

SEGMENTASI WILAYAH KEMISKINAN DI JAWA TENGAH MENGGUNAKAN K-MEANS UNTUK PENENTUAN PRIORITAS INTERVENSI PEMERINTAH

Rizky Ramadhani, Aprilisa Arum Sari, Dwi Hartanti

Teknik Informatika, Universitas Duta Bangsa Surakarta

Jl. Bhayangkara No.55, Tipes, Kec. Serengan, Kota Surakarta, Jawa Tengah 57154

220103073@mhs.udb.ac.id

ABSTRAK

Kemiskinan merupakan salah satu permasalahan pembangunan yang masih menjadi perhatian utama di Provinsi Jawa Tengah. Perbedaan tingkat kemiskinan antar kabupaten/kota menunjukkan adanya ketimpangan kondisi sosial ekonomi yang memerlukan analisis lebih mendalam untuk mendukung pengambilan kebijakan. Permasalahan dalam penelitian ini adalah bagaimana mengelompokkan wilayah berdasarkan tingkat kemiskinan menggunakan indikator sosial ekonomi sehingga diperoleh karakteristik wilayah yang lebih jelas. Penelitian ini bertujuan untuk melakukan segmentasi tingkat kemiskinan pada 35 kabupaten/kota di Provinsi Jawa Tengah menggunakan metode *K-Means Clustering* serta menentukan jumlah cluster optimal menggunakan *Silhouette Coefficient*. Data yang digunakan merupakan data sekunder Badan Pusat Statistik (BPS) tahun 2024 yang meliputi beberapa indikator sosial ekonomi. Tahapan penelitian meliputi *preprocessing data*, normalisasi menggunakan *MinMaxScaler*, proses *clustering* menggunakan algoritma *K-Means*, dan evaluasi menggunakan *Silhouette Coefficient*. Implementasi dilakukan menggunakan bahasa pemrograman *Python* pada Google Colab. Hasil penelitian menunjukkan bahwa jumlah cluster optimal diperoleh pada $K=3$ dengan nilai *Silhouette Score* sebesar 0,308, lebih baik dibandingkan $K=5$ sebesar 0,229. Hasil *clustering* membagi wilayah ke dalam tiga kelompok, yaitu tingkat kemiskinan tinggi, sedang, dan rendah, yang dapat digunakan sebagai dasar penentuan prioritas intervensi pemerintah dalam penanggulangan kemiskinan.

Kata kunci : *Clustering, K-Means, Kemiskinan, Python, Silhouette Coefficient, Jawa Tengah.*

1. PENDAHULUAN

Kemiskinan merupakan salah satu permasalahan pembangunan yang masih menjadi perhatian utama di Provinsi Jawa Tengah. Berdasarkan data Badan Pusat Statistik (BPS) tahun 2024, tingkat kemiskinan di Jawa Tengah masih berada pada kisaran 10%. Namun, kondisi kemiskinan antar kabupaten/kota menunjukkan perbedaan yang cukup signifikan. Wilayah perkotaan seperti Semarang, Surakarta, dan Salatiga memiliki tingkat kemiskinan yang relatif rendah, sedangkan beberapa wilayah kabupaten seperti Brebes, Pemalang, dan Wonosobo memiliki tingkat kemiskinan yang lebih tinggi. Perbedaan tersebut menunjukkan adanya ketimpangan kondisi sosial ekonomi antar wilayah yang dipengaruhi oleh berbagai faktor, seperti pendidikan, kesehatan, tingkat pengangguran, dan kemampuan ekonomi masyarakat. Oleh karena itu, diperlukan analisis yang mampu memberikan gambaran kondisi kemiskinan secara lebih komprehensif untuk mendukung pengambilan keputusan dalam penanggulangan kemiskinan [1], [2].

Permasalahan yang dihadapi adalah bagaimana mengelompokkan wilayah kabupaten/kota di Provinsi Jawa Tengah berdasarkan karakteristik kemiskinan dan indikator sosial ekonomi yang dimiliki. Analisis kemiskinan yang hanya menggunakan satu indikator belum mampu menggambarkan kondisi wilayah secara menyeluruh, sehingga diperlukan pendekatan yang dapat mengelompokkan wilayah berdasarkan kemiripan karakteristiknya. Selain itu, penentuan jumlah kelompok yang optimal juga menjadi tantangan agar hasil pengelompokan dapat

memberikan informasi yang lebih akurat dan bermanfaat [3], [4].

Tujuan penelitian ini adalah mengelompokkan tingkat kemiskinan pada 35 kabupaten/kota di Provinsi Jawa Tengah menggunakan metode *K-Means Clustering*, menentukan jumlah cluster optimal menggunakan metode *Silhouette Coefficient*, serta mengidentifikasi karakteristik masing-masing cluster berdasarkan indikator sosial ekonomi yang digunakan.

Rencana penyelesaian masalah dilakukan dengan menerapkan metode *K-Means Clustering* pada data kemiskinan dan indikator sosial ekonomi yang diperoleh dari Badan Pusat Statistik tahun 2024. Data terlebih dahulu melalui tahap *preprocessing* dan normalisasi agar memiliki skala yang seragam. Selanjutnya dilakukan proses *clustering* menggunakan algoritma *K-Means* dengan beberapa alternatif jumlah cluster. Hasil *clustering* kemudian dievaluasi menggunakan metode *Silhouette Coefficient* untuk menentukan jumlah cluster yang paling optimal. Proses analisis dilakukan menggunakan bahasa pemrograman *Python* pada Google Colab, kemudian hasil pengelompokan diinterpretasikan untuk mengetahui karakteristik masing-masing cluster dan digunakan sebagai dasar dalam memberikan rekomendasi penentuan prioritas penanggulangan kemiskinan di Provinsi Jawa Tengah.

Kebaruan (*novelty*) riset ini terletak pada pemakaian informasi terkini BPS tahun 2024, pelaksanaan pendekatan multi-indikator, pemakaian penilaian *Silhouette Coefficient* buat memastikan jumlah cluster maksimal, dan implementasi analisis

memakai *Python* pada Google Colab. Riset ini bertujuan buat mengelompokkan tingkatan kemiskinan pada 35 kabupaten/ kota di Provinsi Jawa Tengah memakai tata cara *K- Means Clustering* sehingga bisa menciptakan pemetaan daerah kemiskinan yang lebih akurat serta menolong pemerintah dalam memastikan prioritas intervensi penanggulangan kemiskinan secara lebih efisien serta pas sasaran.

2. TINJAUAN PUSTAKA

Penelitian terkait pengelompokan wilayah kemiskinan telah banyak dilakukan dengan berbagai pendekatan algoritma.

Penelitian yang dilakukan oleh [5] mengaplikasikan algoritma *K-Means* untuk mengkategorikan tingkat kemiskinan di Provinsi Jawa Timur. Penelitian ini berhasil menciptakan tiga kelompok utama (rendah, sedang, dan tinggi), tetapi masih memiliki kelemahan karena belum menerapkan metode evaluasi seperti *Silhouette Coefficient* dan variabel yang digunakan relatif terbatas[5].

Penelitian terbaru oleh [6] mempelajari pengelompokan kemiskinan di wilayah Provinsi Jawa Tengah dengan memadukan pendekatan *K-Means* dan *Silhouette Coefficient*. Penilaian menggunakan *Silhouette* dalam penelitian ini berhasil meningkatkan kualitas serta keabsahan hasil pengelompokan, walaupun perhatian utama masih terfokus pada evaluasi metode dan belum menyentuh saran kebijakan yang lebih lengkap[6].

Penelitian ketiga yang dilakukan oleh [7] menitikberatkan pada pengenalan pola kemiskinan dalam konteks Indonesia melalui pendekatan *K-Means* yang berlandaskan indikator sosial ekonomi. Walaupun penelitian tersebut mampu menyuguhkan gambaran mengenai distribusi kemiskinan di tingkat nasional, namun evaluasi kluster tidak diterapkan dan ruang lingkupnya terlalu umum sehingga kurang relevan untuk digunakan dalam pengambilan keputusan kebijakan daerah [7].

Berdasarkan ulasan dari sumber-sumber yang telah disebutkan sebelumnya, penelitian ini dirancang untuk mengatasi kekurangan yang ada dengan mengintegrasikan algoritma *K-Means*, penilaian *Silhouette*, serta penerapan multi-indikator yang berbasis *Python* untuk menghasilkan analisis mengenai karakteristik kluster dan rekomendasi kebijakan yang relevan khususnya di Provinsi Jawa Tengah[6].

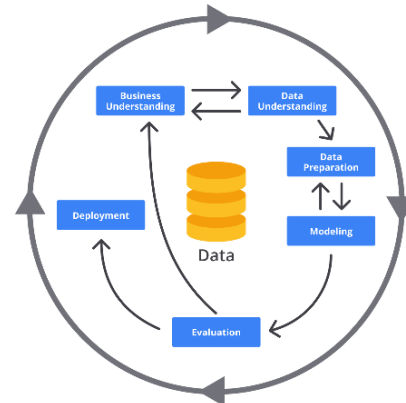
3. METODE PENELITIAN

Penelitian ini menerapkan metode kuantitatif melalui penggalian data, terutama teknik pengelompokan menggunakan algoritma *K-Means*. Metode kuantitatif dipilih dikarenakan studi ini mengelola data numerik yang berbentuk indikator kemiskinan serta sosial ekonomi untuk menciptakan pola pengelompokan area dengan cara yang objektif dan terukur.

Metodologi pengembangan yang digunakan mengacu pada model *CRISP-DM* (*Cross Industry Standard Process for Data Mining*) yang terdiri dari enam tahapan, yaitu:

- Business Understanding*
- Data Understanding*
- Data Preparation*
- Modeling*
- Evaluation*
- Deployment/Interpretation*

Pendekatan ini dipilih karena memberikan alur kerja yang sistematis dalam proses analisis data mining.



Gambar 1. Diagram *CRISP-DM* pada Penelitian

a. Jenis dan Sumber Data

Penelitian ini memanfaatkan jenis data kuantitatif yang berupa data sekunder yang diambil dari Badan Pusat Statistik (BPS) tahun 2024. Data yang digunakan mencakup 35 kabupaten atau kota di Provinsi Jawa Tengah dengan variabel yang meliputi persentase penduduk yang hidup dalam kemiskinan (P0), tingkat pengangguran terbuka (TPT), rata-rata lama pendidikan (RLS), angka harapan hidup (AHH), serta pengeluaran per orang. Data tersebut dipilih karena dapat mencerminkan kondisi kemiskinan secara multidimensi, sehingga dapat memberikan outlook yang lebih menyeluruh dalam proses pengelompokan.

b. Metode Pengumpulan Data

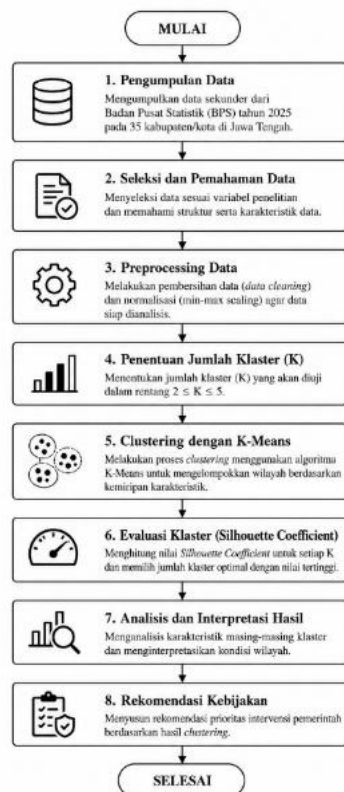
Penelitian ini mengumpulkan data dengan menggunakan metode dokumentasi. Data tersebut berasal dari sumber resmi yang diterbitkan oleh Badan Pusat Statistik (BPS) [1], tersedia di situs web mereka dan dalam laporan statistik tahunan. Proses pengumpulan data meliputi pengunduhan, pemilihan, dan pengelompokan data berdasarkan kebutuhan variabel penelitian. Metode ini dipilih karena datanya dapat diandalkan, konsisten, dan sangat akurat sehingga membantu analisis bekerja dengan sebaik-baiknya.

c. Metode Analisis

Metode analisis yang diterapkan dalam penelitian ini adalah pendekatan data mining dengan teknik pengelompokan menggunakan algoritma *K-Means* [8],[9]. Proses analisis dimulai dengan tahap preprocessing data yang mencakup pembersihan

data (data cleaning) dan normalisasi untuk memastikan keseragaman skala data. Selanjutnya, dilakukan proses pengelompokan dengan menentukan jumlah kluster (K) dalam rentang 2 hingga 5, kemudian data dikelompokkan berdasarkan kedekatan terhadap centroid dengan menggunakan perhitungan jarak Euclidean. Untuk menentukan jumlah kluster yang optimal, diterapkan metode Silhouette Coefficient yang mengukur kualitas hasil pengelompokan berdasarkan tingkat kedekatan dalam kluster dan pemisahan antar kluster. Hasil akhir dari proses ini kemudian dianalisis untuk mengidentifikasi karakteristik masing-masing kluster dan digunakan sebagai dasar dalam penyusunan rekomendasi kebijakan [10].

Pada Gambar 2. Flowchart Metodologi Penelitian yang Menggambarkan Tahapan Pengumpulan Data, Preprocessing, Clustering K-Means, Evaluasi Silhouette, dan Interpretasi Hasil



Gambar 2. Flowchart Clustering

- a. **Pengumpulan Data**
Tahap awal penelitian dimulai dengan pengumpulan data sekunder yang berasal dari Badan Pusat Statistik (BPS) tahun 2024. Data yang dikumpulkan mencakup indikator kemiskinan dan sosial ekonomi di 35 kabupaten/kota di Provinsi Jawa Tengah.
- b. **Seleksi dan Pemahaman Data**
Pada tahap ini, proses seleksi data dilakukan untuk memastikan bahwa data yang digunakan sesuai

dengan variabel penelitian. Selain itu, pemahaman terhadap struktur data, jenis variabel, serta distribusi data juga dilakukan untuk mempermudah proses analisis yang akan datang.

c. **Preprocessing Data**

Tahap *preprocessing* mencakup pembersihan data (data cleaning) untuk menangani data yang tidak lengkap atau tidak konsisten, serta normalisasi data agar semua variabel berada pada skala yang seragam. Ini sangat penting karena algoritma *K-Means* peka terhadap perbedaan skala data.

d. **Penentuan Jumlah Kluster (K)**

Pada tahap ini, jumlah kluster yang akan digunakan dalam proses pengelompokan ditentukan. Nilai K dibatasi dalam rentang 2 hingga 5 untuk memastikan bahwa interpretasi hasil tetap relevan dan mudah untuk dianalisis [11].

e. **Proses Clustering dengan K-Means**

Data yang telah diproses selanjutnya dikelompokkan dengan menggunakan algoritma *K-Means*. Proses ini dilakukan dengan cara menentukan centroid awal, menghitung jarak data terhadap centroid, mengelompokkan data ke kluster yang paling dekat, serta memperbarui centroid hingga mencapai kondisi *konvergen*[12][13].

f. **Evaluasi Kluster (Silhouette Coefficient)**

Untuk menentukan jumlah kluster yang optimal, evaluasi dilakukan dengan menggunakan metode *Silhouette Coefficient*. Metode ini berfungsi untuk menilai kualitas hasil pengelompokan berdasarkan tingkat kedekatan dalam kluster serta jarak antara kluster[14].

g. **Analisis dan Interpretasi Hasil**

Hasil pengelompokan selanjutnya dianalisis untuk mengidentifikasi ciri-ciri dari setiap kluster, seperti kelompok daerah dengan tingkat kemiskinan yang tinggi, sedang, dan rendah.

h. **Penyusunan Rekomendasi**

Akhir dari proses adalah membuat rekomendasi kebijakan yang didasarkan pada hasil analisis kluster untuk digunakan sebagai dasar pengambilan keputusan yang lebih tepat sasaran.

4. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan guna mengategorikan wilayah kabupaten/ kota bersumber pada parameter kemiskinan serta kesejahteraan warga dengan tata cara *K-Means Clustering* lewat tahapan pengolahan informasi berbasis *CRISP-DM* yang mencakup uraian bisnis, uraian informasi, persiapan informasi, pemodelan, penilaian, serta interpretasi. Analisis dicoba dengan menggunakan Google Colab memakai informasi yang terdiri dari tingkatan kemiskinan (P0) serta pengeluaran per kapita yang sudah disesuaikan. Saat sebelum proses pengelompokan diawali, informasi tersebut dinormalisasi terlebih dulu memakai metode *MinMaxScaler* biar segala variabel terletak pada skala yang seragam.

4.1. Normalisasi Data

Normalisasi dilakukan agar tidak ada variabel yang memiliki skala nilai lebih besar mendominasi proses pengelompokan. Teknik *MinMaxScaler* diterapkan untuk mengubah jangkauan data menjadi antara 0 hingga 1.

Persamaan normalisasi *MinMaxScaler* ditunjukkan sebagai berikut:

$$X'_i = \frac{X_i - X_{min}}{X_{max} - X_{min}}$$

Keterangan :

- X'_i = Data hasil normalisasi ke-i
- X_i = Data asli ke-i
- X_{min} = Nilai minimum
- X_{max} = Nilai maximum

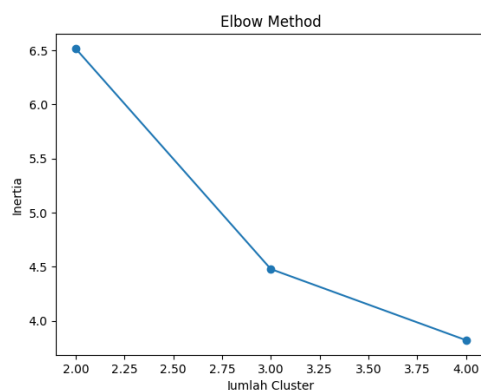
Hasil normalisasi menunjukkan bahwa seluruh data berhasil ditransformasikan ke dalam rentang yang seragam sehingga siap digunakan pada proses clustering.

4.2. Penentuan Jumlah Cluster

Penentuan jumlah cluster dilakukan menggunakan metode Elbow dan evaluasi *Silhouette Score*.

4.3. Elbow Method

Metode Elbow dimanfaatkan buat mengenali jumlah cluster yang sangat sempurna bersumber pada nilai inersia. Inersia mengindikasikan total kuadrat jarak antara informasi serta pusat cluster. Melalui grafik Metode Elbow, bisa dilihat bahwa penurunan inersia mulai melambat setelah mencapai jumlah klaster tertentu. Ini menunjukkan bahwa penambahan jumlah klaster setelah titik tersebut tidak memberikan peningkatan yang berarti terhadap kualitas klaster.



Gambar 3. Elbow Method

4.4. Silhouette Score

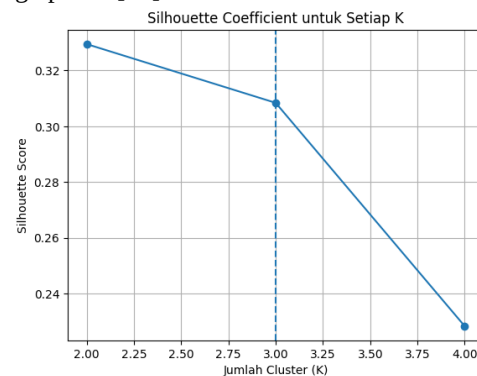
Evaluasi kualitas cluster dilakukan menggunakan *Silhouette Coefficient*. Semakin mendekati nilai 1, maka kualitas cluster semakin baik.

Hasil Pengujian Menunjukkan :

- Silhouette Score K=3 = 0.308
- Silhouette Score K=5 = 0.229

Berdasarkan hasil pengujian, model dengan 3 cluster memiliki kualitas clustering yang lebih baik

dibandingkan model dengan 5 cluster. Selain itu, dilakukan pengujian *Silhouette Score* terhadap beberapa nilai K untuk menganalisis pola perubahan kualitas cluster dan menentukan jumlah cluster yang paling optimal[15].



Gambar 4. Grafik Silhouette Score

Berdasarkan diagram tersebut, nampak bahwa angka *Silhouette Score* cenderung berkurang seiring dengan meningkatnya jumlah cluster. Angka *Silhouette Score* yang paling tinggi menandakan cluster yang paling optimal menurut perhitungan. Namun, dalam penelitian ini, dipilih K=3 sebagai model utama karena memberikan interpretasi yang lebih mewakili kondisi sosial ekonomi daerah, yaitu kategori tinggi, sedang, dan rendah.

4.5. Hasil Clustering K-Means

Penelitian ini menerapkan dua situasi terkait jumlah Cluster, yaitu K=3 dan K=5. *Clustering* dengan K=3 difungsikan sebagai model utama, sementara K=5 digunakan untuk analisis tambahan guna mengamati segmentasi area yang lebih mendalam.

4.6. Hasil Clustering K=3

Hasil rata-rata masing-masing cluster menunjukkan adanya perbedaan karakteristik sosial ekonomi antar wilayah.

Tabel 1. Rata rata Cluster K=3
Berdasarkan hasil clustering di peroleh

No	Cluster	P0	Kategori
1	0	10.070000	Sedang
2	1	6.381429	Rendah
3	2	12.803000	Tinggi

Kategori ditentukan berdasarkan nilai rata-rata P0 pada masing-masing cluster.

4.7. Interpretasi Cluster K=3

a. Cluster 0 (Kategori Sedang)

Cluster 0 memiliki tingkat kemiskinan pada kategori sedang dengan nilai P0 yang berada pada tingkat menengah dibandingkan cluster lainnya. Wilayah pada cluster ini menunjukkan kondisi sosial ekonomi yang relatif stabil, namun masih memerlukan perhatian dalam peningkatan kesejahteraan masyarakat. Pengeluaran per kapita

yang berada pada tingkat menengah menunjukkan bahwa daya beli masyarakat belum sepenuhnya kuat, sehingga intervensi yang tepat untuk cluster ini adalah penguatan ekonomi lokal, pengembangan UMKM, serta peningkatan kualitas sumber daya manusia.

b. Cluster 1 (Kategori Rendah)

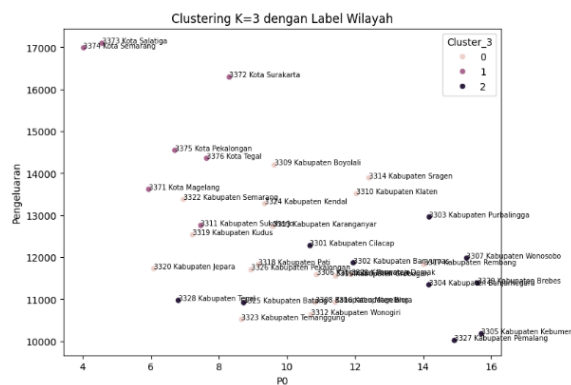
Cluster 1 merupakan kelompok wilayah dengan tingkat kemiskinan paling rendah dan pengeluaran per kapita tertinggi, sehingga menunjukkan bahwa masyarakat pada cluster ini memiliki tingkat kesejahteraan yang relatif baik. Wilayah dalam cluster ini tidak menjadi prioritas utama dalam intervensi pengentasan kemiskinan, namun tetap memerlukan kebijakan untuk menjaga stabilitas ekonomi serta meningkatkan daya saing daerah secara berkelanjutan.

c. Cluster 2 (Kategori Tinggi)

Cluster 2 memiliki nilai P0 tertinggi dibandingkan cluster lainnya serta pengeluaran per kapita yang rendah, yang menunjukkan adanya keterbatasan daya beli masyarakat dan tingginya tingkat kerentanan ekonomi. Oleh karena itu, wilayah dalam cluster ini menjadi prioritas utama dalam intervensi pemerintah melalui program bantuan sosial, peningkatan akses lapangan kerja, pemberdayaan ekonomi masyarakat, serta peningkatan kualitas pendidikan dan kesehatan.

4.8. Visualisasi Clustering K=3

Visualisasi scatter plot digunakan untuk melihat persebaran wilayah berdasarkan tingkat kemiskinan dan pengeluaran per kapita.

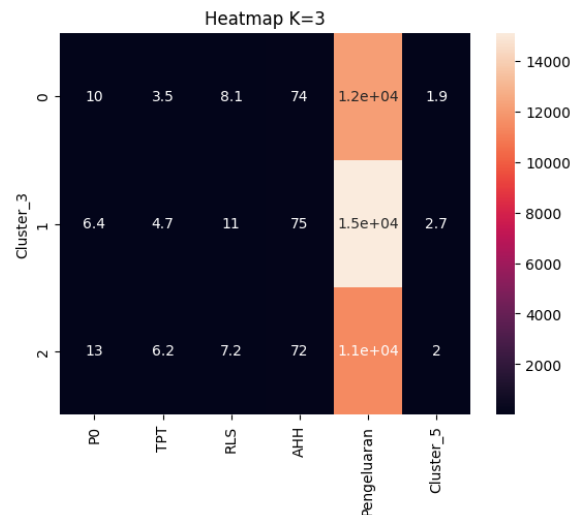


Gambar 5. Scatter Plot K=3

Berdasarkan scatter plot, data berhasil terbagi ke dalam beberapa kelompok dengan karakteristik yang berbeda, di mana wilayah dengan tingkat kemiskinan tinggi cenderung memiliki pengeluaran per kapita yang rendah. Hal ini menunjukkan adanya hubungan negatif antara tingkat kemiskinan dan kemampuan pengeluaran masyarakat.

4.9. Heatmap K=3

Heatmap digunakan untuk melihat perbedaan karakteristik antar cluster secara lebih komprehensif.



Gambar 6. Heatmap K=3

Berdasarkan heatmap, terlihat bahwa cluster dengan tingkat kemiskinan tinggi memiliki nilai pengeluaran yang relatif lebih rendah dibandingkan cluster lainnya. Perbedaan antar cluster tersebut menunjukkan adanya heterogenitas kondisi sosial ekonomi pada masing-masing wilayah.

4.10. Hasil Clustering K=5

Clustering K=5 digunakan untuk memberikan segmentasi wilayah yang lebih rinci.

Tabel 2. Kategori Cluster K=5

No	Cluster	P0	Kategori
1	0	10.9175	Tinggi
2	1	7.3500	Rendah
3	2	12.8030	Sangat Tinggi
4	3	10.7960	Sedang
5	4	5.715	Sangat Rendah

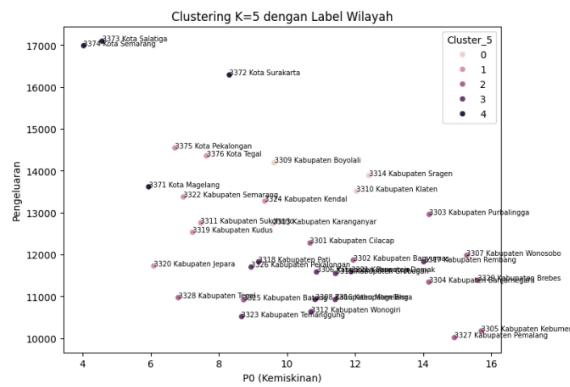
Kategori ditentukan berdasarkan urutan nilai rata-rata P0.

4.11. Interpretasi Cluster K=5

Clustering K=5 memberikan pemetaan area yang lebih rinci dibandingkan K=3, di mana kelompok kategori sangat tinggi merefleksikan daerah yang memiliki tingkat kemiskinan yang parah serta pengeluaran yang minimal. Area yang termasuk dalam kategori ini membutuhkan tindakan struktural berkelanjutan yang meliputi perbaikan infrastruktur, akses terhadap pendidikan, layanan kesehatan, serta pengembangan ekonomi setempat, sementara kelompok kategori rendah dan sangat rendah menunjukkan area dengan kondisi ekonomi yang lebih stabil dan tingkat kesejahteraan rakyat yang cenderung lebih baik.

4.12. Visualisasi Clustering K=5

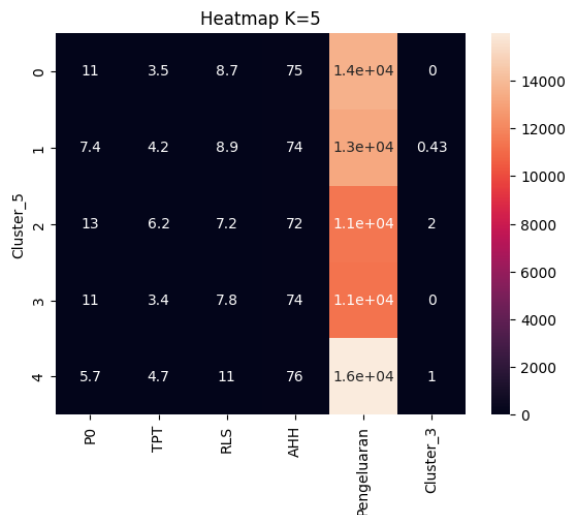
Visualisasi scatter plot K=5 menunjukkan persebaran wilayah yang lebih rinci.



Gambar 7. Scatter Plot K=5

Setiap titik pada scatter plot merepresentasikan satu wilayah, sehingga memudahkan dalam mengidentifikasi posisi relatif masing-masing wilayah berdasarkan tingkat kemiskinan dan pengeluaran per kapita.

4.13. Heatmap K=5



Gambar 8. Heatmap K=5

Visualisasi heatmap pada K=5 memperlihatkan perbedaan karakteristik yang lebih detail pada setiap cluster, di mana cluster dengan kategori sangat tinggi menunjukkan tingkat kemiskinan yang paling besar dibandingkan cluster lainnya. Hasil tersebut mengindikasikan bahwa metode *clustering* dapat digunakan untuk mengidentifikasi pola kondisi sosial ekonomi wilayah secara lebih mendalam dan sistematis.

4.14. Hasil Prioritas Intervensi Pemerintah

Berdasarkan hasil *clustering*, prioritas intervensi pemerintah dapat dibagi menjadi tiga kategori:

Tabel 3. Kategori Prioritas Intervensi

No	Kategori	Prioritas
1	Tinggi/Sangat Tinggi	Prioritas Utama
2	Sedang	Prioritas Menengah
3	Rendah/Sangat Rendah	Prioritas Rendah

Wilayah dengan kategori tinggi dan sangat tinggi menjadi prioritas utama dalam program pengentasan kemiskinan karena memiliki tingkat kerentanan ekonomi yang lebih besar. Sementara itu, wilayah dengan kategori sedang memerlukan intervensi berupa penguatan ekonomi lokal dan peningkatan kualitas sumber daya manusia, sedangkan wilayah dengan kategori rendah dan sangat rendah lebih difokuskan pada upaya menjaga stabilitas ekonomi serta meningkatkan daya saing daerah.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan menggunakan metode K-Means Clustering, diperoleh bahwa metode ini mampu mengelompokkan kabupaten/kota berdasarkan tingkat kemiskinan dan pengeluaran per kapita. Penelitian dilakukan menggunakan pendekatan *CRISP-DM* dengan proses normalisasi data menggunakan *MinMaxScaler* sebelum dilakukan clustering. Hasil evaluasi menunjukkan bahwa model K=3 memiliki nilai *Silhouette Score* sebesar 0.308, lebih baik dibandingkan K=5 sebesar 0.229, sehingga K=3 dipilih sebagai model utama penelitian.

Hasil *clustering* K=3 menghasilkan tiga kategori wilayah, yaitu tinggi, sedang, dan rendah. Wilayah dengan tingkat kemiskinan tinggi cenderung memiliki pengeluaran per kapita yang rendah, sedangkan wilayah dengan tingkat kemiskinan rendah memiliki pengeluaran yang lebih tinggi. Pada *clustering* K=5 diperoleh segmentasi yang lebih rinci, yaitu sangat tinggi, tinggi, sedang, rendah, dan sangat rendah. Hasil penelitian menunjukkan bahwa metode *K-Means* dapat membantu dalam menentukan prioritas intervensi pemerintah secara lebih terarah berdasarkan kondisi sosial ekonomi wilayah.

Berdasarkan hasil penelitian, pemerintah diharapkan dapat memanfaatkan hasil clustering sebagai dasar dalam menentukan prioritas program pengentasan kemiskinan, terutama pada wilayah dengan kategori tinggi dan sangat tinggi. Penelitian selanjutnya disarankan menggunakan variabel yang lebih banyak serta membandingkan metode *K-Means* dengan metode *clustering* lainnya agar hasil yang diperoleh lebih optimal. Selain itu, pengembangan sistem visualisasi berbasis web dan penggunaan data beberapa tahun juga dapat dilakukan untuk mendukung analisis yang lebih mendalam.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik, "Persentase Penduduk Miskin Indonesia Tahun 2024." [Online]. Available: <https://www.bps.go.id>
- [2] K. Setiawan, Kastum, and Y. P. Pratama, "K-Means Clustering Analysis of Poverty Data in Cilacap District," *ijsecs*, vol. 5, no. 1, pp. 53–62, Apr. 2025, doi: 10.35870/ijsecs.v5i1.3759.
- [3] M. W. Sianturi *et al.*, "Penerapan Analisis Cluster K-Means untuk Pengelompokan Kabupaten/Kota di Provinsi Sumatera Utara

- Berdasarkan Indikator Kemiskinan Tahun 2023,” vol. 5, no. 3, 2025.
- [4] G. R. Pratiwi, D. Wahiddin, E. E. Awal, and A. Fauzi, “Klasterisasi Tingkat Kemiskinan Kabupaten/Kota di Indonesia Menggunakan Algoritma K-Means dan K-Medoids,” vol. 21, no. 2, 2024.
- [5] A. Muliawan, D. A. Fauziah, M. Hermansyah, N. A. Prasetyo, and M. Faiz, “Implementation of K-Means Algorithm for Poverty Clustering (Case Study: East Java Province),” *Inside Jurnal Sistem Informatika Cerdas*, vol. 2, no. 2, pp. 5–9, Feb. 2024.
- [6] 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.
- [7] S. Wahyuni, A. Hananto, B. Huda, F. Apriani, and Tukino, “Identifying Regional Patterns of Poverty in Indonesia: a Clustering Approach Using K-Means,” *International Journal of Computer and Information System*, vol. 6, no. 1, pp. 45–52, Feb. 2025.
- [8] Amelia, I. M. Nur, M. Rizky, and S. P. Milasari, “Pengelompokan Tingkat Kemiskinan di Provinsi Jawa Barat dengan Metode K-Means Clustering,” *Journal of Data Insights*, vol. 1, no. 2, pp. 51–61, Dec. 2023.
- [9] I. Tawakal, M. M. Effendi, and A. M. Majid, “ANALISIS TINGKAT KEMISKINAN DENGAN ALGORITMA K-MEANS MENGGUNAKAN RAPIDMINER DI TINGKAT KOTA KABUPATEN DI JAWA TENGAH,” *Journal of Information System Management (JOISM)*, vol. 7, no. 1, pp. 112–119, 2025.
- [10] A. Khalif, A. N. Hasanah, M. H. Ridwan, and B. N. Sari, “Klasterisasi Tingkat Kemiskinan di Indonesia menggunakan Algoritma K-Means,” *generation*, vol. 8, no. 1, pp. 54–62, Mar. 2024, doi: 10.29407/gj.v8i1.21470.
- [11] A. E. Wulandari *et al.*, “Pengelompokan Kemiskinan di Provinsi Sulawesi Selatan Tahun 2023 dengan Metode K-Means Clustering,” *Journal of Statistics and Its Application*, vol. 6, no. 2, Jun. 2023.
- [12] W. S. Agustin, R. Armadhan, S. K. Dini, and H. Murda, “Comparison of K-Means and K-Medoids Methods in Grouping Provinces in Indonesia Based on Economic Development Indicators,” *J. Math. Comput. Stat.*, vol. 9, no. 1, pp. 131–147, Mar. 2026, doi: 10.35580/jmathcosv9n19866.
- [13] N. R. Handitia and A. Sofro, “ANALISIS CLUSTERING INDEKS KEDALAMAN KEMISKINAN DI INDONESIA MENGGUNAKAN K-MEANS,” *Jurnal Ilmiah Matematika*, vol. 13, pp. 347–383, 2025.
- [14] G. Erda, C. Gunawan, and Z. Erda, “GROUPING OF POVERTY IN INDONESIA USING K-MEANS WITH SILHOUETTE COEFFICIENT,” *Pr*, vol. 3, no. 1, pp. 1–6, Jun. 2023, doi: 10.22487/27765660.2023.v3.i1.16435.
- [15] N. N. Fransiska R and D. S. Anggraeni, “Pengelompokan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma K-Means dengan Silhouette Coefficient,” *TEMATIK*, vol. 9, no. 1, pp. 29–35, Jun. 2022, doi: 10.38204/tematik.v9i1.901.