



## PENERAPAN METODE *CLUSTERING AVERAGE LINKAGE* UNTUK MENGELOMPOKKAN KABUPATEN/KOTA DI PROVINSI SUMATERA UTARA BERDASARKAN INDIKATOR KEMISKINAN

Prapti Novitasari<sup>1\*</sup>, Irvana Arovah<sup>2</sup>  
Program Studi Matematika FMIPA Universitas Pamulang<sup>1,2,3</sup>  
praptinv01@gmail.com\*

**Abstrak**– Kemiskinan merupakan permasalahan sosial utama yang ada serta erat berkaitan dengan berbagai sektor. Oleh karena itu, pemerintah diberbagai dunia menerapkan berbagai strategi dalam mengatasi kemiskinan. Salah satu cara mengatasi kemiskinan yaitu dengan membuat kebijakan berdasarkan karakteristik di masing-masing daerah. Metode yang dapat digunakan adalah analisis *cluster average linkage*. Metode ini bertujuan untuk mengelompokkan kabupaten/kota di provinsi Sumatera Utara berdasarkan indikator kemiskinan tahun 2021. Indikator yang digunakan dalam penelitian ini adalah gizi buruk pada balita, jumlah tenaga kesehatan, penduduk yang menggunakan air tidak layak minum, penduduk pengguna PLN dan jumlah pengangguran. Hasil dari penelitian ini didapat 2 *cluster*, diantaranya *cluster 1* terdiri dari 25 kabupaten dan 7 kota sedangkan *cluster 2* terdiri dari 1 kota.

**Kata Kunci** – indikator kemiskinan, analisis *cluster*, *average linkage*,

### I. PENDAHULUAN

Kemiskinan merupakan permasalahan utama dalam berbagai masalah sosial yang ada serta berkaitan dengan berbagai sektor diantaranya kualitas pendidikan, kriminalitas dan sebagainya. Oleh karena itu, pemerintah di berbagai dunia menerapkan berbagai strategi yang berbeda dalam mengatasi angka kemiskinan. Namun sampai saat ini, upaya yang dilakukan pemerintah dengan berbagai macam program dan dukungan disetiap negara terhadap masyarakat miskin belum menunjukkan hasil yang optimal. Salah satu penyebab kegagalan dalam mengurangi

kemiskinan adalah kurang memperhatikan karakteristik secara khusus di masing-masing daerah. Pada kenyataan penyebab disetiap daerah berbeda-beda.

Salah satu penyedia data kemiskinan adalah Badan Pusat Statistik (BPS). Menurut data BPS Sumatera Utara, pada tahun 2020 jumlah penduduk miskin berjumlah 1.268.079 jiwa. Presentase penduduk miskin di Sumatera Utara pada maret 2021 sebesar 9,01 persen. pada Sempتمبر, presentase penduduk miskin mengalami penurunan menjadi 8,49 persen. hal ini membuktikan bahwa pemerintah Sumatera Utara berhasil menangani permasalahan kemiskinan di tengah pademi *covid-19*.

Menurut berita.dom yang ditulis oleh Fatimah Rahmawati per tanggal Januari 2022 dengan judul artikel “Sumatera Utara jadi daerah penurunan kemiskinan nomor 1 di Indonesia, ini komentar warganet”. [1] Merujuk pada sumber yang telah dijelaskan, hal ini membuktikan bahwa beberapa program yang pemerintah Sumatera Utara jalankan sesuai. Namun, banyak komentar yang tidak setuju dengan fakta bahwa Sumatera Utara berhasil menurunkan angka kemiskinan. Salah satu komentar dalam artikel tersebut mengatakan bahwa penilaian tersebut tidak sesuai dengan kondisi di lapangan, kenyatannya tingkat kriminalitas dan pengangguran masih tergolong tinggi. Sehingga dapat disimpulkan masih ada beberapa indikator kemiskinan yang perlu

diperhatikan oleh pemerintah Sumatera Utara mengenai kemiskinan. Menurut Perserikatan Bangsa-Bangsa (PBB) terdapat beberapa indikator yang digunakan sebagai tanda bahwa seseorang dikatakan miskin atau tidak, seperti kurang gizi pada balita, buta huruf, kesehatan yang buruk, pakaian dan perumahan.

Salah satu indikator dari dimensi kesehatan yaitu kurang gizi pada balita. Status gizi yang buruk pada balita akan menimbulkan pengaruh yang sangat menghambat pertumbuhan fisik, mental maupun kemampuan berpikir. Kemiskinan merupakan penyebab pokok gizi buruk pada balita. Semakin kecil pendapatan penduduk semakin tinggi presentase anak yang kekurangan gizi. Di Sumatera Utara prevalensi gizi buruk menyentok angka 25,8 persen. angka tersebut menjadikan provinsi Sumatera Utara sebagai provinsi ke-17 dengan jumlah gizi buruk pada balita terbanyak di Indonesia.

Menurut BPS terdapat dimensi lain yang berpengaruh dalam kemiskinan salah satunya yaitu ketenagakerjaan. Penduduk yang tidak bekerja juga dapat menyebabkan ketidakstabilan sosial sehingga dapat menyebabkan keterganggunya pertumbuhan dan pembangunan ekonomi suatu negara. Tingkat pengangguran terbuka di Sumatera Utara sebesar 6,33 persen pada Agustus 2021 atau sekitar 475 ribu orang. Jika dibandingkan dengan Februari 2021 sebesar 6,01 persen, tingkat pengangguran terbuka di Sumatera Utara mengalami peningkatan.

Berdasarkan beberapa fakta diatas, maka peneliti tertarik melakukan pengelompokan (*clustering*) untuk memahami karakteristik setiap wilayah di Sumatera Utara. Pengelompokan dilakukan agar pemerintah dapat menentukan daerah prioritas dengan karakteristik tertinggi.

Analisis *cluster* merupakan metode statistik yang mengidentifikasi kelompok sampel berdasarkan karakteristik yang relatif sama. [2] Terdapat dua metode cluster diantaranya yaitu non-hirarki dan hirarki *cluster*.

Pada proses pengelompokan metode hirarki *clustering* dilakukan secara bertahap atau terstruktur tetapi banyak *cluster* belum diketahui.[3] pada metode ini dimulai dengan mengelompokkan dua atau lebih data yang memiliki kesamaan yang sangat mirip. Kemudian proses dilanjutkan ke objek lain yang mempunyai kedekatan kedua. Begitu seterusnya sehingga *cluster* tersebut membentuk semacam 'pohon' dimana hirarki yang jelas antar objek, dari yang paling sama karakteristiknya sampai tidak.

Penelitian ini, metode yang digunakan adalah *Average Linkage*. metode *Average Linkage* merupakan metode pengelompokan yang jarak antara dua *cluster* didefinisikan sebagai rata-rata jarak seluruh pasangan yang berada dalam satu *cluster* dengan *cluster* lainnya.[4] Rumus hitung rata-rata jarak pada metode *Average Linkage*:

$$d_{(ij)k} = \frac{\sum_a \sum_b d_{ab}}{N_{ij} N_k}$$

Keterangan:

- $d_{(ij)k}$  = jarak antara *ij* dan cluster *k*
- $d_{ab}$  = jarak antara objek *a* pada cluster *ij* dan objek *b* pada cluster *k*
- $N_{ij}$  = jumlah anggota pada cluster *ij*
- $N_k$  = jumlah anggota pada cluster *k*

Pada penelitian ini dibantu oleh software SPSS 15 dan R *studio*. Berikut Langkah-langkah analisis data pengelompokan kabupaten/kota di Provinsi Sumatera Utara tahun 2021 yaitu:

1. Menentukan dan mengumpulkan variabel-variabel data.
2. Melakukan uji korelasi antar variabel dengan melihat nilai *Varians Inflation Factor(VIF)*.
3. Melakukan uji *Keiser Meyer Olkin (KMO)* untuk kelayakan data.
4. Standarisasi data
5. Menghitung jarak dengan cara mencari *Euclidean Distance*
6. Menghitung jumlah *cluster* terbaik dengan metode *Sillhouette*

7. Menentukan karakteristik dengan melihat nilai-rata-rata pada variabel data yang digunakan
8. Menarik kesimpulan

II. HASIL DAN PEMBAHASAN

1. Uji asumsi

Sebelum melakukan analisis *cluster* asumsi yang dipenuhi antara lain.

a. Kecukupan data

Tabel 1. KMO dan Barlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.527
Bartlett's Test of Sphericity	Approx. Chi-Square	56.567
	Df	10
	Sig.	.000

Tabel 1 menjelaskan bahwa nilai KMO pada data yang digunakan sebesar 0,527, artinya nilai tersebut diatas 0,5 yang berarti data tersebut cukup untuk mewakili populasi sehingga layak digunakan analisis *cluster*. [5]

b. Uji korelasi

Uji korelasi bertujuan untuk mengetahui ada atau tidak hubungan linear dari beberapa variabel. asumsi ini harus dipenuhi karena jika terjadi multikolinearitas pada perhitungan maka metode jarak yang digunakan yaitu metode jarak euclidean.

Tabel 2. Uji VIF

Korelasi	R	R <sup>2</sup>	VIF	Keputusan	Keterangan
X <sub>1</sub> , X <sub>2</sub>	0.0214	0.0005	1.0005	VIF < 10	Terima H <sub>0</sub>
X <sub>1</sub> , X <sub>3</sub>	-0.0285	0.0008	1.0008	VIF < 10	Terima H <sub>0</sub>
X <sub>1</sub> , X <sub>4</sub>	-0.1581	0.0250	1.0256	VIF < 10	Terima H <sub>0</sub>
X <sub>1</sub> , X <sub>5</sub>	-0.0492	0.0024	1.0024	VIF < 10	Terima H <sub>0</sub>
X <sub>2</sub> , X <sub>3</sub>	0.1862	0.0347	1.0359	VIF < 10	Terima H <sub>0</sub>
X <sub>2</sub> , X <sub>4</sub>	-0.2650	0.0702	1.0755	VIF < 10	Terima H <sub>0</sub>
X <sub>2</sub> , X <sub>5</sub>	0.8164	0.6665	2.9987	VIF < 10	Terima H <sub>0</sub>
X <sub>3</sub> , X <sub>4</sub>	-0.6721	0.4517	1.8238	VIF < 10	Terima H <sub>0</sub>
X <sub>3</sub> , X <sub>5</sub>	0.2407	0.0579	1.0615	VIF < 10	Terima H <sub>0</sub>
X <sub>4</sub> , X <sub>5</sub>	-0.3490	0.1218	1.1387	VIF < 10	Terima H <sub>0</sub>

Berdasarkan tabel 2 nilai *Variance Inflation Factors* (VIF) bernilai kurang dari 10, hal ini artinya setiap variabel yang digunakan tidak terjadi multikolinearitas.

c. Standarisasi data

Standarisasi data dapat dilakukan jika pada variabel yang diobservasi terdapat perbedaan ukuran yang cukup besar. Pada penelitian ini data yang digunakan memiliki perbedaan satuan, oleh sebab itu proses standarisasi data melalui transformasi pada data dilakukan. Transformasi dalam bentuk Z-score. Berikut hasil standarisasi:

Tabel 3. Hasil Z-score

ZX <sub>1</sub>	ZX <sub>2</sub>	ZX <sub>3</sub>	ZX <sub>4</sub>	ZX <sub>5</sub>
-0.2238	-0.2786	2.3918	-3.2094	-0.45675
0.2860	-0.2027	0.6551	-0.5715	-0.03282
-0.2848	-0.2662	1.0821	-0.4963	-0.32459
-0.2370	-0.2356	0.9872	0.0697	0.00636
-0.2583	-0.2390	-0.3659	0.6168	-0.44409
5.5060	-0.2050	-0.7655	-0.1334	-0.51382
-0.1919	-0.0543	-0.7188	0.6901	-0.05759
-0.0485	-0.1630	-0.8150	0.3179	0.27857
-0.1255	-0.0917	-1.0796	0.5811	0.19547
-0.2131	-0.2492	-0.5557	0.6225	-0.45188
-0.2131	-0.1415	-0.5243	0.3799	-0.34705
-0.2742	0.5554	-0.9754	0.7090	3.43212
-0.2795	-0.0599	-0.5965	0.3442	0.50165
-0.0326	-0.2843	1.1609	-2.8333	-0.31361
0.1347	-0.2798	-0.5591	0.5698	-0.46613
-0.2769	-0.2956	0.8616	0.3404	-0.53153
-0.2636	-0.2684	1.1990	0.3536	-0.52731
-0.2742	-0.0599	-0.9727	0.6789	-0.09551
0.1480	-0.2231	-0.9520	0.4006	-0.00842
-0.2530	-0.2684	0.4012	-0.7708	-0.36776
-0.0698	-0.2514	0.3838	-0.2255	-0.31502
-0.22906	-0.2333	-0.07192	-0.78957	-0.25316
-0.22906	-0.15284	-0.21158	0.02832	-0.17234
-0.26889	-0.31036	1.69818	-1.50592	-0.4697
-0.26889	-0.31829	0.80611	-1.50968	-0.53456
-0.28217	-0.23103	-0.58912	0.70895	-0.39785
-0.18924	-0.25483	-0.24165	0.70895	-0.33345
-0.26358	-0.01119	-1.08227	0.70895	0.02224
-0.27155	-0.08145	-1.05353	0.70895	-0.27337
0.17451	5.49495	-1.01678	0.69014	4.05298
-0.22906	0.11799	-1.08093	0.70895	-0.13985
-0.21313	-0.19024	1.96814	0.66006	-0.23249
-0.28482	-0.26729	0.63304	0.4476	-0.42871

d. Perhitungan jarak

Rumus perhitungan jarak euclidean:

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

Berikut ringkasan tabel matriks dari jarak Euclidean.

Tabel 4. Hasil Pengukuran Jarak Euclidean

	$D_1$	$D_2$	$D_3$	$D_4$	$D_5$	$D_6$	$D_7$	...	$D_{33}$
$D_1$	0,000	3,228	3,016	3,597	4,717	7,230	5,009	...	4,058
$D_2$	3,228	0,000	0,777	0,893	1,709	5,449	1,931	...	1,235
$D_3$	3,016	0,777	0,000	0,665	1,831	6,093	2,185	...	1,050
$D_4$	3,597	0,893	0,665	0,000	1,528	6,031	1,826	...	0,679
$D_5$	4,717	1,709	1,831	1,528	0,000	5,827	0,564	...	1,014
$D_6$	7,230	5,449	6,093	6,031	5,827	0,000	5,777	...	5,987
$D_7$	5,009	1,931	2,185	1,826	0,564	5,777	0,000	...	1,442
...	...	...	...	...	...	...	...	...	...
$D_{33}$	4,058	1,235	1,050	0,679	1,014	5,987	1,442	...	0,000

e. Proses pengelompokkan dengan metode average linkage

Proses selanjutnya adalah mencari data (objek) yang paling dekat pada matriks jarak *Euclidean*. Kemudian menggabungkannya menjadi sebuah kelompok atau *cluster* baru. Adapun rumus metode *Average Linkage*:

$$d_{(ij)k} = \frac{\sum_a \sum_b d_{ab}}{N_{ij} N_k}$$

Perhitungan ini dilakukan pada data yang paling dekat lalu dilakukan begitu seterusnya sampai menghasilkan matriks baru. Berikut matriks yang dihasilkan.

Tabel 5. Matriks baru

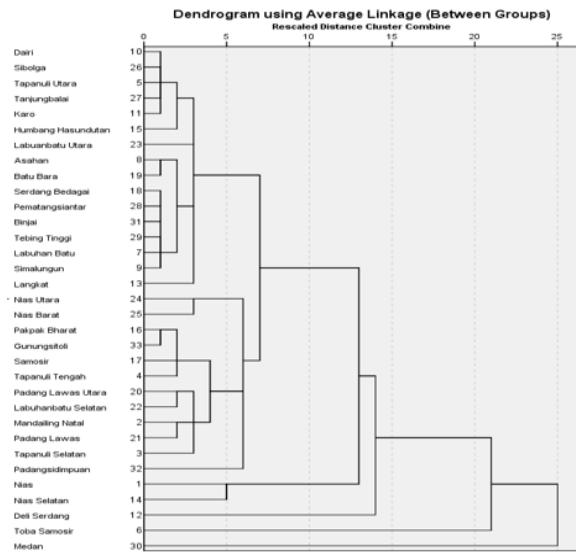
$d_{euclidean}$	1, 14, 2, 21, 20, 22, 3, 4, 16, 33, 17, 32, 24, 25, 5, 27, 10, 26, 11, 15, 23, 7, 18, 28, 31, 29, 9, 8, 19, 13, 6	30
1, 14, 2, 21, 20, 22, 3, 4, 16, 33, 17, 32, 24, 25, 5, 27, 10, 26, 11, 15, 23, 7, 18, 28, 31, 29, 9, 8, 19, 13, 12, 6	0	7,304
30	7,304	0,000

Proses pengelompokkan dapat disajikan dengan menggunakan dendrogram yang memungkinkan penelusuran objek data yang diamati lebih informatif.

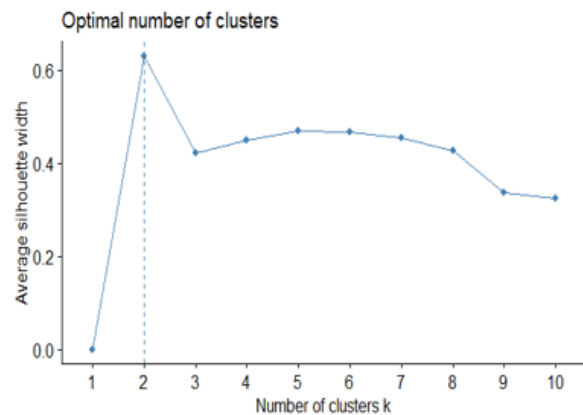
f. Menentukan jumlah anggota *cluster*

Hal utama dalam menentukan analisis cluster adalah menentukan berapa banyak cluster yang terbentuk. sebenarnya dalam menentukan jumlah cluster tidak terdapat aturan secara khusus.

Tabel 6. Dendrogram



Tetapi dapat dilakukan dengan beberapa metode diantaranya adalah *elbow method*, *silhouette method* dan *gap statistic*. [6] Pada penelitian ini menggunakan *silhouette method* dibantu pemograman *R-Studio*.



Gambar 1. Silhouette method

Berdasarkan gambar 1 menghasilkan 2 *cluster* yang ditunjukkan dengan nilai rata-rata tinggi. Dimulai dari *cluster* pertama yaitu Kabupaten Dairi, Kota Sibolga, Kabupaten Tapanuli Utara, Kota Tanjungbalai, Kabupaten Karo, Kabupaten Humbang Hasundutan, Kabupaten Labuanbatu Utara, Kabupaten Asahan, Kabupaten Batu Bara, Kabupaten Serdang Bedagai, Kota Pematang Siantar, Kota Binjai, Kota Tebing Tinggi, Kabupaten Labuhan Batu, Kabupaten Simalungun, Kabupaten Langkat, Kabupaten Nias Utara, Kabupaten Nias Barat, Kabupaten Pakpak Bharat, Kota

Gunungsitoli, Kabupaten Samosir, Kabupaten Tapanuli Tengah, Kabupaten Padang Lawa Utara, Kabupaten Labuhanbatu selatan, Kabupaten Mandailing Natal, Kabupaten Padang Lawas, Kabupaten Tapanuli Selatan, Kota Padangsidempuan, Kabupaten Nias, Kabupaten Nias Selatan, Kabupaten Deli Serdang, Kabupaten Toba samosir. Sedangkan, *cluster* kedua yaitu Kota medan.

g. Karakteristik setiap *cluster*

Variabel	Cluster 1	Cluster 2
Gizi buruk balita	106,219	174
Jumlah tenaga kesehatan	144,344	5145
Pengguna listrik PLN	96,115	99,9
Akses minum tidak layak	16,892	1,2
Pengangguran > 15 tahun	11065,344	121065

Tabel 7. profilisasi metode cluster Average Linkage

Berdasarkan tabel 7 dapat ditentukan karakteristik setiap *cluster* melalui rata-rata setiap *cluster*. Kemudian hasil interpretasi sebagai berikut:

1. Cluster 1 memiliki rata-rata tertinggi pada variabel penduduk yang menggunakan air minum tidak layak. Namun memiliki nilai rata-rata terendah pada variabel gizi buruk, jumlah tenaga kesehatan, penduduk pengguna energi listrik PLN dan jumlah pengangguran >15 tahun.
2. Cluster 2 memiliki rata-rata tertinggi pada variabel gizi buruk, jumlah tenaga kesehatan, penduduk pengguna energi listrik PLN dan jumlah pengangguran >15 tahun. Namun memiliki nilai rata-rata terendah pada variabel penduduk yang menggunakan air minum tidak layak.

### III. KESIMPULAN

Berdasarkan pada hasil yang sudah dijelaskan, maka disimpulkan bahwa pada metode *Average Linkage* dihasilkan 2 *cluster*. Setiap *cluster* memiliki jumlah yang berbeda. *cluster* pertama terdiri dari 25 kabupaten dan 7 kota, sedangkan *cluster* kedua terdiri dari 1 kota.

karakteristik dari masing-masing *cluster* yaitu: *cluster* 1 lebih banyak penduduk yang menggunakan air minum tidak layak, karena pada *cluster 1* memiliki nilai rata-rata ya relatif tinggi dibanding *cluster 2*. Sedangkan pada *cluster 1* paling sedikit kasus gizi buruk pada balita, paling sedikit jumlah tenaga kesehatan, paling sedikit penduduk pengguna energi listrik PLN dan paling sedikit jumlah pengangguran yang berumur >15 tahun, dikarenakan pada *cluster* ini memiliki nilai rata-rata relatif rendah jika dibandingkan dengan *cluster 2*. Pada *cluster 2* paling banyak kasus pada gizi buruk pada balita, paling banyak jumlah tenaga kesehatan, paling banyaknya penduduk yang menggunakan listrik PLN dan paling banyak jumlah pengangguran yang berumur >15 tahun, karena pada *cluster* ini memiliki nilai rata-rata tertinggi dibanding dengan *cluster 1*. Sedangkan pada *cluster* ini paling sedikit penduduk yang menggunakan air minum tidak layak, dikarenakan pada variabel penduduk yang menggunakan air minum tidak layak memiliki nilai rata-rata relatif rendah jika dibandingkan dengan *cluster 1*.

### REFERENSI

- [1] F. Rahmawati, "Sumut Jadi Daerah Penurunan Kemiskinan Nomor 1 di Indonesia, Ini Komentar Warganet." <https://www.merdeka.com/sumut/daerah-sumut-jadi-penurunan-kemiskinan-nomor-1-di-indonesia-ini-komentar-warganet.html> (accessed Oct. 08, 2022).
- [2] U. Putriana and Y. Setyawan, "Metode Cluster Analysis untuk Pengelompokan Kabuoaten/Kota di Provinsi Jawa Tengah berdasarkan Variabel Kemiskinan pada Tahun 2013," p. 15.
- [3] N. Afira and A. W. Wijayanto, "Analisis Cluster dengan Metode Partitioning dan Hierarki pada Data Informasi Kemiskinan Provinsi di Indonesia Tahun 2019," *Komputika J. Sist. Komput.*, vol. 10, no. 2, pp. 101–109, Sep. 2021, doi: 10.34010/komputika.v10i2.4317.
- [4] S. Ningsih, S. Wahyuningsih, and Y. N. Nasution, "Perbandingan Kinerja Metode Complete Linkage dan Average Linkage Dalam Menentukan Hasil

Analisis Cluster (Studi Kasus: Produksi Palawija Provinsi Kalimantan Timur 2014/2015),” vol. 1, no. 1, p. 5, 2016.

- [5] A. Mauludina and Setiawan, “Pengelompokan Kecamatan di Pulau Madura Berdasarkan Sektor Pertanian sebelum dan setelah Berdiri Jembatan Suramadu,” *J. SAINS DAN SENI ITS*, vol. 1, no. 1, Sep. 2012, doi: 10.12962/j23373520.v1i1.582.
- [6] F. A. Dewa and M. T. Jatipaningrum, “SEGMENTASI E-COMMERCE DENGAN CLUSTER K-MEANS DAN FUZZY C-MEANS”.