KLASTERISASI KABUPATEN/KOTA DI JAWA TIMUR BERDASARKAN INDIKATOR KEMISKINAN DAN SOSIAL EKONOMI TAHUN 2025 MENGGUNAKAN ANALISIS *HIERARCHICAL CLUSTERING* DAN K-*MEANS*

Nur Lailatul Nazilah[*1], Nuramaliyah[2]
[1,2]Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas Terbuka, Indonesia

**Penulis korespondensi: nlnazila26@gmail.com*

ABSTRAK

Kemiskinan masih menjadi fokus utama pembangunan ekonomi di Indonesia, termasuk di Provinsi Jawa Timur. Ketimpangan kondisi sosial ekonomi antarkabupaten/kota menjadi fokus utama pemerintah untuk penyusunan kebijakan. Pengelompokkan wilayah berdasarkan karakteristiknya menjadi salah satu solusi untuk permasalahan ini. Penelitian ini bertujuan untuk mengelompokkan kabupatan/kota di Provinsi Jawa Timur menggunakan metode *hierarchical clustering* dan *K-Means clustering* dengan delapan indikator. Hasil penelitian menunjukkan bahwa jumlah klaster optimal sebanyak tiga klaster menghasilkan pola pengelompokan yang relatif serupa, dengan perbedaan hanya pada Kabupaten Gresik. Dua komponen utama mampu menjelaskan 85% variasi data dengan nilai *silhouette* sebesar 0,386 menunjukkan struktur klaster berada pada kategori cukup/*fair structure*. Secara keseluruhan diketahui kabupaten/kota di Jawa Timur memiliki karakteristik kemiskinan dan sosial ekonomi yang berbeda sehingga dapat digunakan sebagai dasar awal dalam penyusunan kebijakan pengentasan kemiskinan yang lebih tepat sasaran.

Kata kunci: kemiskinan, *hierarchical Klastering, K-Means clustering, nilai silhouette.*

1 PENDAHULUAN

Kemiskinan merupakan permasalahan kompleks yang masih menjadi fokus utama pembangunan. Tidak hanya berkaitan erat dengan pendapatan yang rendah, kemiskinan juga berkaitan dengan keterbatasan akses terhadap pendidikan, kesehatan, dan kesempatan kerja. Menurut El Adawiyah *et al.* (2021), kemiskinan dapat diartikan sebagai kondisi dengan standar yang rendah dan keterbatasan sumber daya yang yang dimiliki masyarakat untuk meningkatkan kualitas hidupnya. Di Indonesia, khususnya Provinsi Jawa Timur, kemiskinan masih menjadi isu strategis meskipun menunjukkan tren penurunan. Data Badan Pusat Statistik Provinsi Jawa Timur (2026), menunjukkan persentase penduduk miskin pada September 2025 tercatat 9,3% atau sekitar 3,80 juta jiwa, lebih rendah dibandingkan Maret 2025 sebesar 9,5%. Namun demikian, masih terdapat kesenjangan yang cukup besar antarkabupaten/kota. Kesenjangan tersebut mengindikasikan bahwa ada perbedaan karakteristik sosial ekonomi pada setiap kabupaten/kota. Sehingga diperlukan analisis yang mampu mengidentifikasi karakteristik wilayah berdasarkan indikator kemiskinan dan sosial ekonomi untuk mendukung penyusunan kebijakan.

Salah satu metode yang dapat digunakan untuk mengidentifikasi karakteristik wilayah adalah analisis *clustering*. Analisis *clustering* adalah salah satu metode multivariat untuk mengelompokkan objek berdasarkan kemiripan karakteristik tertentu. Sehingga objek dalam satu kelompok mempunyai karakteristik yang homogen, sedangkan antar kelompok mempunyai karakteristik yang heterogen (Wulandari *et al*. 2025). Metode *clustering* yang umum digunakan adalah *hierarchical clustering* dan K-*Means clustering*. *Hierarchical clustering* membentuk kelompok secara bertahap melalui proses penggabungan objek berdasarkan tingkat kemiripan tertentu, sedangkan metode K-*Means clustering* mengelompokkan objek berdasarkan pusat klaster (*centroid*) melalui proses iteratif.

Penelitian sebelumnya oleh Amelia *et al.* (2025) menggunakan metode K-*Means clustering* untuk pemetaan kemiskinan kabupaten/kota di Indonesia yang menghasilkan 2 klaster optimal. Klaster 1 dengan 470 anggota adalah wilayah dengan persentase penduduk miskin lebih rendah, rata-rata lama sekolah lebih tinggi dan kondisi sosial ekonomi lebih baik, sedangkan Klaster 2 dengan 44 anggota adalah wilayah dengan tingkat kemiskinan signifikan, pendidikan rendah dan minimnya infrastruktur dasar. Selain itu, penelitian juga dilakukan oleh Amalia *et al*. (2025) mengenai pengelompokan kemiskinan di Provinsi Jawa Timur tahun 2024, yang menunjukkan bahwa metode K-*Means* dan K-*Means++* menghasilkan jumlah klaster optimal yang sama, yaitu 3 klaster (struktural, fungsional, dan sedang). Dari kedua metode tersebut diketahui bahwa K-*Means++* memberikan hasil pengelompokan/*clustering* yang lebih optimal dibandingkan dengan K-Means. Meskipun demikian, kedua penelitian tersebut masih berfokus pada pendekatan *non-hierarchical clustering*, sehingga belum memberikan gambaran mengenai perbandingan metode *hierarchical clustering*. Padahal, karakteristik kemiskinan antarkabupaten/kota di Jawa Timur cukup beragam, sehingga hasil pengelompokan dapat dipengaruhi oleh metode yang digunakan. Oleh karena itu, diperlukan penelitian yang membandingkan metode *hierarchical clustering* dan K-*Means clustering* untuk pengelompokan kabupaten/kota di Jawa Timur yang lebih optimal berdasarkan indikator kemiskinan dan sosial ekonomi. Penelitian ini diharapkan mampu memberikan gambaran karakteristik wilayah secara lebih komprehensif sebagai dasar dalam pengambilan kebijakan penanggulangan kemiskinan yang lebih tepat sasaran.

2 METODE PENELITIAN

2.1 Data dan Sumber Data

Data yang digunakan dalam penelitian ini diperoleh dari Badan Pusat Statistik Provinsi Jawa Timur yang bersumber dari *website* resmi https://jatim.bps.go.id. Unit analisis dalam penelitian ini adalah 38 kabupaten/kota di Provinsi Jawa Timur dengan variabel penelitian sebagai berikut:

Tabel 1. Variabel Penelitian

Variabel	Nama Variabel
X_1	Persentase penduduk miskin (persen)
X_2	Garis kemiskinan (Rp/kapita/bulan)
X_3	Indeks kedalaman kemiskinan
X_4	Indeks keparahan kemiskinan
X_5	Indeks pembangunan manusia
X_6	Tingkat pengangguran terbuka (persen)
X_7	Rata-rata lama sekolah (tahun)
X_8	Pengeluaran per kapita (Rp/kapita)

Sumber: Badan Pusat Statistik Provinsi Jawa Timur 2025

2.2 Standardisasi Data

Standardisasi Z-*score* disebut juga dengan standardized values/scores adalah transformasi data untuk mengubah angka pada data menjadi suatu nilai standar dengan tujuan menyetarakan skala antar variabel agar berkontribusi seimbang saat analisis (Hair *et al.*, 2019). Standardisasi data dilakukan dengan menggunakan persamaan berikut:

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

keterangan:
$Z =$ nilai standardisasi
$X =$ nilai asli
$\mu =$ rata – rata

$\sigma =$ standar deviasi

2.3 Analisis *Hierarchical Clustering*

Hierarchical clustering adalah metode *clustering* yang membentuk klaster secara bertahap melalui proses penggabungan objek berdasarkan ukuran jarak tertentu. Keunggulan metode ini adalah mampu menampilkan proses pengelompokan dalam bentuk *dendrogram* sehingga memudahkan interpretasi hubungan kedekatan antarobjek maupun penentuan jumlah klaster optimal (Mulyana *et al.*, 2024). Salah satu metode yang digunakan dalam *hierarchical clustering* adalah metode *Ward* yang membentuk klaster dengan meminimalkan variasi dalam klaster (*within-cluster variance*). Ukuran kemiripan yang umum digunakan adalah *euclidean distance* yang dihitung menggunakan persamaan berikut:

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

keterangan:
$x_i =$ objek ke-*i* pada *x* (*i* = 1, 2, …, *n*)
$y_i =$ objek ke-*i* pada *y* (*i* = 1, 2, …, *n*)
$n =$ jumlah objek

2.4 Analisis K-*Means Clustering*

K-*Means clustering* termasuk salah satu metode klaster non-hierarki yang banyak digunakan karena memiliki proses komputasi yang relatif cepat, sederhana, dan efektif untuk mengelompokkan data berukuran besar (Yulisasih *et al.*, 2024). K-*Means clustering* mampu menghasilkan pengelompokan wilayah yang cukup baik karena setiap klaster dibentuk berdasarkan kedekatan terhadap pusat atau *centroid* (Mulyana *et al*, 2024). *Centroid* dihitung menggunakan rata-rata seluruh anggota klaster dengan menggunakan persamaan berikut:

$$c_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i \tag{3}$$

keterangan:
$c_j =$ *centroid* klaster ke-*j*
$n_j =$ jumlah anggota pada klaster ke-*j*
$x_i =$ data anggota klaster

2.5 *Silhoutte Score*

Silhoutte score merupakan salah satu metode evaluasi yang digunakan untuk mengukur kualitas hasil *clustering*. *Silhoutte score* memiliki rentang nilai $-1 \leq s(i) \leq 1$, dimana semakin mendekati nilai 1 menunjukkan bahwa objek sangat sesuai dengan klasternya dan memiliki jarak yang jauh terhadap klaster lain. Menurut Hasan (2024), *silhouette score* dapat digunakan sebagai metode evaluasi untuk membandingkan kualitas hasil *clustering* dari beberapa algoritma. *Silhouette score* dihitung dengan menggunakan persamaan berikut:

$$s(i) = \frac{b(i) - a(i)}{max(a(i), b(i))} \tag{4}$$

keterangan:
$s(i) =$ *silhouette score* untuk data ke-*i*
$a(i) =$ rata – rata jarak antara data ke ke-*i* dan semua titik dalam klasternya
$b(i) =$ rata – rata jarak antara data ke ke-*i* dan semua titik pada klaster terdekat

2.6 Langkah Analisis

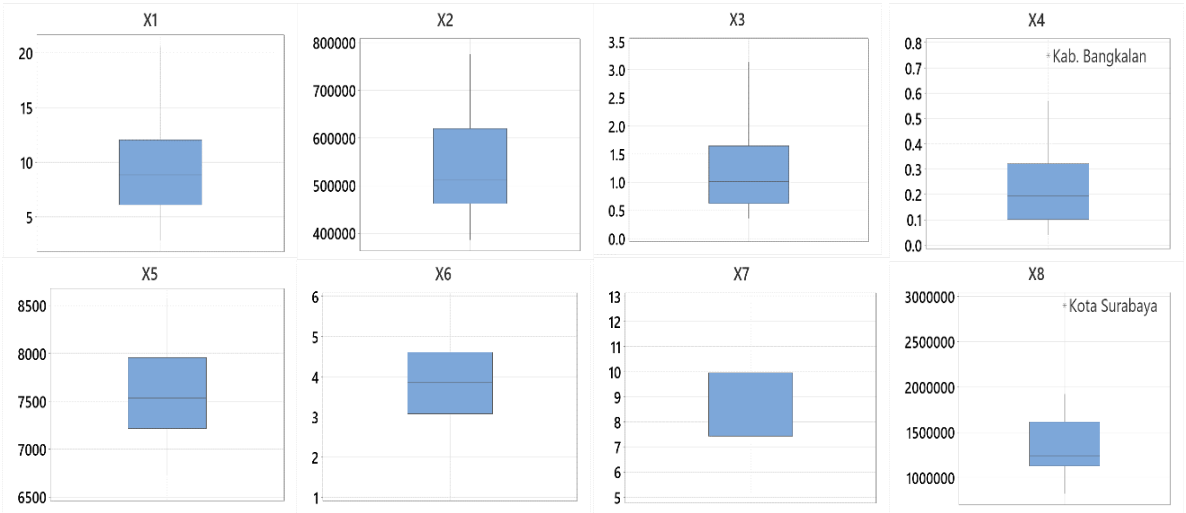
Langkah-langkah analisis dalam penelitian ini adalah sebagai berikut:

1. Mengumpulkan data indikator kemiskinan dan sosial ekonomi menurut kabupaten/kota di Provinsi Jawa Timur tahun 2025;
2. Melakukan analisis statistika deskriptif terhadap seluruh variable;
3. Melakukan standardisasi data variabel penelitian;
4. Melakukan analisis *hierarchical clustering* dengan tahapan sebagai berikut:
   a. Menghitung jarak antar kabupaten/kota menggunakan *euclidean distance* sebagai dasar pengukuran kemiripan objek
   b. Melakukan penggabungan klaster menggunakan *ward's method*
   c. Menentukan jumlah klaster optimal berdasarkan *dendrogram*
   d. Mengidentifikasi jumlah dan anggota klaster berdasarkan jumlah klaster optimal yang terbentuk
   e. Melakukan validasi hasil klaster untuk melihat konsistensi struktur pengelompokkan melalui visualisasi PCA
5. Melakukan analisis K-*Means clustering* dengan tahapan sebagai berikut:
   a. Menentukan jumlah klaster optimal menggunakan *elbow method*
   b. Menghitung jarak antara setiap objek dan setiap *centroid* menggunakan *euclidean distance*
   c. Mengelompokkan data ke klaster dengan *centroid* terdekat
   d. Menghitung ulang rata-rata *centroid* untuk kelompok yang baru terbentuk
   e. Mengulang proses hingga posisi *centroid* stabil atau iterasi maksimum tercapai
   f. Mengidentifikasi jumlah dan anggota klaster berdasarkan jumlah klaster optimal yang terbentuk
   g. Melakukan validasi hasil klaster untuk melihat konsistensi struktur pengelompokkan melalui visualisasi PCA
6. Membandingkan hasil pengelompokkan kabupaten/kota menggunakan *hierarchical clustering* dan K-*Means clustering* menggunakan *silhouette score*;
7. Menyimpulkan hasil yang diperoleh dari analisis yang telah dilakukan.

3 HASIL DAN PEMBAHASAN

3.1 Analisis Statistika Deskriptif

Analisis statistika deskriptif menunjukkan gambaran secara umum data yang digunakan pada penelitian ini. Berdasarkan analisis diperoleh hasil sebagai berikut:



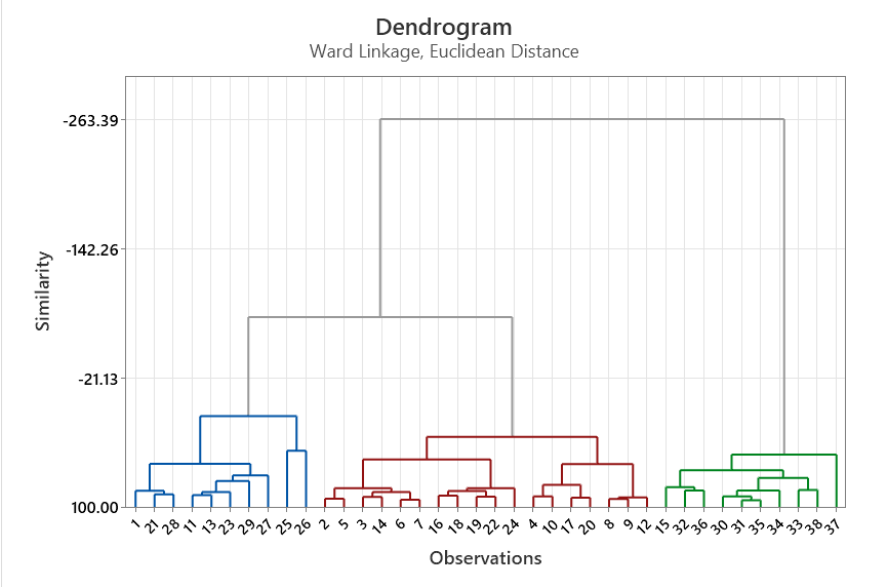
Gambar 1. Boxplot Indikator Kemiskinan dan Sosial Ekonomi

Gambar 1 menunjukkan bahwa variabel X_1, X_5, X_6, dan X_7 memiliki sebaran cukup baik dengan nilai median yang berada hampir di tengah, yang berarti distribusi data cenderung simetris. Selain itu, tidak terdapat data *outlier*. Variabel X_2 dan X_3 menyebar cukup luas. Nilai median mendekati kuartil bawah, mengindikasikan distribusi data cenderung ke kanan (*positively skewed*), artinya sebaran data cukup luas di bagian atas. Selain itu, tidak terdapat data *outlier*. Begitu pula pada variabel X_4 dan X_8 memiliki distribusi data cenderung miring ke kanan, dan terdapat data *outlier*, yaitu Kabupaten Bangkalan dan Kota Surabaya yang memiliki nilai lebih tinggi dibandingkan kabupaten/kota lain.

3.2 Analisis *Hierarchical Clustering*

3.2.1 Penentuan Jumlah Klaster

Jumlah klaster optimal ditentukan berdasarkan interpretasi *dendrogram*. Titik yang dipilih tepat sebelum terjadi lonjakan *agglomeration coefficient* yang besar agar diperoleh anggota yang homogen dalam klaster dan heterogen antarklaster (Everitt *et al.*, 2011).



Gambar 2. *Dendrogram Hierarchical Clustering*

Gambar 2 menunjukkan berdasarkan *dendrogram* dengan metode *Ward Linkage* pemotongan dilakukan pada ketinggian (*height*) sekitar -50 hingga -70, yaitu sebelum terjadi lonjakan penggabungan yang cukup besar pada tahap pembentukan 3 klaster, sehingga jumlah klaster optimal adalah sebanyak 3 klaster.

3.2.2 Hasil *Hierarchical Clustering*

Berdasarkan intepretasi *dendrogram* diperoleh jumlah klaster optimal sebanyak 3 klaster dengan jumlah anggota masing – masing klaster sebagai berikut:

Tabel 2. Jumlah Anggota *Hierarchical Clustering*

Klaster	Jumlah Anggota	Persentase
1	10	26,32%
2	18	47,36%
3	10	26,32%

Tabel 2 menunjukkan bahwa jumlah anggota klaster 1 sebanyak 10 kabupaten/kota, jumlah klaster 2 sebanyak 18 kabupaten/kota, dan jumlah klaster 3 sebanyak 10 kabupaten/kota. Seluruh kabupaten/kota yang ada di Provinsi Jawa Timur yaitu 38

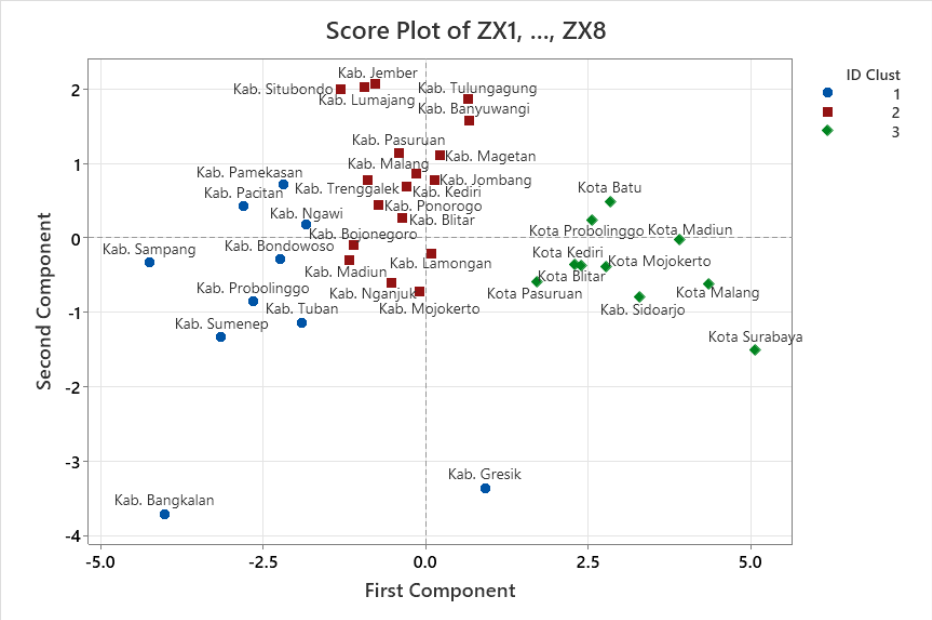
kabupaten/kota sudah masuk ke dalam keanggotaan klaster. Masing – masing anggota klaster tersebut adalah sebagai berikut:

Tabel 3. Anggota *Hierarchical Clustering*

Klaster 1	Klaster 2	Klaster 3
1. Kab. Pacitan	1. Kab. Ponorogo	1. Kab. Sidoarjo
2. Kab. Bondowoso	2. Kab. Trenggalek	2. Kota Kediri
3. Kab. Probolinggo	3. Kab. Tulungagung	3. Kota Blitar
4. Kab. Ngawi	4. Kab. Blitar	4. Kota Malang
5. Kab. Tuban	5. Kab. Kediri	5. Kota Probolinggo
6. Kab. Bangkalan	6. Kab. Malang	6. Kota Pasuruan
7. Kab. Sampang	7. Kab. Lumajang	7. Kota Mojokerto
8. Kab. Pamekasan	8. Kab. Jember	8. Kota Madiun
9. Kab. Sumenep	9. Kab. Banyuwangi	9. Kota Surabaya
10. Kab. Gresik	10. Kab. Situbondo	10. Kota Batu
	11. Kab. Pasuruan	
	12. Kab. Mojokerto	
	13. Kab. Jombang	
	14. Kab. Nganjuk	
	15. Kab. Madiun	
	16. Kab. Magetan	
	17. Kab. Bojonegoro	
	18. Kab. Lamongan	

3.2.3 Visualisasi PCA Hasil *Hierarchical Clustering*

Delapan indikator kemiskinan dan sosial ekonomi direduksi menjadi 2 komponen utama mampu menjelaskan 85% variasi data (PC1 = 65,2%; PC2 = 19,8%) dengan visualisasi sebagai berikut:



Gambar 3. Visualisasi PCA Hasil *Hierarchical Clustering*

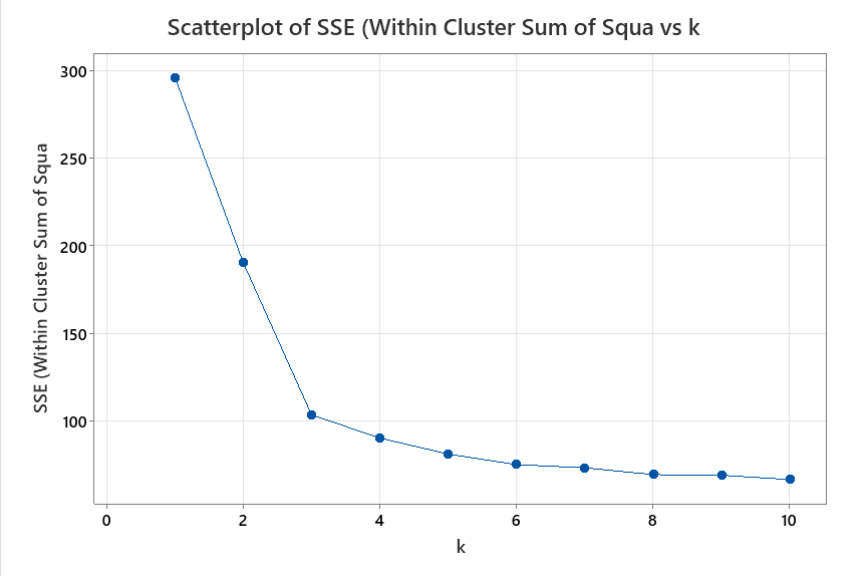
Gambar 3 menunjukkan bahwa klaster 1 beranggotakan kabupaten/kota dengan tingkat kemiskinan tinggi dan kondisi sosial ekonomi rendah. Pada klaster ini terdapat anggota yang berada cukup jauh dari anggota lainnya, yaitu Kabupaten Bangkalan dan Kabupaten Gresik,

sehingga dapat dikatakan kedua kabupaten ini memiliki karakteristik yang paling berbeda di antara anggota klaster tersebut. Perbedaan karakteristik tersebut yaitu pada indeks keparahan kemiskinan (X_4) yang jauh di atas kabupaten/kota lain. Klaster 2 beranggotakan kabupaten/kota dengan tingkat kemiskinan dan kondisi sosial ekonomi sedang. Seluruh anggotanya relatif berdekatan, menunjukkan karakteristik kabupaten/kota pada klaster ini cukup seimbang dengan kemiripan yang cukup tinggi. Sedangkan klaster 3 beranggotakan kabupaten/kota dengan tingkat kemiskinan rendah dan kondisi sosial ekonomi tinggi. Pada klaster ini, Kota Surabaya paling menonjol, menandakan adanya karakteristik khusus yang lebih *outlier*, yaitu pendapatan per kapita (X_8) yang jauh di atas kabupaten/kota lain.

3.3 Analisis K-*Means Clustering*

3.3.1 Penentuan Nilai K

Jumlah klaster optimal ditentukan dengan *Elbow Method* yaitu memplot nilai *Within Cluster Sum of Squares* (WCSS) terhadap jumlah klaster K. Berdasarkan analisis diperoleh hasil berikut:



Gambar 4. Penentuan Nilai K (*Elbow Method)*

Gambar 4 menunjukkan bahwa nilai WCSS konsisten menurun, dimana penurunan sangat tajam dari K=1 hingga K=3 menunjukkan bahwa penambahan jumlah klaster pada tahap tersebut mampu menaikkan homogenitas data secara signifikan. Namun pada K =3 penurunan nilai WCSS semakin melandai di mana titik siku (*elbow point*) terlihat pada K=3. Oleh karena itu, jumlah klaster optimal yang digunakan ditentukan sebanyak 3 klaster.

3.3.2 Hasil K-*Means Clustering*

Analisis K-*Means clustering* dengan menggunakan 3 klaster diperoleh hasil pengelompokkan sebagai berikut:

Tabel 4. Jumlah Anggota Klaster

Klaster	Jumlah Anggota	Persentase
1	9	23.68%
2	11	28.95%
3	18	47.37%

Tabel 4 menunjukkan bahwa klaster 1 terdiri dari 9 kabupaten/kota, klaster 2 terdiri dari 11 kabupaten/kota, dan klaster 3 terdiri dari 18 kabupaten/kota. Total 38 kabupaten/kota yang ada di Provinsi Jawa Timur telah masuk ke dalam keanggotaan klaster. Anggota masing – masing klaster tersebut adalah sebagai berikut:

Tabel 5. Anggota K-*Means Clustering*

Klaster 1	Klaster 2	Klaster 3
1. Kab. Pacitan	1. Kab. Sidoarjo	1. Kab. Ponorogo
2. Kab. Bondowoso	2. Kab. Gresik	2. Kab. Trenggalek
3. Kab. Probolinggo	3. Kota Kediri	3. Kab. Tulungagung
4. Kab. Ngawi	4. Kota Blitar	4. Kab. Blitar
5. Kab. Tuban	5. Kota Malang	5. Kab. Kediri
6. Kab. Bangkalan	6. Kota Probolinggo	6. Kab. Malang
7. Kab. Sampang	7. Kota Pasuruan	7. Kab. Lumajang
8. Kab. Pamekasan	8. Kota Mojokerto	8. Kab. Jember
9. Kab. Sumenep	9. Kota Madiun	9. Kab. Banyuwangi
	10. Kota Surabaya	10. Kab. Situbondo
	11. Kota Batu	11. Kab. Pasuruan
		12. Kab. Mojokerto
		13. Kab. Jombang
		14. Kab. Nganjuk
		15. Kab. Madiun
		16. Kab. Magetan
		17. Kab. Bojonegoro
		18. Kab. Lamongan

Tabel 5 menunjukkan bahwa berdasarkan analisis K-*Means clustering* diperoleh hasil pengelompokan yang mirip dengan analisis *hierarchical clustering*, meskipun terdapat satu perbedaan pengelompokan, yaitu Kabupaten Gresik pada klaster 2. Hasil pengelompokan kedua metode tersebut menunjukkan struktur klaster yang relatif konsisten, meskipun kualitas pemisahan antarklaster masih berada pada kategori cukup.

3.3.3 Hasil *Centroid* K-*Means Clustering*

Hasil *centroid* K-*Means clustering* dihitung berdasarkan skala asli dari variabel yang digunakan, di mana nilai terbesar menunjukkan variabel yang dominan pada klaster tersebut. Hasil *centroid* K-*Means clustering* adalah sebagai berikut:

Tabel 6. Hasil *Centroid* K-*Means Clustering*

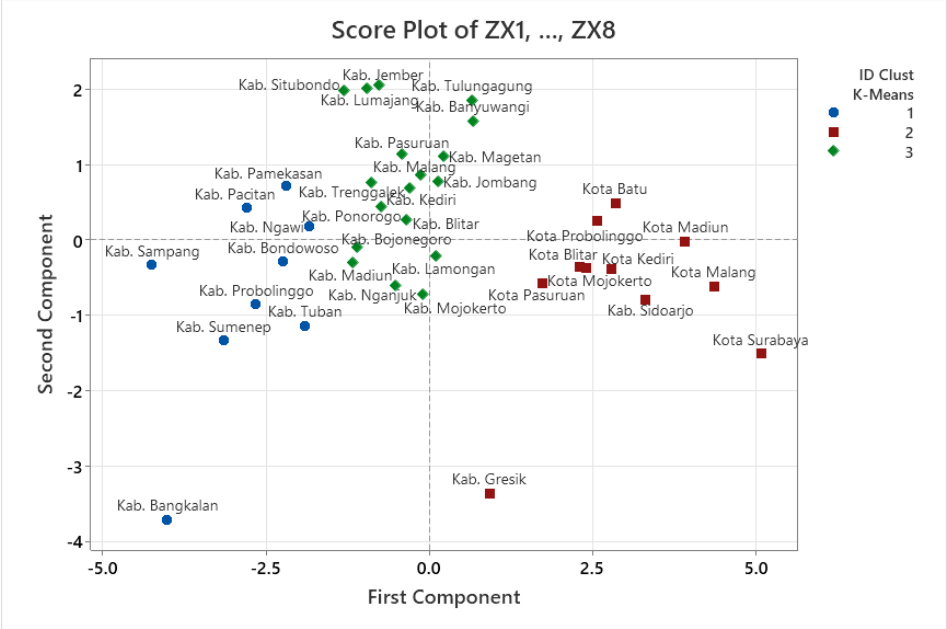
Variabel	Klaster 1	Klaster 2	Klaster 3
X_1	**11.3936**	4.7517	8.1100
X_2	476,901.4545	**638,993.6667**	566,976.6000
X_3	**1.3214**	0.6833	1.0050
X_4	**0.2495**	0.1717	0.2120
X_5	7,323.7273	**8,328.5000**	7,795.2000
X_6	3.3632	**4.9283**	4.0070
X_7	7.5641	**10.9417**	9.3030
X_8	1.12407×10^6	$\mathbf{2.06156 \times 10^6}$	1.54394×10^6

Sumber: *Output Software Minitab*

Tabel 6 menunjukkan bahwa klaster 1 mempunyai 3 variabel dengan nilai *centroid* tertinggi, yaitu variabel X_1, X_3, dan X_4, sedangkan variabel lainnya mempunyai nilai *centroid* terendah dibandingkan dengan klaster lainnya. Kondisi ini memperlihatkan bahwa pada klaster 1 mempunyai tingkat kemiskinan tinggi dan kondisi sosial ekonomi rendah. Klaster 2 mempunyai 5 variabel dengan nilai *centroid* tertinggi, yaitu X_2, X_5, X_6, X_7, dan X_8, sedangkan variabel lainnya mempunyai nilai *centroid* terendah dibandingkan dengan klaster lainnya. Kondisi ini menunjukkan bahwa pada klaster 2 mempunyai tingkat kemiskinan rendah dan kondisi sosial ekonomi tinggi. Sementara itu, klaster 3 mempunyai nilai *centroid* di antara kedua klaster untuk seluruh variabel. Kondisi ini menunjukkan bahwa pada klaster 3 mempunyai tingkat kemiskinan dan kondisi sosial ekonomi yang sedang.

3.3.4 Visualisasi PCA Hasil K-*Means Clustering*

Hasil dari visualisasi PCA berdasarkan K-*Means clustering* adalah sebagai berikut:



Gambar 5. Visualisasi PCA Hasil K-*Means Clustering*

Gambar 5 menunjukkan bahwa klaster 1 beranggotakan kabupaten/kota dengan tingkat kemiskinan tinggi dan kondisi sosial ekonomi rendah. Pada klaster ini terdapat anggota yang cukup jauh dari anggota lainnya, yaitu Kabupaten Bangkalan yang mempunyai karakteristik yang paling berbeda di antara anggota klaster 1. Perbedaan karakteristik tersebut yaitu pada indeks keparahan kemiskinan (X_4) yang jauh di atas kabupaten/kota lain. Klaster 2 beranggotakan kabupaten/kota dengan tingkat kemiskinan rendah dan kondisi sosial ekonomi tinggi. Pada klaster ini, Kabupaten Gresik dan Kota Surabaya tampak menjadi kabupaten/kota yang cukup menonjol dibandingkan dengan anggota lainnya, menandakan adanya karakteristik khusus yang lebih *outlier*, yaitu pendapatan per kapita (X_8) yang jauh di atas kabupaten/kota lain. Serta klaster 3 beranggotakan kabupaten/kota dengan tingkat kemiskinan dan keadaan sosial ekonomi sedang yang tampak bergerombol di tengah grafik. Seluruh anggotanya relatif berdekatan, menunjukkan karakteristik kabupaten/kota pada klaster ini cukup seimbang dengan kemiripan yang cukup tinggi.

Baik *hierarchical clustering* dan K-*Means clustering* menghasilkan jumlah klaster optimal yang sama, yaitu 3 klaster. Hanya saja, terdapat perbedaan pada pengelompokan anggota, yaitu Kabupaten Gresik pada *hierarchical clustering* masuk ke klaster 1, sedangkan pada K-*Means clustering* masuk ke klaster 2. Terlihat pada Gambar 5, Kabupaten Gresik berada cukup terpisah dari kabupaten/kota lain atau berada di zona *borderline* dengan

karakteristik yang berbeda, sehingga penggantian metode *clustering* dapat menjadi penyebab perpindahan klaster. Kasus serupa pernah terjadi pada penelitian yang dilakukan oleh Mulyana *et al*. (2024) yang menyatakan bahwa perbedaan hasil *clustering* dikarenakan kedua metode mempunyai algoritma dengan pendekatan yang berbeda, di mana *hierarchical clustering* menghasilkan klaster dengan bentuk yang lebih kompleks, sedangkan K-*Means clustering* menghasilkan klaster yang lebih bulat dan terpisah.

3.4 Perbandingan Metode *Hierarchical Clustering* dengan K-*Means Clustering*

Perbandingan metode *hierarchical clustering* dan K-*means clustering* dilakukan menggunakan *silhouette score* untuk mengukur tingkat kemiripan objek dalam satu klaster. Hasil analisis perbandingan kedua metode tersebut sebagai berikut:

Tabel 7. Nilai *Silhouette*

Metode	Nilai *Silhouette*
Hierarchical clustering	0.3861498
K-*Means clustering*	0.3861498

Sumber: *Output Software RStudio*

Tabel 7 menunjukkan bahwa nilai *silhouette* yang diperoleh dari metode *hierarchical clustering* dan K-*Means clustering* adalah sama, yaitu 0.3861498. Nilai tersebut berada pada rentang $0.26 \leq s(i) \leq 0.50$, yang menunjukkan bahwa kualitas klaster yang terbentuk berada pada kategori cukup baik (*fair structure*), di mana objek dalam satu klaster memiliki kemiripan yang cukup, namun pemisahan antarklaster belum terlalu kuat (Utami *et al.*, 2023). Kesamaan nilai *silhouette* pada kedua metode menunjukkan bahwa hasil pengelompokan relatif konsisten dan struktur data yang terbentuk cukup stabil meskipun menggunakan algoritma *clustering* yang berbeda, meskipun terdapat satu perbedaan pengelompokan yaitu Kabupaten Gresik.

4 KESIMPULAN

Berdasarkan hasil analisis pengelompokkan kabupaten/kota di Provinsi Jawa Timur tahun 2025 terhadap delapan indikator kemiskinan dan sosial ekonomi menggunakan metode *hierarchical clustering* dan K-*Means clustering*, diperoleh kesimpulan bahwa kedua metode menunjukkan pola klaster yang hampir sama. Jumlah klaster optimal sebanyak tiga, dengan perbedaan pengelompokan hanya pada satu anggota, yaitu Kabupaten Gresik. Kedua metode tersebut mendapatkan nilai *silhouette* yang sama, yaitu 0.3861498 (kategori cukup baik), menunjukkan bahwa objek dalam satu klaster memiliki tingkat kemiripan yang cukup meskipun pemisahan antarklaster belum terlalu kuat. Masing-masing klaster memiliki karakteristik kemiskinan dan sosial ekonomi yang berbeda. Oleh karena itu, Pemerintah Provinsi Jawa Timur diharapkan dapat merumuskan kebijakan dan program penanggulangan kemiskinan yang disesuaikan dengan karakteristik masing-masing klaster. Kabupaten/kota yang tergabung dalam klaster dengan tingkat kemiskinan tinggi dan kondisi sosial ekonomi rendah perlu menjadi prioritas utama dalam penyusunan program pembangunan. Sementara itu, kabupaten/kota yang tergolong dalam klaster dengan tingkat kemiskinan rendah dan kondisi sosial ekonomi tinggi diharapkan dapat mempertahankan dan meningkatkan kualitas pembangunan daerah melalui penguatan sektor ekonomi produktif, peningkatan kualitas sumber daya manusia, dan pemerataan kesejahteraan masyarakat.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Program Studi Statistika Universitas Terbuka, dosen pembimbing, serta Badan Pusat Statistik Provinsi Jawa Timur yang telah menyediakan data dalam penyusunan artikel ilmiah ini. Serta kepada keluarga dan seluruh pihak yang telah

memberikan dukungan, motivasi, dan bantuan sehingga penelitian dapat terselesaikan dengan baik.

DAFTAR PUSTAKA

Amalia, S., Nikmah, A., & Siniwi, L.M. (2025). Pengelompokkan Kemiskinan di Provinsi Jawa Timur Tahun 2024 dengan Pendekatan Algoritma K-Menas dan K-Means ++. *MUTIARA: Multidiciplinary Scientifict Journal*, 3(8), 815-823. https://mutiara.al-makkipublisher.com/index.php/al/article/view/407.

Amelia, M., Faqih, A., & Rinaldi, A. R. (2025). Penerapan Metode *K-Means Clustering* dalam Pemetaan Kemiskinan Kabupaten/Kota di Indonesia untuk Perencanaan Kebijakan yang Tepat. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(2), 437-444. https://doi.org/10.23960/jitet.v13i2.6231.

Badan Pusat Statistik Provinsi Jawa Timur. Persentase Penduduk Miskin Menurut Kabupaten/Kota di Jawa Timur (Persen) Tahun 2025. Retrieved April 22, 2026, from jatim.bps.go.id.https://jatim.bps.go.id/id/statistics-table/2/NDk3IzI=/persentase-penduduk-miskin-menurut-kabupaten-kota-di-jawa-timur.html.

El Adawiyah, S., Hermanto, A., Yasya, W., Kristanti R., & Chrisye, M. (2021). Akses terhadap Sumber Daya Alam pada Kemiskinan dan Ketahanan Pangan. *Sosio Informa*, 7(2), 172-185. https://ejournal.poltekesos.ac.id/index.php/Sosioinforma/article/view/2664.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). John Wiley & Sons.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage Learning.

Hasan, Y. (2024). Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil *Cluster* K-*Means* dan DBSCAN. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(3S1). https://ejournal.ust.ac.id/index.php/KAKIFIKOM/article/view/3938

Mulyana, A. A. R., Juwita, A. R., Siregar, A. M., & Fauzi, A. (2024). Penerapan Algoritma K-*means Clustering dan Hierarchical Clustering* dalam Mengelompokkan Data Pengangguran di Karawang. *Jurnal Algoritma*, 21(2), 209–220. https://jurnal.itg.ac.id/index.php/algoritma/article/view/2155.

Utami, I. T., Suryaningrum, F., & Ispriyanti, D. (2023). K-*Means Cluster Count Optimization with Silhouette Index Validation and Davies Bouldin Index (Case Study: Coverage of Pregnant Women, Childbirth, And Postpartum Health Services in Indonesia in 2020).* BAREKENG*: Jurnal Ilmu Matematika dan Terapan*, 17(2), 707-716. https://doi.org/10.30598/barekengvol17iss2pp0707-0716.

Wulandari, Elisha. A., Baso, Andi M. Alfin., Fajri, Belia Nurul., Kalondeng, Anisa., Islamiyati, Anna., Pannu, Abdullah., Fadil, Muhammad., Vallarino, Alfian Akbar., & Rahman, Anugrah N, 1. (2025). Pengelompokan Kemiskinan di Provinsi Sulawesi Selatan Tahun 2023 dengan Metode K-*Means Clustering*. *ESTIMASI: Journal of Statistics and its Application*, 7(2), 281-285. https://journal.unhas.ac.id/index.php/ESTIMASI/article/view/45824/13232.

Yulisasih, B. N., Herman., Sunardi., & Yuliansyah, H. (2024). *Evaluation of K-Means Clustering Using Silhouette Score Method on Customer Segmentation*. *ILKOM Jurnal Ilmiah.* 16(3), 330-342. https://jurnal.fikom.umi.ac.id/index.php/ILKOM/article/view/2325?utm.