

# **Analisis *Clustering K-Means* dan *K-Medoids* terhadap Faktor-Faktor Kemiskinan untuk Karakterisasi Kemiskinan Kabupaten/Kota di Indonesia**

**Data Dives**

**Farhan Akhtar G., Maryesta Apriliani S., Rachmania Azzahra S.**

## **Abstrak**

Kemiskinan adalah permasalahan dunia yang kompleks. Topik kemiskinan bahkan menjadi urutan pertama dari 17 Tujuan SDGs (*Sustainable Development Goals*). Kemiskinan di Indonesia tidak mudah untuk ditangani karena dipengaruhi oleh banyak faktor. Untuk dapat memahami fenomena kemiskinan di Indonesia, penelitian ini menerapkan model *clustering* untuk mengelompokkan kota/kabupaten di Indonesia berdasarkan faktor-faktor penyebab kemiskinan. Jenis data yang digunakan adalah data sekunder yang digunakan mencakup berbagai indikator sosial-ekonomi yang diduga dapat mempengaruhi kemiskinan, seperti rata-rata lama sekolah, pengeluaran perkapita, tingkat pengangguran terbuka dan lain sebagainya. Analisis yang dilakukan berhasil menghasilkan mengelompokkan kabupaten/desa dengan metode *clustering K-Means* serta dengan PCA menjadi 3 klaster dengan *silhouette score* dan Davies-Bouldin Index (DBI) sebesar 0.3272382 dan 1.2323586 yang cukup baik dalam. Hasil analisis menunjukkan terdapat 3 klaster. Dilakukan karakterisasi pada 3 klaster dan beberapa rekomendasi diusulkan berdasarkan karakteristik setiap klaster sehingga dapat diberikan rekomendasi yang lebih spesifik dan terperinci.

**Kata kunci:** *clustering*, k-means, k-medoids, kemiskinan

## **1. Latar Belakang**

Kemiskinan adalah fenomena multidimensional global yang dipengaruhi oleh berbagai faktor sosial, ekonomi, maupun demografi, dan telah menjadi isu krusial di banyak negara di dunia. Sebagai bagian dari komitmen global untuk meningkatkan kesejahteraan masyarakat, SDGs (*Sustainable Development Goals*) menempatkan pengentasan kemiskinan sebagai tujuan pertama dari 17 tujuan sasaran global yang harus dicapai. Hal ini menunjukkan bahwa kemiskinan masih menjadi tantangan utama di seluruh dunia (Leal Filho et al., 2021), di mana upaya pengentasan kemiskinan ini menjadi landasan bagi tercapainya tujuan

pembangunan lainnya. Kemiskinan di Indonesia masih menjadi tantangan besar meskipun angka kemiskinan telah menurun dalam beberapa tahun terakhir. Berdasarkan data Badan Pusat Statistik (BPS) tahun 2024, persentase penduduk miskin tercatat sebesar 9,03%, tetapi angka ini tidak merata di seluruh wilayah. Hill (2021) menyatakan bahwa kesenjangan antarwilayah tetap menjadi masalah signifikan dan masih banyak daerah yang kesulitan menghadapi kemiskinan. Kondisi ini menunjukkan bahwa diperlukan analisis yang lebih mendalam dan berdasarkan data untuk memahami faktor-faktor penyebab kemiskinan di setiap wilayah.

**Tabel 1.** Rangkuman hasil kajian literatur mengenai faktor kemiskinan di Indonesia

No	Referensi	Faktor Kemiskinaan
1	Azis, dkk., (2023)	Jumlah penduduk, indeks pembangunan manusia, tingkat pengangguran, dan produk domestik bruto
2	Purba, dkk., (2023)	Literasi, akses listrik rumah tangga, Produk Domestik Regional Bruto (PDRB)
3	Sinurat (2023)	Indeks pembangunan manusia
4	Fitri (2020)	Tingkat pengangguran terbuka dan rata-rata pengeluaran per kapita sebulan untuk makanan
5	Prawoto & Basuki (2022)	Pengeluaran daerah
6	Hendayanti & Nurhidayati (2020)	Rata-rata lama sekolah dan harapan lama sekolah

Beberapa peneliti sudah melakukan analisis terkait faktor kemiskinan di indonesia. Tabel 1 di bawah ini merupakan ringkasan dari hasil kajian literatur yang telah dilakukan. Azis, dkk., (2023) menyatakan bahwa jumlah penduduk, indeks pembangunan manusia, tingkat pengangguran, dan produk domestik bruto berpengaruh secara signifikan pada tingkat kemiskinan. Sementara itu, Sinurat (2023) menyatakan bahwa indeks pembangunan manusia memiliki hubungan negatif dengan tingkat kemiskinan, menunjukkan bahwa peningkatan kualitas pembangunan manusia dapat membantu mengurangi kemiskinan. Prawoto & Basuki (2022) mendapatkan bahwa justru pengeluaran daerah memiliki hubungan positif terhadap tingkat kemiskinan, yang disebabkan oleh ketidakefektifan

program pemerintah dalam mengatasi masalah tersebut. Purba, dkk., (2023) mendapatkan bahwa dana alokasi khusus (DAK) yang diharapkan dapat mendukung pengentasan kemiskinan ternyata tidak menunjukkan efek signifikan di berbagai wilayah. Lebih lanjut, Fitri (2020) menyoroti bahwa tingkat pengangguran terbuka serta rata-rata pengeluaran per kapita untuk makanan adalah faktor yang mempengaruhi indeks kemiskinan. Hendayanti & Nurhidayati (2020) menyimpulkan bahwa rata-rata lama sekolah dan harapan lama sekolah berpengaruh terhadap kemiskinan.

## 2. Tujuan Penulisan

Kemiskinan di Indonesia merupakan masalah kompleks yang dipengaruhi oleh berbagai faktor. Persentase penduduk miskin dalam kabupaten/kota merupakan indikator langsung yang mengukur proporsi penduduk yang hidup di bawah garis kemiskinan. Rata-rata lama sekolah penduduk yang berusia lebih dari 15 tahun menjadi salah satu faktor penting karena pendidikan yang lebih baik meningkatkan peluang untuk mendapatkan kesempatan hidup yang lebih baik pula. Selain itu, pengeluaran per kapita, yang mencerminkan kemampuan daya beli masyarakat, juga sangat mempengaruhi kemiskinan. Daerah dengan pengeluaran per kapita yang rendah cenderung memiliki tingkat kemiskinan yang lebih tinggi (Wulandari & Pratama, 2022). Selanjutnya, Indeks Pembangunan Manusia (IPM) yang mengukur mengenai sumber daya dan kesejahteraan masyarakat dari berbagai aspek, diantaranya pendidikan, kesehatan, dan ekonomi, juga signifikansi berpengaruh terhadap kemiskinan (Mukhtar & Saptono, 2019). Usia harapan hidup juga memiliki signifikansi negatif terhadap kemiskinan di Indonesia (Wulandari & Pratama, 2022). Persentase penduduk miskin memiliki rumus, yaitu:

$$P_{\alpha} = \frac{1}{n} \sum_{i=1}^q \left[ \frac{z - y_i}{z} \right]^{\alpha}$$

Keterangan:

$\alpha = 0$

$z$  = garis kemiskinan

$y_i$  = rata-rata pengeluaran per kapita sebulan penduduk yang berada dibawah garis kemiskinan ( $i = 1, 2, 3, \dots, q$ ),  $y_i < z$ .

$q$  = banyaknya penduduk yang berada dibawah garis kemiskinan.

$n$  = jumlah penduduk.

Akses terhadap sanitasi layak dan air minum layak juga mempengaruhi kemiskinan, karena keduanya merupakan aspek penting dari kesehatan masyarakat (Adhitya, Prabawa, & Kencana 2022). Tingkat pengangguran terbuka juga berdampak signifikan terhadap kemiskinan (Mukhtar & Saptono 2019). Semakin tinggi pengangguran, semakin banyak orang yang tidak memiliki pendapatan tetap. Tingkat partisipasi angkatan kerja mencerminkan proporsi penduduk yang aktif secara ekonomi. Muttaqin dan Anwar (2023) menemukan bahwa tingkat partisipasi angkatan kerja signifikan berpengaruh secara negatif terhadap kemiskinan di Indonesia. Terakhir, Produk Domestik Regional Bruto (PRDB) atas dasar harga konstan menurut pengeluaran menunjukkan aktivitas ekonomi di suatu wilayah. Wilayah dengan PDRB yang lebih tinggi umumnya memiliki ekonomi yang lebih maju, yang dapat menciptakan lebih banyak lapangan kerja dan meningkatkan pendapatan masyarakat, sehingga menurunkan angka kemiskinan (Azizah & Aisyah 2023).

Untuk memahami kemiskinan, diperlukan pembahasan yang lebih spesifik dan tepat sasaran, mengingat kondisi di berbagai kota dan kabupaten di Indonesia sangat beragam. Analisis yang mempertimbangkan karakteristik setiap daerah akan lebih mampu mengidentifikasi rekomendasi yang sesuai dengan kebutuhan dan tantangan spesifik masing-masing wilayah.

### **3. Metode Penelitian**

#### **3.1. Data**

Data yang digunakan merupakan data sekunder yang berasal dari website kaggle.com. Data berisikan variabel-variabel yang terkait dengan faktor penyebab kemiskinan. Terdapat sepuluh variabel terkait kemiskinan yang terdapat pada dataset. Kesepuluh variabel ini merupakan faktor-faktor dari kemiskinan.

**Tabel 2.** Variabel Penelitian

Variabel	Keterangan
X1	Persentase penduduk miskin dalam kabupaten/kota
X2	Rata-rata lama sekolah penduduk yang berusia lebih dari 15 tahun
X3	Pengeluaran perkapita (dalam satuan ribuan per orang per tahun)
X4	Indeks pembangunan manusia
X5	Usia harapan hidup
X6	Persentase rumah tangga yang memiliki akses sanitasi yang layak
X7	Persentase rumah tangga yang memiliki akses air minum layak
X8	Tingkat pengangguran terbuka
X9	Tingkat partisipasi angkatan kerja
X10	PDRB atas dasar harga konstan menurut pengeluaran (dalam satuan rupiah)

### **3.2. Metode *Clustering***

Pengklasteran (*clustering*) merupakan salah satu teknik penting dalam analisis data yang digunakan untuk mengelompokkan data berdasarkan kesamaan nilai-nilai yang dimilikinya (Buulolo, 2020). Teknik ini memungkinkan data dalam setiap kelompok (*klaster*) memiliki kemiripan yang signifikan. Dalam praktiknya, terdapat berbagai algoritma yang digunakan dalam pengklasteran, termasuk *K-Means*, *K-Medoids*, *K-Nearest Neighbor*, dan lainnya. Algoritma-algoritma ini digunakan untuk mengelompokkan data berdasarkan pola kemiripan tertentu sehingga dapat memberikan pemahaman yang lebih baik terhadap data yang kompleks (Buulolo, 2020).

### **3.3. *K-Means Clustering***

*K-Means Clustering* mengelompokkan data berdasarkan kesamaan dengan langkah-langkah berikut: (a) menetapkan jumlah klaster (*k*), (b) menginisialisasi pusat klaster (*centroid*) secara acak, (c) menghitung jarak Euclidean antara tiap data dan *centroid*, (d) mengklasifikasikan data ke klaster dengan *centroid* terdekat, (e) memperbarui *centroid* berdasarkan rata-rata data dalam klaster, (f) mengulangi proses hingga perubahan posisi *centroid* minimal atau mencapai batas iterasi. Hasil akhir adalah pengelompokan data ke dalam klaster yang relevan. *K-Means* dipilih karena cepat dan efisien dalam mengelompokkan data besar (Butsianto & Mayangwulan 2020).

### **3.4. *K-Medoids Clustering***

*K-Medoids Clustering* adalah algoritma pengelompokan yang mirip dengan *K-Means*, namun dengan beberapa perbedaan utama. Pada *K-Medoids*, setiap klaster diwakili oleh objek yang sebenarnya dari dataset, yang disebut "medoid", sedangkan *K-Means* menggunakan rata-rata (centroid) dari semua titik data dalam klaster. *K-Medoids* lebih tahan terhadap outlier atau data yang ekstrem dibandingkan *K-Means* (Khan, dkk 2023).

### **3.5. Evaluasi *Clustering***

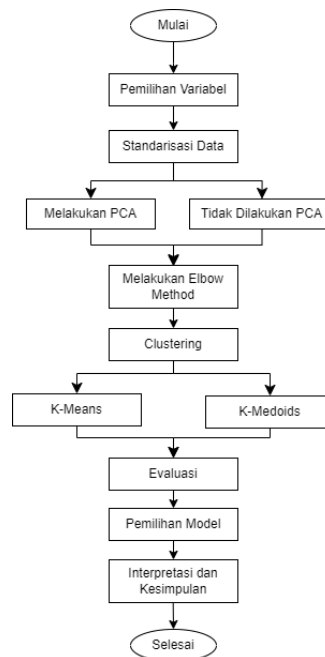
#### **3.5.1. *Silhouette Score***

Skor *Silhouette* digunakan sebagai metrik untuk mengevaluasi kualitas dan efektivitas pengelompokan, terutama untuk mengukur sejauh mana setiap objek sesuai dalam kelompoknya. Rentang skor *Silhouette* berkisar antara 0 hingga 1, di mana nilai yang lebih rendah menunjukkan bahwa sebuah objek lebih mendekati kelompok yang tidak sesuai. Objek dalam kelompok yang tepat diharapkan memiliki skor *Silhouette* yang mendekati 1 (Atira & Sari 2023).

#### **3.5.2. *Davies-Bouldin Index (DBI)***

*Davies-Bouldin Index (DBI)* adalah metrik yang digunakan untuk mengevaluasi kualitas hasil *clustering* yang sering disebut reliabilitas klasifikasi indeks (Link, dkk 2023). *DBI* mengukur rata-rata kesamaan antara setiap klaster dengan klaster lain, dengan mempertimbangkan rasio antara jarak *between-class* dan jarak *within-class* (Sopyan, dkk 2022). Nilai *DBI* yang lebih rendah menunjukkan klaster yang lebih terpisah secara jelas dan lebih kompak, yang berarti hasil *clustering* lebih baik.

### 3.6. Tahapan Penelitian



**Gambar 1.** *Flowchart Tahapan Penelitian*

#### 3.6.1. Pemilihan Variabel untuk *Clustering*

Pemilihan variabel untuk proses *clustering* sangat penting agar hasil *clustering* lebih akurat. Dalam penelitian ini, terdapat 10 variabel yang terkait dengan faktor-faktor kemiskinan. Salah satu pertimbangan utama dalam memilih variabel adalah adanya multikolinearitas. Jika dua atau lebih variabel saling berkorelasi kuat, hal ini dapat menyebabkan bias dalam proses kluster. Variabel yang merupakan hasil perhitungan dari variabel lain, seperti indeks yang dihitung dari beberapa variabel lain, tidak akan digunakan dalam clustering karena dapat menciptakan redundansi informasi.

#### 3.6.2. Standarisasi Data

Standarisasi data adalah salah satu proses penting dalam *clustering*, terutama ketika variabel-variabel yang digunakan memiliki satuan atau skala pengukuran yang berbeda-beda. Perbedaan skala ini, jika tidak diatasi, dapat menghasilkan kluster yang kurang akurat. Dengan melakukan standarisasi, semua variabel diubah ke skala yang sama, sehingga setiap variabel memiliki bobot yang setara dalam memengaruhi pembentukan kluster.

#### 3.6.3. *Principal Component Analysis* (PCA) pada Variabel Penelitian

Dalam proses *clustering*, terdapat 10 variabel yang akan digunakan, dan jumlah ini tergolong cukup banyak. Banyaknya variabel ini dapat menimbulkan masalah yang dikenal sebagai "The Curse of Dimensionality", di mana peningkatan dimensi dari

data dapat menyebabkan performa *clustering* menurun. Untuk mengatasi masalah ini, akan diterapkan *Principal Component Analysis (PCA)* untuk mereduksi dimensi tanpa mengorbankan terlalu banyak informasi penting hilang dari dataset. Di bagian akhir, kami akan membandingkan hasil *clustering* antara data yang sudah dilakukan *PCA* dengan data tanpa dilakukan *PCA*.

#### 3.6.4. Menentukan Banyak Klaster dengan Elbow Method

*Elbow method* digunakan untuk menentukan jumlah klaster yang optimal dalam proses *clustering*. Dalam hal ini, *Elbow Method* diterapkan pada *K-Means Clustering* dan *K-Medoids Clustering* untuk mengidentifikasi jumlah klaster yang sesuai. Metode ini menghitung inersia atau total jarak antara data dan pusat klasternya untuk berbagai nilai  $k$  (jumlah klaster), kemudian grafik inersia dianalisis untuk menemukan titik "siku" yang menandakan penurunan signifikan, yang dianggap sebagai jumlah klaster optimal. Setelah banyak klaster diperoleh, perbandingan akan dilakukan untuk mengevaluasi apakah ada perbedaan hasil *clustering* dari jumlah klaster optimal kedua metode tersebut.

#### 3.6.5. Evaluasi Model

Evaluasi model dilakukan untuk membandingkan beberapa aspek penting dalam proses *clustering*, yaitu penggunaan *PCA* atau tanpa *PCA*, metode *K-Means Clustering* atau *K-Medoids Clustering*, serta jumlah klaster yang optimal. Evaluasi ini akan menggunakan *Silhouette Score* dan *Davies-Bouldin Index (DBI)* sebagai metrik penilaian.

#### 3.6.6. Interpretasi *Clustering*

Setelah model *clustering* terbaik dipilih, langkah selanjutnya adalah melakukan interpretasi terhadap hasil *clustering*. Masing-masing klaster akan dianalisis untuk memahami karakteristik unik yang dimiliki. Berdasarkan karakteristik unik setiap klaster, akan diberikan rekomendasi yang lebih spesifik terkait upaya penanganan kemiskinan di setiap klaster. Rekomendasi ini diharapkan dapat memberikan solusi yang lebih tepat sasaran dan terpersonalisasi, mengingat kota dan kabupaten di Indonesia memiliki kondisi yang sangat beragam.

### 4. Hasil dan Analisis

#### 4.1.1. Pemilihan Variabel untuk *Clustering*

Diputuskan untuk tidak menggunakan Indeks Pembangunan Manusia (IPM) dalam *clustering* karena IPM merupakan hasil perhitungan beberapa variabel lain, seperti pendidikan dan harapan hidup, yang sudah tercakup dalam variabel lainnya. Penggunaannya dapat menyebabkan redundansi informasi dan mengurangi efektivitas model. Selain itu, hasil analisis menunjukkan nilai *Variance Inflation Factor (VIF)* untuk IPM cukup tinggi, yang mengindikasikan adanya



multikolinearitas. Oleh karena itu, IPM dikeluarkan dari variabel *clustering* untuk menghindari bias pada interpretasi *clustering*.

**Tabel 3.** VIF untuk Setiap Variabel Penelitian

Variabel	VIF
X1	2.248691
X2	9.619108
X3	8.627558
X4	37.019905
X5	4.501827
X6	2.314261
X7	1.650604
X8	2.167504
X9	2.038465
X10	1.584445

#### 4.1.2. Principal Component Analysis (PCA) pada Data Variabel Penelitian

Setelah standarisasi data dilakukan, tahap selanjutnya adalah menerapkan *Principal Component Analysis (PCA)* untuk mereduksi dimensi data. Penggunaan *PCA* bertujuan untuk mengurangi masalah *curse of dimensionality*. Hasil analisis menunjukkan bahwa 5 *principal components* mampu menjelaskan 85,64% variansi kumulatif dari data. Oleh karena itu, diputuskan untuk menggunakan 5 komponen utama ini dalam proses clustering.

**Tabel 4.** Summary untuk *Principal Components*

Component	Standard Deviation	Proportion of Variance	Cumulative Proportion
PC1	2.1415	0.5096	0.5096
PC2	0.9974	0.1105	0.6201
PC3	0.9439	0.0990	0.7191
PC4	0.81319	0.07347	0.79257
PC5	0.75791	0.06382	0.85640
PC6	0.67578	0.05074	0.90714
PC7	0.59381	0.03918	0.94632
PC8	0.51646	0.02964	0.97595
PC9	0.46520	0.02405	1.00000

#### 4.1.3. Menentukan Banyak Klaster dengan *Elbow Method*

Berdasarkan hasil *Elbow Method*, didapatkan bahwa jumlah klaster optimal untuk data tanpa *PCA* adalah 3 untuk *K-Means* dan 4 untuk *K-Medoids*. Hasil yang sama ditemukan pada data dengan *PCA*. Selanjutnya, akan dilakukan perbandingan performa antara jumlah klaster 3 dan 4 untuk kedua metode *clustering* tersebut.

**Tabel 5.** Ringkasan Hasil *Elbow Method*

Clustering	Banyak Klaster	
	Data Tanpa PCA	Data dengan PCA
K-Means	3	3
K-Medoids	4	4

#### 4.1.4. Evaluasi Model

**Tabel 6.** Evaluasi Tiap Model *Clustering*

Model		Silhouette Score		Davies-Bouldin Index (DBI)	
		3 Klaster	4 Klaster	3 Klaster	4 Klaster
K-Means	Tanpa PCA	0.2746607	0.190466	1.3116897	1.644842
	PCA	0.3272382	0.3055112	1.2323586	1.021469
K-Medoids	Tanpa PCA	0.1331244	0.149379	2.0925584	1.815592
	PCA	0.2027242	0.201983	2.1777466	1.66518

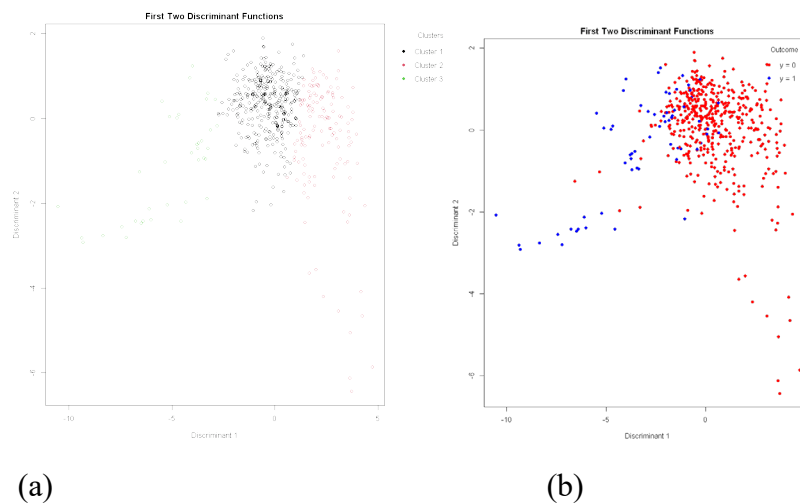
**Tabel 7.** Evaluasi dengan PCA dan Tanpa PCA

Clustering	Rata-Rata Silhouette Score		Rata-Rata DBI	
	Tanpa PCA	Dengan PCA	Tanpa PCA	Dengan PCA
K-Means	0.23256335	0.2880871	1.47826585	1.1269138
K-Medoids	0.1412517	0.2023536	1.9540752	1.9214633

Tabel di atas menunjukkan hasil evaluasi model clustering menggunakan *Silhouette Score* dan *Davies-Bouldin Index (DBI)* untuk berbagai kombinasi metode dan jumlah klaster, baik dengan maupun tanpa *PCA*. Berdasarkan perbandingan rata-rata dari kedua metrik tersebut, terlihat bahwa model *clustering* dengan *PCA* menghasilkan performa yang lebih baik dibandingkan tanpa *PCA*. Selanjutnya, dilakukan perbandingan antara *K-Means* dan *K-Medoids*. Hasil evaluasi menunjukkan bahwa *K-Means* memiliki performa yang lebih baik berdasarkan

kedua metrik evaluasi. Terakhir, untuk menentukan jumlah kluster yang optimal, diperoleh bahwa model dengan 3 kluster memberikan hasil evaluasi yang lebih baik dibandingkan 4 kluster untuk dapat diinterpretasi karena nilai *Silhouette Score* yang lebih besar. Oleh karena itu, model akhir yang akan digunakan untuk interpretasi adalah *K-Means* dengan 3 kluster dan *PCA*.

#### 4.1.5. Karakteristik *Klaster*



**Gambar 2.** (a) Hasil dari *K-Means Clustering* dengan 3 Kluster dan *PCA* (b)

Visualisasi dengan Informasi Klasifikasi Kemiskinan pada Data

**Tabel 8.** Rata-Rata untuk Setiap Variabel Penelitian

Klaster	X1	X2	X3	X5	X6	X7	X8	X9	X10
1	29.6	5.48	5938	65.3	35.8	54.9	1.96	81.3	1.84e6
2	7.46	9.98	13256	72.7	89.5	96.3	7.95	65.6	5.40e7
3	12.3	8.14	9621	68.9	76.8	83.9	4.23	69.7	1.12e7

**Tabel 9.** Informasi Hasil Clustering dengan Informasi Klasifikasi Kemiskinan

Klaster	Persentase Jumlah Kab/Kota dengan Status Miskin (%)	Jumlah Kab/Kota dengan Status Miskin	Jumlah Kab/Kota dengan Status Tidak Miskin	Jumlah Kab/Kota
1	81.1	30	7	37
2	0	0	137	137
3	9.41	32	308	340

Berikut adalah hasil visualisasi untuk *K-Means* dengan 3 klaster dan *PCA*. Berdasarkan tabel karakteristik rata-rata kemiskinan variabel pada masing-masing klaster, ditunjukkan bahwa klaster 1 adalah klaster dengan tingkat kemiskinan yang bernilai sangat tinggi. Hal ini disebabkan karena nilai persentase penduduk miskin adalah yang paling tinggi dibandingkan dengan klaster lainnya. Di sisi lain, berdasarkan ketiga variabel tersebut, Klaster 2 adalah klaster dengan tingkat kemiskinan yang sangat rendah. Klaster 2 mempunyai karakteristik dengan persentase rata-rata lama sekolah (X2) yang paling tinggi dibandingkan dengan klaster lainnya. Pada klaster 2 juga memiliki karakteristik angka tingkat pengangguran terbuka yang lebih tinggi dibandingkan dengan beberapa klaster lainnya. Hal ini menunjukkan bahwa pada provinsi yang tercakup dalam klaster 2 memiliki kualitas pendidikan yang baik dengan dilihat dari X2. Pendidikan yang berkualitas pada provinsi tersebut dapat mendukung penurunan tingkat kemiskinan. Tetapi, persentase pengangguran pada klaster ini perlu ditinjau lebih lanjut karena dengan angka yang cukup tinggi dibandingkan dengan klaster lainnya, dapat memicu peningkatan tingkat kemiskinan. Klaster 2 memiliki karakteristik dengan X8 yang paling tinggi dibandingkan dengan klaster lainnya. Selain itu, nilai X2 juga tertinggi dibandingkan dengan klaster lainnya. Indikasi dari karakteristik ini menunjukkan bahwa kualitas pendidikan pada klaster 2 memiliki kriteria yang cukup sehingga klaster 2 tergolong dalam tingkat kemiskinan yang rendah. Masalah pengangguran akan menjadi tantangan dengan angka yang cukup tinggi. Diperlukan kebijakan pemerintah untuk dapat memberikan pelatihan kerja dan memperluas lapangan pekerjaan kepada masyarakat sehingga dapat meningkatkan tenaga kerja yang terampil dan mampu bersaing dengan lainnya. Klaster 3 memiliki karakteristik dengan tingkat kemiskinan yang tertinggi kedua. Pada provinsi yang termasuk dalam klaster ini menunjukkan bahwa rata-rata masyarakat telah mendapatkan pendidikan yang cukup dan memiliki tingkat pengangguran yang sedang.

## **5. Kesimpulan dan Saran**

### **5.1. Kesimpulan**

Model *clustering* dengan kombinasi *K-Means*, 3 *klaster* dan *PCA* memberikan performa terbaik berdasarkan *Silhouette Score* dan *Davies-Bouldin Index*. Hasil analisis lebih lanjut menunjukkan bahwa klaster 1 merupakan klaster dengan tingkat kemiskinan yang sangat tinggi yang dapat terlihat pada persentase penduduk miskin yang paling tinggi diantara semua klaster sehingga diperlukan peningkatan program bantuan sosial dan pemberdayaan ekonomi sebagai upaya mengurangi tingkat kemiskinan yang sangat tinggi, klaster 2 memiliki tingkat kemiskinan yang sangat rendah dengan didukung oleh kualitas pendidikan yang baik, namun memiliki tingkat pengangguran yang tinggi sehingga diperlukan penanganan untuk mencegah peningkatan kemiskinan seperti memperluas akses lapangan kerja, dan klaster 3 adalah klaster dengan tingkat kemiskinan tertinggi kedua dengan pendidikan yang memadai, namun dengan tingkat pengangguran dalam kategori sedang sehingga diperlukan peningkatan program pendidikan dan pelatihan keterampilan.

### **5.2. Saran**

Untuk penelitian selanjutnya, dapat digunakan variabel-variabel lain yang dapat lebih menggambarkan karakterisasi kemiskinan di Indonesia sehingga hasil yang didapat dapat memberikan berbagai macam *insight*. Selain itu, variasi penggunaan metode standarisasi, metode mencari jumlah klaster, metode *clustering*, maupun metrik evaluasi dapat dieksplorasi lebih jauh untuk didapatkan karakterisasi kabupaten/kota di Indonesia yang lebih baik.

## **6. Daftar Pustaka**

Adhitya, B., Prabawa, A., & Kencana, H. (2022). Analisis Pengaruh Pendidikan, Kesehatan, Sanitasi dan Rata-Rata Jumlah Anggota Keluarga Per Rumah Tangga terhadap Kemiskinan di Indonesia. *Ekonomis: Journal of Economics and Business*, 6(1), 288-295.

- Atira, A., & Sari, B. N. (2023). Penerapan Silhouette Coefficient, Elbow Method dan Gap Statistics untuk Penentuan Klaster Optimum dalam Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Kebahagiaan. *Jurnal Ilmiah Wahana Pendidikan*, 9(17), 76-86.
- Azis, I., Sumertajaya, I. M., Purwaningsih, S. S., & Tjahjawati, S. S. (2023). Penentuan faktor kemiskinan Indonesia menggunakan regresi logistik. *Journal of Mathematics, Computations, and Statistics*, 6(1), 61–65. <https://www.ojs.unm.ac.id/jmathcos>
- Buulolo, E. (2020). Data Mining Untuk Perguruan Tinggi (Edisi 1). Yogyakarta: Deepublish. [Online].
- Butsianto, S., & Mayangwulan, N. T. (2020). Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering. *J. Nas. Komputasi dan Teknol. Inf*, 3(3).
- Leal Filho, W., Lovren, V. O., Will, M., Salvia, A. L., & Frankenberger, F. (2021). Poverty: A central barrier to the implementation of the UN Sustainable Development Goals. *Environmental Science & Policy*, 125, 96-104.
- Hendayanti, N. P. N., & Nurhidayati, M. (2020). Regresi Logistik Biner dalam Penentuan Ketepatan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi-Provinsi di Indonesia. *Sainstek: Jurnal Sains dan Teknologi*, 12(2), 63-70.
- Hill, H. (2021). What's happened to poverty and inequality in Indonesia over half a century?. *Asian Development Review*, 38(1), 68-97.
- Khan, A. S. S., Fatekurohman, M., & Dewi, Y. S. (2023). Perbandingan Algoritma K-Medoids Dan K-Means Dalam Pengelompokan Kecamatan Berdasarkan Produksi Padi Dan Palawija Di Jember. *Jurnal Statistika dan Komputasi*, 2(2), 67-75.
- Ling, X., Tu, Q., Jin, M., Wang, W., Cui, Y., & Zhu, J. (2022, December). Research on Distributed Resource Aggregation Technology Based on Hierarchical Agglomerative Clustering Analysis. In *Journal of Physics: Conference Series* (Vol. 2399, No. 1, p. 012034). IOP Publishing.

- Mukhtar, S., & Saptono, A. (2019). Analisis pengaruh indeks pembangunan manusia dan tingkat pengangguran terbuka terhadap kemiskinan di Indonesia. *Ecoplan*, 2(2), 77-89.
- Muttaqin, M., & Anwar, K. (2023). Pengaruh Inflasi, Tingkat Pengangguran Terbuka, dan Tingkat Partisipasi Angkatan Kerja terhadap Tingkat Kemiskinan di Indonesia. *Jurnal Aplikasi Ilmu Ekonomi*, 2(2), 83-95.
- Prawoto, N., & Basuki, A. T. (2022). Factors Affecting Poverty in Indonesia: A Panel Data Approach. *Quality-Access to Success*, 23(186), 156-161.
- Purba, J. T., Budiono, S., Hariandja, E. S., & Pramono, R. (2023). Sustainability Strategy to Alleviate Poverty Through Education, Energy, GRDP, and Special Funds: Evidence from Indonesia. *International Journal of Sustainable Development & Planning*, 18(5).
- Sinurat, R. P. P. (2023). Analisis faktor-faktor penyebab kemiskinan sebagai upaya penanggulangan kemiskinan di Indonesia. *Jurnal Registratie*, 5(2), 87-103.
- Sopyan, Y., Lesmana, A. D., & Juliane, C. (2022). Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Klaster Terbaik Kasus Perceraian di Kabupaten Kuningan. *Building of Informatics, Technology and Science (BITS)*, 4(3), 1464-1470.
- Sugiharti, L., Purwono, R., Esquivias, M. A., & Jayanti, A. D. (2022). Poverty Dynamics in Indonesia: The Prevalence and Causes of Chronic Poverty. *Journal of Population & Social Studies*, 30.
- Wulandari, I., & Pratama, A. A. N. (2022). Analisis Pengaruh Dana ZIS (Zakat, Infak, Sedekah), Pertumbuhan Ekonomi, Angka Harapan Hidup, Rata-Rata Lama Sekolah, Dan Pengeluaran Perkapita Terhadap Tingkat Kemiskinan Di Indonesia Periode 2010-2021. *Jurnal Ilmiah Ekonomi Islam*, 8(3), 3301-3309.