Pemetaan Profil Kemiskinan Global Menggunakan K-Means dan Average Linkage

Revina Simara
Sistem Informasi, Fakultas Teknologi Informasi, Universitas Merdeka Malang
e-mail: revinasimara283@gmail.com [1]

Kata Kunci:

K-Means
Average Linkage
Profil Kemiskinan Global
Clustering
PIP World Bank

ABSTRAK

Kemiskinan global memiliki karakteristik berbeda antarwilayah sehingga diperlukan pengelompokan berdasarkan beberapa indikator. Penelitian ini bertujuan membandingkan metode *K-Means* dan *Average Linkage* dalam memetakan karakteristik kemiskinan global. Data PIP World Bank periode 2005–2023 terdiri atas 1.200 observasi dengan lima indikator, yaitu *Headcount*, *Poverty Gap*, *Poverty Severity*, *Gini Ratio*, dan PPP. Analisis meliputi *preprocessing*, normalisasi *Min-Max*, pembentukan lima klaster, dan evaluasi menggunakan *R-square*. Hasil menunjukkan *K-Means* lebih unggul pada *Headcount* (0,842), *Poverty Gap* (0,867), *Poverty Severity* (0,822), dan *Gini Ratio* (0,651), sedangkan *Average Linkage* unggul pada PPP (0,608). Dengan demikian, *K-Means* lebih representatif dalam memetakan heterogenitas kemiskinan global.

Keyword:

K-Means
Average Linkage
Poverty Global Profile
Clustering
PIP World Bank

ABSTRACT

Global poverty exhibits regional variations, necessitating classification based on specific indicators. This study aims to compare the K-Means and Average Linkage methods in mapping global poverty characteristics. The dataset, sourced from the World Bank's PIP (2005–2023), comprises 1,200 observations across five indicators: Headcount, Poverty Gap, Poverty Severity, Gini Ratio, and PPP. The analysis involved data preprocessing, Min-Max normalization, the formation of five clusters, and evaluation using R-squared values. The results indicate that K-Means performed better for Headcount (0.842), Poverty Gap (0.867), Poverty Severity (0.822), and Gini Ratio (0.651), whereas Average Linkage performed better for PPP (0.608). Consequently, K-Means is more representative for mapping the heterogeneity of global poverty.

Copyright © SENASIF 2026
ISSN: 2598-0076

I. PENDAHULUAN

Kemiskinan adalah tidak adanya kesejahteraan, seseorang atau kelompok dipandang miskin karena tingkat kesadaran terhadap kesejahteraan di bawah rata-rata yang ditetapkan. Kemiskinan dikategorikan menjadi 2 yaitu kemiskinan berdasarkan penyebab dan kemiskinan secara konseptual. Kemiskinan berdasarkan penyebabnya dibagi menjadi dua yaitu, kemiskinan struktural yang terjadi karena ketidakberdayaan terhadap tatanan sosial sedangkan kemiskinan kultural dipengaruhi oleh faktor adat atau budaya yang membelenggu.

Fenomena ekonomi globalisasi sejak tahun 1970 telah menjadi perbincangan serius antara PBB dan World Bank diantaranya permasalahan tentang kemiskinan. Salah satu penyedia data kemiskinan global yaitu World Bank, data mengenai tingkat kemiskinan bersumber dari Web resmi World Bank yang mencakup wilayah *East Asia* & *Pasific*, *Europe* dan *Central Asia*, *Latin America* & *Caribbean*, *Middle East* & *North Africa*, *Other High Income Countries*, *South Asia*, *Sub-Saharan Africa*. Selain berkaitan dengan rendahnya pendapatan, kemiskinan juga mencakup keterbatasan akses terhadap berbagai layanan dasar, rendahnya kualitas hidup, serta minimnya kesempatan untuk meningkatkan kesejahteraan [1].

Hasil pengumpulan data kemiskinan global dari PIP World Bank menetapkan pada tahun 2024 sebanyak 692 ribu juta jiwa hidup berada di bawah garis kemiskinan $2,15 per kapita per hari atau setara dengan Rp. 36.118,57. Keadaan ekonomi dunia sempat mengalami guncangan kemiskinan ekstrem terparah yang terjadi pada tahun 1820 dengan jumlah mencapai 75% sementara itu di tahun 2018 kemiskinan ekstrem global menurun diperkirakan sebanyak 18% [2].

Pengukuran tingkat kemiskinan secara global menggunakan metode membandingkan tingkat harga barang dan jasa melalui konsep *Purchasing Power Parity*. Teori konsep ekonomi *Purchasing Power Parity* (PPP) menyatakan bahwa dalam jangka panjang nilai tukar mata uang mempengaruhi harga yang mengakibatkan berfluktuasinya indeks harga masing-masing negara. Menurut teori ini, jika negara A memiliki harga barang dan jasa yang lebih tinggi dibandingkan negara B, maka mata uang negara B harus terdepresiasi terhadap mata uang negara A untuk menyeimbangkannya [3].

Dalam pendekatan statistik, kemiskinan dibedakan menjadi kemiskinan relatif yang didasarkan pada distribusi pendapatan dalam suatu wilayah, dan kemiskinan absolut yang ditentukan berdasarkan standar minimum kebutuhan hidup [1]. Indeks kedalaman kemiskinan atau *The Poverty Depth* (P1) adalah nilai indeks yang menghitung ukuran jumlah kemiskinan melalui rata-rata kesenjangan dari banyaknya pengeluaran penduduk miskin terhadap garis kemiskinan yang ditentukan. Nilai indeks mengukur seberapa jauh individu yang hidup di dalam garis kemiskinan.

Copyright © SENASIF 2026
ISSN: 2598-0076

Indeks keparahan kemiskinan atau *The Poverty Severity Index* (P2) yaitu informasi mengenai distribusi pengeluaran diantara masyarakat miskin [4]. Keparahan kemiskinan menggambarkan seberapa jauh pengeluaran kelompok miskin terbesar di bawah garis kemiskinan. Kenaikan nilai indeks mempengaruhi nilai kesenjangan diantara kelompok miskin. Indeks gini rasio adalah sebuah metode yang digunakan sebagai alat ukur ketimpangan pendapatan atau kekayaan dalam suatu negara [5], [6].

Data yang digunakan bersifat sintetis dan dibentuk secara simulatif berdasarkan rentang dan pola yang menyerupai data riil dari PIP *World Bank*. Peneliti mendapatkan data kemiskinan global bersumber dari web Penelitian ini berpusat pada beberapa langkah, diantaranya:

1. Jumlah penduduk miskin: Total populasi yang hidup dibawah garis kemiskinan dengan ketetapan PPP $2, 15 per hari.
2. Kesenjangan kemiskinan: Selisih rata-rata antara garis kemiskinan dan jenis pengeluaran penduduk miskin, sebagai persentase dari garis kemiskinan.
3. Tingkat keparahan kemiskinan: Mengukur ketimpangan di antara penduduk miskin tersebut.
4. *Purchasing Power Parity* (PPP): Indikator daya beli yang mengoreksi pendapatan terhadap perbedaanharga antar negara.
5. Gini rasio: Indikator ketimpangan distribusi pendapatan dengan rentang nilai antara 0 (pemerataan sempurna) hingga 1 (ketimpangan sempurna).

Capaian tahun 2030 *Sustainable Development Goals* (SDGs) memiliki sasaran untuk menghapus kemiskinan ekstrem di dunia. Kemiskinan ekstem adalah kondisi dimana kesejahteraan penduduk berada di bawah garis kemiskinan ekstrem US $1,9 per-kapita per-hari yang dihitung berdasarkan PPP (*Purchasing Power Parity*). Kemisikinan ekstrem diukur dengan *absolute poverty measure* secara konsisten antar negara dan antar waktu [7].

Indeks kedalaman kemiskinan atau *The Poverty Depth* (P1) adalah nilai indeks yang menghitung ukuran jumlah kemiskinan melalui rata-rata kesenjangan dari banyaknya pengeluaran penduduk miskin terhadap garis kemiskinan yang ditentukan. Nilai indeks mengukur seberapa jauh individu yang hidup di dalam garis kemiskinan. Indeks keparahan kemiskinan atau *The Poverty Severity Index* (P2) yaitu informasi mengenai distribusi pengeluaran diantara masyarakat miskin terbesar di bawah garis kemiskinan. Kenaikan nilai indeks mempengaruhi nilai kesenjangan diantara kelompok miskin. Indeks gini rasio adalah sebuah metode yang digunakan sebagai alat ukur ketimpangan pendapatan atau kekayaan dalam suatu negara [5], [6].

Copyright © SENASIF 2026

ISSN: 2598-0076

Clustering merupakan salah satu teknik data mining yang dapat digunakan untuk menemukan kelompok objek yang memiliki karakteristik serupa. K-Means merupakan metode non-hierarki yang mempartisi data berdasarkan kedekatan terhadap centroid, sedangkan Average Linkage merupakan metode hierarki yang menggabungkan kelompok berdasarkan rata-rata jarak antarobjek Tujuan penelitian adalah menghasilkan profil kelompok kemiskinan global berdasarkan lima indikator, membandingkan struktur klaster yang dibentuk K-Means dan Average Linkage, serta menentukan metode yang lebih representatif berdasarkan nilai R-square.

II. METODE

Metode normalisasi Min-Max digunakan untuk mengubah skala data ke dalam rentang tertentu, seperti [0,1] atau [-1,1], dengan tujuan menghindari dominasi variabel yang memiliki nilai lebih besar terhadap variabel lainnya [9]. Terdapat beberapa metode yang umum digunakan dalam normalisasi data, antara lain Min-Max Normalization, Z-score Normalization, dan Decimal Scaling Normalization, yang masing-masing memiliki pendekatan berbeda dalam menyesuaikan skala data [10]. Formulasi matematis untuk normalisasi *Min-Max* yang digunakan adalah sebagai berikut:

$$X_t = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Keterangan: X_t = Nilai data baru hasil normalisasi; X = Nilai data asli (observasi saat ini); X_{min} = Nilai minimum dari keseluruhan observasi pada variabel tersebut; X_{max} = Nilai maksimum dari keseluruhan observasi pada variabel tersebut. Proses ini tidak mengubah informasi atau pola dasar dalam data, melainkan hanya menyesuaikan representasi nilainya sehingga analisis dapat dilakukan secara lebih optimal dan hasil model menjadi lebih akurat [8].

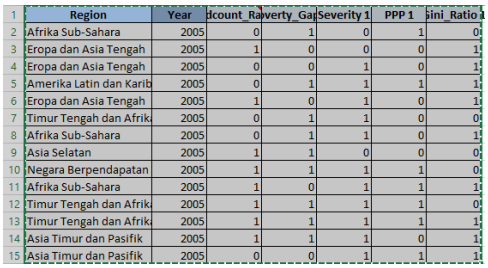
Seluruh variabel dinormalisasi ke rentang 0–1 menggunakan metode Min-Max Scaling untuk menyetarakan skala antar-variabel. Langkah ini diperlukan karena variabel PPP memiliki rentang yang jauh lebih besar dibanding variabel lainnya, sehingga dapat mendominasi perhitungan jarak jika tidak dinormalisasi.

III. HASIL DAN PEMBAHASAN

Penanganan Isblank

Hasil pemeriksaan data menunjukkan bahwa tidak terdapat nilai kosong (NA) pada seluruh variabel kemiskinan yang dianalisis. Hal ini mempercepat proses analisis karena tidak diperlukan teknik imputasi, eliminasi kasus, ataupun penyesuaian nilai.

Copyright © SENASIF 2026
ISSN: 2598-0076



	Region	Year	dcount_Ra	verty_Gap	Severity 1	PPP 1	ini_Ratio
2	Afrika Sub-Sahara	2005	0	1	0	1	0
3	Eropa dan Asia Tengah	2005	1	0	0	0	1
4	Eropa dan Asia Tengah	2005	0	0	1	0	1
5	Amerika Latin dan Karib	2005	0	1	1	1	1
6	Eropa dan Asia Tengah	2005	1	0	1	0	1
7	Timur Tengah dan Afrik	2005	0	1	1	0	0
8	Afrika Sub-Sahara	2005	0	1	1	0	1
9	Asia Selatan	2005	1	1	0	0	0
10	Negara Berpendapatan	2005	1	1	1	1	0
11	Afrika Sub-Sahara	2005	1	0	1	1	1
12	Timur Tengah dan Afrik	2005	1	1	1	1	0
13	Timur Tengah dan Afrik	2005	1	1	1	1	1
14	Asia Timur dan Pasifik	2005	1	1	1	0	1
15	Asia Timur dan Pasifik	2005	0	0	1	1	1

Gambar 1. Contoh Isblank

Pengecekan *Outlier*

Deteksi *outlier* dilakukan menggunakan pendekatan *interquartile range* (IQR). Berdasarkan hasil perhitungan pada variabel headcount, poverty gap, poverty severity, gini, dan PPP, ditemukan batas bawah (*lower bound*) dan batas atas (upper bound) untuk setiap variabel. Hasilnya menunjukkan bahwa terdapat 203 observasi yang masuk kategori outlier.

Tabel 1. Hasil Pengecekan Outlier

Parameter	*headcount*	*Poverty gap*	*Poverty severity*	*Gini*	ppp
Q1	0.010	0.004	0.002	0.328	1.283
Q3	0.306	0.107	0.054	0.428	172.492
IQR	0.296	0.103	0.053	0.100	171.209
Lower	-0.434	-0.150	-0.077	0.178	-255.531
Upper	0.750	0.262	0.133	0.578	429.306
Outlier	203				

Normalisasi Data

Tabel 1. Hasil Normalisasi data

Country, Year	*headcount*	*poverty_gap*	*poverty_severity*	Gini	PPP
Albania, 2005	0.038	0.016	0.008	0.215	0.13869
Albania, 2008	0.012	0.005	0.003	0.197	0.13869
Albania, 2012	0.026	0.015	0.010	0.167	0.13869
Albania, 2014	0.053	0.025	0.013	0.331	0.13869
Albania, 2015	0.016	0.005	0.001	0.277	0.13869
…	…	…	…	…	…
Zimbabwe, 2017	0.649	0.572	0.474	0.615	0.0007
Zimbabwe, 2019	0.715	0.737	0.708	0.787	0.0007

Untuk mengatasi permasalahan tersebut, penelitian ini menggunakan metode Normalisasi *Min-Max* (*Min-Max Scaling*). Metode ini secara linier mengubah rentang seluruh data asli menjadi skala baru yang seragam, yaitu berada pada interval 0 hingga 1 (0–1).

Hasil normalisasi menunjukkan bahwa nilai setiap variabel kini berada pada rentang 0 hingga 1, sehingga algoritma K-Means maupun Average Linkage dapat bekerja secara adil. Contohnya, negara Albania pada beberapa tahun memiliki nilai kemiskinan yang rendah (0.01–0.05), sedangkan Zimbabwe menunjukkan nilai kemiskinan tinggi (0.60–0.78), dan kontras tersebut tetap terjaga setelah proses normalisasi.

Copyright © SENASIF 2026
ISSN: 2598-0076

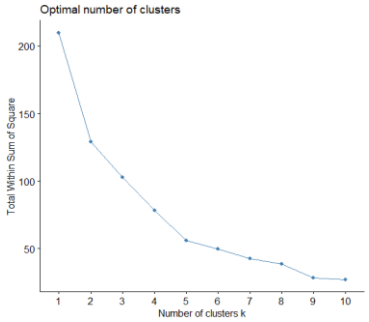
Uji Korelasi

Tabel 2. Hasil Uji Korelasi

Variabel		C1	C2	C3	C4	C5
Headcount	*r*	1	0.976	0.925	0.341	0.283
	Sig.		0.000	0.000	0.000	0.000
Poverty Gap	*r*	0.976	1	0.984	0.381	0.272
	Sig.	0.000		0.000	0.000	0.000
Poverty Severity	*r*	0.925	0.984	1	0.412	0.249
	Sig.	0.000	0.000		0.000	0.000
Gini	*r*	0.341	0.381	0.412	1	0.116
	Sig.	0.000	0.000	0.000		0.000
PPP	*r*	0.283	0.272	0.249	0.116	1
	Sig.	0.000	0.000	0.000	0.000	

***Clustering* K-Means**

Metode *clustering* ialah metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karateristik antara satu dengan data yang lainnya [11]. Penentuan jumlah klaster optimal langkah awal dalam metode K-Means adalah menentukan jumlah klaster yang paling representatif. Penggunaan metode *K-Means* akan membagi data untuk mengemlompokkan objek berdasarkan nilai (k) yang telah ditentukan [12]. Grafik WSS menunjukkan penurunan yang tajam dari k = 1 hingga k = 5, kemudian penurunan mulai melambat secara signifikan pada k = 6 dan seterusnya mendatar hingga k = 10. Pola ini mengindikasikan bahwa titik "elbow" berada pada k = 5, karena pada titik tersebut penurunan WSS tidak lagi memberikan pengurangan yang berarti terhadap total variasi data. Oleh karena itu, lima klaster dipilih sebagai jumlah klaster optimal dalam analisis K-Means.



Gambar 2. Grafik Hasil Penentuan Jumlah Cluster (Metode WSS)

Tabel 4. Hasil Initial Cluster Center K-means

Variabel	Cluster				
	1	2	3	4	5
Headcount	0.24	0.93	0.00	0.05	0.87
Poverty Gap	0.39	1.00	0.00	0.06	0.72
Poverty Severity	0.57	1.00	0.00	0.07	0.56
Gini	1.00	0.60	0.01	0.80	0.13
PPP	0.00	0.80	0.00	1.00	0.00

Copyright © SENASIF 2026
ISSN: 2598-0076

Setelah menetapkan *k*=5, algoritma melakukan inisialisasi pusat klaster awal (initial cluster centers) yang disajikan pada Tabel 4.4. Nilai-nilai pada tabel ini berada dalam skala 0-1. Sebagai contoh, pada inisialisasi awal, Klaster 2 memiliki titik pusat Poverty Gap sebesar 1.00, yang berarti titik awalnya berada tepat pada wilayah dengan kesenjangan kemiskinan tertinggi (maksimum) dalam dataset. Sebaliknya, Klaster 3 diinisialisasi pada titik-titik nilai minimum (0.00).

Tabel 5. Hasil Final *Cluster Center*

Variabel	*Cluster*				
	1	**2**	**3**	**4**	**5**
Headcount	0.07	0.68	0.02	0.11	0.32
Poverty Gap	0.05	0.63	0.01	0.09	0.29
Poverty Severity	0.05	0.58	0.02	0.08	0.27
Gini	0.57	0.48	0.23	0.66	0.63
PPP	0.03	0.38	0.05	0.90	0.04

Proses perbaikan posisi sentroid dilakukan secara iteratif untuk mencapai presisi pengelompokan. Hasil akhir dari proses iterasi adalah terbentuknya pusat klaster final yang disajikan pada Tabel 5. Interpretasi nilai pada tabel ini didasarkan pada posisi relatif terhadap rentang data (0 = terendah, 1 = tertinggi):

a. Klaster 1: Memiliki nilai sentroid Gini Ratio sebesar 0.57. Ini menunjukkan bahwa tingkat ketimpangan di klaster ini berada sedikit di atas nilai tengah (median) rentang data. Namun, indikator kemiskinan lainnya (Headcount, Gap, Severity) sangat rendah (mendekati 0.05 - 0.07), mengindikasikan kemiskinan absolut yang minim.
b. Klaster 2: Secara konsisten memiliki nilai sentroid yang tinggi pada indikator kemiskinan utama. Nilai Headcount (0.68), Gap (0.63), dan Severity (0.58) menunjukkan bahwa rata-rata wilayah di klaster ini memiliki tingkat kemiskinan yang mendekati nilai maksimum data (sekitar 60-70% dari rentang data terburuk).
c. Klaster 3: Menunjukkan nilai sentroid yang sangat mendekati 0.00 pada hampir seluruh variabel (misalnya Headcount 0.02 dan Gap 0.01). Hal ini mengindikasikan bahwa anggota klaster ini merupakan wilayah-wilayah dengan nilai kemiskinan dan ketimpangan yang berada pada batas minimum (terbaik) dalam populasi penelitian.
d. Klaster 4: Memiliki karakteristik unik dengan nilai Purchasing Power Parity (PPP) sebesar 0.90. Artinya, wilayah dalam klaster ini memiliki standar biaya hidup/daya beli yang mendekati nilai maksimum dalam dataset. Selain itu, klaster ini juga memiliki nilai Gini Ratio tertinggi (0.66) dibandingkan klaster lain, menandakan ketimpangan pendapatan yang lebar.
e. Klaster 5: Merupakan klaster dengan tingkat keparahan menengah ke atas. Nilai sentroid Headcount (0.32) dan Severity (0.27) berada di posisi sekitar sepertiga dari rentang data

Copyright © SENASIF 2026

ISSN: 2598-0076

maksimum, menempatkannya sebagai klaster dengan kondisi kemiskinan terburuk kedua setelah Klaster 2.

Tabel 6. Karakteristik *Cluster*

Variabel	*Cluster*					ANOVA	
	1	2	3	4	5	F	Sig.
Headcount	4.5%	46.7%	1.3%	7.7%	22.3%	1968.04	0.000
poverty_gap	1.4%	16.3%	0.4%	2.2%	7.4%	2402.10	0.000
poverty_severity	0.7%	7.7%	0.2%	1.0%	3.6%	1693.58	0.000
Gini	0.43	0.40	0.31	0.46	0.45	684.70	0.000
PPP	12.91	139.50	20.17	327.17	16.93	509.04	0.000

Validitas pengelompokan diuji menggunakan ANOVA untuk memastikan bahwa kelima klaster yang terbentuk benar-benar memiliki karakteristik yang berbeda secara signifikan. Berdasarkan Tabel 6, kolom Sig. menunjukkan nilai 0.000 untuk kelima variabel (Headcount, Poverty Gap, Poverty Severity, Gini, dan PPP). Nilai ini jauh di bawah taraf signifikansi $\alpha = 0.05$, yang membuktikan bahwa terdapat perbedaan karakteristik yang signifikan secara statistik antar klaster pada seluruh variabel pembentuk.Nilai F-hitung terbesar ditemukan pada variabel Poverty Gap (F = 2402.10) dan Headcount (F = 1968.04). Tingginya nilai F ini mengindikasikan bahwa kedua variabel tersebut merupakan pembeda (diskriminator) yang paling kuat dalam pembentukan struktur klaster kemiskinan ini. Berdasarkan rata-rata nilai riil pada Tabel 7, karakteristik spesifik dari masing-masing klaster dapat diinterpretasikan sebagai berikut:

a. Klaster 1 (Klaster Kemiskinan Rendah - PPP Rendah): Wilayah dalam klaster ini memiliki tingkat kemiskinan (headcount) yang relatif rendah, yaitu rata-rata 4.5%. Meskipun bukan yang terendah, angka ini jauh lebih baik dibandingkan Klaster 2 dan 5. Ciri khas klaster ini adalah nilai PPP terendah (12.91), yang mengindikasikan wilayah dengan biaya hidup atau standar harga yang relatif murah, dengan tingkat ketimpangan (Gini 0.43) yang moderat.
b. Klaster 2 (Klaster Kemiskinan Ekstrem): Ini merupakan klaster dengan kondisi terburuk (lagging regions). Rata-rata 46.7% penduduknya hidup di bawah garis kemiskinan, dengan kedalaman kemiskinan (gap) mencapai 16.3% dan tingkat keparahan (severity) 7.7%. Wilayah ini memerlukan intervensi kebijakan yang paling intensif dan mendesak.
c. Klaster 3 (Klaster Sejahtera): Klaster ini merepresentasikan wilayah dengan performa terbaik. Tingkat kemiskinan sangat minim (1.3%), dengan kedalaman dan keparahan mendekati nol (0.4% dan 0.2%). Selain itu, klaster ini memiliki rasio Gini terendah (0.31), mencerminkan distribusi pendapatan yang paling merata dibandingkan klaster lainnya.
d. Klaster 4 (Klaster Disparitas & Biaya Hidup Tinggi): Klaster ini memiliki anomali yang unik. Tingkat kemiskinannya tergolong rendah (7.7%), namun memiliki nilai PPP yang sangat ekstrem yaitu 327.17 (tertinggi). Hal ini disertai dengan rasio Gini tertinggi (0.46). Profil ini umumnya mencerminkan wilayah perkotaan maju atau pusat ekonomi di mana biaya hidup sangat tinggi dan kesenjangan antara si kaya dan si miskin sangat lebar.

Copyright © SENASIF 2026
ISSN: 2598-0076

e. Klaster 5 (Klaster Kemiskinan Tinggi): Klaster ini berada pada "lapis kedua" kemiskinan. Dengan headcount 22.3%, wilayah ini memiliki hampir seperempat penduduk miskin. Tingkat keparahan kemiskinan (3.6%) dan kesenjangan (7.4%) juga signifikan, menempatkannya sebagai prioritas penanganan kedua setelah Klaster 2.

Berdasarkan karakteristik di atas, dilakukan pelabelan (labeling) untuk setiap klaster guna memudahkan identifikasi kebijakan. Distribusi anggota klaster disajikan dalam tabel berikut:

Tabel 7. Jumlah Anggota Cluster dan Penamaan *Cluster*

Cluster	**Penamaan**	**Jumlah Anggota**
1	Klaster kemiskinan rendah	353
2	Klaster Kemiskinan Ekstrem	58
3	Klaster Sejahtera	917
4	Klaster Disparitas Tinggi	32
5	Klaster Kemiskinan Tinggi	115

Analisis distribusi anggota menunjukkan pola yang menarik. Mayoritas populasi penelitian (917 anggota) tergabung dalam Klaster 3 (Sejahtera). Hal ini merupakan indikasi positif bahwa sebagian besar wilayah observasi telah memiliki tingkat kemiskinan yang rendah dan pemerataan yang baik. Sebaliknya, Klaster 2 (Kemiskinan Ekstrem) dan Klaster 4 (Disparitas Tinggi) memiliki jumlah anggota yang paling sedikit (masing-masing 58 dan 32).

Average Linkage Clustering

Menjelaskan bahwa *clustering* merupakan teknik pengelompokan data yang dilakukan berdasarkan kesamaan karakteristik antar objek, sehingga data dalam satu kelompok memiliki tingkat kemiripan yang tinggi [13]. Berbeda dengan metode *K-Means* yang mempartisi data secara langsung ke dalam jumlah kelompok yang ditentukan di awal, analisis *Average Linkage* menggunakan pendekatan hierarki aglomeratif (agglomerative hierarchical clustering). Pada metode ini melakukan proses secara pencarian pasangan dengan melihat rata-rata jarak setiap observasinya [14].

Metode ini dipilih untuk memberikan gambaran alami mengenai struktur data kemiskinan tanpa adanya bias penentuan jumlah klaster di awal, serta untuk membandingkan konsistensi pengelompokan dengan hasil metode partisi sebelumnya. Proses ini terus diulang hingga posisi *centroid* stabil, sehingga menghasilkan pembentukan cluster yang optimal beserta pusat cluster akhirnya [15]. Langkah awal dalam analisis hierarki adalah memastikan validitas data yang diproses. Tingkat validitas data mencapai 100% tanpa adanya *missing values*, yang mengindikasikan bahwa seluruh dataset siap untuk dihitung jarak Euclidean kuadratnya

Copyright © SENASIF 2026

ISSN: 2598-0076

Tabel 8. *Agglomeration Schedule*

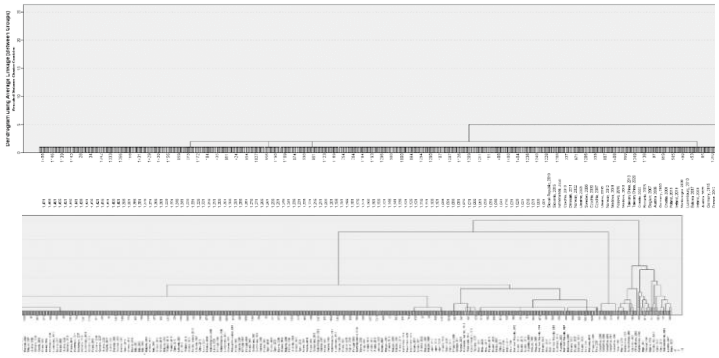
Stage	*Cluster Combined*		*Coefficients*	*Stage Cluster First Appears*		**Next Stage**
	Cluster 1	*Cluster 2*		*Cluster 1*	*Cluster 2*	
1	1460	1461	0.000	0	0	187
2	1247	1248	0.000	0	0	90
3	1232	1233	0.000	0	0	91
4	974	976	0.000	0	0	679
5	541	547	0.000	0	0	44
…	…	…	…	…	…	…
1473	1	270	1.081	1468	1455	1474
1474	1	14	1.230	1473	1472	0

Tabel Agglomeration *Schedule* merekam proses bertahap algoritma dalam menggabungkan 1.475 wilayah observasi, dimulai dari pasangan yang paling identik hingga menyatu menjadi satu kelompok besar, sehingga interpretasi dapat difokuskan pada perbedaan antara fase homogenitas awal dan fase lonjakan heterogenitas pada tahap akhir. Pada fase awal (Stage 1–5), kolom Coefficients menunjukkan nilai konsisten 0.000, menandakan bahwa wilayah-wilayah yang digabungkan memiliki karakteristik kemiskinan yang benar-benar identik sehingga tidak terdapat jarak antar kasus, mencerminkan homogenitas tinggi pada level mikro.

Sebaliknya, pada fase akhir (Stage 1473–1474), lonjakan koefisien menjadi indikator utama untuk menentukan jumlah klaster optimal berdasarkan prinsip stopping rule, di mana rumus penentukannya adalah $k=(N-\text{Stage})+1$ dengan $N=1.475$. Pada Stage 1474, algoritma menggabungkan Cluster 1 dan Cluster 14 menjadi satu klaster tunggal dengan koefisien tertinggi 1.230, sementara pada Stage 1473 algoritma menggabungkan Cluster 1 dan Cluster 270 dan membentuk tiga klaster dengan koefisien 1.081. Lonjakan ekstrem terjadi pada transisi dari Stage 1472 ke 1473, di mana koefisien meningkat tajam dari 0.617 menjadi 1.081 (selisih > 0.4), jauh lebih besar dibandingkan kenaikan dari Stage 1473 ke 1474 yang hanya sekitar 0.15.

Hal ini menunjukkan bahwa penggabungan pada Stage 1473 sudah bersifat "paksa" karena menyatukan kelompok yang berbeda jauh, sehingga proses aglomerasi idealnya dihentikan sebelum titik lonjakan tersebut, yaitu pada Stage 1472. Dengan demikian, jumlah klaster yang paling optimal berdasarkan Agglomeration Schedule adalah empat klaster, dihitung dari $1.475-1.472+1=4$, karena pada tahap ini struktur homogenitas masih terjaga sebelum varians meningkat drastis pada tahap penggabungan berikutnya.

Copyright © SENASIF 2026
ISSN: 2598-0076



Gambar 3. *Hasil Dendogram*

Interpretasi Visual Dendrogram Konfirmasi atas jumlah klaster ini diperkuat oleh visualisasi dendrogram (Gambar 3). Grafik pohon ini menunjukkan struktur percabangan yang merepresentasikan jarak antar-klaster. Jika dilakukan pemotongan garis *(cut-off)* secara horizontal pada skala jarak (rescaled distance cluster combine) di kisaran 5 hingga 10, garis potong tersebut akan menabrak garis vertikal utama yang membentuk 5 cabang klaster yang terpisah.

Kelima cabang ini memiliki panjang garis vertikal yang cukup signifikan sebelum menyatu kembali di level yang lebih tinggi, yang menandakan bahwa kelima kelompok tersebut (yang terdiri dari satu klaster besar dan beberapa klaster kecil/outlier) memiliki karakteristik unik yang saling terpisah satu sama lain.Dengan demikian, baik berdasarkan perhitungan koefisien aglomerasi maupun visualisasi dendrogram, ditetapkan bahwa 4 atau 5 klaster merupakan jumlah pengelompokan yang paling representatif untuk memetakan pola kemiskinan di wilayah penelitian.

Tabel 9. Perbandingan *Cluster*

Variabel	***4 Cluster***			***5 Cluster***		
	ANOVA		***R Square***	***ANOVA***		***R Square***
	F	**Sig.**		**F**	**Sig.**	
headcount	653.34	0.000	0.640	846.34	0.000	0.633
poverty_gap	854.73	0.000	0.699	1076.88	0.000	0.687
poverty_severity	787.44	0.000	0.682	985.84	0.000	0.668
gini	29.86	0.000	0.075	39.62	0.000	0.075
PPP	768.18	0.000	0.676	758.96	0.000	0.608

Berdasarkan hasil perbandingan statistik pada Tabel 9, solusi 5 klaster menunjukkan kinerja yang lebih superior dalam membedakan karakteristik kemiskinan dibandingkan solusi 4 klaster. Hal ini terlihat dari peningkatan nilai F-hitung pada ANOVA, yang menunjukkan daya beda klaster terhadap varians dalam kelompok. Pada solusi 5 klaster, nilai F untuk indikator utama kemiskinan meningkat signifikan, seperti Headcount Index yang naik dari 653.34 menjadi 846.34, Poverty Gap dari 854.73 menjadi 1076.88, dan Poverty Severity dari 787.44 menjadi 985.84.

Peningkatan ini menegaskan bahwa pembagian menjadi 5 klaster lebih mampu menangkap nuansa perbedaan tingkat kedalaman dan keparahan kemiskinan secara lebih sensitif. Meskipun seluruh variabel pada kedua skenario sama-sama menunjukkan nilai signifikansi 0.000 (< 0.05),

Copyright © SENASIF 2026

ISSN: 2598-0076

tingginya nilai F pada solusi 5 klaster memberikan dasar yang lebih kuat untuk menetapkan 5 klaster sebagai model akhir dalam analisis *Average Linkage.*

Tabel 3. Jumlah Anggota dan Karakteristik *Cluster*

Variabel	*Cluster*					ANOVA	
	1	2	3	4	5	F	Sig.
headcount	4.3%	32.2%	53.8%	43.6%	3.5%	1968.04	0.000
poverty_gap	1.2%	9.5%	20.3%	15.6%	1.2%	2402.10	0.000
poverty_severity	0.6%	3.9%	10.1%	7.6%	0.6%	1693.58	0.000
gini	0.35	0.39	0.41	0.42	0.47	684.70	0.000
PPP	17.72	201.09	213.33	17.36	355.54	509.04	0.000

Berdasarkan hasil validasi di atas, profil lengkap dari kelima klaster yang terbentuk disajikan dalam tabel karakteristik (Tabel 10). Nilai rata-rata pada tabel ini menggambarkan kondisi riil dari setiap kelompok, yang memungkinkan dilakukannya pelabelan (profiling) sebagai berikut:

Analisis Profil Klaster:

a. Cluster 1 (Wilayah Sejahtera & Merata - Low Poverty, Low Inequality): Klaster ini merepresentasikan wilayah dengan tingkat kemiskinan rendah (4.3%) dan biaya hidup yang sangat terjangkau (PPP 17.72). Keunggulan utama klaster ini adalah tingkat ketimpangan yang paling rendah (Gini 0.35) dibandingkan kelompok lain, mencerminkan pemerataan ekonomi yang baik meskipun dalam standar harga rendah.
b. Cluster 2 (Wilayah Rentan dengan Biaya Menengah): Wilayah ini memiliki tingkat kemiskinan yang cukup tinggi (32.2%), dengan standar biaya hidup yang berada di level menengah (PPP 201.09). Kedalaman kemiskinan (9.5%) menunjukkan bahwa penduduk miskin di wilayah ini berada cukup jauh di bawah garis kemiskinan, memerlukan intervensi ekonomi yang substansial.
c. Cluster 3 (Wilayah Kemiskinan Ekstrem - The Poorest): Ini adalah klaster dengan kondisi terburuk. Tingkat kemiskinan mencapai angka tertinggi yaitu 53.8%, yang berarti lebih dari separuh penduduknya miskin. Indikator keparahan kemiskinan (severity) juga tertinggi (10.1%), menandakan ketimpangan yang parah di antara penduduk miskin itu sendiri.
d. Cluster 4 (Wilayah Kemiskinan Tinggi - Biaya Rendah): Klaster ini memiliki kemiripan dengan Cluster 3 dalam hal tingginya angka kemiskinan (43.6%), namun memiliki karakteristik biaya hidup/PPP yang sangat rendah (17.36), setara dengan Cluster 1. Ini mengindikasikan wilayah tertinggal yang mungkin terisolasi secara ekonomi, di mana harga-harga murah namun daya beli masyarakat tetap sangat minim.
e. Cluster 5 (Wilayah Maju/Perkotaan Ketimpangan Tinggi - High Inequality, High Cost): Klaster ini memiliki tingkat kemiskinan terendah secara absolut (3.5%), bahkan lebih rendah dari Cluster 1. Namun, klaster ini memiliki standar biaya hidup (PPP) tertinggi (355.54) dan rasio Gini tertinggi (0.47). Profil ini sangat khas untuk wilayah perkotaan maju atau pusat

Copyright © SENASIF 2026
ISSN: 2598-0076

pertumbuhan ekonomi, di mana kemiskinan minim tetapi kesenjangan sangat lebar serta biaya hidup mahal.

Perbandingan Metode K-Means vs Average Linkage

Evaluasi kinerja metode pengelompokan dilakukan dengan membandingkan nilai R Square dari masing-masing teknik, karena indikator ini menunjukkan proporsi varians data yang mampu dijelaskan oleh model klaster. Semakin mendekati angka 1, semakin baik kemampuan metode tersebut dalam membentuk kelompok yang homogen di dalam klaster dan heterogen antar klaster.

Tabel 4. Hasil perbandingan metode *K-Means* vs *Average Linkage*

Variabel	R Square	
	K-means	**Average Linkage**
headcount	0.842	0.633
poverty_gap	0.867	0.687
poverty_severity	0.822	0.668
Gini	0.651	0.075
PPP	0.581	0.608

Secara khusus, K-Means menunjukkan performa yang jauh lebih unggul pada indikator utama kemiskinan. Variabel Poverty Gap mencatat nilai R^2 tertinggi sebesar 0.867, jauh melebihi hasil Average Linkage yang hanya mencapai 0.687. Tren ini juga tampak pada Headcount Index (0.842 vs 0.633) dan Poverty Severity (0.822 vs 0.668), mengindikasikan bahwa K-Means lebih efektif dalam memisahkan wilayah berdasarkan intensitas kemiskinan. Perbedaan terbesar muncul pada Gini Ratio, di mana K-Means mencapai nilai R^2 sebesar 0.651, sementara Average Linkage hanya memperoleh 0.075, menunjukkan bahwa metode hierarki tersebut kurang mampu menangkap variasi ketimpangan pendapatan karena struktur pengelompokan yang sensitif terhadap nilai ekstrem.

Meski demikian, Average Linkage hanya unggul tipis pada variabel Purchasing Power Parity (PPP) dengan nilai R^2 sebesar 0.608, sedikit lebih tinggi dari K-Means yang sebesar 0.581. Keunggulan kecil ini tidak mengimbangi kelemahan Average Linkage pada variabel lain yang lebih penting dalam analisis kemiskinan. Secara keseluruhan, K-Means menjadi metode yang lebih tepat digunakan dalam penelitian ini karena mampu menghasilkan klaster yang lebih representatif dan konsisten, terutama pada tiga indikator utama kemiskinan dengan nilai R^2 di atas 0.8. Dengan demikian, interpretasi hasil dan rekomendasi kebijakan lebih akurat bila didasarkan pada struktur klaster yang dihasilkan oleh metode K-Means.

IV. SIMPULAN

Untuk penelitian selanjutnya, beberapa pengembangan dapat dilakukan. Pertama, disarankan untuk mengintegrasikan metode validasi klaster tambahan seperti Silhouette Coefficient,

Copyright © SENASIF 2026
ISSN: 2598-0076

Dunn Index, atau Gap Statistic untuk memperkuat penentuan jumlah klaster optimal. Kedua, penggunaan metode klaster alternatif seperti K-Medoids, DBSCAN, atau Hierarchical Clustering dapat dibandingkan dengan K-Means untuk melihat apakah ada algoritma yang lebih sesuai dengan karakteristik dataset kemiskinan global yang kompleks. Ketiga, indikator kemiskinan dapat diperluas ke arah Multidimensional Poverty Index (MPI) agar analisis mencerminkan kondisi riil secara lebih komprehensif. Keempat, pemodelan dapat ditingkatkan dengan melakukan preprocessing lanjutan seperti standardisasi robust, deteksi outlier, serta eksplorasi teknik PCA untuk mereduksi dimensi. Terakhir, implementasi metode berbasis machine learning ini sebaiknya dikombinasikan dengan analisis kualitatif sehingga hasil klaster dapat lebih mudah dipahami dan diinterpretasikan oleh pembuat kebijakan.

DAFTAR RUJUKAN

[1] J. P. R. C. Ranasinghe, A. C. I. D. Karunarathna, dan H. M. J. . Herath, "After Corona (COVID-19) impacts on global poverty and recovery of tourism based service economies: An appraisal," *Int. J. Tour. Hosp.*, vol. 1, no. 1, hal. 52–64, 2021, doi: 10.51483/IJTH.1.1.2021.52-64.

[2] M. Moatsos, "Global poverty: A Review of Measurement, Levels, and Trends in a Historical Perspective," *J. Econ. Surv.*, hal. 706–766, 2024, doi: 10.1111/joes.12644.

[3] R. Satya dan M. A. Wulan, "PENGARUH INFLASI DAN SUKU BUNGA TERHADAP NILAI PERUSAHAAN (Bukti Empiris dari Sektor Manufaktur)," no. October, 2024.

[4] R. M. Wahid dan S. N. Sarfiah, "Analysis of Factors Affecting the Poverty Gap Index in Aceh Province 2017-2019," *Res. Horiz.*, vol. 1, no. 6, hal. 217–228, 2021, doi: 10.54518/rh.1.6.2021.217-228.

[5] S. S. Angkat dan S. Saharuddin, "Pengaruh Indeks Gini Rasio, Indeks Kemahalan Konstruksi, Pengeluaran PerkapitaTerhadap Indeks Pembangunan Manusia (IPM)," *J. Ekon. Reg. Unimal*, vol. 6, no. 2, hal. 13, 2024, doi: 10.29103/jeru.v6i2.14582.

[6] V. Bantika, O. L. S. Benu, dan G. H. M. Kapantow, "Faktor-Faktor yang Mempengaruhi Ketimpangan Distribusi Pendapatan di Sulawesi Utara," *COCOS (Jurnal Ilm. Fak. Pertan. Univ. Sam Ratulangi)*, vol. 6, no. 17, hal. 1–33, 2015.

[7] N. Taufiq, "Penciri Kemiskinan Ekstrem di 35 Kabupaten Prioritas Penanganan Kemiskinan Ekstrem," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, hal. 895–904, 2022, doi: 10.34123/semnasoffstat.v2022i1.1258.

[8] M. Sholeh, D. Andayati, dan R. Y. Rachmawati, "Data Mining Model Klasifikasi Menggunakan Algoritma K-Nearest Neighbor dengan Normalisasi untuk Prediksi Penyakit Diabetes," *J. TeIKa*, vol. 12, no. 2, hal. 77–87, 2022.

[9] P. J. M. Ali, "Investigating the Impact of Min-Max Data Normalization on the Regression Performance of K-Nearest Neighbor with Different Similarity Measurements," *ARO-THE*

Copyright © SENASIF 2026
ISSN: 2598-0076

Sci. J. KOYA Univ., vol. 10, no. 1, 2022, doi: 10.14500/aro.10955.

[10] M. R. Kusnaidi, T. Gulo, dan S. Aripin, "Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal," *J. Comput. Syst. Informatics*, vol. 3, no. 4, hal. 330–338, 2022, doi: 10.47065/josyc.v3i4.2112.

[11] R. A. Ananda, "Clustering Menggunakan Algoritma K-Means untuk Mengelompokan Data Perjudian Berdasarkan Wilayah di Kota Binjai (Studi Kasus : Pengadilan Negeri Binjai)," no. 4, 2024.

[12] F. Nurdiyansyah dan I. Akbar, "Implementasi Algoritma K-Means untuk Menentukan Persediaan Barang pada Poultry Shop," *J. Teknol. dan Manaj. Inform.*, vol. 7, no. 2, hal. 86–94, 2021, doi: 10.26905/jtmi.v7i2.6377.

[13] T. M. Ghazal, M. Z. Hussain, R. A. Said, dan A. Nadeem, "Performances of K-Means Clustering Algorithm with Different Distance Metrics," *Intell. Autom. Soft Comput.*, vol. 30, no. 2, 2021, doi: 10.32604/iasc.2021.019067.

[14] W. Widyawati, W. L. Y. Saptomo, dan Y. R. W. Utami, "Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan," *J. Ilm. SINUS*, vol. 18, no. 1, hal. 75, 2020, doi: 10.30646/sinus.v18i1.448.

[15] S. M. Miraftabzadeh, C. G. Colombo, M. Longo, dan F. Foiadelli, "K-Means and Alternative Clustering Methods in Modern Power Systems," *IEEE Access*, vol. 11, no. November, hal. 119596–119633, 2023, doi: 10.1109/ACCESS.2023.3327640.

Copyright © SENASIF 2026

ISSN: 2598-0076