

# **Penerapan Algoritma K-Medoids untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Kemiskinan: Akses Perumahan dan Kesejahteraan**

**Akhfan Ikhram<sup>1</sup>, Sani Mutia<sup>2</sup>, Purnama Akbar<sup>3\*</sup>, Choerul Umam<sup>4</sup>**

<sup>1,2,3</sup> Program Studi Statistika, Universitas Islam Indonesia, Jalan Kaliurang KM 14.5, Kabupaten Sleman, Daerah Istimewa Yogyakarta, 55583, Indonesia

<sup>4</sup> Badan Pusat Statistik Provinsi Daerah Istimewa Yogyakarta, Jalan Brawijaya, Tamantirto, Kasihan, Bantul, Daerah Istimewa Yogyakarta, 55183, Indonesia

\*Corresponding author: [246111202@uii.ac.id](mailto:246111202@uii.ac.id)



**P-ISSN:** 2986-4178  
**E-ISSN:** 2988-4004

## **Riwayat Artikel**

Dikirim: 11 Maret 2025  
Direvisi: 07 Mei 2025  
Diterima: 07 Mei 2025

## **ABSTRAK**

Tantangan dalam mengentaskan kemiskinan di Indonesia masih sangat besar, terutama disebabkan oleh kesenjangan yang signifikan antarwilayah. Penelitian ini menganalisis pengelompokan kemiskinan di Indonesia menggunakan algoritma K-Medoids dengan data tahun 2024 dari Badan Pusat Statistik (BPS). Indikator yang digunakan meliputi persentase rumah tangga yang memiliki akses terhadap air minum layak, sanitasi layak, dan hunian layak. Analisis deskriptif menunjukkan rata-rata akses air minum layak sebesar 87%, sanitasi layak 81,14%, dan rumah layak huni 61,66%. Metode *Elbow* menentukan jumlah klaster optimal sebanyak tiga. Klaster 1 mencakup 19 provinsi dengan tingkat kemiskinan rendah, ditandai akses yang tinggi terhadap kebutuhan dasar. Klaster 2 terdiri dari 18 provinsi dengan tingkat kemiskinan sedang dan akses infrastruktur yang cukup baik. Klaster 3 hanya mencakup Papua Pegunungan dengan akses sangat terbatas dan tingkat kemiskinan sangat tinggi. Hasil penelitian menyoroti ketimpangan signifikan antara wilayah timur dan barat Indonesia. Temuan ini menekankan perlunya kebijakan pengentasan kemiskinan berbasis wilayah yang berfokus pada peningkatan infrastruktur dasar di daerah tertinggal guna mendukung pembangunan yang lebih inklusif dan pencapaian *Sustainable Development Goals* (SDGs).

**Kata Kunci:** K-Medoids, pengelompokan kemiskinan, air minum layak, sanitasi layak, hunian layak.

## **ABSTRACT**

The challenge of eradicating poverty in Indonesia remains substantial, primarily due to significant regional disparities. This study analyzes poverty clustering in Indonesia using the K-Medoids algorithm with 2024 data from the Indonesian Central Bureau of Statistics (BPS). The indicators used include the percentage of households with access to safe drinking water, proper sanitation, and adequate housing. Descriptive analysis shows the average access to safe drinking water is 87%, proper sanitation 81,14%, and adequate housing 61,66%. The elbow method determined the optimal number of clusters as three.

Cluster 1 consists of 19 provinces with low poverty levels and high access to basic needs. Cluster 2 comprises 18 provinces with moderate poverty levels and sufficient infrastructure access. Cluster 3 includes only Papua Pegunungan, which has limited access and is in extreme poverty. The results highlight significant regional disparities between eastern and western Indonesia. These findings emphasize the need for region-based poverty alleviation policies to improve basic infrastructure in underdeveloped areas to support more inclusive development and achieve the Sustainable Development Goals (SDGs).

**Keywords:** **K-Medoids, poverty clustering, safe drinking water, proper sanitation, adequate housing.**

## 1. Pendahuluan

Indonesia menargetkan untuk menjadi negara yang maju, mandiri, adil, dan makmur pada tahun 2045, yang dikenal sebagai visi Indonesia Emas 2045. Aspirasi ini sejalan dengan agenda Pembangunan Berkelanjutan (SDGs), yang berfokus pada upaya mengurangi kemiskinan, meningkatkan kualitas pendidikan, serta menekan ketimpangan sosial. Salah satu tujuan utama SDGs adalah menghapus kemiskinan dalam segala bentuknya di seluruh dunia. Pengentasan kemiskinan juga menjadi indikator tujuan dari pembangunan ekonomi yang inklusif. Selain itu juga, merujuk pada RPJPN 2025-2045 target kemiskinan di Indonesia adalah 0% pada tahun 2045 [1].

Namun, tantangan pengentasan kemiskinan di Indonesia masih besar. Pada September 2024, persentase penduduk miskin di Indonesia tercatat sebesar 8,57%, mengalami sedikit penurunan dari Maret 2024 yang berada di angka 9,03%. Namun, kesenjangan antarwilayah masih signifikan, terutama di wilayah timur Indonesia. Provinsi dengan tingkat kemiskinan tertinggi adalah Papua Tengah dengan 27,60%, diikuti oleh Papua Pegunungan sebesar 29,66% dan Papua sebesar 18,09%. Sebaliknya, beberapa provinsi mencatat angka kemiskinan yang lebih rendah. Bali memiliki persentase penduduk miskin terendah sebesar 3,8%, disusul oleh Kalimantan Selatan sebesar 4,02% [2]. Kondisi kemiskinan di Indonesia tidak hanya berkaitan dengan rendahnya penghasilan, tetapi juga dengan terbatasnya akses terhadap infrastruktur dasar yang sangat penting untuk meningkatkan kualitas hidup masyarakat.

Salah satu indikator penting dalam menilai kesejahteraan masyarakat adalah akses terhadap perumahan yang layak yang diukur dengan adanya kemudahan terhadap akses terhadap air minum layak, sanitasi, dan hunian layak. Infrastruktur sanitasi yang memadai, termasuk pasokan air bersih dan sistem pengelolaan limbah, berperan penting dalam meningkatkan kualitas hidup penduduk [3]. Infrastruktur ini sangat berperan dalam meningkatkan kualitas hidup, yang pada akhirnya dapat membantu mengurangi tingkat kemiskinan secara signifikan. Menurut data BPS, pada tahun 2024, persentase rumah tangga yang memiliki akses terhadap hunian layak mencapai 65,25%. Namun, terdapat disparitas yang signifikan antarprovinsi. Misalnya, DI Yogyakarta memiliki persentase akses hunian layak sebesar 86,68%, sementara Papua Pegunungan hanya mencapai 4,44% [4].

Untuk mengatasi masalah ini, penting untuk mengelompokkan wilayah-wilayah berdasarkan tingkat akses hunian layak dan karakteristiknya. Meskipun secara nasional terdapat peningkatan persentase akses hunian layak, masih banyak wilayah yang belum

memenuhi standar tersebut. Oleh karena itu, pengelompokan daerah perlu didasarkan pada kesamaan kondisi kemiskinan serta akses terhadap fasilitas dasar, seperti air minum layak, sanitasi layak, dan hunian layak. Metode statistika yang dinilai cukup efektif untuk mengatasi hal ini adalah algoritma *K-Medoids* yang bekerja dengan cara memecah data set menjadi kelompok-kelompok [5]. Metode ini diperlukan untuk mengelompokkan data yang memiliki karakteristik serupa, sehingga memudahkan dalam menentukan daerah yang membutuhkan perhatian lebih. Kelebihan dari metode ini, mampu mengatasi kelemahan dari metode *K-Means* yang sensitif terhadap *outlier* [6]. Penelitian terdahulu menunjukkan *K-Means* kurang efektif pada data outlier dengan kualitas klaster yang rendah ditunjukkan oleh nilai Silhouette 0,542 dan Dunn 0,04 setelah 22,98% data terdeteksi sebagai outlier [7].

Penelitian yang dilakukan oleh [8] membandingkan dua algoritma, *K-Means* dan *K-Medoids* untuk mengelompokkan daerah miskin di Indonesia. Dalam penelitian ini, *K-Means* dianggap lebih efektif. Sementara itu, *K-Medoids* digunakan untuk mengelompokkan data kemiskinan di Kabupaten/Kota di Pulau Kalimantan dengan indikator seperti rata-rata lama sekolah, angka harapan hidup, pengeluaran per kapita, dan persentase rumah tangga miskin [9]. Penelitian lain yang dilakukan oleh [10] menggunakan *K-Medoids* untuk mengelompokkan 34 provinsi di Indonesia berdasarkan indikator kemiskinan, seperti indeks kedalaman kemiskinan, indeks keparahan kemiskinan, TPT, RLS, UHH, IPM, PDRB, dan sebagainya. Sebagian besar penelitian sebelumnya masih terbatas pada beberapa variabel utama dan belum secara menyeluruh mempertimbangkan faktor-faktor seperti akses terhadap air minum layak, sanitasi layak, serta hunian layak sebagai komponen penting dalam analisis kemiskinan. Padahal infrastruktur berperan dalam penanggulangan kemiskinan dengan meningkatkan akses bagi masyarakat miskin dan efektivitas intervensi pemerintah. Akses yang lebih baik dapat menurunkan biaya hidup, meningkatkan pendapatan, dan membuka peluang ekonomi bagi kaum miskin [11].

Penelitian ini bertujuan untuk mengelompokkan provinsi di Indonesia menggunakan indikator kemiskinan berupa akses terhadap air minum layak, sanitasi layak, dan hunian layak dengan menggunakan metode *K-Medoids*. Pendekatan ini tidak hanya bertujuan untuk memahami dinamika sosial-ekonomi di tingkat provinsi, tetapi juga menyoroti pentingnya pembangunan infrastruktur dasar sebagai strategi untuk mengurangi kemiskinan. Hasil dari penelitian ini diharapkan dapat memberikan wawasan baru bagi para pembuat kebijakan untuk menyusun intervensi yang lebih efektif dan tepat sasaran dalam mengentaskan kemiskinan di Indonesia.

## 2. Metodologi Penelitian

### 2.1 Populasi dan Variabel Penelitian

Populasi yang digunakan dalam penelitian adalah seluruh provinsi di Indonesia yang terdiri dari 38 provinsi yang ada di Indonesia pada tahun 2024.

### 2.2 Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS). Data tersebut berisi informasi mengenai akses terhadap air minum layak, sanitasi layak, dan hunian layak yang mencakup seluruh provinsi di Indonesia untuk tahun 2024.

### 2.3 Definisi Operasional Variabel Penelitian

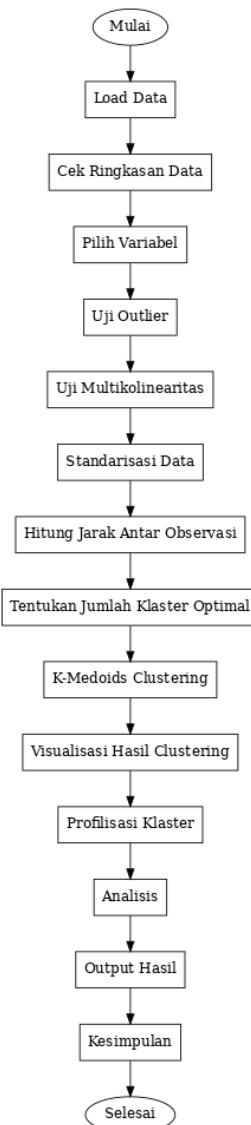
Variabel yang digunakan dalam penelitian ini mencakup tiga indikator kemiskinan yang dipilih berdasarkan relevansinya terhadap isu kemiskinan. Ketiga variabel tersebut adalah:

**Tabel 1** Variabel Penelitian

Variabel	Deskripsi	Keterangan
X1	Persentase rumah tangga yang memiliki akses terhadap air minum layak.	Persen (%)
X2	Persentase rumah tangga yang memiliki akses terhadap sanitasi layak.	Persen (%)
X3	Persentase rumah tangga yang memiliki akses terhadap hunian layak.	Persen (%)

#### 2.4 Tahapan Penelitian

Pada penelitian ini akan dilakukan analisis pengelompokan menggunakan metode *Elbow* dan *K-Medoids*. Adapun prosedur penelitian dapat dilihat pada diagram alir (*flowchart*) pada Gambar 1.



**Gambar 1.** *Flowchart* Penelitian

#### 2.4.1 Statistika Deskriptif

Menurut [12], statistika deskriptif merupakan salah satu metode statistika yang berkaitan untuk pengumpulan dan penyajian sehingga dapat memberikan informasi yang berguna. Memberikan informasi yang berguna disini dimaksudkan bahwa penyajiannya akan memudahkan pembaca atau pengguna informasi untuk dapat membaca dan memanfaatkan data secara lebih mudah. Beberapa bentuk deskriptif dari data dapat dibuat berdasarkan kebutuhan dan optimalnya sebuah informasi. Peneliti dapat menentukan jenis deskripsi seperti apa yang akan dibuat sehingga data dapat menggambarkan informasi yang berguna dan mudah dicerna.

#### 2.4.2 Multikolinieritas

Multikolinearitas adalah suatu kondisi dimana terjadi korelasi antara variabel bebas atau antar variabel bebas tidak bersifat saling bebas. Besaran (*quality*) yang dapat digunakan untuk mendeteksi adanya multikolinearitas adalah faktor inflasi ragam (*Variance Inflation Factor / VIF*). VIF digunakan sebagai kriteria untuk mendeteksi multikolinearitas pada regresi linier yang melibatkan lebih dari dua variabel bebas. Nilai VIF lebih besar dari 10 mengidentifikasi adanya masalah multikolinearitas yang serius [13]. Uji multikolinieritas berdasarkan nilai VIF yang diperoleh dari persamaan berikut [14].

$$VIF = \frac{1}{R_i^2} \quad (1)$$

dengan  $R_i^2$  merupakan koefisien determinasi dari regresi antara variabel independen yang diuji terhadap semua variabel independen lainnya.

#### 2.4.3 Clustering

Menurut [15], *clustering* merupakan suatu proses pengelompokan *record*, observasi, atau mengelompokkan kelas yang memiliki kesamaan objek. Perbedaan *clustering* dengan klasifikasi yaitu tidak adanya variabel target dalam melakukan pengelompokan pada proses *clustering*. *Clustering* sering dilakukan sebagai langkah awal dalam proses data mining. Kualitas hasil *clustering* sangat bergantung pada metode yang dipakai. Metode *clustering* juga harus dapat mengukur kemampuannya sendiri dalam usaha untuk menemukan suatu pola tersembunyi pada data yang sedang diteliti. Terdapat berbagai metode yang dapat digunakan untuk mengukur nilai kesamaan antar objek-objek yang dibandingkan, salah satunya ialah dengan *euclidean distance*. *Euclidean distance* menghitung jarak dua buah point dengan mengetahui nilai dari masing-masing atribut pada kedua poin tersebut.

Penentuan jumlah klaster optimal menggunakan metode *Elbow*. Metode ini memanfaatkan grafik yang menggambarkan hubungan antara jumlah klaster ( $k$ ) dan total *within-cluster sum of squares* (WCSS), atau bisa disebut inertia. Jumlah klaster dipilih dengan mengamati grafik, pada titik mana terjadi penurunan yang signifikan, berbentuk lengkungan yang tajam. Tujuan metode ini yaitu memilih nilai  $k$  yang terkecil dan mempunyai nilai internal yang rendah. Penentuan jumlah *cluster* yang optimal diidentifikasi dengan mempertimbangkan perbandingan perhitungan SEE pada setiap nilai *cluster*, peningkatan jumlah *cluster* akan membentuk siku, sehingga semakin besar nilai  $k$ , nilai SEE akan semakin kecil. Analisa metode *Elbow* digunakan untuk menentukan jumlah *cluster* optimal rumus metode *Elbow* sebagai berikut [16].

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} ||x_i - C_k||^2 \quad (2)$$

Keterangan:

$k$  = Jumlah klaster

$x_i$  = Nilai atribut dari data ke- $i$

$C_k$  = Jumlah klaster  $i$  pada klaster ke- $k$

$\|\cdot\|$  = Jarak *Euclidian*

#### 2.4.4 K-Medoids Clustering

Algoritma *K-Medoids* atau sering disebut juga dengan algoritma PAM (*Partitioning Around Medoid*) dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw, dimana merupakan algoritma yang mirip dengan *K-Means* karena kedua algoritma ini *partitional* yang memecah dataset menjadi kelompok-kelompok [17]. Perbedaan dari algoritma *K-Means* dengan algoritma *K-Medoids* terletak pada penentuan pusat *cluster*, dimana algoritma *K-Means* menggunakan nilai rata-rata(*mean*) dari setiap *cluster* sebagai pusat *cluster* dan algoritma *K-Medoids* menggunakan objek data sebagai perwakilan (*medoids*) sebagai pusat *cluster*. *K-Medoids* diterapkan untuk mengatasi kelemahan algoritma *K-Means* yang rentan terhadap pengaruh penculan (*outlier*). Hal ini terjadi karena objek-objek *outlier* memiliki jarak yang jauh atau karakteristik yang berbeda dari sebagian besar data lainnya, sehingga jika dimasukkan dalam suatu klaster, dapat mengubah nilai rata-rata (*mean*) klaster tersebut secara signifikan.

Berikut adalah tahapan dalam melakukan *K-Medoids* [18]:

- Menentukan banyaknya *cluster* yang akan dibentuk menggunakan metode *Elbow*.
- Mengalokasikan setiap data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan (3).

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

Keterangan:

$d(x,y)$  = Jarak *Euclidean* antara dua titik  $x$  dan  $y$

$x_i$  = Komponen ke- $i$  dari vektor  $x$

$y_i$  = Komponen ke- $i$  dari vektor  $y$

$n$  = Jumlah observasi

- Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *medoids* baru.
- Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoids* baru.
- Hitung total simpangan ( $S$ ) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika  $S < 0$ , maka tukar objek dengan data *cluster* untuk membentuk sekumpulan  $k$  objek baru sebagai *medoids*.
- Ulangi langkah (c) sampai (e) hingga tidak terjadi perubahan medoids, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

### 3. Hasil dan Pembahasan

#### 3.1. Statistika Deskriptif

Statistika deskriptif dilakukan untuk memberikan gambaran mengenai tiga variabel yang terkait dengan kualitas hidup, yaitu persentase air minum layak, sanitasi layak, dan hunian layak. Tabel ini menampilkan nilai minimum (*min*), kuartil pertama (*Q1*), median, rata-rata (*mean*), kuartil ketiga (*Q3*), dan maksimum (*max*) untuk masing-masing variabel. Tujuan dari analisis ini adalah untuk memahami distribusi data secara lebih mendalam, sehingga pola dan variasi antar variabel dapat terlihat dengan jelas. Data yang digunakan terdiri dari 38 sampel untuk setiap variabel, memberikan informasi yang

representatif untuk menggambarkan kualitas akses terhadap air minum, sanitasi, dan hunian yang layak di Indonesia.

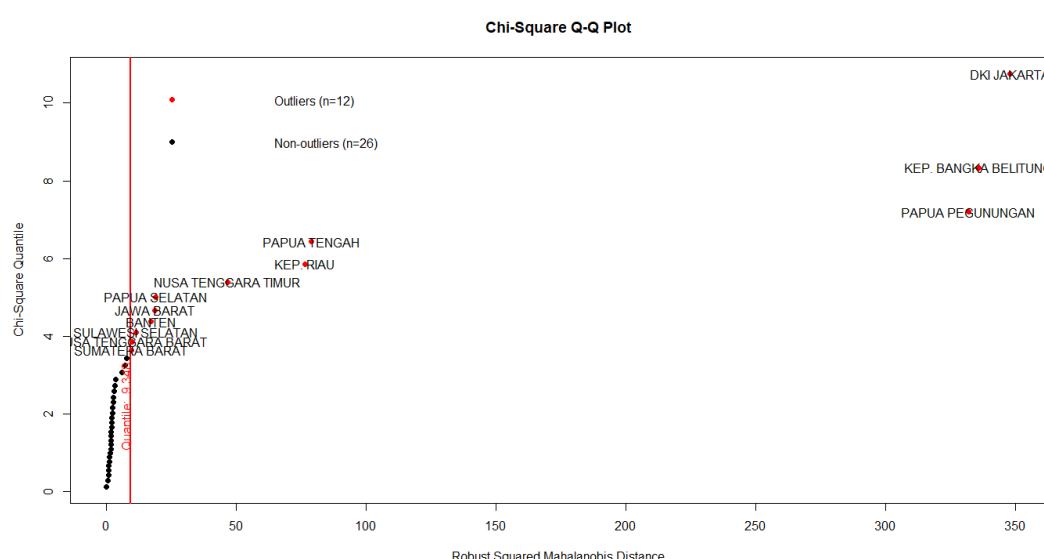
**Tabel 2** Statistik Deskriptif

	Persentase Air Minum Layak (%)	Persentase Sanitasi Layak (%)	Persentase Hunian Layak (%)
Min	30,64	12,61	4,44
Q1	82,22	79,38	58,23
Median	89,02	83,36	65,32
Mean	87	81,14	61,66
Q3	94,65	87,18	71,83
Max	99,96	96,83	86,68
N	38	38	38

Berdasarkan Tabel 2, persentase rumah tangga yang memiliki akses terhadap air minum layak di Indonesia pada tahun 2024 berkisar antara 30,64% hingga 99,96%, dengan rata-rata 87%. Kuartil pertama (Q1) menunjukkan 25% daerah memiliki akses air minum layak sebesar 82,22%, sementara kuartil ketiga (Q3) berada di 94,65%, dan median tercatat pada 89,02%. Untuk persentase sanitasi layak, nilai minimum adalah 12,61% dan maksimum 96,83%, dengan rata-rata 81,14%. Kuartil pertama dan ketiga berada pada 79,38% dan 87,18%, sementara median tercatat 83,36%. Sedangkan untuk persentase hunian layak, nilai minimum adalah 4,44% dan maksimum 86,68%, dengan rata-rata 61,66%. Kuartil pertama dan ketiga berada di angka 58,23% dan 71,83%, dengan median di 65,32%. Setiap variabel memiliki 38 data yang lengkap, tanpa ada nilai yang hilang (NA), sehingga analisis ini memberikan gambaran yang jelas mengenai kualitas akses terhadap air minum, sanitasi, dan hunian yang layak di Indonesia.

### 3.2. Pengecekan Data Outlier

Pengecekan data *outlier* dilakukan menggunakan metode *quan*, yang didasarkan pada perhitungan jarak dan dibandingkan dengan nilai kuantil *Chi-Square*. Metode ini memungkinkan identifikasi observasi yang memiliki perbedaan signifikan dibandingkan dengan data lainnya, berdasarkan hubungan multivariat di antara variabel. Melalui pendekatan ini, hasil analisis dapat memberikan gambaran yang lebih akurat tentang keberadaan *outlier* dalam dataset, sehingga memungkinkan langkah-langkah penanganan yang tepat untuk menjaga kualitas dan validitas data.



**Gambar 2.** Grafik *Outlier*

Hasil analisis dengan *Chi-Square Q-Q Plot* pada Gambar 2 menunjukkan terdapat 12 *outlier*, dengan *outlier* terjauh yaitu DKI Jakarta, Kepulauan Bangka Belitung, dan Papua Pegunungan, yang memiliki jarak *Mahalanobis* tinggi, mengindikasikan perbedaan signifikan dalam karakteristik variabel dibandingkan 26 observasi lainnya. Data *outlier* tidak dapat dihilangkan karena penelitian ini menggunakan analisis klaster, sehingga penghapusan data tersebut dapat mempengaruhi akurasi hasil klaster yang dihasilkan. Oleh karena itu, metode klaster yang tepat diterapkan untuk mengatasi keberadaan *outlier* dalam analisis ini yaitu *K-Medoids*. Sebelum itu, akan dilakukan pengujian korelasi antar variabel (multikolinieritas).

### 3.3 Uji Multikolinieritas

Dalam analisis klaster *non-hirarki* dengan metode *K-Medoids*, diperlukan pengecekan multikolinieritas untuk memastikan bahwa variabel independen tidak memiliki hubungan linear yang kuat satu sama lain. Multikolinieritas dapat dievaluasi menggunakan nilai *Variance Inflation Factor* (VIF). Suatu variabel dikatakan tidak mengalami multikolinieritas jika nilai  $VIF < 10$ . Berikut adalah hasil perhitungan VIF untuk variabel yang digunakan.

**Tabel 3** Hasil Uji Multikolinieritas

Model Regresi	Variabel Bebas	VIF
X1 ~ X2 + X3	X2	2,296
X1 ~ X2 + X3	X3	2,296
X2 ~ X1 + X3	X1	2,006
X2 ~ X1 + X3	X3	2,006
X3 ~ X1 + X2	X1	2,834
X3 ~ X1 + X2	X2	2,834

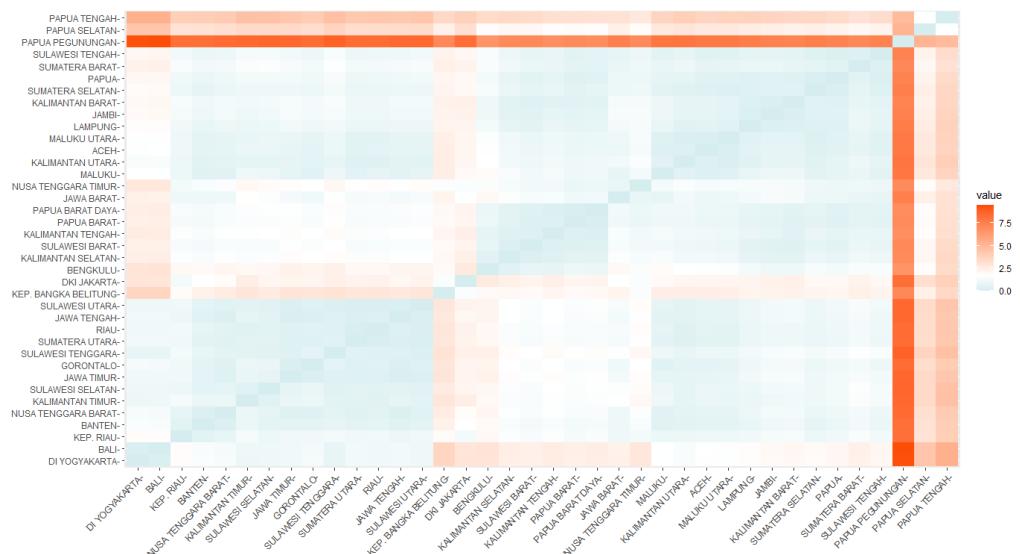
Berdasarkan Tabel 3, seluruh nilai VIF berada di bawah 10, sehingga dapat disimpulkan bahwa tidak terjadi multikolinieritas dalam model ini.

### 3.4 Analisis *K-Medoids*

Sebelum melakukan analisis *K-Medoids*, dilakukan standarisasi untuk memastikan bahwa semua variabel memiliki skala yang sama, sehingga tidak ada variabel yang mendominasi dalam analisis klaster. Dalam penelitian ini, data distandarisasi menggunakan metode *z-score normalization* agar memiliki rata-rata 0 dan standar deviasi 1.

#### 3.2.1. Jarak Antar Observasi

Jarak antar observasi dihitung untuk mengetahui sejauh mana kesamaan atau perbedaan antar individu dalam dataset. Visualisasi jarak menggunakan *heatmap* memberikan gambaran tentang pola kedekatan antar observasi sebelum dilakukan pengelompokan. Berikut adalah tampilan *heatmap* dari jarak antar observasi atau provinsi.



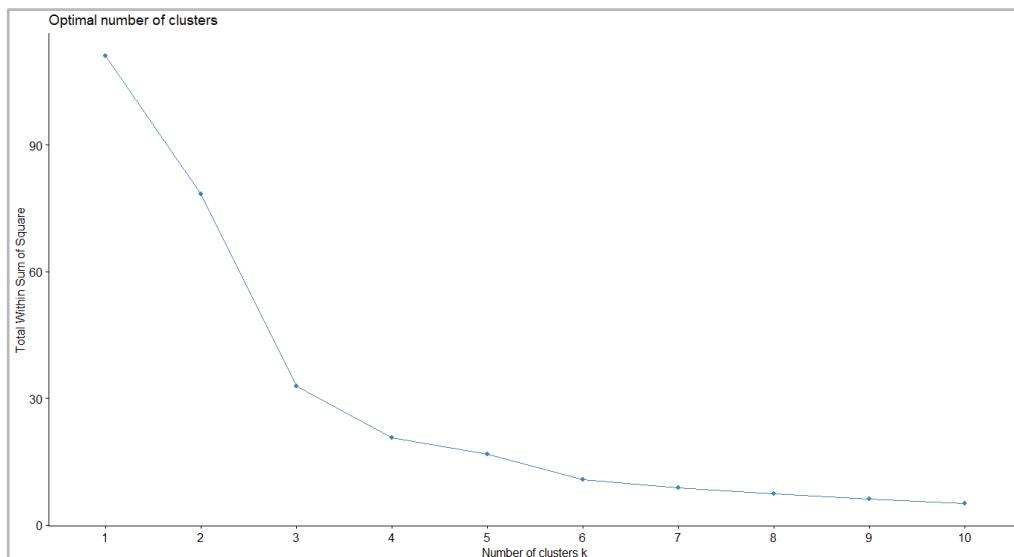
Gambar 3. Heatmap Jarak

Berdasarkan Gambar 3, *heatmap* tersebut menggambarkan jarak kemiskinan antar provinsi di Indonesia berdasarkan nilai tertentu yang ditampilkan dengan gradasi warna, di mana jarak yang kecil atau tingkat kemiskinan yang mirip ditandai dengan warna biru, sedangkan jarak yang besar atau perbedaan signifikan dalam tingkat kemiskinan ditunjukkan dengan warna merah. Berdasarkan analisis jarak kemiskinan antar provinsi di Indonesia, terlihat bahwa provinsi-provinsi di wilayah Papua, seperti Papua Tengah, Papua Selatan, dan Papua Pegunungan, memiliki tingkat kemiskinan yang mirip satu sama lain, ditunjukkan oleh nilai hubungan yang tinggi. Namun, provinsi-provinsi ini menunjukkan perbedaan yang signifikan dengan provinsi di luar wilayah timur, seperti Jawa dan Bali, yang memiliki tingkat kemiskinan lebih rendah. Provinsi-provinsi di Pulau Jawa, Sumatera, dan Kalimantan cenderung memiliki jarak kemiskinan yang lebih kecil, menandakan tingkat kemiskinan yang lebih seragam. Sementara itu, Bali dan DI Yogyakarta tampaknya memiliki pola kemiskinan yang berbeda dari provinsi lain, yang mungkin mencerminkan kondisi ekonomi yang lebih baik atau karakteristik unik wilayah tersebut. Secara keseluruhan, hasil ini menunjukkan adanya kesenjangan kemiskinan yang besar di Indonesia, khususnya antara wilayah timur dan barat, sehingga kebijakan pengentasan kemiskinan perlu disesuaikan dengan kebutuhan dan karakteristik tiap wilayah.

### 3.2.2. Pembentukan Cluster

Untuk menentukan jumlah klaster yang optimal, digunakan metode *Elbow* dengan pendekatan *Within-Cluster Sum of Squares* (WSS). Metode ini membantu

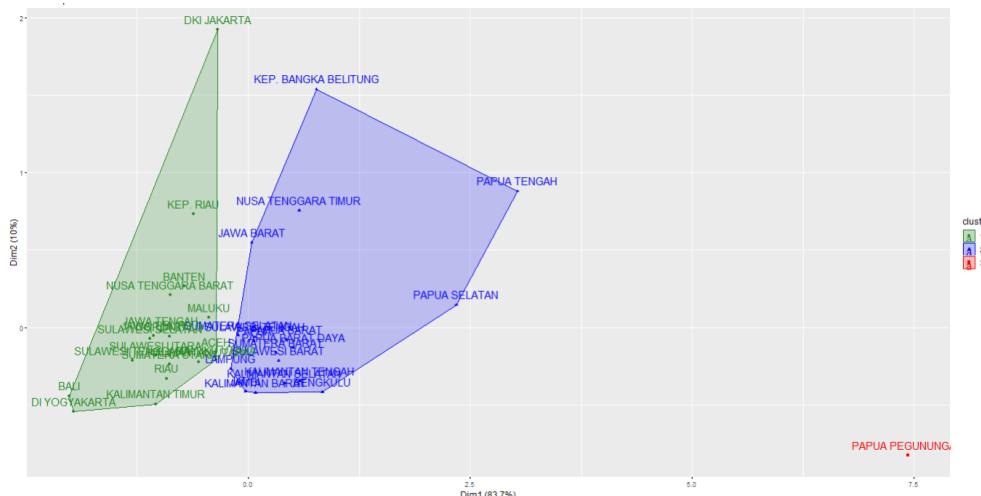
menemukan titik optimal di mana penambahan jumlah klaster tidak lagi memberikan pengurangan yang signifikan dalam variabilitas dalam klaster.



**Gambar 4.** Penentuan  $k$  Optimal Metode *Elbow*

Berdasarkan Gambar 4, hasil visualisasi metode *Elbow* menunjukkan bahwa nilai WSS mengalami penurunan yang tajam dari  $k = 1$  hingga  $k = 3$ . Setelah  $k = 3$ , penurunan nilai WSS menjadi lebih landai, menunjukkan bahwa penambahan jumlah klaster tidak memberikan peningkatan signifikan dalam pemisahan data. Titik di mana kurva mulai melandai ini disebut sebagai *Elbow point* dan digunakan sebagai dasar untuk menentukan jumlah klaster optimal. Berdasarkan hasil ini, dipilih  $k = 3$  sebagai jumlah klaster yang optimal. Dengan jumlah klaster ini, data dapat dikelompokkan tanpa mengalami overfitting atau kehilangan informasi penting yang diperlukan dalam analisis.

Setelah menentukan jumlah klaster optimal sebanyak tiga klaster, analisis klasterisasi dilakukan menggunakan metode *K-Medoids*. Metode ini dipilih karena lebih tahan terhadap *outlier* dibandingkan *K-Means*, serta menggunakan medoid sebagai pusat klaster. Proses klasterisasi dilakukan dengan menggunakan jarak Euclidean untuk mengelompokkan data berdasarkan kemiripan karakteristik antar wilayah. Hasil klaster divisualisasikan untuk melihat pola pengelompokan yang terbentuk. Visualisasi ini memberikan gambaran distribusi masing-masing klaster berdasarkan karakteristik yang digunakan dalam analisis. Berikut adalah hasil dari visualisasi klaster dengan metode *K-Medoids*.



**Gambar 5. Cluster Plot**

Hasil klasterisasi pada Gambar 5, menunjukkan distribusi provinsi di Indonesia berdasarkan klaster yang terbentuk. Setiap warna dalam grafik merepresentasikan kelompok yang berbeda, dengan wilayah yang memiliki karakteristik serupa tergabung dalam klaster yang sama. Dari hasil visualisasi, terlihat bahwa data terbagi menjadi tiga klaster utama. Klaster 1 (hijau) adalah provinsi dengan tingkat kemiskinan rendah seperti DKI Jakarta dan Kepulauan Riau. Klaster 2 (biru) memiliki cakupan lebih luas dengan provinsi seperti Jawa Barat dan Papua Tengah, yang menunjukkan variasi karakteristik yang lebih besar dengan tingkat kemiskinan sedang. Sementara itu, klaster 3 (merah) hanya terdiri dari Papua Pegunungan, yang memiliki kemiskinan paling tinggi dibandingkan klaster lainnya.

### 3.2.3. Hasil Pengelompokan

Setelah dilakukan proses klasterisasi menggunakan metode *K-Medoids*, provinsi di Indonesia berhasil dikelompokkan ke dalam tiga klaster berdasarkan kesamaan karakteristiknya. Berdasarkan Gambar 6, dapat dilihat pada peta bahwa setiap klaster terdiri atas sejumlah anggota provinsi. Klaster pertama beranggotakan 19 provinsi, klaster kedua terdiri atas 18 provinsi, sedangkan klaster ketiga memiliki 1 provinsi sebagai anggotanya. Rincian persebaran ersebaran provinsi pada masing-masing klaster adalah sebagai berikut.



**Gambar 6. Peta Cluster Provinsi di Indonesia Berdasarkan Indikator Kemiskinan**

Daerah yang berwarna hijau merupakan klaster 1, yang menunjukkan provinsi dengan tingkat kemiskinan yang rendah. Daerah yang berwarna biru merupakan klaster 2, yang menunjukkan tingkat kemiskinan yang sedang. Sementara itu, daerah yang berwarna merah merupakan klaster 3, yang menunjukkan tingkat kemiskinan yang tinggi.

### 3.2.4. Profilisasi

Setelah dilakukan klasterisasi berdasarkan tingkat kemiskinan, masing-masing klaster dianalisis lebih lanjut dengan melihat karakteristiknya melalui tiga indikator utama, yaitu persentase rumah tangga yang memiliki akses terhadap sumber air minum layak (X1), persentase rumah tangga yang memiliki akses terhadap sanitasi layak (X2), dan persentase rumah tangga yang memiliki akses terhadap hunian yang layak (X3). Hasil profilisasi dapat dilihat pada Tabel 5 berikut.

**Tabel 4** Profilisasi

Tingkat Kemiskinan	X1	X2	X3
Klaster 1	94,03	88,07	70,43
Klaster 2	82,71	77,64	55,59
Klaster 3	30,64	12,61	4,44

Berdasarkan Tabel 4, profilisasi tiga klaster berdasarkan tiga komponen (X1, X2, dan X3) sesuai dengan tingkatannya, yaitu kemiskinan tinggi, kemiskinan sedang, dan kemiskinan rendah. Klaster 1 menunjukkan provinsi dengan tingkat kemiskinan yang rendah, tercermin dari tingginya akses terhadap kebutuhan dasar seperti air minum layak (94,03%), sanitasi layak (88,07%), dan hunian layak (70,43%). Akses yang tinggi ini menunjukkan bahwa provinsi-provinsi dalam klaster ini memiliki infrastruktur yang memadai, sehingga penduduknya cenderung memiliki kualitas hidup yang lebih baik dan lebih sedikit yang tergolong dalam kategori miskin.

Sementara itu, klaster 2 menunjukkan tingkat kemiskinan yang sedang. Meskipun akses terhadap air minum layak (82,71%), sanitasi layak (77,64%), dan hunian layak (55,59%) masih cukup baik, namun beberapa provinsi di klaster ini masih menghadapi keterbatasan dalam hal akses layanan dasar. Provinsi seperti Jawa Barat, Sumatera Selatan, Lampung, dan Papua Barat memerlukan perhatian lebih dalam pengembangan infrastruktur dasar. Upaya untuk meningkatkan akses terhadap kebutuhan tersebut di provinsi-provinsi ini akan membantu menurunkan tingkat kemiskinan yang lebih tinggi dibandingkan dengan klaster 1.

Klaster 3 menunjukkan tingkat kemiskinan yang tinggi, dengan hanya 30,64% akses terhadap air minum layak, 12,61% untuk sanitasi layak, dan 4,44% untuk hunian layak. Kondisi ini mencerminkan keterbatasan ekstrem dalam kebutuhan dasar, yang berhubungan langsung dengan tingkat kemiskinan yang sangat tinggi. Provinsi dalam klaster ini, yaitu Papua Pegunungan. Oleh karena itu, provinsi dalam klaster 3 membutuhkan perhatian khusus dalam bentuk kebijakan yang berfokus pada penyediaan infrastruktur dasar untuk mengurangi tingkat kemiskinan yang tinggi.

#### 4. Kesimpulan

Penelitian ini berhasil mengelompokkan 38 provinsi di Indonesia ke dalam tiga klaster berdasarkan tingkat kemiskinan menggunakan metode *K-Medoids*. Analisis mempertimbangkan tiga indikator utama yaitu akses terhadap air minum layak, sanitasi layak, dan hunian layak. Klaster 1 mencakup 19 provinsi dengan tingkat kemiskinan rendah, ditandai dengan akses yang tinggi terhadap kebutuhan dasar. Klaster 2 terdiri dari 18 provinsi dengan tingkat kemiskinan sedang, di mana akses terhadap kebutuhan dasar cukup baik, namun memerlukan peningkatan infrastruktur untuk mengurangi kesenjangan. Sementara itu, klaster 3 hanya terdiri dari satu provinsi, yaitu Papua Pegunungan, yang memiliki tingkat kemiskinan sangat tinggi dan akses yang sangat terbatas terhadap kebutuhan dasar.

Hasil penelitian ini menunjukkan adanya kesenjangan signifikan antara wilayah timur dan barat Indonesia, di mana wilayah timur, khususnya Papua, memiliki tingkat kemiskinan yang jauh lebih tinggi dibandingkan wilayah barat dilihat dari akses terhadap air bersih, sanitasi layak, dan hunian layak. Persentase akses air bersih, sanitasi layak, dan hunian layak yang tinggi merepresentasikan tingkat kemiskinan yang rendah, sedangkan persentase akses yang rendah mencerminkan tingkat kemiskinan yang tinggi. Kesimpulan ini menggaris bawahi perlunya pendekatan berbasis wilayah dalam kebijakan pengentasan kemiskinan. Program intensif diperlukan untuk wilayah dengan tingkat kemiskinan tinggi seperti klaster 3, sementara wilayah dengan tingkat kemiskinan rendah dan sedang (klaster 1 dan 2) dapat difokuskan pada peningkatan keberlanjutan akses terhadap air minum layak, sanitasi layak, hunian layak, dan program preventif. Penelitian ini diharapkan dapat menjadi acuan bagi pemerintah dalam merancang kebijakan pengentasan kemiskinan yang lebih efektif dan terfokus.

#### 5. Daftar Pustaka

- [1] I. F. Budiman, "Rekomendasi Kebijakan Perpajakan untuk Mendukung Pertumbuhan Ekonomi Berkualitas untuk Mewujudkan Pembangunan Ekonomi yang Inklusif di Indonesia," *Jurnal Ilmiah Neraca Ekonomi Bisnis Manajemen Akuntansi*, vol. 3, no. 1, pp. 590–600, Sep. 2024.
- [2] Badan Pusat Statistik. (2025, Mei 6). *Persentase Penduduk Miskin (P0) Menurut Provinsi dan Daerah (Persen)* [Online]. Tersedia: <https://www.bps.go.id/id/statistics-table/2/MTkyIzI=/persentase-penduduk-miskin--p0--menurut-provinsi-dan-daerah--persen-.html>.
- [3] Anisa, S. N., Aulia, S., Indah, A., Arif, M., & Panorama, M. (2024). Analisis Peran Infrastruktur Dalam Pertumbuhan Ekonomi Pembangunan Di Kota Palembang. *Jurnal Publikasi Ekonomi dan Akuntansi (JUPEA)*, vol. 4, no. 1, pp. 36–54, Januari 2024, doi: <https://doi.org/10.51903/jupea.v4i1.2435>.
- [4] Badan Pusat Statistik. (2025, Januari 15). *Persentase Penduduk Miskin September 2024 Turun Menjadi 8,57 Persen* [Online]. Tersedia: <https://www.bps.go.id/id/pressrelease/2025/01/15/2401/persentase-penduduk-miskin-september-2024-turun-menjadi-8-57-persen-.html>.
- [5] S. Tuhipatussania, S. Erniwati, and Z. Mutaqin, "Perbandingan metode agglomerative hierarchical clustering dan metode K-Medoids dalam pengelompokan data titik panas kebakaran hutan di Indonesia," *Journal Computer and Technology*, vol. 2, no. 1, pp. 21–38, 2024, doi: <https://doi.org/10.69916/comtechno.v2i1.146>.
- [6] Qomariyah and M. U. Siregar, "Comparative Study of K-Means Clustering Algorithm and K-Medoids Clustering in Student Data Clustering," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 7, no. 2, pp. 91–99, May 2022, doi: <https://doi.org/10.14421/jiska.2022.7.2.91-99>.
- [7] M. P. A. Ariawan, N. P. Sastra and I. M. Sudarma, "Clustering Data Remunerasi PNS Menggunakan Metode," *Majalah Ilmiah Teknologi Elektro*, vol. 19, no. 1, pp. 33-40, 2020.

- [8] N. T. Luchia, H. Handayani, F. S. Hamdi, D. Erlangga, and S. F. Octavia, "Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 2, pp. 35–41, Sep. 2022, doi: <https://doi.org/10.57152/malcom.v2i2.422>.
- [9] M. Yafi, R. Goejantoro, and A. Trii, "Pengelompokan Algoritma K-Medoids Dengan Principal Component Analysis (PCA) (Studi Kasus: Kabupaten/Kota di Pulau Kalimantan Berdasarkan Indikator Kemiskinan Tahun 2021)," *Prosiding Seminar Nasional Matematika dan Statistika*, vol. 3, no. 01, pp. 183–195, 2021.
- [10] F. Zahra, A. Khalif, and B. N. Sari, "Pengelompokan Tingkat Kemiskinan di Setiap Provinsi di Indonesia Menggunakan Algoritma K-Medoids," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, 2024.
- [11] H. Atmaja and K. Mahalli, "Pengaruh Peningkatan Infrastruktur terhadap Pertumbuhan Ekonomi di Kota Sibolga," *unpublished*, 2015.
- [12] L. D. Martias, "Statistika deskriptif sebagai kumpulan informasi," *FIHRIS: Jurnal Ilmu Perpustakaan dan Informasi*, vol. 16, no. 1, pp. 40–59, 2021, doi: <https://doi.org/10.14421/fhrs.2021.161.40-59>.
- [13] M. Sriningsih, D. Hatidja, and J. D. Prang, "Penanganan multikolinearitas dengan menggunakan analisis regresi komponen utama pada kasus impor beras di Provinsi Sulut," *Jurnal Ilmiah Sains*, vol. 18, no. 1, p. 18, Jul. 2018. [Online]. Available: <https://doi.org/10.35799/jis.18.1.2018.19396>
- [14] A. F. Hidayatullah, D. Saputra, F. Inarah, I. Evita, M. Fadillah, and L. Harsyiah, "Analisis regresi komponen utama untuk mengatasi multikolinearitas pada faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia," *JSN Jurnal Sains Natural*, vol. 2, no. 1, pp. 19–24, Feb. 2024, doi: <https://doi.org/10.35746/jsn.v2i1.497>.
- [15] B. Wira, A. E. Budianto, and A. S. Wiguna, "Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang," *RAINSTEK: Jurnal Terapan Sains & Teknologi*, vol. 1, no. 3, pp. 53–68, Sep. 2019, doi: <https://doi.org/10.21067/jtst.v1i3.3046>.
- [16] N. Syahfitri, E. Budianita, A. Nazir, and I. Afrianty, "Pengelompokan Produk Berdasarkan Data Persediaan Barang Menggunakan Metode Elbow dan K-Medoid," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 3, pp. 1668–1675, Dec. 2023, doi: <https://doi.org/10.30865/klik.v4i3.1525>.
- [17] L. Kaufman, and P. J. Rousseeuw, *Partitioning Around Medoids (Program PAM)*, Wiley Series in Probability and Statistics, Hoboken, NJ, USA: John Wiley & Sons, Inc., 1990.
- [18] I. J. Panjaitan, T. Wulandari, and B. Susetyo, "Perbandingan Hasil Analisis Cluster K-Means dan K-Medoids untuk Pemetaan Mutu SMK," *Jurnal Bayesian: Jurnal Ilmiah Statistika dan Ekonometrika*, vol. 4, no. 1, pp. 63–75, 2024, doi: <https://doi.org/10.46306/bay.v4i1.77>.