskinan yang relatif tinggi dan tingkat pengangguran terbuka (TPT) yang cenderung sedang hingga rendah. Kondisi ini mencerminkan dominasi wilayah non-perkotaan yang masih terfokus pada sektor primer sehingga terjadi keterbatasan aktivitas ekonomi dan kesempatan kerja di wilayah tersebut. Selain itu, kepadatan penduduk yang relatif rendah menunjukkan bahwa wilayah ini cenderung bersifat terpencar dan tidak terpusat pada kawasan

perkotaan. Dari sisi *gini ratio*, ketimpangan pendapatan dapat terjadi akibat distribusi ekonomi yang belum merata, meskipun skalanya tidak setinggi wilayah perkotaan. Karakteristik tersebut berkaitan dengan SDG 1 (Tanpa Kemiskinan), karena tingkat kemiskinan yang tinggi masih menjadi permasalahan utama, serta SDG 8 (Pekerjaan Layak dan Pertumbuhan Ekonomi) yang berfokus pada penciptaan lapangan kerja dan peningkatan produktivitas. Sementara itu, tingkat ketimpangan yang rendah menunjukkan bahwa SDG 10 (Berkurangnya Kesenjangan) belum menjadi permasalahan dominan, meskipun tetap membutuhkan perhatian dalam jangka panjang.

Anggota klaster dua terdiri atas Kota Sibolga dan Kota Medan menunjukkan karakteristik wilayah yang sangat kontras dengan klaster satu. Kepadatan penduduk yang sangat tinggi, tingkat pengangguran yang meningkat, serta ketimpangan pendapatan yang lebih besar dibanding wilayah lain, mencerminkan karakter wilayah perkotaan dengan konsentrasi aktivitas ekonomi dan arus urbanisasi yang tinggi. Meskipun aktivitas ekonomi lebih berkembang, tingginya persaingan pasar kerja dapat meningkatkan jumlah pengangguran, dan distribusi. Di sisi lain, distribusi pendapatan yang tidak merata turut memicu ketimpangan yang lebih besar. Kondisi ini berkaitan dengan SDG 8 (Pekerjaan Layak dan Pertumbuhan Ekonomi) dalam hal mengatasi pengangguran, SDG 10 (Berkurangnya Kesenjangan) berhubungan dengan kesenjangan pendapatan, dan SDG 11 (Kota dan Permukiman yang Berkelanjutan) yang menyoroti tantangan kepadatan penduduk dan perencanaan kota.

4 KESIMPULAN

Penelitian ini menyimpulkan bahwa kabupaten/kota di Sumatera Utara dapat dikelompokkan dalam dua klaster kesejahteraan berdasarkan persentase kemiskinan, tingkat pengangguran terbuka, kepadatan penduduk, dan *gini ratio*, dengan metode hierarki *Complete Linkage* dan *Single Linkage* sebagai pendekatan paling optimum dibandingkan *Average Linkage* maupun K-Means, ditunjukkan oleh nilai *Silhouette Score* sebesar 0,4821. Klaster 1 merupakan kelompok terbesar dengan 31 wilayah yang mencerminkan kondisi sosial ekonomi rendah dan terkonsentrasi di wilayah kepulauan yang relatif terisolasi, sementara Klaster 2 hanya terdiri dari 2 wilayah, yaitu Kota Medan dan Kota Sibolga, yang mewakili karakteristik sosial ekonomi tinggi namun tetap menghadapi tantangan ketimpangan pendapatan yang lebih besar. Temuan ini mengindikasikan bahwa kebijakan pembangunan berbasis klaster sangat diperlukan guna mengetahui pola ketimpangan antarwilayah dan meningkatkan kesejahteraan masyarakat di Sumatera Utara secara lebih tepat sasaran. Pemerintah Provinsi Sumatera Utara disarankan untuk fokus pada kebijakan pemerataan pembangunan, memprioritaskan peningkatan infrastruktur dasar dan konektivitas guna mengurangi keterisolasian wilayah kepulauan, serta menerapkan kebijakan pemerataan pendapatan di wilayah perkotaan agar pertumbuhan ekonomi tidak hanya dirasakan oleh sebagian kelompok masyarakat. Penelitian ini diharapkan dapat memberikan gambaran bagi pemerintah Provinsi Sumatera Utara terkait pola ketimpangan kesejahteraan antar wilayah serta menjadi dasar perumusan kebijakan pembangunan untuk mendukung pencapaian *Sustainable Development Goals* (SDGs), khususnya SDG 1 (Tanpa Kemiskinan), SDG 8 (Pekerjaan Layak dan Pertumbuhan Ekonomi), SDG 10 (Berkurangnya Kesenjangan), serta SDG 11 (Kota dan Permukiman yang Berkelanjutan).

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada seluruh anggota tim yang telah berjuang bersama, saling mendukung, dan berkontribusi dengan sepuh hati sehingga penelitian dapat terlaksana dan paper ini dapat tersusun dengan baik.

DAFTAR PUSTAKA

- Amalia, C. I., Fitria, & Sitorus, M. (2025). Analisa dan Pengelompokan Kota/Kabupaten di Indonesia Berdasarkan Aspek Sosial Ekonomi Menggunakan Algoritma K-Means. *Jurnal Ilmu Komputer, Sistem Informasi Dan Teknologi Informasi (Innotech)*, 2(1), 9–18. <https://ejournal.cyber-univ.ac.id/index.php/innotech/article/view/71>
- Arfindi, I. Si., & Rahmawati, F. (2025). Analisis Klaster Kesejahteraan Masyarakat Kabupaten / Kota Di Jawa Timur Menggunakan Metode Hierarki Dan K-means. *Ilmu Kesejahteraan Sosial*, 14(2), 44–51. <https://doi.org/10.15408/empati>
- Badan Pusat Statistik Provinsi Sumatera Utara. (2025). *Provinsi Sumatera Utara Dalam Angka 2025*. <https://sumut.bps.go.id/id/publication/2025/02/28/eb64d25aa56a75d579156d6a/provinsi-sumatera-utara-dalam-angka-2025.html>
- Badan Pusat Statistik Provinsi Sumatera Utara. (2026). *Gini Ratio Sumatera Utara Menurut Kabupaten/Kota, 2024*. <https://sumut.bps.go.id/id/statistics-table/2/NDY3IzI=/gini-ratio-sumatera-utara-menurut-kabupaten-kota.html>
- Darmawan, S. (2023). *Statistik deskriptif* (Cet. 1). Yogyakarta: Gava Media.
- Ghozali, I. (2016). *Aplikasi Analisis Multivariate SPSS 23* (8th ed.). Semarang : Badan Penerbit Universitas Diponegoro.
- Hanifah, S., & Primandari, A. H. (2023). Implementasi Metode K-Means Clustering dalam Pengelompokan Kabupaten/ Kota di Provinsi NTB Berdasarkan Indikator Pendidikan. *Emerging Statistics and Data Science Journal*, 1(3), 378–393. <https://doi.org/10.20885/esds.vol1.iss.3.art44>
- Ibrahim, M. Z. N., & Marlina, R. R. (2025). Perbandingan K-Means Clustering dan Fuzzy C-Means Clustering untuk Mengelompokkan Kabupaten/Kota Berdasarkan Faktor Tingkat Pengangguran di Jawa Barat Tahun 2023. *Bandung Conference Series: Statistics*, 5(2), 391–400. <https://doi.org/10.29313/bcss.v5i2.19756>
- Intyas, C. A., Susilo, E., & Indrayani, E. (2022). *Modal Sosial dan Kemiskinan Nelayan*. Malang : Universitas Brawijaya Press.
- Kristanti, B. T., Junaidi, A., & Mandyartha, E. P. (2024). Implementasi K-means Clustering Dalam Segmentasi Pelanggan Berdasarkan Usia, Pendapatan, Dan Model Rfm (Studi Kasus: Lantikya Store Jombang). *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3), 2099–2112. <https://doi.org/10.23960/jitet.v12i3.4677>
- Lindrianti, N. F. (2024). Pengaruh Pendapatan Asli Daerah (PAD) dan Alokasi Dana Dari Pusat Terhadap Tingkat Kemiskinan Di Provinsi di Yogyakarta Tahun 2015-2020. *Diponegoro Journal of Economics*, 11(1), 46–56. <https://doi.org/10.14710/djoe.32822>
- Mahmudan, A. (2020). Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering (Pengelompokan Kabupaten/Kota di Jawa Tengah Berdasarkan Kasus COVID-19 Menggunakan K-Means Clustering). *Jurnal Matematika, Statistika, Dan Komputasi*, 17(1), 1–13. <https://doi.org/10.20956/jmsk.v>
- Malakar, I. M. (2023). Conceptualizing central tendency and dispersion in applied statistics. *Cognition*, 5(1), 50–62. <https://doi.org/10.3126/cognition.v5i1.55408>
- Meytriana, A., Marsyalia Azahra, G., Muyasaroh Khasanah, K., Artika, P. A., & Hotman, H. (2025). Peran Ekonomi Politik Dalam Menanggulangi Kemiskinan dan Pengangguran Di Indonesia. *Journal Of Business, Economics, And Finance*, 3(2), 478–493. <https://doi.org/10.70437/benefit.v3i2.1259>
- Ramadhan, A. (2024). Pengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan IPM dan Rasio Gini Pada Tahun 2023 Menggunakan Clustering K-Means. *Jurnal Kecerdasan*

- Buatan, Komputasi Dan Teknologi Informasi*, 5(2), 174–183.
<https://doi.org/10.33650/coreai.v5i2.10593>
- Rompas, G., Paendong, M., & Salaki, D. T. (2023). Penerapan analisis cluster hierarki dalam pengelompokan kecamatan berdasarkan produksi buah-buahan di Kabupaten Minahasa Selatan. *D'Cartesian*, 12(2), 41–44. <https://doi.org/10.35799/dc.12.2.2023.51621>
- Sinaga, E., Lubis, T. A., Andy, E., Situmorang, O., Harahap, A. S., Perbendaharaan, D. J., & Keuangan, K. (2022). Dampak Program Perlindungan Sosial terhadap Kemiskinan di Sumatra Utara. *Jurnal Manajemen Perbendaharaan*, 3(2), 116–131. <https://doi.org/10.33105/jmp.v3i2.416>.
- Vichi, M., Cavicchia, C., & Groenen, P. J. F. (2022). Hierarchical means clustering. *Journal of Classification*, 39(3), 553–577. <https://doi.org/10.1007/s00357-022-09419-7>
- Wahyuni, S., & Jatmiko, Y. A. (2019). Pengelompokan Kabupaten/Kota di Pulau Jawa Berdasarkan Faktor-Faktor Kemiskinan dengan Pendekatan Average Linkage Hierarchical Clustering. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 10(1), 1–8. <https://doi.org/10.34123/jurnalasks.v10i1.197>
- Wahyuni, W. (2024). Kemiskinan Multidimensi di Provinsi Sulawesi Selatan Tahun 2020. *Jurnal Ekonomi Dan Pendidikan*, 20(1), 57–67. <https://doi.org/10.21831/jep.v20i1.43720>