

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

## Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Artikel penelitian

# Klasterisasi Wilayah Kemiskinan Jawa Tengah Menggunakan K-Means Berbasis Indikator Sosial-Ekonomi

Moh. Fachri Alif<sup>a</sup>, Amiq Fahmi<sup>b\*</sup>

<sup>a</sup> Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia

<sup>b</sup> Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia

### INFORMASI ARTIKEL

#### Sejarah Artikel:

Diterima Redaksi: 05 September 2025

Revisi Akhir: 30 Desember 2025

Diterbitkan Online: 01 Januari 2026

### KATA KUNCI

Klasterisasi Wilayah Kemiskinan;  
Indikator Sosial Ekonomi;  
Algoritma K-Means;  
Metode Elbow;  
Principal Component Analysis (PCA).

### KORESPONDENSI

E-mail: [amiq.fahmi@dsn.dinus.ac.id](mailto:amiq.fahmi@dsn.dinus.ac.id)\*

### A B S T R A C T

Kemiskinan merupakan isu multidimensi yang berdampak signifikan terhadap kualitas pembangunan wilayah, khususnya di Provinsi Jawa Tengah yang menempati urutan ketiga secara nasional. Meskipun data sosial ekonomi tersedia secara melimpah dan terbuka, pemanfaatannya untuk segmentasi wilayah serta perumusan kebijakan berbasis data, informasi, dan pengetahuan masih terbatas. Penelitian ini bertujuan untuk melakukan klasterisasi kabupaten/kota di Provinsi Jawa Tengah menggunakan algoritma K-Means, berdasarkan tujuh indikator utama sosial ekonomi: Indeks Pembangunan Manusia (IPM), proporsi penduduk miskin ekstrem (Prioritas 1), sangat miskin (Prioritas 2), pengeluaran per kapita, upah minimum kabupaten/kota (UMK), tingkat pengangguran terbuka, dan jumlah rumah tidak layak huni (RTLH). Data yang digunakan merupakan data sekunder tahun 2024 yang diperoleh dari Badan Pusat Statistik dan Dinas Sosial Provinsi Jawa Tengah dan telah melalui proses normalisasi. Penentuan jumlah kluster optimal dilakukan dengan metode Elbow, sedangkan Principal Component Analysis (PCA) digunakan sebagai teknik visualisasi. Pendekatan kuantitatif dan interpretatif ini memastikan bahwa kluster yang terbentuk bersifat optimal secara statistik, mudah dijelaskan secara visual, dan relevan untuk ditindaklanjuti dalam kebijakan. Hasil analisis menunjukkan terbentuknya tiga kluster wilayah dengan karakteristik sosial ekonomi yang berbeda secara signifikan, yaitu wilayah berkembang, wilayah transisi, dan wilayah prioritas pengentasan kemiskinan. Temuan hasil penelitian ini, yang mengintegrasikan multi-indikator sosial ekonomi dengan pendekatan visual dan analitis, mampu menghasilkan segmentasi wilayah yang lebih akurat dan aplikatif bagi penyusunan kebijakan pembangunan wilayah yang lebih berkeadilan, dengan penekanan pada intervensi intensif terhadap kabupaten/kota yang memiliki tingkat kemiskinan ekstrem dan sangat miskin.

## 1. PENDAHULUAN

Kemiskinan merupakan persoalan kompleks yang memberikan dampak luas terhadap pembangunan sosial dan ekonomi, terutama di wilayah dengan tingkat ketimpangan yang tinggi seperti Provinsi Jawa Tengah [1]. Meskipun pemerintah telah melaksanakan berbagai program, seperti Program Keluarga Harapan (PKH), Bantuan Pangan Non Tunai (BPNT), serta bantuan pembangunan rumah layak huni secara masif, efektivitas pelaksanaannya belum sepenuhnya optimal [2]. Hal ini disebabkan oleh lemahnya akurasi dalam distribusi bantuan, yang masih memerlukan evaluasi dan penyesuaian berbasis kondisi nyata di lapangan. Oleh karena itu, pendekatan berbasis data

sangat diperlukan guna mengidentifikasi wilayah prioritas pengentasan kemiskinan secara objektif, transparan, dan efisien [3].

Penelitian-penelitian sebelumnya umumnya menggunakan metode regresi untuk memprediksi tingkat kemiskinan [4], [5], [6]. Namun, pendekatan tersebut belum mampu menghasilkan segmentasi wilayah berdasarkan kesamaan karakteristik sosial ekonomi, padahal segmentasi semacam ini sangat penting dalam perumusan kebijakan berbasis kewilayahan. Menurut beberapa studi [2], [7], [8], [9], metode *unsupervised learning*, seperti *K-Means Clustering* menjadi solusi yang semakin banyak digunakan karena kemampuannya dalam mengelompokkan data tidak berlabel berdasarkan kemiripan atribut. Atribut yang

digunakan antara lain Indeks Pembangunan Manusia (IPM), tingkat pengangguran terbuka, upah minimum kabupaten/kota (UMK), pengeluaran per kapita, serta indeks kedalaman kemiskinan berdasarkan kategori Prioritas (P), yang mencakup kemiskinan ekstrem (P1), sangat miskin (P2), dan miskin (P3) [1], [2], [3].

Studi yang dilakukan oleh [7] menunjukkan bahwa algoritma K-Means efektif dalam menganalisis dan mengelompokkan wilayah berdasarkan indikator sosial ekonomi di tingkat kota/kabupaten, sehingga mendukung penerapannya dalam analisis spasial kemiskinan. Penggunaan K-Means melalui platform RapidMiner telah banyak diimplementasikan untuk segmentasi wilayah administratif, termasuk pada 35 kabupaten/kota di Provinsi Jawa Tengah. Evaluasi hasil kluster dapat dilakukan dengan mengukur nilai Davies-Bouldin Index (DBI), yang merepresentasikan kualitas pemisahan antar kluster secara kuantitatif. Penelitian oleh [10] membandingkan algoritma K-Means dan Average Linkage dalam klusterisasi kemiskinan, dan menyimpulkan bahwa K-Means lebih unggul dari segi stabilitas dan efisiensi. Sementara itu, penelitian [11] mengintegrasikan metode McQuitty Linkage dan K-Means untuk mengelompokkan kabupaten di Jawa Tengah berdasarkan indikator sosial ekonomi. Penelitian lain oleh [12] menggunakan algoritma Fuzzy C-Means dan berhasil membentuk lima kluster kemiskinan dengan karakteristik sosial ekonomi yang berbeda secara signifikan. Lebih lanjut, studi oleh [13] dan [14] mengklasifikasikan seluruh provinsi di Indonesia berdasarkan indikator kemiskinan, tingkat keparahan kemiskinan, tingkat literasi, lama sekolah, dan tingkat pengangguran, menggunakan pendekatan hierarchical agglomerative clustering, yang menghasilkan tiga kategori wilayah: rendah, sedang, dan tinggi.

Munandar [15] menerapkan algoritma clustering di Provinsi Banten untuk memetakan wilayah yang rentan terhadap kemiskinan. Sepriyanti et al. [9] menggunakan metode K-Means dalam klusterisasi wilayah miskin di Provinsi Riau, yang berhasil membagi kabupaten/kota ke dalam tiga kelompok berdasarkan indikator sosial ekonomi yang relevan. Zahra et al. [16] mengadopsi pendekatan klasifikasi data sosial berbasis opini publik. Di sisi lain, studi oleh [14] berhasil melakukan klusterisasi kabupaten/kota di Pulau Jawa dengan menggunakan hierarchical clustering berbasis faktor kemiskinan. Meskipun studi-studi tersebut menunjukkan efektivitas metode clustering dalam segmentasi wilayah miskin, sebagian besar masih terbatas pada visualisasi hasil kluster tanpa eksplorasi karakteristik tiap kelompok secara sistematis. Selain itu, hanya sedikit penelitian yang memasukkan variabel tambahan seperti jumlah rumah tidak layak huni (RTLH) dan upah minimum kabupaten/kota (UMK) sebagai indikator penting dalam proses klusterisasi. Menurut [17], validasi terhadap metode yang digunakan juga belum banyak diperkuat melalui pendekatan visual berbasis Principal Component Analysis (PCA) maupun pengujian kestabilan kluster. Oleh karena itu, diperlukan strategi klusterisasi yang lebih komprehensif dan berbasis multi-indikator guna menghasilkan segmentasi wilayah yang akurat serta rekomendasi kebijakan yang aplikatif dan berkeadilan.

Tujuan dari penelitian ini adalah: (1) mengelompokkan kabupaten/kota di Provinsi Jawa Tengah berdasarkan indikator sosial ekonomi menggunakan algoritma K-Means; (2)

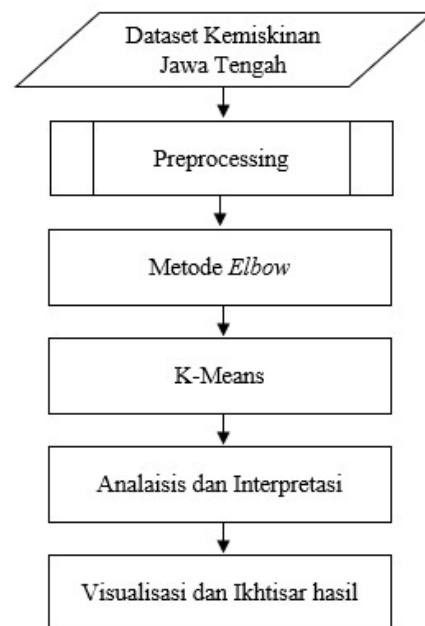
mengidentifikasi karakteristik masing-masing kluster sebagai dasar penentuan wilayah prioritas dalam program pengentasan kemiskinan; dan (3) merumuskan rekomendasi kebijakan berbasis hasil segmentasi wilayah guna mendukung perencanaan dan intervensi pembangunan yang lebih tepat sasaran dan berkeadilan.

Adapun kebaruan dalam penelitian ini terletak pada integrasi indikator sosial ekonomi tambahan yang jarang digunakan secara simultan, seperti Upah Minimum Kabupaten/Kota (UMK) dan jumlah Rumah Tidak Layak Huni (RTLH), ke dalam model klusterisasi berbasis algoritma K-Means. Selain itu, validasi hasil kluster dilakukan secara visual melalui pendekatan Principal Component Analysis (PCA), yang dipadukan dengan analisis deskriptif mendalam untuk menginterpretasikan karakteristik masing-masing kluster secara sistematis. Pendekatan ini memberikan kontribusi nyata dalam penguatan pengambilan keputusan berbasis data (evidence-based policy), khususnya dalam identifikasi wilayah prioritas pembangunan di Provinsi Jawa Tengah secara lebih akurat, transparan, dan aplikatif.

## 2. METODE

### 2.1. Rancangan Penelitian

Penelitian ini dirancang untuk mengidentifikasi pengelompokan wilayah prioritas pengentasan kemiskinan di Provinsi Jawa Tengah secara objektif dan berbasis data. Untuk mencapai tujuan tersebut, algoritma K-Means Clustering digunakan sebagai metode analisis segmentatif yang mampu membagi wilayah ke dalam kluster homogen berdasarkan karakteristik sosial ekonomi yang relevan. Proses implementasi algoritma K-Means Clustering digambarkan melalui presentasi diagram Gambar 1.



Gambar 1. Diagram Alur

Gambar 1 menyajikan diagram alur sistematis yang menggambarkan tahapan implementasi algoritma K-Means Clustering dalam penelitian ini. Diagram tersebut mencakup proses utama, mulai dari pengumpulan dataset dan tahap pre-

processing (meliputi normalisasi data dan ekstraksi fitur), penentuan jumlah kluster optimal menggunakan metode Elbow, hingga analisis serta interpretasi hasil pengelompokan wilayah prioritas pengentasan kemiskinan di Provinsi Jawa Tengah. Tahapan akhir berupa visualisasi hasil klasterisasi, yang berfungsi sebagai representasi komprehensif atas prosedur teknis yang diterapkan dalam studi ini. Setiap tahapan penerapan metode K-Means Clustering pada Gambar 1 dijelaskan secara rinci sebagai berikut:

#### 1. Pengumpulan Data

Langkah awal adalah mengumpulkan data dari sumber resmi, yaitu :

- Badan Pusat Statistik (BPS): menyediakan data kemiskinan, pengeluaran per kapita, dan Indeks Pembangunan Manusia (IPM).
- Dinas Sosial Provinsi Jawa Tengah: menyajikan data rumah tidak layak huni (RTLH) serta program bantuan sosial lainnya.
- Portal Data Terbuka Pemerintah (data.go.id) digunakan untuk melengkapi data jika ada variabel yang belum tersedia di sumber utama.

Data dikumpulkan dalam format tabular (.CSV atau .XLSX), yang memuat informasi dari 35 kabupaten/kota di Jawa Tengah.

#### 2. Pra-pemrosesan Data

Data yang telah dikumpulkan perlu diproses agar siap digunakan dalam model machine learning. Tahapan ini meliputi:

- Pemilihan fitur (feature selection), hanya variabel kuantitatif dan relevan dengan indikator sosial ekonomi yang digunakan.
- Pembersihan data (data cleaning), menghapus atau mengisi nilai kosong (missing values) dengan nilai rata-rata (mean) atau median.
- Normalisasi data dilakukan dengan menggunakan metode *StandardScaler* agar seluruh variabel memiliki skala yang seragam. Normalisasi ini penting karena K-Means sangat sensitif terhadap perbedaan skala antar fitur.

#### 3. Penentuan Jumlah Cluster (Metode Elbow)

Sebelum algoritma K-Means dijalankan, perlu ditentukan terlebih dahulu jumlah cluster optimal (k) menggunakan metode Elbow.

- Metode Elbow dilakukan dengan memvisualisasikan nilai inertia atau jumlah kuadrat jarak dalam kluster (*Within-Cluster Sum of Squares/ WCSS*) untuk setiap jumlah kluster (k), mulai dari 1 hingga nilai maksimum tertentu. Nilai WCSS tersebut dapat dihitung menggunakan rumus (1).

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} (x - \mu_i)^2 \quad (1)$$

Formula (1) menghitung total kuadrat jarak antara setiap titik data x dalam kluster  $C_i$  terhadap centroid-nya  $\mu_i$ , untuk seluruh kluster  $i=1$  hingga k. Komponen  $\|x - \mu_i\|^2$  merepresentasikan jarak Euclidean kuadrat dari titik ke pusat kluster. Semakin kecil hasil perhitungan WCSS, maka anggota kluster semakin dekat terhadap pusatnya, menandakan bahwa kluster tersebut lebih kompak dan homogen.

- Elbow point merujuk pada titik pada grafik WCSS (Within-Cluster Sum of Squares) di mana penurunan nilai mulai melambat secara signifikan. Titik ini biasanya menjadi indikator jumlah kluster yang paling optimal. Sebagai contoh, jika penurunan WCSS yang tajam terjadi dari  $k = 2$  ke  $k = 3$  dan setelahnya cenderung stabil, maka jumlah kluster yang optimal kemungkinan berada pada  $k = 3$ .

#### 4. Implementasi K-Means

Setelah k ditentukan, dilakukan proses klasterisasi menggunakan algoritma K-Means. Implementasi dilakukan dengan pustaka scikit-learn dalam bahasa pemrograman Python, dengan parameter:

- `n_clusters = 3`
  - `init = 'k-means++'`
  - `random_state = 42`
- Langkah-langkah algoritma K-Means meliputi :
- Inisialisasi centroid awal secara acak (k-means++).
  - Menghitung jarak Euclidean dari setiap data ke centroid. Dapat menggunakan formula (2).

$$d(x, y) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \quad (2)$$

Rumus (2) digunakan untuk menghitung jarak antara data x dan pusat kluster y. Semakin kecil nilai jarak tersebut, semakin dekat data ke centroid, sehingga semakin besar kemungkinan data tersebut termasuk dalam kluster yang bersangkutan. Perhitungan ini merupakan komponen inti dalam algoritma K-Means, yang berfungsi untuk menentukan pembagian kluster secara iteratif berdasarkan kemiripan atribut.

- Mengelompokkan data ke centroid terdekat.
- Memperbarui posisi centroid berdasarkan rata-rata data dalam cluster.
- Mengulangi proses hingga posisi centroid tidak berubah (konvergen).

#### 5. Analisis Interpretasi

Setelah proses klasterisasi selesai, hasilnya dianalisis menggunakan pendekatan deskriptif. Setiap kluster diidentifikasi berdasarkan nilai rata-rata dari sejumlah indikator sosial ekonomi, seperti proporsi penduduk miskin ekstrem (P1), sangat miskin (P2), Indeks Pembangunan Manusia (IPM), dan pengeluaran per kapita. Berdasarkan karakteristik dominan dari masing-masing kluster, dilakukan pelabelan interpretatif guna memudahkan pemahaman terhadap pola segmentasi yang terbentuk. Sebagai contoh, kluster dengan nilai IPM tinggi dan pengeluaran per kapita yang besar dapat dilabeli sebagai “kluster sejahtera,” sedangkan kluster dengan nilai indikator yang rendah dikategorikan sebagai “kluster rentan secara sosial ekonomi.”

- Cluster 0: wilayah dengan tingkat kemiskinan dan pengangguran tinggi, IPM rendah (wilayah sangat miskin).
- Cluster 1: wilayah dengan kondisi sedang atau transisi.
- Cluster 2: wilayah dengan indikator sosial ekonomi lebih baik (wilayah berkembang).

Hasil interpretasi ini menjadi dasar strategis dalam mengidentifikasi wilayah yang membutuhkan intervensi pengentasan kemiskinan secara lebih intensif dan tepat sasaran.

## 6. Visualisasi dan Kesimpulan

Tahapan akhir adalah menyajikan hasil klusterisasi dalam bentuk visual dan menyusun simpulan serta rekomendasi.

- a) Visualisasi : dilakukan menggunakan scatter plot berbasis dua komponen utama hasil Principal Component Analysis (PCA), untuk menampilkan sebaran antar cluster secara dua dimensi.
- b) Kesimpulan dan rekomendasi kebijakan disusun berdasarkan karakteristik masing-masing kluster, yang diidentifikasi melalui analisis indikator sosial ekonomi. Hasil ini dapat menjadi bahan pertimbangan bagi pemerintah daerah dalam merumuskan perencanaan anggaran sosial yang lebih tepat sasaran, bagi pengambil kebijakan nasional dalam menentukan strategi intervensi wilayah yang berbasis data, serta bagi lembaga non-pemerintah yang bergerak di bidang pemberdayaan masyarakat untuk merancang program yang sesuai dengan kebutuhan lokal.

## 2.2. Sumber Data dan Teknik Pengumpulan Data

Penelitian ini memanfaatkan data sekunder yang diperoleh dari sumber-sumber resmi pemerintah yang valid. Seluruh data dikumpulkan dari instansi yang memiliki otoritas dalam penyediaan informasi statistik dan sosial ekonomi daerah. Adapun sumber data yang digunakan berasal dari:

- 1) Badan Pusat Statistik (BPS) Provinsi Jawa Tengah, yang menyediakan data makro terkait kemiskinan yang meliputi ukuran kedalaman (P1) dan keparahan kemiskinan (P2), serta estimasi rata-rata pengeluaran individu per kapita.
- 2) Dinas Sosial Provinsi Jawa Tengah, yang memberikan data terkait jumlah rumah tidak layak huni (RTLH) dan informasi tambahan mengenai distribusi bantuan sosial.
- 3) Portal data.go.id, yang menyediakan data terbuka terkait Penelitian memanfaatkan indikator makro seperti IPM, tingkat pengangguran terbuka, dan upah minimum regional sebagai dasar segmentasi wilayah. Data dikumpulkan dalam bentuk tabular (CSV dan Excel), kemudian disatukan dan disesuaikan untuk mencakup 35 kabupaten/kota di Provinsi Jawa Tengah sebagai objek penelitian. Teknik pengumpulan data dilakukan dengan metode dokumentasi dan pengunduhan langsung dari situs resmi masing-masing instansi.

## 2.3. Variabel Penelitian

Penelitian ini menggunakan sejumlah variabel kuantitatif sebagai dasar segmentasi wilayah. Adapun variabel yang digunakan dalam proses klusterisasi meliputi:

- 1) P1 (Indeks Kedalaman Kemiskinan)
- 2) P2 (Indeks Keparahannya Kemiskinan)
- 3) Pengeluaran per Kapita (Rp)
- 4) Upah Minimum Kabupaten/Kota (UMK)
- 5) Jumlah Rumah Tidak Layak Huni (RTLH)
- 6) Tingkat Pengangguran Terbuka (TPT) (%)
- 7) Indeks Pembangunan Manusia (IPM)

Seluruh variabel yang digunakan dalam penelitian ini bersifat numerik dan memiliki satuan yang berbeda-beda. Oleh karena itu, diperlukan tahap normalisasi guna menyeragamkan skala antar

variabel, sehingga proses klusterisasi dapat dilakukan secara adil dan proporsional.

## 2.4. Teknik Pra-Pemrosesan Data

Sebelum data dianalisis menggunakan algoritma K-Means Clustering, dilakukan beberapa tahapan pra-pemrosesan guna memastikan kualitas dan kesiapan data. Tahapan tersebut meliputi:

1. Seleksi Variabel (Feature Selection), Hanya variabel yang relevan dengan aspek kemiskinan dan kesejahteraan sosial yang dipertahankan, guna memastikan fokus analisis sesuai dengan tujuan penelitian.
2. Penanganan Missing Value, dimana data yang memiliki nilai kosong diperiksa secara menyeluruh. Jika proporsinya kecil dan tidak signifikan, data tersebut dihapus. Namun, jika jumlahnya cukup besar dan berdampak terhadap distribusi, dilakukan pengisian menggunakan nilai rata-rata atau median.
3. Z-Score standardization, yaitu proses transformasi data dilakukan agar setiap variabel memiliki nilai rata-rata nol dan deviasi standar satu. Langkah ini bertujuan untuk menyetarakan skala antar variabel sebelum analisis lanjutan, mengingat algoritma K-Means sangat sensitif terhadap perbedaan skala antar fitur.

## 2.5. Penentuan Jumlah Cluster

Jumlah kluster optimal (k) ditentukan menggunakan Metode Elbow, yaitu pendekatan grafis yang memplot jumlah kluster terhadap nilai inertia atau Within-Cluster Sum of Squares (WCSS). Metode ini bertujuan untuk mengidentifikasi titik perubahan (elbow point) di mana penambahan jumlah kluster tidak lagi memberikan penurunan signifikan terhadap nilai WCSS. Adapun langkah-langkah pelaksanaannya adalah sebagai berikut:

- 1) Menjalankan K-Means dengan berbagai nilai k (misal dari 1 sampai 10).
- 2) Mencatat nilai inertia untuk tiap k.
- 3) Membuat plot inertia terhadap k.
- 4) Menentukan titik "tekukan" (elbow point), yaitu saat penurunan inertia mulai tidak signifikan.

Pada penelitian ini, nilai optimal diperoleh saat k = 3, yang ditunjukkan dengan penurunan signifikan inertia hingga titik tersebut.

## 2.6. Implementasi K-Means Clustering

Proses klusterisasi dilakukan menggunakan algoritma K-Means, yang termasuk dalam kategori unsupervised learning dan umum digunakan untuk segmentasi data berdasarkan kemiripan fitur. Implementasi algoritma ini dilakukan dengan memanfaatkan pustaka Python scikit-learn, menggunakan parameter sebagai berikut:

- 1) `n_clusters = 3`
- 2) `init = 'k-means++'`
- 3) `n_init = 10`
- 4) `max_iter = 300`
- 5) `random_state = 42`

Langkah kerja algoritma K-Means :

- 1) Inisialisasi centroid awal.
- 2) Mengelompokkan data ke centroid terdekat menggunakan jarak Euclidean.

- 3) Memperbarui centroid berdasarkan rata-rata data dalam cluster.
- 4) Mengulangi proses sampai centroid stabil (konvergen).

### 2.7. Teknik Analisis Hasil

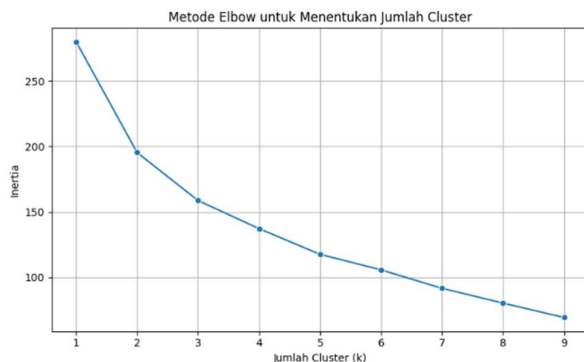
Setelah proses K-Means Clustering selesai, dilakukan beberapa tahapan analisis untuk mengevaluasi dan menginterpretasikan hasil klasterisasi secara komprehensif, yaitu:

- a) Deskripsi tiap cluster : masing-masing cluster dianalisis berdasarkan rata-rata nilai setiap indikator.
- b) Labeling cluster: cluster diberi interpretasi makna kebijakan, seperti “Wilayah Prioritas”, “Wilayah Transisi”, dan “Wilayah Berkembang”.
- c) Visualisasi: hasil cluster divisualisasikan menggunakan Principal Component Analysis (PCA) dalam bentuk scatter plot 2D.
- d) Pemetaan (opsional): data hasil clustering dapat dipetakan kembali ke peta geografis kabupaten/kota untuk keperluan analisis spasial.

## 3. HASIL

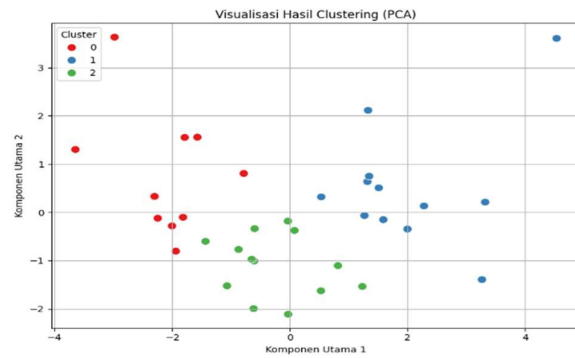
### 3.1. Hasil Clustering Wilayah Kabupaten/Kota

Proses K-Means Clustering berhasil membagi 35 kabupaten/kota di Provinsi Jawa Tengah ke dalam tiga kelompok, berdasarkan analisis terhadap tujuh variabel sosial ekonomi. Tahapan ini diawali dengan penentuan jumlah kluster optimal menggunakan metode Elbow, yang menunjukkan nilai ideal pada  $k = 3$ .



Gambar 2. Visualisasi Metode Elbow

Gambar 2 merupakan grafik Elbow Method yang digunakan untuk menentukan jumlah kluster optimal dalam proses K-Means Clustering. Grafik ini menunjukkan hubungan antara jumlah kluster ( $k$ ) dan nilai inertia atau *Within-Cluster Sum of Squares* (WCSS). Semakin kecil nilai inertia, semakin kompak dan homogen anggota kluster. Berdasarkan grafik, penurunan nilai inertia terlihat sangat signifikan hingga  $k = 3$ , kemudian mulai melandai setelahnya. Titik tersebut dikenal sebagai elbow point dan dianggap sebagai jumlah kluster yang paling optimal. Dengan demikian, pemilihan nilai  $k = 3$  dalam proses klasterisasi pada penelitian ini memiliki dasar yang kuat secara matematis maupun visual.



Gambar 3. Visualisasi Hasil Clustering

Gambar 3 merupakan visualisasi hasil klasterisasi menggunakan pendekatan *Principal Component Analysis* (PCA). PCA digunakan untuk mereduksi dimensi data dari tujuh variabel sosial ekonomi menjadi dua komponen utama, sehingga memungkinkan visualisasi dalam bidang dua dimensi. Dalam grafik tersebut, setiap titik merepresentasikan satu kabupaten/kota di Provinsi Jawa Tengah, dengan pewarnaan yang menunjukkan hasil pengelompokan oleh algoritma K-Means. Visualisasi ini menunjukkan bahwa objek data membentuk tiga kelompok yang cukup terpisah secara visual, sehingga memperkuat validitas hasil klasterisasi. Kejelasan pemisahan antar kluster juga mengindikasikan bahwa variabel sosial ekonomi yang digunakan telah cukup representatif dalam membedakan karakteristik wilayah secara statistik dan spasial.

### 3.2. Interpretasi Tiap Cluster

Setelah proses klasterisasi dilakukan dan jumlah cluster optimal ditetapkan sebanyak tiga kelompok, masing-masing kabupaten/kota dalam Provinsi Jawa Tengah dikelompokkan ke dalam salah satu dari tiga cluster tersebut. Setiap cluster merepresentasikan wilayah-wilayah dengan karakteristik sosial ekonomi yang relatif serupa. Untuk memberikan gambaran lebih jelas mengenai hasil pengelompokan tersebut, berikut disajikan tabel yang menampilkan label cluster yang diperoleh oleh masing-masing kabupaten/kota. Informasi ini menjadi dasar dalam menganalisis dan menafsirkan karakteristik khas dari tiap kelompok wilayah serta merumuskan arah kebijakan pembangunan yang sesuai.

Tabel 1. Hasil Pengelompokan Wilayah

| Kabupaten       | Cluster |
|-----------------|---------|
| Banjarnegara    | 1       |
| Banyumas        | 1       |
| Batang          | 1       |
| Blora           | 2       |
| Boyolali        | 2       |
| Brebes          | 1       |
| Cilacap         | 1       |
| Demak           | 1       |
| Grobogan        | 1       |
| Jepara          | 2       |
| Karanganyar     | 2       |
| Kebumen         | 1       |
| Kendal          | 0       |
| Klaten          | 2       |
| Kota Magelang   | 0       |
| Kota Pekalongan | 0       |
| Kota Salatiga   | 0       |
| Kota Semarang   | 0       |



|                       |   |
|-----------------------|---|
| <b>Kota Surakarta</b> | 0 |
| <b>Kota Tegal</b>     | 0 |
| <b>Kudus</b>          | 0 |
| <b>Magelang</b>       | 1 |
| <b>Pati</b>           | 2 |
| <b>Pekalongan</b>     | 2 |
| <b>Pemalang</b>       | 1 |
| <b>Purbalingga</b>    | 1 |
| <b>Purworejo</b>      | 1 |
| <b>Rembang</b>        | 2 |
| <b>Semarang</b>       | 0 |
| <b>Sragen</b>         | 2 |
| <b>Sukoharjo</b>      | 0 |
| <b>Tegal</b>          | 2 |
| <b>Temanggung</b>     | 2 |
| <b>Wonogiri</b>       | 2 |
| <b>Wonosobo</b>       | 2 |

Tabel 2 menyajikan hasil pengelompokan 35 kabupaten/kota di Provinsi Jawa Tengah ke dalam tiga klaster yang terbentuk melalui proses K-Means Clustering. Setiap baris dalam tabel merepresentasikan satu daerah beserta nomor klaster yang dihasilkannya. Penomoran klaster ditentukan secara otomatis oleh algoritma dan tidak mencerminkan urutan kualitas, melainkan hanya membedakan kelompok berdasarkan kemiripan pola data. Sebagai ilustrasi, kabupaten/kota yang tergabung dalam Klaster 0 menunjukkan kesamaan dalam hal Indeks Pembangunan Manusia (IPM) yang tinggi, pengeluaran per kapita yang besar, serta tingkat pengangguran yang rendah. Sebaliknya, daerah-daerah yang masuk dalam Klaster 2 umumnya memiliki IPM rendah, tingkat pengangguran tinggi, dan jumlah rumah tidak layak huni (RTLH) yang besar, sehingga dikategorikan sebagai wilayah prioritas pengentasan kemiskinan. Tabel ini berperan penting dalam menjembatani hasil pemodelan kuantitatif dengan identifikasi wilayah secara geografis, serta menjadi dasar pertimbangan dalam perumusan kebijakan pembangunan yang berbasis bukti dan lebih tepat sasaran.

#### 4. PEMBAHASAN

Hasil pengelompokan wilayah kabupaten/kota di Provinsi Jawa Tengah ke dalam tiga klaster menunjukkan adanya perbedaan yang nyata dalam karakteristik sosial ekonomi antar wilayah. Perbedaan ini tercermin dari variasi nilai Indeks Pembangunan Manusia (IPM), pengeluaran per kapita, tingkat pengangguran terbuka, serta indikator kedalaman dan keparahan kemiskinan (P1 dan P2). Klaster yang tergolong sebagai wilayah prioritas (Klaster 2) umumnya memiliki kombinasi indikator yang mengindikasikan tingkat kerentanan ekonomi dan sosial yang tinggi. Temuan ini sejalan dengan hasil penelitian [1], yang menyatakan bahwa wilayah dengan IPM rendah dan tingkat pengangguran tinggi cenderung masuk ke dalam kategori kemiskinan parah.

Dibandingkan dengan penelitian [3], yang mengelompokkan wilayah di Jawa Tengah menggunakan McQuitty Linkage dan K-Means, penelitian ini menambahkan variabel Upah Minimum Kabupaten/Kota (UMK) dan jumlah rumah tidak layak huni (RTLH) sebagai indikator tambahan. Hasilnya menunjukkan bahwa UMK memiliki korelasi erat dengan pengeluaran per kapita dan menjadi pembeda penting dalam struktur klasterisasi,

memperkuat argumen bahwa faktor ekonomi regional harus diperhatikan dalam desain kebijakan sosial.

Secara teoritis, penelitian ini memperkuat pendekatan multidimensional poverty yang menekankan bahwa kemiskinan tidak hanya dilihat dari sisi pendapatan, tetapi juga dari sisi akses terhadap pendidikan, kesehatan, dan perumahan. Ini sejalan dengan konsep human development oleh UNDP, serta pendekatan capability oleh Amartya Sen yang menilai kesejahteraan dari sejauh mana seseorang dapat mengembangkan potensi hidupnya. Dari sisi kebijakan, segmentasi wilayah ke dalam tiga kelompok memungkinkan pemerintah daerah untuk menyusun strategi pembangunan yang lebih terarah. Cluster prioritas memerlukan pendekatan intervensi menyeluruh yang mencakup peningkatan akses pendidikan dan kesehatan, pembangunan infrastruktur dasar, serta penciptaan lapangan kerja produktif. Cluster transisi berpotensi menjadi wilayah maju jika mendapatkan intervensi yang tepat, sedangkan Cluster berkembang dapat dijadikan sebagai role model pembangunan daerah. Hasil ini juga relevan dalam konteks implementasi pendekatan berbasis bukti (evidence-based policy), di mana sumber daya pembangunan disalurkan berdasarkan kondisi objektif wilayah, bukan semata-mata pada pertimbangan administratif atau politik. Penelitian ini memberikan kontribusi praktis bagi perencanaan program-program sosial seperti bantuan rumah tidak layak huni, pelatihan kerja, dan alokasi Dana Alokasi Khusus (DAK) sektoral berbasis indikator wilayah.

Dengan demikian, diskusi ini tidak hanya menegaskan validitas pendekatan K-Means dalam klasifikasi wilayah kemiskinan, tetapi juga menunjukkan bahwa pengelompokan berbasis data mampu mendukung pembangunan daerah yang lebih inklusif dan tepat sasaran.

#### 5. KESIMPULAN

Berdasarkan hasil penelitian, metode K-Means Clustering berhasil mengelompokkan 35 wilayah administratif (kabupaten/kota) di Jawa Tengah menjadi tiga klaster utama berdasarkan tujuh indikator sosial ekonomi, yaitu Indeks Pembangunan Manusia (IPM), P1, P2, pengeluaran per kapita, Upah Minimum Kabupaten/Kota (UMK), tingkat pengangguran terbuka, serta jumlah rumah tidak layak huni (RTLH). Segmentasi ini memberikan gambaran yang lebih jelas mengenai perbedaan kondisi sosial ekonomi antarwilayah, sehingga memudahkan dalam mengidentifikasi kelompok wilayah yang relatif sejahtera, sedang berkembang, maupun yang masih menghadapi tantangan besar dalam pengentasan kemiskinan.

Hasil klasterisasi menunjukkan tiga kategori utama yang cukup bermakna. Cluster 0 (Wilayah Berkembang) mencerminkan daerah dengan indikator sosial ekonomi relatif baik. Cluster 1 (Wilayah Transisi) menggambarkan daerah dengan kondisi menengah yang berpotensi berkembang lebih baik atau justru mengalami penurunan jika tidak dikelola dengan tepat. Sementara itu, Cluster 2 (Wilayah Prioritas) menandai daerah dengan indikator kemiskinan tinggi yang membutuhkan perhatian khusus serta intervensi kebijakan yang lebih mendalam. Pembagian ini penting sebagai dasar pertimbangan dalam merumuskan strategi pembangunan yang lebih tepat sasaran.

Meskipun pendekatan ini memiliki keunggulan karena memanfaatkan data terbuka, mudah direplikasi, serta menghasilkan output numerik dan visual yang mendukung kebijakan berbasis data, terdapat pula beberapa keterbatasan. Penelitian ini belum mempertimbangkan faktor spasial dan geografis, variabel sosial budaya, serta akses layanan dasar seperti kesehatan, pendidikan, dan infrastruktur. Selain itu, penamaan klaster masih bersifat deskriptif tanpa validasi kualitatif. Oleh karena itu, penelitian selanjutnya disarankan untuk memperkaya variabel, mengombinasikan clustering dengan metode klasifikasi atau prediksi, serta mengembangkan sistem pendukung keputusan berbasis dashboard interaktif atau integrasi dengan peta Sistem Informasi Geografis (SIG).

## DAFTAR PUSTAKA

- [1] R. Kurnia Desita, A. Fahmi, A. Rohmani, and MY. T. Sulistyono, "Agglomerative Hierarchical Clustering For Regional Grouping In Central Java Based On Welfare Indices," *Jurnal Pilar Nusa Mandiri*, vol. 21, no. 1, pp. 108–116, Mar. 2025, doi: [10.33480/pilar.v21i1.6445](https://doi.org/10.33480/pilar.v21i1.6445).
- [2] C. Tacharri, A. Rohmani, and A. Fahmi, "Strategic Clustering of Poverty Areas in Central Java Using K-Means and Silhouette Evaluation," *Sinkron*, vol. 9, no. 2, pp. 895–904, May 2025, doi: [10.33395/sinkron.v9i2.14734](https://doi.org/10.33395/sinkron.v9i2.14734).
- [3] M. A. Tsani and A. Fahmi, "Implementation of DBSCAN Algorithm for Grouping Poverty Levels in Central Java Province," *Jurnal Sistem Komputer dan Informatika (JSON) Hal*, vol. 6, no. 4, pp. 131–140, 2025, doi: [10.30865/json.v6i4](https://doi.org/10.30865/json.v6i4).
- [4] G. W. Kusuma and I. Y. Wulansari, "Analisis kemiskinan dan kerentanan kemiskinan dengan Regresi Ridge, LASSO, dan Elastic-Net di Provinsi Jawa Tengah tahun 2017," in *Seminar Nasional Official Statistics*, 2019, pp. 503–513. Accessed: Jul. 29, 2025. [Online]. Available: <https://prosiding.stis.ac.id/index.php/semnasoffstat/article/view/189>.
- [5] S. Fungsional, S. Muda, B. P. Statistik, and K. Cilacap, "(2022) e-issn, 2620-6099."
- [6] M. Marizal *et al.*, "Pemodelan angka kematian bayi di Indonesia menggunakan Geographically Weighted Regression (GWR) dan Mixed Geographically Weighted Regression (MGWR) (Modeling infants mortality rate in Indonesia using Geographically Weighted Regression (GWR) and Mixed Geographically Weighted Regression (MGWR))." [Online]. Available: <https://jurnal.unej.ac.id/index.php/MIMS/index>
- [7] I. Tawakal, M. Effendi, and A. Maulana Majid, "Analisis Tingkat Kemiskinan Dengan Algoritma K-Means Menggunakan Rapidminer Di Tingkat Kota Kabupaten Di Jawa Tengah," 2025.
- [8] I. P. Putra and A. Fadhillah, "Perbandingan Metode K-Means dan Hierarchical Clustering dalam Pengelompokan Data Penduduk Miskin di Kabupaten Cianjur," *LANCAH: Jurnal Inovasi dan Tren*, vol. 3, no. 1, Apr. 2025, doi: [10.35870/ljit.v3i1.4028](https://doi.org/10.35870/ljit.v3i1.4028).
- [9] N. Sepriyanti, R. Sani Nahampun, M. H. Zikri, I. Ambarani, and A. Rahmadyan, "SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat Implementation of K-Means Clustering to Group Poverty Levels in Riau Province Penerapan K-Means Clustering Untuk Mengelompokkan Tingkat Kemiskinan di Provinsi Riau." [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas>
- [10] D. Widyadhana, R. B. Hastuti, I. Kharisudin, and F. Fauzi, "Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 4, pp. 584–594, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [11] Y. D. Ramadani and B. Winarno, "Implementasi Metode Mcquitty Linkage dan K-Means Clustering dengan Pendekatan Rule-of-Thumb pada Data Kemiskinan di Jawa Tengah," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 6, pp. 751–756, 2023, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [12] R. D. Faturahman and N. Hidayati, "Implementasi Fuzzy C-Means Dalam Pengelompokan Tingkat Kemiskinan Pada Kabupaten/Kota Di Provinsi Jawa Tengah," *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 1, pp. 137–149, Jan. 2025, doi: [10.29100/jipi.v10i1.5747](https://doi.org/10.29100/jipi.v10i1.5747).
- [13] E. Widodo, P. Ermayani, L. N. Laila, and A. T. Madani, "Pengelompokan Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Analisis Hierarchical Agglomerative Clustering (Indonesian Province Grouping Based on Poverty Level Using Hierarchical Agglomerative Clustering Analysis)."
- [14] A. Putrian Wijaya, A. Salma Kinanthi, D. Yanawati, S. Syamsidar, and E. Widodo, "Pengelompokan Kabupaten/Kota di Pulau Jawa Berdasarkan Faktor Kemiskinan Menggunakan Metode Hierarchical Clustering," *Jurnal Sains dan Manajemen*, vol. 13, no. 1, 2025.
- [15] Tb. A. Munandar, "Penerapan Algoritma Clustering Untuk Pengelompokan Tingkat Kemiskinan Provinsi Banten," *JSiI (Jurnal Sistem Informasi)*, vol. 9, no. 2, pp. 109–114, Sep. 2022, doi: [10.30656/jsii.v9i2.5099](https://doi.org/10.30656/jsii.v9i2.5099).
- [16] F. Zahra, A. Khalif, B. N. Sari, U. S. Karawang, J. H. Ronggo Waluyo, and T. Timur, "Pengelompokan Tingkat Kemiskinan Di Setiap Provinsi Di Indonesia Menggunakan Algoritma K-Medoids," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, pp. 2830–2862, doi: [10.23960/jitet.v12i2.4199](https://doi.org/10.23960/jitet.v12i2.4199).
- [17] R. H. Bhahari and K. Kusnawi, "Clustering Analysis of Socio-Economic Districts/Cities In East Java Province Using PCA And Hierarchical Clustering Methods," *sinkron*, vol. 8, no. 4, pp. 2242–2251, Oct. 2024, doi: [10.33395/sinkron.v8i4.14078](https://doi.org/10.33395/sinkron.v8i4.14078).

## NOMENKLATUR

Nomenklatur disertai arti dari semua persamaan matematika ataupun nomenklatur lain di alam artikel, dituliskan pada bagian ini.

- $p(x,y)$  yaitu Jarak Euclidean antara titik data xxx dan centroid yyy.

- $E$  yaitu Nilai *Within-Cluster Sum of Squares (WCSS)*, yaitu total kuadrat jarak seluruh titik data dalam kluster terhadap centroid-nya.
- $J_c$  yaitu Fungsi objektif K-Means, yaitu jumlah total kuadrat jarak semua titik data ke centroid terdekat (sama dengan nilai minimum WCSS dari semua kluster).
- $x$  yaitu Vektor data yang merepresentasikan suatu kabupaten/kota berdasarkan indikator sosial ekonomi.
- $y/\mu_i$  yaitu Pusat kluster (centroid) dari kluster ke- $i$ .
- $k$  yaitu Jumlah kluster yang ditentukan dalam proses K-Means.
- $C_i$  yaitu Himpunan anggota dari kluster ke- $i$ .
- $n$  yaitu Dimensi data (jumlah indikator sosial ekonomi yang digunakan, misalnya IPM, P1, P2, UMK, dll.).
- $x_j, y_j$  yaitu Nilai pada dimensi ke- $j$  dari vektor data  $xxx$  dan centroid  $yyy$ .

## BIODATA PENULIS



Moh. Fachri Alif

Mahasiswa Program Studi Teknik Informatika di Universitas Dian Nuswantoro, Semarang, Jawa Tengah. Ia lahir di Pemalang, Jawa Tengah, pada tanggal 9 April 2004. Saat ini,

Fachri aktif dalam pengembangan kompetensi di bidang teknologi informasi, khususnya dalam pemrograman, sistem cerdas, dan analisis data. Sebagai bagian dari komunitas akademik, ia berkomitmen untuk mengembangkan solusi digital yang relevan dan berdampak bagi masyarakat.



Dr. Amiq Fahmi, S. Kom., M. Kom.

Amiq Fahmi adalah dosen di Program Studi Sistem Informasi, Universitas Dian Nuswantoro, Semarang. Ia aktif dalam berbagai forum ilmiah dan memiliki minat riset di bidang sistem informasi cerdas, data mining, machine learning,

serta kecerdasan buatan untuk pendukung keputusan. Selain mengajar dan meneliti, ia juga berpengalaman dalam pengembangan serta penyusunan kurikulum institusional. Karyanya berfokus pada pemanfaatan teknologi untuk solusi berbasis data yang aplikatif dan berdampak.