

PERBANDINGAN METODE K-MEANS DAN K-MEDOIDS DALAM PENGELOMPOKAN KABUPATEN/KOTA DI PROVINSI JAWA TIMUR BERDASARKAN INDIKATOR TINGKAT KESEJAHTERAAN

Shella Mayasyah^{1*}, Emeylia Safitri²

^{1,2}Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas Terbuka, Tangerang Selatan,
Indonesia

*Penulis korespondensi: shellamayasyah@gmail.com

ABSTRAK

Penyebaran tingkat kesejahteraan yang merata di berbagai daerah menjadi dasar penting untuk mempercepat pembangunan nasional yang adil bagi setiap daerah. Penelitian ini dilatarbelakangi oleh perlunya pemetaan wilayah yang akurat guna mengatasi ketimpangan kondisi sosial-ekonomi di tingkat daerah. Jawa Timur menjadi salah satu provinsi dengan tingkat kesejahteraan yang cukup beragam. Adapun tujuan utama dari studi ini adalah untuk mengelompokkan kabupaten dan kota di Provinsi Jawa Timur berdasarkan indikator kesejahteraan tahun 2025, sekaligus menguji efektivitas antara metode *K-Means* dan *K-Medoids*. Data yang digunakan mencakup berbagai indikator kesejahteraan seperti indeks pembangunan manusia, tingkat pengangguran terbuka, jumlah penduduk miskin, penduduk yang memiliki jaminan kesehatan BPJS PBI, rata-rata upah/gaji bersih sebulan, dan rata-rata lama sekolah usia >25. Temuan penelitian ini bahwa kedua metode menghasilkan jumlah kluster optimal yang sama yaitu 2 kluster, dengan komposisi keanggotaan wilayah yang sepenuhnya identik. Kluster 1 terdiri atas 27 wilayah yang seluruhnya merupakan kabupaten, sedangkan Kluster 2 terdiri atas 11 wilayah. Kluster 1 yang didominasi oleh wilayah kabupaten menunjukkan tingkat kesejahteraan yang relatif lebih rendah pada sebagian besar indikator tingkat kesejahteraan. Sedangkan, kluster 2 dapat dikategorikan sebagai kelompok wilayah dengan tingkat kesejahteraan yang lebih tinggi.

Kata Kunci: *K-Means*, *K-Medoids*, Jawa Timur, Kesejahteraan, *Clustering*.

1 PENDAHULUAN

Pembangunan daerah yang berkelanjutan di Indonesia saat ini menitikberatkan pada upaya pemerataan kesejahteraan masyarakat di seluruh wilayah kabupaten dan kota. Provinsi Jawa Timur, sebagai salah satu wilayah strategis di Pulau Jawa, masih mengalami kendala terkait ketimpangan kondisi sosial ekonomi di berbagai wilayahnya (Jatipaningrum et al., 2022). Sejalan dengan Setiawan & Hakim (2013) tingkat kesejahteraan ini tidak hanya diukur melalui satu dimensi, melainkan mencakup berbagai indikator seperti indikator yang berfokus pada pembangunan manusia, tingkat kemiskinan, pengangguran terbuka, hingga akses terhadap fasilitas dasar (Mahroji & Nurkhasanah, 2019). Identifikasi kelompok wilayah dengan karakteristik kesejahteraan yang serupa menjadi krusial agar pemerintah daerah dapat merancang program yang lebih spesifik dan efektif (Saputri & Arianto, 2023).

Dalam studi mengenai pengelompokan data, analisis kluster merupakan teknik yang paling umum digunakan untuk mengidentifikasi objek yang memiliki kemiripan sifat dalam satu grup. Beberapa algoritma partisi seperti *K-Means* banyak diaplikasikan karena kesederhanaannya dalam menangani data besar (Thamrin & Wijayanto, 2021). Namun, *K-Means* memiliki kelemahan yaitu sangat sensitif terhadap data pencilan (*outlier*) yang seringkali ditemukan dalam data indikator kesejahteraan daerah (Hoerunnisa et al., 2024). Sebagai alternatif, algoritma *K-Medoids* atau *Partitioning Around Medoids* (PAM) hadir

dengan pendekatan yang lebih kokoh (*robust*) terhadap pencilon karena menggunakan objek nyata (*medoid*) sebagai pusat klaster dibandingkan nilai rata-rata (*centroid*) (Ikhrum et al., 2025).

Beberapa penelitian terdahulu telah mencoba membandingkan efektivitas kedua metode ini dalam berbagai kasus pembangunan. Studi oleh Saputri & Arianto (2023) menunjukkan bahwa pemilihan algoritma yang tepat sangat bergantung pada distribusi data dan keberadaan outlier untuk menghasilkan evaluasi wilayah yang akurat. Sementara itu, penelitian Thamrin & Wijayanto (2021) menekankan bahwa pemetaan tingkat kesejahteraan di Pulau Jawa memerlukan validasi yang ketat agar hasil pengelompokan dapat mencerminkan kondisi lapangan yang sebenarnya. Meskipun kedua metode telah banyak dikaji, penerapan perbandingannya secara spesifik pada data indikator kesejahteraan Provinsi Jawa Timur tahun 2025 masih memerlukan eksplorasi lebih lanjut untuk melihat konsistensi performanya.

Berdasarkan latar belakang tersebut dan sejalan dengan penelitian yang dilakukan oleh Wahyuni & Wulandari (2022), penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Jawa Timur ke dalam klaster-klaster berdasarkan kemiripan indikator kesejahteraan tahun 2025. Melalui perbandingan metode *K-Means* dan *K-Medoids*, diharapkan dapat ditemukan algoritma terbaik yang memberikan hasil pengelompokan paling optimal. Hasil dari penelitian ini diproyeksikan menjadi rujukan ilmiah bagi Pemerintah Provinsi Jawa Timur dalam menentukan prioritas pembangunan wilayah, sehingga kebijakan yang diambil dapat lebih tepat sasaran guna menekan angka ketimpangan sosial di setiap daerah (Jatipaningrum et al., 2022).

2 METODE

2.1 Data dan Sumber Data

Objek dalam penelitian ini mencakup seluruh wilayah kabupaten/kota di Provinsi Jawa Timur yang terdiri dari 29 kabupaten dan 9 kota. Data yang digunakan adalah data sekunder indikator tingkat kesejahteraan masyarakat tahun 2025 yang bersumber dari publikasi resmi Badan Pusat Statistik (BPS) Provinsi Jawa Timur, khususnya publikasi Provinsi Jawa Timur Dalam Angka 2026 dan Statistik Daerah Provinsi Jawa Timur 2025 (BPS 2025a; BPS 2025b, BPS 2026). Analisis ini melibatkan enam variabel indikator. Variabel-variabel yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Satuan
X_1	Indeks Pembangunan Manusia	Indeks
X_2	Tingkat Pengangguran Terbuka	Persen (%)
X_3	Jumlah Penduduk Miskin	Ribu Jiwa
X_4	Penduduk yang Memiliki Jaminan Kesehatan BPJS PBI	Persen (%)
X_5	Rata-Rata Upah/Gaji Bersih Sebulan	Rupiah (Rp)
X_6	Rata-rata Lama Sekolah Usia >25	Tahun

Seluruh data variabel ditransformasikan terlebih dahulu menggunakan rumus *Z-score* untuk mengeliminasi bias akibat perbedaan satuan pengukuran (Saputri & Arianto, 2023).

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Keterangan:

- z : Nilai standardisasi
- x : Nilai asli data
- μ : Nilai rata-rata variabel
- σ : Simpangan baku

2.2 Prosedur Penelitian

Tahapan penelitian ini meliputi tahapan sebagai berikut:

1. Pengumpulan data yaitu penyusunan data terhadap 29 kabupaten dan 9 kota di Provinsi Jawa Timur, berdasarkan indikator tingkat kesejahteraan yang bersumber dari Badan Pusat Statistik (BPS) Provinsi Jawa Timur.
2. Analisis deskriptif yaitu melakukan karakteristik awal data untuk melihat gambaran umum dari setiap indikator kesejahteraan di 38 kabupaten/kota Provinsi Jawa Timur.
3. Prapemrosesan data yaitu melakukan pemeriksaan data seperti deteksi *outlier* dan transformasi data (standarisasi/*z-score*).
4. Penentuan dan validitas jumlah klaster optimal yaitu menentukan jumlah klaster (k) optimal dengan menggunakan pendekatan *silhouette method*.
5. Analisis *clustering* dengan metode *K-Means* yaitu mengimplementasikan algoritma *K-Means* untuk mengelompokkan 38 kabupaten/kota di Jawa Timur berdasarkan jumlah klaster (k) optimal yang telah ditetapkan.
6. Analisis *clustering* dengan metode *K-Medoids* yaitu mengimplementasikan algoritma *K-Medoids* sebagai pembanding untuk mengelompokkan 38 kabupaten/kota di Jawa Timur menggunakan jumlah klaster (k) optimal yang telah ditetapkan.
7. Perbandingan dan evaluasi hasil klaster yaitu membandingkan kinerja dan kualitas hasil pengelompokan dari kedua metode berdasarkan nilai indeks validitas yang dihasilkan, guna menentukan metode mana yang paling baik (Luthfi & Wijayanto, 2021).

2.3 Metode *K-Means*

Algoritma *K-Means* mengelompokkan data berdasarkan jarak terdekat ke pusat kluster atau *centroid* (Wahyudi et al., 2020). Proses ini dilakukan untuk meminimalkan *Within-Klaster Sum of Squares* (WCSS). Jarak antar objek dihitung menggunakan rumus *Euclidean Distance* (Ikhrum et al., 2025):

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

Keterangan:

- $d(x, y)$: Jarak *Euclidean* antara objek x dan objek y
 n : Jumlah dimensi variabel
 i : Indeks dimensi atau variabel ($1, 2, \dots, n$)
 x_i : Nilai koordinat objek x pada dimensi ke- i
 y_i : Nilai koordinat objek y pada dimensi ke- i

K-Means dipilih karena efisiensinya dalam pengelompokan data dengan volume yang cukup besar (Luchia et al., 2022).

2.4 Metode *K-Medoids*

Metode *K-Medoids* atau *Partitioning Around Medoids* (PAM) digunakan sebagai metode pembanding karena memiliki ketahanan yang lebih baik terhadap keberadaan pencilan (*outlier*). Berbeda dengan *K-Means* yang menggunakan rata-rata sebagai pusat klaster, *K-Medoids* menentukan pusat klaster berdasarkan salah satu objek yang terdapat dalam data (medoid), sehingga pengaruh nilai ekstrem terhadap hasil pengelompokan dapat dikurangi (Ikhrum et al., 2025).

2.5 *Silhouette Coefficient* (SC)

Untuk menentukan jumlah klaster (k) yang paling optimal, penelitian ini menggunakan metode *Silhouette Coefficient* (SC). *Silhouette Coefficient* (SC) digunakan untuk mengukur seberapa mirip suatu objek dengan kelompoknya sendiri dibandingkan dengan kelompok lain. Nilai SC yang mendekati 1 menunjukkan pengelompokan yang sangat baik. Nilai ini dihitung dengan menggunakan rumus (Saputri & Arianto, 2023):

$$S_i = \frac{b_i - a_i}{\max\{a_i, b_i\}} \quad (3)$$

Keterangan:

- S_i : Nilai *silhouette coefficient* untuk objek i
 a_i : Rata-rata jarak objek i ke semua objek lain dalam klaster yang sama (*cohesion*)
 b_i : Rata-rata jarak objek i ke semua objek yang ada di klaster terdekat (*separation*)

3 HASIL DAN PEMBAHASAN

3.1 Analisis Deskriptif

Data yang digunakan dalam penelitian ini terdiri atas enam indikator kesejahteraan masyarakat pada 38 kabupaten/kota di Provinsi Jawa Timur tahun 2025. Enam indikator tersebut antara lain, IPM (X_1), TPT (X_2), jumlah penduduk miskin (X_3), persentase penduduk penerima BPJS PBI (X_4), rata-rata upah bersih per bulan (X_5), dan rata-rata lama sekolah penduduk usia 25 tahun ke atas (X_6). Data tersebut disajikan pada Tabel 2.

Tabel 2. Data Variabel Penelitian

Kab /Kota	X_1	X_2	X_3	X_4	X_5	X_6
Kab Malang	74,45	5,00	235,63	31,80	2.616.018	7,98
Kab Sidoarjo	83,35	5,75	107,57	34,78	3.947.614	10,94
Kab Sampang	67,23	2,44	213,98	61,02	3.155.720	5,19
Kab Pamekasan	71,64	1,33	118,52	67,53	2.021.722	7,29
Kab Sumenep	70,54	1,64	188,48	88,95	2.266.004	6,24
...	...	...	...	...	...	...
Kota Kediri	82,71	3,76	17,93	63,16	2.861.745	11,09
Kota Mojokerto	82,35	4,66	6,77	55,00	3.296.766	11,44
Kota Madiun	85,12	4,29	6,96	62,70	2.888.690	12,12
Kota Surabaya	85,65	4,84	105,09	64,77	3.963.069	11,15
Kota Batu	80,35	3,53	6,22	57,64	2.982.673	9,95

Berdasarkan data indikator kesejahteraan kabupaten/kota di Provinsi Jawa Timur, terlihat adanya ketidaksetaraan yang cukup besar antarwilayah. Pada indikator Indeks Pembangunan Manusia (IPM), Kota Surabaya mencatat nilai tertinggi sebesar 85,65, sedangkan Kabupaten Sampang memiliki nilai terendah yaitu 67,23. Untuk indikator Tingkat Pengangguran Terbuka (TPT), angka tertinggi terdapat di Kabupaten Sidoarjo sebesar 5,75%, sementara yang terendah berada di Kabupaten Pamekasan sebesar 1,33%. Dari sisi jumlah penduduk miskin, Kabupaten Malang mencatat jumlah terbesar yaitu 235,63 ribu jiwa, sedangkan Kota Batu memiliki jumlah penduduk miskin paling sedikit yaitu 6,22 ribu jiwa. Pada indikator persentase jumlah penduduk penerima BPJS PBI (Penerima Bantuan Iuran), wilayah tertinggi terdapat di Kabupaten Sumenep sebesar 88,95%, sementara yang terendah tercatat di Kabupaten Malang sebesar 31,80%. Untuk rata-rata upah sebulan, Kota Surabaya menempati posisi tertinggi dengan nilai Rp. 3.963.069, sedangkan Kabupaten Pamekasan memiliki rata-rata upah terendah sebesar Rp. 2.021.722. Sementara itu, indikator rata-rata lama sekolah penduduk usia 25 tahun keatas menunjukkan capaian tertinggi di Kota Madiun sebesar 12,12 tahun, sedangkan capaian terendah terdapat di Kabupaten Sampang dengan rata-rata 5,19 tahun.

Secara umum, wilayah perkotaan cenderung menunjukkan tingkat pembangunan manusia, besar upah bersih perbulan, dan lama sekolah yang lebih tinggi dibandingkan wilayah kabupaten. Sebaliknya, beberapa wilayah kabupaten masih menghadapi tantangan yang lebih besar terkait kemiskinan, pendidikan, dan akses terhadap kesejahteraan masyarakat. Untuk memperoleh gambaran umum mengenai karakteristik data yang digunakan dalam penelitian, dilakukan analisis statistika deskriptif terhadap seluruh variabel penelitian. Hasil analisis

statistika deskriptif yang meliputi nilai minimum, maksimum, rata-rata, dan standar deviasi disajikan pada Tabel 3.

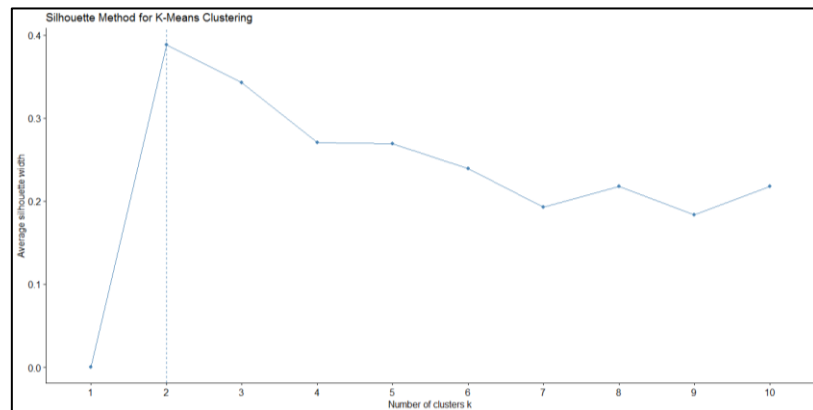
Tabel 3. Statistika Deskriptif Variabel Penelitian

Variabel	Minimum	Maksimum	Rata-rata	Standar Deviasi
X_1	67,23	85,65	76,0645	4,76139
X_2	1,33	5,75	3,7797	1,11276
X_3	6,22	235,63	101.9971	64,19300
X_4	31,80	88,95	54,7874	14,45305
X_5	2.021.722	3.963.069	2.719.172,92	489.645,518
X_6	5,19	12,12	8,5550	1,70245

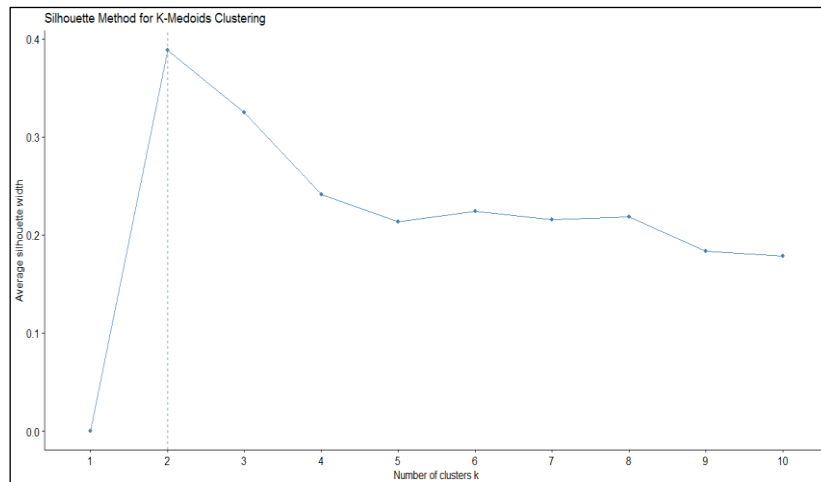
Selanjutnya dilakukan standardisasi data untuk menyeragamkan skala antar variabel yang memiliki satuan dan rentang nilai berbeda. Seluruh variabel ditransformasikan ke dalam skala yang sama untuk mengurangi pengaruh perbedaan satuan pengukuran dan mencegah dominasi variabel tertentu dalam analisis. Dalam penelitian ini, standarisasi dilakukan dengan metode *z-score normalization* agar memiliki rata-rata 0 dan standar deviasi 1 (Ikhran et al., 2025).

3.2 Penentuan dan Validasi Jumlah Kluster Optimal

Pada penelitian yang dilakukan oleh Thamrin & Wijayanto (2021), tahap pertama dalam melakukan analisis kluster adalah menentukan jumlah kluster (k) optimal yang akan digunakan untuk mengelompokkan data, baik pada metode *K-Means* maupun *K-Medoids*. Pada penelitian ini, pendekatan pertama yang digunakan untuk melihat kecenderungan jumlah kluster terbaik adalah dengan menggunakan *silhouette method*. Berdasarkan visualisasi *silhouette method*, didapatkan plot grafik seperti yang disajikan pada Gambar 1 dan Gambar 2.



Gambar 1. Plot *Silhouette Method* untuk *K-Means Clustering*



Gambar 2. Plot *Silhouette Method* untuk *K-Medoids Clustering*

Berdasarkan plot *silhouette method* pada Gambar 1 dan Gambar 2, terlihat bahwa nilai rata-rata koefisien *silhouette* tertinggi diperoleh pada jumlah kluster (k) sebesar 2. Hasil tersebut menunjukkan bahwa struktur kluster yang paling baik dan paling terpisah terbentuk ketika data dikelompokkan ke dalam 2 kluster.

3.3 Analisis *Clustering* dengan Metode *K-Means*

Setelah diperoleh jumlah kluster optimal sebanyak dua kluster ($k = 2$), proses pengelompokan dilakukan menggunakan metode *K-Means*. Metode *K-Means* mengelompokkan kabupaten/kota berdasarkan kemiripan karakteristik kesejahteraan yang diukur melalui enam indikator. Hasil pengelompokan kabupaten/kota di Provinsi Jawa Timur menggunakan metode *K-Means* disajikan pada Tabel 4.

Tabel 4. Hasil *Clustering* dengan Metode *K-Means*

Klaster	Jumlah Anggota	Kabupaten/Kota
1	27	Kabupaten Pacitan, Kabupaten Ponorogo, Kabupaten Trenggalek, Kabupaten Tulungagung, Kabupaten Blitar, Kabupaten Kediri, Kabupaten Malang, Kabupaten Lumajang, Kabupaten Jember, Kabupaten Banyuwangi, Kabupaten Bondowoso, Kabupaten Situbondo, Kabupaten Probolinggo, Kabupaten Pasuruan, Kabupaten Mojokerto, Kabupaten Jombang, Kabupaten Nganjuk, Kabupaten Madiun, Kabupaten Magetan, Kabupaten Ngawi, Kabupaten Bojonegoro, Kabupaten Tuban, Kabupaten Lamongan, Kabupaten Bangkalan, Kabupaten Sampang, Kabupaten Pamekasan, Kabupaten Sumenep
2	11	Kabupaten Sidoarjo, Kabupaten Gresik, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu

Berdasarkan Tabel 4, hasil klusterisasi dengan metode *K-Means* menghasilkan dua kelompok. Klaster 1 terdiri atas 27 kabupaten, sedangkan Klaster 2 terdiri atas 11 kabupaten/kota. Secara umum, Klaster 2 didominasi oleh wilayah perkotaan dan daerah penyangga pusat pertumbuhan ekonomi di Jawa Timur, seperti Kabupaten Sidoarjo, Kabupaten Gresik, Kota Surabaya, Kota Malang, dan Kota Batu. Sementara itu, Klaster 1 sebagian besar terdiri atas wilayah kabupaten yang tersebar di wilayah selatan, timur, dan Pulau Madura.

Untuk mengetahui karakteristik masing-masing klaster, dilakukan analisis terhadap nilai *centroid* yang diperoleh dari hasil pengelompokan *K-Means*. Nilai *centroid* menunjukkan rata-rata setiap indikator pada masing-masing klaster. Hasil perhitungan *centroid* disajikan pada Tabel 5.

Tabel 5. *Centroid* Metode *K-Means*

Klaster	IPM	TPT	Jumlah Penduduk Miskin	Penduduk Penerima BPJS PBI	Rata-Rata Upah Bersih Sebulan	Rata-rata Lama Sekolah Usia >25
1	73,54	3,40	126,51	52,92	2.528.820	7,64
2	82,26	4,71	41,83	59,36	3.186.403	10,79

Berdasarkan nilai *centroid* pada Tabel 5, terlihat bahwa Klaster 2 memiliki nilai Indeks Pembangunan Manusia (IPM), persentase kepemilikan jaminan kesehatan BPJS PBI, rata-rata upah/gaji bersih sebulan, dan rata-rata lama sekolah yang lebih tinggi dibandingkan Klaster 1. Selain itu, jumlah penduduk miskin pada Klaster 2 jauh lebih rendah dibandingkan Klaster 1. Kondisi tersebut menunjukkan bahwa wilayah yang tergabung dalam Klaster 2 memiliki tingkat kesejahteraan yang relatif lebih baik.

Sebaliknya, Klaster 1 ditandai dengan nilai IPM, rata-rata lama sekolah, dan rata-rata pendapatan yang lebih rendah serta jumlah penduduk miskin yang lebih tinggi. Oleh karena itu, Klaster 1 dapat dikategorikan sebagai kelompok wilayah dengan tingkat kesejahteraan relatif lebih rendah dibandingkan Klaster 2. Meskipun demikian, tingkat pengangguran terbuka pada Klaster 2 tercatat lebih tinggi dibandingkan Klaster 1. Hal ini menunjukkan bahwa wilayah dengan tingkat kesejahteraan yang lebih baik tidak selalu memiliki tingkat pengangguran yang lebih rendah. Fenomena tersebut dapat disebabkan karena tingginya tingkat pendidikan serta tuntutan biaya hidup di perkotaan menyebabkan tenaga kerja lebih berhati-hati dalam memilih pekerjaan, dengan mempertimbangkan tingkat upah yang mampu mendukung kesejahteraan mereka.

3.4 Analisis *Clustering* dengan Metode *K-Medoids*

Sebagai metode pembanding terhadap algoritma *K-Means*, penelitian ini juga menerapkan metode *K-Medoids* yang dikenal memiliki tingkat kestabilan yang lebih baik serta lebih tahan terhadap pengaruh data pencilan (*outlier*) (Farissa et al., 2021). Berbeda dengan *K-Means* yang menggunakan nilai rata-rata (*centroid*) sebagai pusat klaster, *K-Medoids* menentukan pusat klaster berdasarkan objek data yang paling representatif (*medoid*), sehingga hasil pengelompokan cenderung tidak mudah dipengaruhi oleh nilai ekstrem. Berdasarkan hasil penentuan dan validasi jumlah klaster optimal, diperoleh jumlah klaster terbaik sebanyak dua klaster. Oleh karena itu, proses pengelompokan menggunakan metode *K-Medoids* dilakukan dengan membagi seluruh kabupaten/kota di Provinsi Jawa Timur ke dalam dua kelompok berdasarkan karakteristik indikator kesejahteraan yang digunakan dalam penelitian. Adapun hasil pengelompokan kabupaten/kota di Provinsi Jawa Timur menggunakan metode *K-Medoids* secara rinci disajikan pada Tabel 5.

Tabel 6. Hasil *Clustering* dengan Metode *K-Medoids*

Klaster	Jumlah Anggota	Kabupaten/Kota
1	27	Kabupaten Pacitan, Kabupaten Ponorogo, Kabupaten Trenggalek, Kabupaten Tulungagung, Kabupaten Blitar, Kabupaten Kediri, Kabupaten Malang, Kabupaten Lumajang, Kabupaten Jember, Kabupaten Banyuwangi, Kabupaten Bondowoso, Kabupaten Situbondo, Kabupaten Probolinggo, Kabupaten Pasuruan, Kabupaten Mojokerto, Kabupaten

2	11	Jombang, Kabupaten Nganjuk, Kabupaten Madiun, Kabupaten Magetan, Kabupaten Ngawi, Kabupaten Bojonegoro, Kabupaten Tuban, Kabupaten Lamongan, Kabupaten Bangkalan, Kabupaten Sampang, Kabupaten Pamekasan, Kabupaten Sumenep
		Kabupaten Sidoarjo, Kabupaten Gresik, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu

Berdasarkan Tabel 6, metode *K-Medoids* menghasilkan dua kelompok wilayah, yaitu Klaster 1 yang terdiri atas 27 kabupaten dan Klaster 2 yang terdiri atas 11 kabupaten/kota. Hasil pengelompokan yang diperoleh menunjukkan pola yang sama dengan metode *K-Means*, dimana Klaster 2 didominasi oleh wilayah perkotaan dan pusat pertumbuhan ekonomi, sedangkan Klaster 1 didominasi oleh wilayah kabupaten.

Pusat klaster pada metode *K-Medoids* disebut *medoid*, yaitu objek aktual dalam data yang paling merepresentasikan karakteristik masing-masing klaster. *Medoid* yang diperoleh dari hasil pengelompokan disajikan pada Tabel 7. Kabupaten Banyuwangi merupakan objek yang paling merepresentasikan karakteristik Klaster 1, sedangkan Kota Mojokerto merupakan objek yang paling merepresentasikan karakteristik Klaster 2.

Tabel 7. *Medoid* Setiap Klaster pada Metode *K-Medoids*

Variabel	Klaster 1	Klaster 2
IPM	75,2	82,4
TPT	3,94	4,66
Jumlah Penduduk Miskin (Ribuan)	100	6,77
Penduduk Penerima BPJS PBI	51,5	55
Rata-Rata Upah Bersih Sebulan	2.737.734	3.296.766
Rata-rata Lama Sekolah Usia >25	7,84	11,44

Berdasarkan Tabel 7, Klaster 2 memiliki nilai IPM, rata-rata lama sekolah, persentase kepemilikan jaminan kesehatan BPJS PBI, dan rata-rata upah/gaji bersih sebulan yang lebih tinggi dibandingkan Klaster 1. Selain itu, jumlah penduduk miskin pada Klaster 2 juga jauh lebih rendah dibandingkan Klaster 1. Oleh karena itu, Klaster 2 dapat diidentifikasi sebagai kelompok wilayah dengan tingkat kesejahteraan relatif tinggi, sedangkan Klaster 1 merupakan kelompok wilayah dengan tingkat kesejahteraan relatif rendah.

3.5 Perbandingan Metode *K-Means* dan *K-Medoids*

Setelah dilakukan pengelompokan menggunakan metode *K-Means* dan *K-Medoids*, langkah selanjutnya adalah membandingkan hasil yang diperoleh dari kedua metode tersebut. Perbandingan dilakukan berdasarkan hasil pengelompokan kabupaten/kota serta nilai indeks validitas internal yang meliputi *Dunn Index* (DN), *Davies-Bouldin Index* (DB), dan *Calinski-Harabasz Index* (CH). Evaluasi lanjutan menggunakan tiga indeks validitas kluster ini dilakukan untuk memperoleh jumlah klaster yang benar-benar optimal pada masing-masing metode *non-hierarchical clustering*, yaitu *K-Means* dan *K-Medoids*. Hasil perhitungan ketiga indeks validitas tersebut untuk variasi jumlah klaster $k = 2$ hingga $k = 5$ disajikan pada Tabel 8.

Tabel 8. Perhitungan Nilai *Dunn Index* (DN), *Davies-Bouldin Index* (DB), dan *Calinski-Harabasz Index* (CH)

Jumlah Kelompok	<i>K-Means</i>			<i>K-Medoids</i>		
	DN	DB	CH	DN	DB	CH

k = 2	0,3617705	0,977551	25,93164	0,3617705	0,977551	25,93164
k = 3	0,2905053	1,080990	23,09316	0,2905053	1,162677	22,39428
k = 4	0,2210676	1,143190	20,11939	0,1228250	1,220925	19,30537
k = 5	0,2757615	1,087199	20,05582	0,1228250	1,253082	15,94796

Berdasarkan Tabel 8, metode *K-Means* dan *K-Medoids* menghasilkan nilai *Dunn Index*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index* yang sama pada nilai $k=2$. Nilai *Dunn Index* sebesar 0,36177 menunjukkan bahwa kedua metode memiliki tingkat pemisahan antar klaster yang baik. Sementara itu, nilai *Davies-Bouldin Index* sebesar 0,97755 yang mendekati nol menunjukkan bahwa klaster yang terbentuk memiliki tingkat homogenitas yang cukup baik. Selain itu, nilai *Calinski-Harabasz Index* sebesar 25,9316 menunjukkan bahwa struktur klaster yang dihasilkan memiliki kualitas yang baik dalam memaksimalkan variasi antar klaster dan meminimalkan variasi di dalam klaster.

Kesamaan nilai indeks validitas tersebut menunjukkan bahwa metode *K-Means* dan *K-Medoids* menghasilkan kualitas pengelompokan yang setara pada data indikator kesejahteraan kabupaten/kota di Provinsi Jawa Timur. Dengan kata lain, kedua metode mampu mengidentifikasi struktur data yang sama dan menghasilkan pembagian wilayah yang konsisten. Berdasarkan hasil perbandingan tersebut, tidak terdapat perbedaan yang signifikan antara metode *K-Means* dan *K-Medoids* dalam mengelompokkan kabupaten/kota di Provinsi Jawa Timur berdasarkan indikator kesejahteraan. Oleh karena itu, kedua metode dapat digunakan sebagai alternatif yang sama baiknya dalam proses *Clustering* pada penelitian ini.

Berdasarkan hasil *clustering*, metode *K-Means* dan *K-Medoids* menghasilkan komposisi anggota klaster yang identik, yaitu Klaster 1 yang terdiri atas 27 kabupaten dan Klaster 2 yang terdiri atas 11 kabupaten/kota. Selain itu, karakteristik masing-masing klaster yang ditunjukkan oleh nilai rata-rata indikator kesejahteraan juga menunjukkan pola yang sama. Klaster 2 memiliki nilai IPM, rata-rata lama sekolah, persentase kepemilikan BPJS PBI, dan rata-rata upah/gaji yang lebih tinggi, serta jumlah penduduk miskin yang lebih rendah dibandingkan Klaster 1. Dengan demikian, kedua metode secara konsisten mengelompokkan wilayah dengan tingkat kesejahteraan relatif tinggi dan rendah.

4 KESIMPULAN

Berdasarkan hasil analisis *clustering* terhadap kabupaten/kota di Provinsi Jawa Timur menggunakan metode *K-Means* dan *K-Medoids*, diperoleh jumlah klaster optimal sebanyak dua klaster ($k = 2$). Hasil pengelompokan menunjukkan bahwa metode *K-Means* dan *K-Medoids* menghasilkan komposisi anggota klaster yang identik. Klaster 1 terdiri atas 27 wilayah yang seluruhnya merupakan kabupaten, sedangkan klaster 2 terdiri atas 11 wilayah yang didominasi oleh kota-kota di Jawa Timur serta dua kabupaten yang memiliki tingkat perkembangan ekonomi dan sosial yang relatif lebih tinggi, yaitu Kabupaten Sidoarjo dan Kabupaten Gresik. Temuan ini menunjukkan adanya pola pemisahan yang cukup jelas antara wilayah kabupaten dan wilayah perkotaan dalam hal tingkat kesejahteraan masyarakat. Klaster 2 dapat dikategorikan sebagai kelompok wilayah dengan tingkat kesejahteraan yang lebih tinggi. Hal ini ditunjukkan oleh nilai Indeks Pembangunan Manusia (IPM), rata-rata lama sekolah, persentase kepemilikan jaminan kesehatan BPJS PBI, serta rata-rata upah atau gaji bersih sebulan yang cenderung lebih tinggi, disertai jumlah penduduk miskin yang lebih rendah dibandingkan Klaster 1. Sebaliknya, Klaster 1 yang didominasi oleh wilayah kabupaten menunjukkan tingkat kesejahteraan yang relatif lebih rendah pada sebagian besar indikator yang digunakan dalam penelitian. Secara keseluruhan, penelitian ini menunjukkan bahwa terdapat perbedaan tingkat kesejahteraan yang cukup nyata antara wilayah kabupaten dan wilayah perkotaan di Provinsi Jawa Timur. Sebagian besar kabupaten tergolong dalam klaster

dengan tingkat kesejahteraan yang relatif lebih rendah, sedangkan kota-kota serta wilayah penyangga industri seperti Sidoarjo dan Gresik tergolong dalam klaster dengan tingkat kesejahteraan yang lebih tinggi. Kesamaan hasil yang diperoleh dari metode *K-Means* dan *K-Medoids* juga mengindikasikan bahwa struktur data yang digunakan cukup stabil sehingga pola pengelompokan yang terbentuk dapat dianggap konsisten dan representatif dalam menggambarkan kesenjangan kesejahteraan antarwilayah di Provinsi Jawa Timur.

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

- Badan Pusat Statistik Provinsi Jawa Timur. (2025a). *Provinsi Jawa Timur dalam angka 2025* (Vol. 48). BPS Provinsi Jawa Timur.
- Badan Pusat Statistik Provinsi Jawa Timur. (2025b). *Statistik daerah Provinsi Jawa Timur 2025* (Vol. 15). BPS Provinsi Jawa Timur.
- Badan Pusat Statistik Provinsi Jawa Timur. (2026). *Provinsi Jawa Timur dalam angka 2026* (Vol. 49). BPS Provinsi Jawa Timur.
- Farissa, R. A., Mayasari, R., & Umaidah, Y. (2021). Perbandingan algoritma K-Means dan K-Medoids untuk pengelompokan data obat dengan silhouette coefficient. *Journal of Applied Informatics and Computing (JAIC)*, 5(2), 109–116. <https://doi.org/10.30871/jaic.v5i1.3237>
- Hoerunnisa, A., Dwilestari, G., Dikananda, F., Sunana, H., & Pratama, D. (2024). Komparasi algoritma K-Means dan K-Medoids dalam analisis pengelompokan daerah rawan kriminalitas di Indonesia. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1). <https://doi.org/10.36040/jati.v8i1.8249>
- Ikhrum, A., Mutia, S., Akbar, P., & Umam, C. (2025). Penerapan algoritma K-Medoids untuk pengelompokan provinsi di Indonesia berdasarkan indikator kemiskinan: Akses perumahan dan kesejahteraan. *Emerging Statistics and Data Science Journal*, 3(2).
- Jatipaningrum, M. T., Azhari, S. E., & Suryowati, K. (2022). Pengelompokan kabupaten dan kota di Provinsi Jawa Timur berdasarkan tingkat kesejahteraan dengan metode K-Means dan Density-Based Spatial Clustering of Applications with Noise. *Jurnal Derivat*, 9(1). <https://doi.org/10.31316/j.derivat.v9i1.2832>
- Luchia, N. T., Handayani, H., Hamdi, F. S., Erlangga, D., & Octavia, S. F. (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.
- Luthfi, E., & Wijayanto, A. W. (2021). Analisis perbandingan metode hierarchical, K-Means, dan K-Medoids clustering dalam pengelompokan indeks pembangunan manusia Indonesia. *Inovasi*, 17(4), 761–773.
- Mahroji, D., & Nurkhasanah, I. (2019). Pengaruh indeks pembangunan manusia terhadap tingkat pengangguran di Provinsi Banten. *Jurnal Ekonomi-Qu*, 9(1).
- Saputri, F. W., & Arianto, D. B. (2023). Perbandingan performa algoritma K-Means, K-Medoids, dan DBSCAN dalam penggerombolan provinsi di Indonesia berdasarkan indikator kesejahteraan masyarakat. *Jurnal Aplikasi Statistika dan Komputasi Statistik*, 17(2).
- Setiawan, M. B., & Hakim, A. (2013). Indeks pembangunan manusia Indonesia. *Jurnal Economia*, 9(1).
- Thamrin, N., & Wijayanto, A. W. (2021). Analisis cluster dengan menggunakan hard clustering dan soft clustering untuk pengelompokan tingkat kesejahteraan kabupaten/kota di Pulau

- Jawa. *Indonesian Journal of Statistics and Its Applications*, 5(1), 141–160.
<https://doi.org/10.29244/ijsa.v5i1p141-160>
- Wahyudi, M., Masitha, Saragih, R., & Solikhun. (2020). *Data mining: Penerapan algoritma K-Means clustering dan K-Medoids clustering*. Yayasan Kita Menulis.
- Wahyuni, I., & Wulandari, S. P. (2022). Pemetaan kabupaten/kota di Jawa Timur berdasarkan indikator kesejahteraan rakyat menggunakan analisis cluster hierarki. *Jurnal Sains dan Seni ITS*, 11(1).