PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN INDIKATOR KEMISKINAN MENGGUNAKAN METODE K-MEANS DAN K-MEDOIDS

Saskia Kayla Zahfarina[1*], Emeylia Safitri [2]
[1,2]Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas Terbuka, Tangerang Selatan, Indonesia

**Penulis korespondensi: saskiakayla07@gmail.com*

ABSTRAK

Kemiskinan merupakan isu kompleks yang tetap menjadi tantangan utama bagi pembangunan di Indonesia, sehingga pemetaan profil kemiskinan yang akurat di tingkat provinsi sangat diperlukan untuk mendukung perumusan kebijakan yang tepat sasaran. Penelitian ini bertujuan untuk membandingkan kinerja metode *K-Means* dan *K-Medoids* dalam mengelompokkan 38 provinsi di Indonesia berdasarkan lima indikator utama kemiskinan yang dirilis oleh Badan Pusat Statistik (BPS) per Maret 2024, meliputi garis kemiskinan, persentase penduduk miskin (p0), indeks kedalaman kemiskinan (p1), indeks keparahan kemiskinan (p2), dan pengeluaran per kapita. Analisis data dimulai dengan normalisasi data, sedangkan penentuan jumlah klaster optimal dilakukan melalui *Elbow Method* yang menghasilkan nilai k=3. Kualitas klaster diuji menggunakan *Davies-Bouldin Index* (DBI) dan *silhouette index*. Hasil penelitian menunjukkan bahwa metode *K-Means* membentuk tiga klaster kemiskinan yang terdiri dari klaster rendah (12 provinsi), sedang (20 provinsi), dan tinggi (6 provinsi) dengan nilai DBI sebesar 0,7448 dan *silhouette score* sebesar 0,4290. Di sisi lain, metode *K-Medoids* menghasilkan pengelompokan masing-masing sebanyak 13, 17, dan 8 provinsi dengan nilai DBI sebesar 0,7884 dan *Silhouette score* sebesar 0,3948. Meskipun tingkat keselarasan antara kedua metode mencapai 92,1% (35 dari 38 provinsi), metode *K-Means* terbukti menghasilkan kualitas klaster yang sedikit lebih baik. Selain itu, seluruh provinsi di wilayah Papua secara konsisten berada dalam klaster kemiskinan tertinggi, yang menegaskan perlunya perhatian lebih di wilayah Indonesia bagian timur.

Kata kunci: K-Means, K-Medoids, Clustering, Kemiskinan, Provinsi.

1 PENDAHULUAN

Kemiskinan merupakan salah satu permasalahan sosial dan ekonomi paling mendasar yang secara konsisten menjadi prioritas utama dalam agenda pembangunan nasional Indonesia (BPS, 2023). Fenomena kemiskinan tidak hanya direpresentasikan oleh rendahnya tingkat pendapatan, melainkan juga oleh keterbatasan akses terhadap pendidikan yang berkualitas, layanan kesehatan yang memadai, serta infrastruktur dasar yang layak (Rahmawati & Hadi, 2022). Dalam perspektif yang lebih luas, kemiskinan dipandang sebagai isu multidimensional yang memerlukan pendekatan holistik dan terintegrasi dalam penanganannya (Wulandari et al., 2023). Kondisi ini kian diperparah oleh adanya ketimpangan struktural yang bersumber dari perbedaan kapasitas sumber daya manusia serta ketidakmerataan akses terhadap hasil pembangunan antarwilayah (Putri & Ardiansyah, 2022). Oleh karena itu, pemahaman yang mendalam mengenai karakteristik kemiskinan pada tingkat regional menjadi prasyarat fundamental dalam merumuskan kebijakan penanggulangan kemiskinan yang efektif dan tepat sasaran (Ningrum et al., 2023). Berdasarkan data Survei Sosial Ekonomi Nasional (Susenas) per Maret 2024, tingkat kemiskinan nasional tercatat menurun menjadi 9,03 persen dengan jumlah penduduk miskin sebanyak 25,22 million jiwa, yang sekaligus melanjutkan tren

penurunan sejak September 2022 (BPS, 2024). Perbaikan indikator ini didorong oleh kuatnya aktivitas ekonomi domestik serta implementasi program bantuan sosial pemerintah yang dilakukan secara sistematis. Meskipun demikian, pencapaian tersebut masih jauh dari kondisi ideal, terutama jika mempertimbangkan adanya disparitas kemiskinan yang tajam antarwilayah yang memerlukan perhatian kebijakan yang lebih terdiferensiasi.

Heterogenitas kemiskinan antarprovinsi di Indonesia terlihat sangat nyata berdasarkan data terbaru, di mana pada Maret 2024, Provinsi Papua Pegunungan mencatat persentase penduduk miskin tertinggi sebesar 32,97 persen, diikuti oleh Papua Tengah sebesar 29,76 persen (BPS Papua, 2024). Sebaliknya, Provinsi Bali mencatat tingkat kemiskinan yang jauh lebih rendah, yaitu sebesar 3,83 persen. Kesenjangan yang mencapai hampir 30 persen ini menegaskan perlunya kebijakan yang terdiferensiasi di setiap wilayah berdasarkan karakteristik dan profil kemiskinan spesifik yang mendasarinya (Tangke & Dzikrullah, 2023). Pengelompokan wilayah berdasarkan kesamaan karakteristik kemiskinan terbukti dapat meningkatkan efektivitas alokasi anggaran serta ketepatan sasaran program pembangunan (Maulana & Kurniawan, 2023). Lebih lanjut, analisis kemiskinan yang komprehensif tidak dapat hanya bertumpu pada persentase penduduk miskin (p0). Indeks kedalaman kemiskinan (p1) yang mengukur rata-rata kesenjangan pengeluaran penduduk miskin terhadap garis kemiskinan, serta indeks keparahan kemiskinan (p2) yang menggambarkan ketimpangan pengeluaran di antara penduduk miskin, memiliki peran yang sama pentingnya dalam memberikan gambaran yang utuh mengenai realitas kemiskinan di suatu daerah. Oleh karena itu, integrasi ketiga indikator tersebut mutlak diperlukan untuk memperoleh pemahaman yang lebih komprehensif mengenai dimensi kemiskinan di tingkat provinsi.

Analisis klister menggunakan pendekatan berbasis data yang objektif untuk mengelompokkan wilayah berdasarkan kesamaan karakteristik kemiskinannya (Luchia et al., 2022). Teknik ini memungkinkan perancangan kebijakan yang lebih akurat karena mempertimbangkan kesamaan profil kemiskinan antarwilayah secara sistematis (Mayasari & Nugraha, 2023). Penerapan analisis klaster dalam studi kemiskinan regional terbukti mampu menghasilkan pengelompokan yang lebih informatif, sehingga menjadi dasar perencanaan pembangunan yang lebih baik dibandingkan metode klasifikasi(Prabowo & Santoso, 2022). Di antara berbagai metode klasterisasi yang tersedia, *K-Means* dan *K-Medoids* merupakan algoritma partisi yang paling sering digunakan dalam penelitian sosio-ekonomi. *K-Means* bekerja dengan meminimalkan jarak antara titik data dan pusat klaster berupa nilai rata-rata (*centroid*), sedangkan *K-Medoids* menggunakan medoid yaitu titik data aktual yang paling representatif sebagai pusat klaster (Sinaga & Yang, 2020). Penggunaan *medoid* ini membuat *K-Medoids* lebih resisten terhadap nilai-nilai ekstrem (*outliers*) yang umumnya ditemukan dalam distribusi data kemiskinan antarwilayah (Amrulloh et al., 2022). Karakteristik tersebut menjadi sangat relevan dalam konteks data kemiskinan di Indonesia yang memiliki tingkat ketimpangan persebaran sangat tinggi (Tawakal et al., 2025).

Sejumlah penelitian terdahulu menunjukkan relevansi dan efektivitas kedua metode ini dalam konteks analisis kemiskinan. Luchia et al. (2022) dalam kajiannya membandingkan kinerja *K-Means* dan *K-Medoids* pada data kemiskinan di Indonesia dan menemukan bahwa *K-Medoids* menghasilkan pengelompokan yang lebih stabil secara statistik. Hidayat et al. (2023) mengimplementasikan *K-Medoids* untuk pengelompokan kabupaten di Provinsi Aceh berdasarkan faktor-faktor determinan kemiskinan dan menghasilkan klaster yang bermakna serta dapat diinterpretasikan secara kebijakan. Fauzi (2024) secara eksplisit membandingkan kedua algoritma pada data kemiskinan kabupaten/kota di Indonesia dan menyimpulkan bahwa pemilihan metode *clustering* berpengaruh secara signifikan terhadap kualitas dan interpretabilitas hasil pengelompokan. Dalam hal validasi kualitas klaster, *Davies-Bouldin Index* (DBI) dan *Silhouette Index* merupakan dua indeks yang paling sering digunakan. DBI mengukur rasio dispersi dalam klaster terhadap jarak antar *centroid* klaster, sementara

Silhouette Index mengukur derajat kesesuaian penempatan setiap titik data dalam klasternya masing-masing. Hasan (2024) dan Amrulloh et al. (2022) menegaskan bahwa penggunaan kedua indeks secara bersamaan memberikan evaluasi yang lebih komprehensif dan reliabel dibandingkan dengan mengandalkan satu indeks tunggal. Pendekatan evaluasi ganda ini secara khusus direkomendasikan untuk data indicator kemiskinan yang memiliki distribusi tidak homogen dan berpotensi mengandung nilai pencilan (Astuti et al., 2026). Oleh karena itu, penelitian ini akan membandingkan metode *K-Means* dan *K-Medoids* dalam mengelompokkan provinsi di Indonesia berdasarkan indikator kemiskinan.

2 METODE

2.1 Sumber Data

Penelitian ini menggunakan data sekunder dari Berita Resmi Statistik BPS No. 56/07/Th. XXVII tanggal 1 Juli 2024 tentang Profil Kemiskinan di Indonesia Maret 2024 (Badan Pusat Statistik, 2024). Data mencakup 38 provinsi termasuk empat provinsi pemekaran Papua (Papua Pegunungan, Papua Tengah, Papua Selatan, Papua Barat Daya).

Tabel 1. Variabel Penelitian

Variabel	Keterangan
X1 - P0 (%)	Persentase Penduduk Miskin terhadap total penduduk
X2 - Garis Kemiskinan	Nilai pengeluaran minimum kebutuhan dasar (Rp/kap/bln)
X3 - P1	Indeks Kedalaman Kemiskinan
X4 - P2	Indeks Keparahan Kemiskinan
X5 - Pengeluaran/kap	Rata-rata pengeluaran per kapita per bulan (Rp ribu)

2.2 Analisis Klaster

Analisis klaster atau pengelompokan merupakan suatu teknik analisis data *unsupervised learning* yang bertujuan untuk mempartisi sekumpulan objek ke dalam beberapa kelompok, sedemikian rupa sehingga objek-objek dalam klaster yang sama memiliki tingkat kemiripan yang lebih tinggi dibandingkan dengan objek di klister lain. Tingkat kemiripan antarobjek tersebut umumnya diukur menggunakan *euclidean distance* atau *manhattan distance* (Sinaga & Yang, 2020). Metode klasterisasi telah terbukti efektif sebagai alat analisis dalam berbagai studi regional berbasis data, khususnya untuk mengidentifikasi pola dan struktur tersembunyi dalam distribusi indikator sosial dan ekonomi (Bahauddin, Fatmawati, & Sari, 2021). Keunggulan utama dari pendekatan ini adalah kemampuannya untuk menghasilkan pengelompokan secara objektif dan berbasis data tanpa memerlukan asumsi awal mengenai jumlah atau komposisi kelompok tersebut (Luchia et al., 2022). Dalam konteks analisis data kemiskinan, penerapan analisis klaster membantu mengidentifikasi provinsi-provinsi yang memiliki kemiripan profil kemiskinan secara otomatis tanpa membutuhkan label kategori yang ditentukan sebelumnya. Pendekatan ini sangat bermanfaat ketika peneliti tidak memiliki pengetahuan awal yang cukup mengenai struktur pengelompokan data tersebut (Luchia et al., 2022; Bahauddin, Fatmawati, & Sari, 2021). Lebih lanjut, hasil pengelompokan berbasis klaster ini dapat digunakan sebagai landasan empiris untuk menyusun indeks kemiskinan komposit yang lebih representatif pada tingkat regional (Mayasari & Nugraha, 2023).

2.3 Algoritma K-Means Clustering

K-Means merupakan salah satu algoritma *clustering* yang paling banyak digunakan dalam berbagai bidang aplikasi, termasuk analisis data sosial dan ekonomi. Algoritma ini pertama kali diperkenalkan oleh MacQueen pada tahun 1967 dan selanjutnya dikembangkan

oleh berbagai peneliti dari lintas disiplin ilmu. *K-Means* bekerja dengan meminimalkan nilai *Within-Cluster Sum of Squares* (WCSS), yaitu jumlah kuadrat jarak dari setiap titik data ke *centroid* klaster yang menjadi tempat penalannya (Sinaga & Yang, 2020). Keunggulan utama algoritma ini terletak pada efisiensi komputasinya yang tinggi, menjadikannya pilihan yang relevan untuk analisis dataset skala menengah seperti data kemiskinan antarprovinsi (Fauzi, 2024). Rumus jarak *euclidean* yang digunakan dalam *K-Means* adalah:

$$d(x, c) = \sqrt{[\Sigma(xi - ci)^2]} \quad (1)$$

Keterangan: d = jarak *euclidean*; x = vector data; c = vektor *centroid* klaster; i = indeks variabel.

Kelebihan utama *K-Means* adalah kemudahan implementasinya, menjadikannya algoritma yang banyak dipilih dalam penelitian terapan. Meskipun demikian, *K-Means* memiliki sejumlah keterbatasan yang perlu diperhatikan, di antaranya: sensitivitas terhadap nilai pencilan (*outlier*) akibat penggunaan rata-rata sebagai *centroid*, ketergantungan hasil pada inisialisasi *centroid* awal, serta asumsi bahwa klaster berbentuk sferis dengan ukuran yang relatif seragam (Fauzi, 2024; Tawakal et al., 2025). Keterbatasan ini secara khusus relevan dalam analisis data kemiskinan yang seringkali memiliki distribusi yang sangat tidak merata dan mengandung nilai ekstrem (Sinaga & Yang, 2020).

2.4 Algoritma *K-Medoids Clustering* (PAM)

K-Medoids merupakan varian dari *K-Means* yang menggunakan medoid yaitu titik data aktual yang paling representatif dalam suatu klister sebagai pusat klaster, sebagai alternatif dari penggunaan nilai rata-rata pada *K-Means*. Algoritma *K-Medoids* yang paling dikenal dan banyak diterapkan adalah *Partitioning Around Medoids* (PAM), yang pertama kali diperkenalkan oleh Kaufman dan Rousseeuw pada tahun 1990 (Hidayat et al., 2023). Penggunaan titik data aktual sebagai *medoid* membuat algoritma ini lebih tahan terhadap gangguan nilai pencilan yang kerap dijumpai dalam data empiris kemiskinan.

Keunggulan utama *K-Medoids* dibandingkan *K-Means* terletak pada sifat *robust* terhadap *outlier* pada data. Karena medoid merupakan titik data aktual yang ada dalam dataset, satu nilai pencilan yang ekstrem tidak akan menggeser pusat klaster secara dramatis sebagaimana yang dapat terjadi pada *centroid K-Means*. Di samping itu, *medoid* memiliki tingkat interpretabilitas yang lebih tinggi karena merepresentasikan objek nyata dari dataset, sehingga memudahkan diinterpretasikan (Amrulloh et al., 2022; Tangke & Dzikrullah, 2023). Keunggulan interpretatif ini menjadikan *K-Medoids* sangat sesuai untuk analisis kemiskinan yang berorientasi pada implikasi kebijakan (Luchia et al., 2022).

Kelemahan utama *K-Medoids* dibandingkan *K-Means* terletak pada kompleksitas komputasinya yang lebih tinggi, sehingga algoritma ini cenderung kurang efisien apabila diterapkan pada dataset dengan jumlah observasi yang sangat besar. Namun demikian, untuk dataset berukuran sedang seperti data 38 provinsi Indonesia, kompleksitas ini tidak menjadi kendala berarti dalam praktiknya (Hidayat et al., 2023).

2.5 Normalisasi Data (*Z-Score Standardization*)

Normalisasi data merupakan tahap pra-pemrosesan yang sangat krusial dalam analisis *clustering*, terutama ketika variabel-variabel yang dianalisis memiliki skala pengukuran dan satuan yang berbeda-beda. Tanpa normalisasi, variabel dengan skala yang jauh lebih besar akan mendominasi perhitungan jarak *euclidean* dan memperkecil kontribusi variabel-variabel lain dalam proses pembentukan klaster (Fauzi, 2024). Kondisi ini sangat relevan dalam penelitian ini mengingat variabel penelitian mencakup besaran yang sangat beragam, mulai dari persentase P0 hingga nilai garis kemiskinan dalam satuan rupiah per kapita per bulan. Standardisasi *z-score* dipilih sebagai metode normalisasi karena mampu mempertahankan

struktur distribusi data asli sekaligus menyetarakan kontribusi setiap variabel dalam analisis (Astuti et al., 2026).

Metode normalisasi yang digunakan dalam penelitian ini adalah Z-Score Standardization, dengan rumus:

$$Z = (X - \mu) / \sigma \quad (2)$$

Keterangan: Z = nilai terstandarisasi; X = nilai observasi; μ = rata-rata variabel; σ = standar deviasi variabel.
Hasil normalisasi Z-Score menghasilkan distribusi dengan mean=0 dan standar deviasi=1 untuk setiap variabel, sehingga seluruh variabel memiliki bobot kontribusi yang setara dalam perhitungan jarak Euclidean.

2.6 *Elbow Method* untuk Penentuan Jumlah Klaster Optimal

Elbow Method merupakan teknik yang digunakan secara luas untuk menentukan jumlah klaster optimal (k) dalam analisis clustering. Metode ini mengevaluasi nilai *Within-Cluster Sum of Squares* (WCSS) untuk berbagai nilai k secara berturut-turut, kemudian mengidentifikasi titik siku (elbow) sebagai nilai k di mana penambahan jumlah klaster tidak lagi memberikan penurunan WCSS yang signifikan secara proporsional (Astuti et al., 2026). Penentuan jumlah klaster yang tepat merupakan hal krusial, karena pemilihan k yang tidak sesuai dapat menghasilkan klaster yang terlalu luas atau terlalu sempit sehingga mempengaruhi interpretasi hasil (Sinaga & Yang, 2020). Dalam penelitian dengan data kemiskinan, *Elbow Method* terbukti mampu menghasilkan penentuan k yang stabil dan dapat direplikasi (Bahauddin et al., 2021).
WCSS didefinisikan sebagai:

$$WCSS = \Sigma \Sigma ||x_i - c_k||^2 \quad (3)$$

Keterangan: x*i* = titik data ke-*i;* c_k = centroid klaster ke-k.
Titik elbow diidentifikasi sebagai nilai k di mana laju penurunan WCSS mulai melambat secara signifikan. Penambahan klaster melampaui titik *elbow* tidak lagi proporsional dengan peningkatan homogenitas di dalam klaster yang diperoleh.

2.7 Indeks Validasi Klaster

Kualitas klaster yang terbentuk dievaluasi secara komprehensif menggunakan dua indeks validasi internal, yaitu *Davies-Bouldin Index* (DBI) dan *silhouette index*. Penggunaan dua indeks secara bersamaan direkomendasikan untuk memperoleh gambaran evaluasi yang lebih menyeluruh dan tidak bergantung pada satu perspektif pengukuran saja (Hasan, 2024). *Davies-Bouldin Index* (DBI) mengukur rasio rata-rata dispersi dalam klaster (*intra-cluster scatter*) terhadap jarak antar *centroid* klaster (*inter-cluster separation*). DBI yang lebih kecil mengindikasikan klaster yang lebih valid dan lebih terpisah satu sama lain. DBI=0 merepresentasikan klaster yang sempurna. Rumus DBI:

$$DBI = (1/k) \Sigma \max_{i \neq j} [(s_i + s_j) / d(c_i,c_j)] \quad (4)$$

Keterangan: s_i = rata-rata jarak titik data dalam klaster *i* ke centroidn; d(ci,cj) = jarak antara *centroid* klaster *i* dan *j* (Amrulloh et al., 2022; Hasan, 2024).
Silhouette index mengukur seberapa tepat setiap titik data ditempatkan dalam klasternya dibandingkan dengan klaster terdekat lainnya. Nilai Silhouette berkisar antara −1 hingga +1. Nilai mendekati +1 berarti titik data sangat sesuai dengan klasternya, nilai mendekati 0 berarti titik data berada di perbatasan antara dua klaster, dan nilai negatif menunjukkan titik data mungkin salah klaster. Rumus Silhouette untuk satu titik data:

$$s(i) = [b(i) - a(i)] / \max\{a(i), b(i)\} \quad (5)$$

Keterangan: a(*i*) = rata-rata jarak titik *i* ke semua titik lain dalam klaster yang sama; b(*i*) = rata-rata jarak terkecil dari titik *i* ke klaster terdekat lainnya. Interpretasi *silhouette*: >0,70=struktur

kuat; 0,51-0,70=struktur sedang; 0,26-0,50 = struktur lemah (Hasan, 2024; Amrulloh et al., 2022).

2.8 Tahapan Analisis Data

Penelitian ini dilakukan dengan beberapa tahapan analisis data secara sistematis, sebagaimana diuraikan berikut:

1. Pengumpulan data indikator kemiskinan untuk 38 provinsi di Indonesia .
2. Eksplorasi data melalui analisis statistik deskriptif untuk mengidentifikasi karakteristik pemusatan, penyebaran, dan rentang data tiap variabel.
3. Normalisasi data menggunakan *z-score standardization* agar seluruh variabel memiliki skala dan bobot kontribusi yang setara.
4. Penentuan jumlah klaster optimal (k) menggunakan *Elbow Method* pada berbagai nilai k.
5. Penerapan algoritma *K-Means clustering* dengan k optimal.
6. Penerapan algoritma K-Medoids dengan k optimal.
7. Validasi klaster yang dihasilkan kedua metode menggunakan *Davies-Bouldin Index* (DBI) dan *silhouette index*.
8. Perbandingan hasil pengelompokan kedua metode serta interpretasi hasil.

3 HASIL DAN PEMBAHASAN

3.1 Statistik Deskriptif

Sebelum dilakukan proses *clustering*, dilakukan analisis statistik deskriptif untuk memahami karakteristik distribusi data indikator kemiskinan 38 provinsi Indonesia. Statistik deskriptif ini memberikan gambaran awal mengenai pemusatan, penyebaran, dan rentang data.

Tabel 2. Statistik Deskriptif Variabel Penelitian

Variabel	Minimum	Maksimum	Rata-rata	Median	Std. Deviasi
Garis Kem. (Rp/kap)	379.854	793.804	573.202	549.116	108.703
P0 (%)	3,83	32,97	11,09	10,13	6,77
P1	0,52	8,24	2,01	1,59	1,68
P2	0,1	2,98	0,53	0,36	0,58
Pengeluaran/kap (Rp,ribu)	789	2.456	1.331	1.301	350

Berdasarkan Tabel 2, nilai rata-rata P0 (persentase penduduk miskin) sebesar 11,09% dengan standar deviasi 6,77 mengindikasikan tingkat variabilitas yang tinggi antar provinsi. Rentang P0 yang sangat lebar dari 3,83% (Bali) hingga 32,97% (Papua Pegunungan) menunjukkan adanya heterogenitas ekstrem dalam distribusi kemiskinan antar wilayah di Indonesia. Kondisi ini secara empiris mendukung perlunya pendekatan *clustering* untuk memetakan provinsi dengan profil kemiskinan yang mirip. Indeks P1 dan P2 juga menunjukkan sebaran yang cukup lebar, mengindikasikan bahwa perbedaan antar provinsi tidak hanya pada proporsi penduduk miskin, namun juga pada kedalaman dan keparahan kemiskinannya.

3.2 Penentuan Jumlah Klaster Optimal

Jumlah klaster optimal ditentukan menggunakan *Elbow Method* dengan mengevaluasi nilai *Within-Cluster Sum of Squares* (WCSS) untuk berbagai nilai k (Sinaga & Yang, 2020). Tabel 3 menyajikan nilai WCSS untuk k=2 hingga k=7.

Tabel 3. WCSS

k = 2	k = 3	k = 4	k = 5	k = 6	k = 7
102,0703	59,1762	36,7970	25,1153	20,0029	16,7594
Penurunan 43,1%	Penurunan 42,0%	Penurunan 37,9%	Penurunan 31,7%	Penurunan 20,3%	Penurunan 16,2%

Berdasarkan Tabel 3, penurunan WCSS yang paling signifikan terjadi dari k=2 ke k=3, yakni sebesar 42,0%. Setelah k=3, laju penurunan WCSS mulai melambat secara konsisten. Titik siku (elbow) yang paling jelas tampak pada k=3, sehingga tiga klaster ditetapkan sebagai jumlah klaster optimal.

3.3 Hasil *K-Means Clustering*

Tabel 4, menyajikan nilai rata-rata skala asli untuk setiap klaster yang terbentuk.

Tabel 4. Nilai Rata-rata per Klaster

Klaster	Kategori	Garis Kem. (Rp)	P0 (%)	P1	P2	Pengeluar./kap (Rp rb)	n
1	Kemiskinan Rendah	655.214	5,34	0,76	0,16	1.731	12
2	Kemiskinan Sedang	491.830	11,00	1,85	0,45	1.212	20
3	Kemiskinan Tinggi	680.414	22,87	5,06	1,56	928	6

Klaster 1 (kemiskinan rendah) memiliki rata-rata P0 terendah sebesar 5,34% dengan pengeluaran per kapita tertinggi (Rp1.731.000/bulan), mencerminkan kondisi kesejahteraan relatif lebih baik. Klaster 2 (kemiskinan sedang) merupakan kelompok terbesar (n=20) dengan P0 rata-rata 11,00%, berada di sekitar nilai rata-rata nasional. Klaster 3 (kemiskinan tinggi) menunjukkan P0 rata-rata yang sangat tinggi (22,87%) disertai P1 dan P2 yang jauh di atas klaster lain, mengindikasikan kemiskinan yang tidak hanya luas tetapi juga dalam dan parah, dengan pengeluaran per kapita terendah (Rp928.000/bulan).

Tabel 5. Keanggotaan Provinsi per Klaster Metode K-Means

Klaster	Kategori & Jumlah	Provinsi Anggota
1	Kemiskinan Rendah (n = 12)	Sumatera Utara, Sumatera Barat, Riau, Kep. Bangka Belitung, Kepulauan Riau, DKI Jakarta, Banten, Bali, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara
2	Kemiskinan Sedang (n = 20)	Aceh, Jambi, Sumatera Selatan, Bengkulu, Lampung, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, NTB, NTT, Kalimantan Barat, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara
3	Kemiskinan Tinggi (n = 6)	Papua Barat, Papua Barat Daya, Papua, Papua Selatan, Papua Tengah, Papua Pegunungan

Klaster 1 yang beranggotakan 12 provinsi didominasi oleh provinsi-provinsi di wilayah Kalimantan, Kepulauan Riau, dan Pulau Jawa bagian utara-barat yang memiliki basis ekonomi kuat. Klaster 2 merupakan kelompok terbesar dengan 20 provinsi, mencakup sebagian besar

Pulau Jawa, Sumatera bagian selatan, dan sebagian Sulawesi. Klaster 3 terdiri dari 6 provinsi Papua yang secara struktural menghadapi tantangan kemiskinan paling serius.

3.4 Hasil *K-Medoids Clustering*

K-Medoids diimplementasikan menggunakan algoritma PAM. *K-Medoids* menggunakan titik data aktual (*medoid*) sebagai pusat klaster, sehingga lebih *robust* terhadap *outlier* dibandingkan *K-Means*.

Tabel 6. Medoid Terpilih Metode K-Medoids per Klaster

Klaster	Kategori	Medoid (Provinsi Representatif)	Garis Kem. (Rp)	P0 (%)	P1	P2	Pengeluar./kap
1	Kemiskinan Rendah	Riau	624.870	6,45	0,97	0,22	1.589
2	Kemiskinan Sedang	Lampung	503.217	10,69	1,64	0,37	1.198
3	Kemiskinan Tinggi	Papua	612.548	17,26	3,89	1,18	901

Provinsi Riau terpilih sebagai medoid Klaster 1 karena paling mewakili profil provinsi dengan kemiskinan rendah, ditandai P0=6,45% dan pengeluaran per kapita Rp1.589.000. Lampung menjadi medoid Klaster 2 dengan P0=10,69%, berada tepat di sekitar nilai tengah klaster sedang. Provinsi Papua terpilih sebagai medoid Klaster 3 dengan P0=17,26%.

Tabel 7. Keanggotaan Provinsi per Klaster - K-Medoids

Klaster	Medoid dan Jumlah	Provinsi Anggota
1	Riau (n=13)	Sumatera Utara, Sumatera Barat, Riau, Kep. Bangka Belitung, Kepulauan Riau, DKI Jakarta, Banten, Bali, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara
2	Lampung (n =17)	Aceh, Jambi, Sumatera Selatan, Bengkulu, Lampung, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, NTB, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara
3	Papua (n=8)	NTT, Maluku, Papua Barat, Papua Barat Daya, Papua, Papua Selatan, Papua Tengah, Papua Pegunungan

K-Medoids menghasilkan pembagian 13 provinsi di klaster 1, 17 provinsi di klaster 2, dan 8 provinsi di klaster 3. Dibandingkan K-*Means*, terdapat perbedaan keanggotaan pada tiga provinsi yaitu NTT, Maluku, dan Kalimantan Barat. NTT dan Maluku yang pada *K-Means* termasuk Klaster 2, oleh *K-Medoids* dimasukkan ke Klaster 3, sedangkan Kalimantan Barat berpindah dari Klaster 2 ke Klaster 1. Perbedaan ini terjadi karena *K-Medoids* tidak terpengaruh oleh *outlier* ekstrem Papua Tengah dan Papua Pegunungan dalam menentukan batas klaster.

3.5 Perbandingan dan Evaluasi Metode *Clustering*

Kualitas klaster dievaluasi menggunakan dua indeks validasi yaitu *Davies-Bouldin Index* (DBI) dan *silhouette index*. DBI mengukur rasio rata-rata dispersi dalam klaster terhadap jarak antar klaster, nilai lebih kecil lebih baik. *Silhouette index* mengukur seberapa tepat setiap titik data ditempatkan dalam klasternya, nilai mendekati 1 lebih baik.

Tabel 8. Perbandingan Metode Berdasarkan Indeks Validasi

Metode	Jumlah Klaster	*Davies-Bouldin Index* (DBI)	*Silhouette Index*

K-Means	3	0,7448	0,4290
K-Medoids	3	0,7884	0,3948

Berdasarkan kedua indeks validasi, *K-Means* menghasilkan kualitas klaster yang lebih baik. DBI *K-Means* lebih rendah dari *K-Medoids*, mengindikasikan klaster metode *K-Means* lebih valid dan terpisah lebih baik. *Silhouette index* metode *K-Means* juga lebih tinggi dari *K-Medoids*. Hasil ini menunjukkan bahwa pada penelitian ini, pengaruh *outlier* (Papua Tengah dan Papua Pegunungan) justru membantu *K-Means* membentuk klaster yang lebih terpisah, sementara *K-Medoids* yang lebih konservatif terhadap *outlier* tetapi kurang terdiferensiasi secara keseluruhan.

Tabel 9. Perbandingan Hasil *Clustering K-Means* dan *K-Medoids*

No.	Provinsi	*K-Means*	*K-Medoids*
1	Aceh	K2 (Sedang)	K2 (Sedang)
2	Sumatera Utara	K1 (Rendah)	K1 (Rendah)
3	Sumatera Barat	K1 (Rendah)	K1 (Rendah)
4	Riau	K1 (Rendah)	K1 (Rendah)
5	Jambi	K2 (Sedang)	K2 (Sedang)
6	Sumatera Selatan	K2 (Sedang)	K2 (Sedang)
7	Bengkulu	K2 (Sedang)	K2 (Sedang)
8	Lampung	K2 (Sedang)	K2 (Sedang)
9	Kep. Bangka Belitung	K1 (Rendah)	K1 (Rendah)
10	Kepulauan Riau	K1 (Rendah)	K1 (Rendah)
11	DKI Jakarta	K1 (Rendah)	K1 (Rendah)
12	Jawa Barat	K2 (Sedang)	K2 (Sedang)
13	Jawa Tengah	K2 (Sedang)	K2 (Sedang)
14	DI Yogyakarta	K2 (Sedang)	K2 (Sedang)
15	Jawa Timur	K2 (Sedang)	K2 (Sedang)
16	Banten	K1 (Rendah)	K1 (Rendah)
17	Bali	K1 (Rendah)	K1 (Rendah)
18	NTB	K2 (Sedang)	K2 (Sedang)
19	NTT	K2 (Sedang)	K3 (Tinggi)
20	Kalimantan Barat	K2 (Sedang)	K1 (Rendah)

No.	Provinsi	*K-Means*	*K-Medoids*
21	Kalimantan Tengah	K1 (Rendah)	K1 (Rendah)
22	Kalimantan Selatan	K1 (Rendah)	K1 (Rendah)
23	Kalimantan Timur	K1 (Rendah)	K1 (Rendah)
24	Kalimantan Utara	K1 (Rendah)	K1 (Rendah)
25	Sulawesi Utara	K2 (Sedang)	K2 (Sedang)
26	Sulawesi Tengah	K2 (Sedang)	K2 (Sedang)
27	Sulawesi Selatan	K2 (Sedang)	K2 (Sedang)
28	Sulawesi Tenggara	K2 (Sedang)	K2 (Sedang)
29	Gorontalo	K2 (Sedang)	K2 (Sedang)
30	Sulawesi Barat	K2 (Sedang)	K2 (Sedang)
31	Maluku	K2 (Sedang)	K3 (Tinggi)
32	Maluku Utara	K2 (Sedang)	K2 (Sedang)
33	Papua Barat	K3 (Tinggi)	K3 (Tinggi)
34	Papua Barat Daya	K3 (Tinggi)	K3 (Tinggi)
35	Papua	K3 (Tinggi)	K3 (Tinggi)
36	Papua Selatan	K3 (Tinggi)	K3 (Tinggi)
37	Papua Tengah	K3 (Tinggi)	K3 (Tinggi)
38	Papua Pegunungan	K3 (Tinggi)	K3 (Tinggi)

Berdasarkan Tabel 9, terdapat 3 provinsi yang berbeda keanggotaan klasternya, yaitu NTT (*K-Means*: klaster 2, *K-Medoids*: klaster 3), Maluku (*K-Means*: Klaster 2, *K-Medoids*: Klaster 3), dan Kalimantan Barat (*K-Means*: klaster 2, *K-Medoids*: Klaster 1). Tingkat kesesuaian antar metode mencapai 92,1% (35 dari 38 provinsi), mengindikasikan adanya kemiripan dari kedua pendekatan dalam mengelompokkan profil kemiskinan provinsi-provinsi di Indonesia.

4 KESIMPULAN

Berdasarkan analisis klaster metode *K-Means* menghasilkan distribusi kelompok yang terdiri dari 12 provinsi pada klaster kemiskinan rendah, 20 provinsi pada klaster sedang, dan 6 provinsi pada klaster tinggi, sedangkan metode *K-Medoids* menghasilkan pengelompokan masing-masing sebanyak 13, 17, dan 8 provinsi. Tingkat keselarasan antara kedua metode ini mencapai 92,1%. Metode *K-Means* menghasilkan kualitas klaster yang sedikit lebih baik dengan nilai *Davies-Bouldin Index* (DBI) sebesar 0,7448 dan *silhouette index* sebesar 0,4290, dibandingkan dengan metode *K-Medoids* yang menghasilkan nilai DBI sebesar 0,7884 dan *silhouette score* sebesar 0,3948. Hal ini terjadi karena terdapat nilai-nilai ekstrem (*outliers*),

seperti Provinsi Papua Tengah dan Papua Pegunungan, justru membantu metode *K-Means* dalam membentuk pemisahan yang lebih valid antar-klaster.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada seluruh pihak yang telah memberikan dukungan, bantuan, dan masukan selama proses pelaksanaan penelitian hingga penyusunan artikel ini.

DAFTAR PUSTAKA

Amrulloh, K., Pudjiantoro, T. H., Sabrina, P. N., & Hadiana, A. (2022). Comparison Between Davies-Bouldin Index and Silhouette Coefficient Evaluation Methods in Retail Store Sales Transaction Data Clusterization Using K-Medoids Algorithm. *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI), 5*(1).

Astuti, N. T., Sunardi, & Fadlil, A. (2026). Optimalisasi Pemetaan Klaster Kependudukan Provinsi di Indonesia Menggunakan Elbow Method dan Davies-Bouldin Index. *Jurnal Informatika Polinema, 12*(2), 295–304. https://doi.org/10.33795/jip.v12i2.9162

Bahauddin, A., Fatmawati, A., & Sari, F. P. (2021). Analisis Clustering Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means. *Jurnal Manajemen Informatika dan Sistem Informasi (MISI), 4*(1), 1–8. https://doi.org/10.36595/misi.v4i1.216

Badan Pusat Statistik. (2024). Profil Kemiskinan di Indonesia Maret 2024 (BRS No. 56/07/Th. XXVII, 1 Juli 2024). Jakarta: BPS Indonesia. https://www.bps.go.id

Fauzi, I. A. (2024). Klasterisasi Tingkat Kemiskinan Kabupaten/Kota di Indonesia Menggunakan Algoritma K-Means dan K-Medoids. *Jurnal Algoritma, 21*(2), 197–208. https://doi.org/10.33364/algoritma/v.21-2.1788

Hasan, Y. (2024). Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means dan DBSCAN. *Jurnal Informatika dan Teknik Elektro Terapan (JITET), 12*(3S1). https://doi.org/10.23960/jitet.v12i3S1.5001

Hidayat, F. P., Putra, R. P., Alfitrah, M. D., & Widodo, E. (2023). Implementasi Clustering K-Medoids dalam Pengelompokan Kabupaten di Provinsi Aceh Berdasarkan Faktor yang Mempengaruhi Kemiskinan. Indonesian *Journal of Applied Statistics, 5*(2), 121. https://doi.org/10.13057/ijas.v5i2.55080

Luchia, N. T., Handayani, H., & Hamdi, F. S. (2022). Perbandingan K-Means dan K-Medoids pada Pengelompokan Data Miskin di Indonesia. MALCOM*: Indonesian Journal of Machine Learning and Computer Science, 2*(2), 35–41. https://doi.org/10.57152/malcom.v2i2.422

Mayasari, S. N., & Nugraha, J. (2023). Implementasi K-Means Cluster Analysis untuk Mengelompokkan Kabupaten/Kota Berdasarkan Data Kemiskinan di Provinsi Jawa Tengah Tahun 2022. KONSTELASI: *Konvergensi Teknologi dan Sistem Informasi, 3*(2), 317–329.

Maulana, R., & Kurniawan, A. (2023). Pengelompokan Kabupaten/Kota Berdasarkan Indikator Kemiskinan sebagai Dasar Diferensiasi Kebijakan Pembangunan Daerah. *Jurnal Ekonomi dan Pembangunan Indonesia, 23(*2), 145–162. https://doi.org/10.21002/jepi.v23i2.1547

Ningrum, S. A., Wahyudi, T., & Prasetyo, E. (2023). Analisis Karakteristik Kemiskinan Multidimensi dan Implikasinya terhadap Kebijakan Pengentasan Kemiskinan di

Indonesia. Jurnal Perencanaan Pembangunan*: The Indonesian Journal of Development Planning, 7*(1), 78–96. https://doi.org/10.36574/jpp.v7i1.304

Prabowo, H., & Santoso, D. B. (2022). Perbandingan Metode Clustering untuk Identifikasi Profil Kemiskinan Regional di Indonesia: Pendekatan K-Means, Hierarchical, dan DBSCAN. *Jurnal Ilmu Statistika dan Komputasi*, *3*(1), 21–35. https://doi.org/10.21107/jsk.v3i1.12834

Putri, D. A., & Ardiansyah, F. (2022). Ketimpangan Pembangunan dan Kemiskinan Struktural di Indonesia: Tinjauan terhadap Disparitas Antarwilayah Periode 2018–2022. *Jurnal Ekonomi Pembangunan, 20*(2), 110–128. https://doi.org/10.29259/jep.v20i2.15732

Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-Means Clustering Algorithm. IEEE Access, 8, 80716–80727. https://doi.org/10.1109/ACCESS.2020.2988796

Tangke, N. R., & Dzikrullah, A. A. (2023). Implementasi K-Medoids Clustering pada Kemiskinan Ekstrem di Provinsi Maluku. *Emerging Statistics and Data Science Journal, 1*(3), 415–424.

Tawakal, I., Effendi, M. M., & Majid, A. M. (2025). Analisis Tingkat Kemiskinan dengan Algoritma K-Means Menggunakan RapidMiner di Tingkat Kota Kabupaten di Jawa Tengah. *Journal of Information System Management (JOISM), 7*(1), 112–119.