CLUSTERING WILAYAH PULAU JAWA BERDASARKAN INDIKATOR SOSIAL EKONOMI MENGGUNAKAN METODE *K-MEANS*

Khalimatus Sa'diyah, Kinthana Diaz Primadhieta, Harun Al Rosyid
Teknik Informatika, Universitas Negeri Surabaya
Jl. Ketintang, Ketintang, Kec. Gayungan, Kota Surabaya, Jawa Timur 60231
khalimatus.22080@mhs.unesa.ac.id,

ABSTRAK

Pulau Jawa memiliki peran strategis dalam pembangunan nasional karena konsentrasi penduduk yang tinggi, namun masih menghadapi kesenjangan kondisi sosial ekonomi antar wilayah. Perbedaan tingkat pendidikan, kesehatan, pengangguran, dan kemiskinan menunjukkan perlunya analisis berbasis data untuk mendukung perencanaan pembangunan yang lebih tepat sasaran. Permasalahan utama dalam pembangunan wilayah adalah belum adanya pengelompokan kabupaten/kota yang mampu merepresentasikan karakteristik sosial ekonomi secara komprehensif. Oleh karena itu, penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Pulau Jawa berdasarkan indikator sosial ekonomi menggunakan metode K-Means Clustering. Data yang digunakan merupakan data sekunder Badan Pusat Statistik (BPS) tahun 2023 yang mencakup 119 kabupaten/kota dengan lima indikator, yaitu Tingkat Pengangguran Terbuka (TPT), Indeks Kedalaman Kemiskinan (P1), Indeks Pembangunan Manusia (IPM), Angka Harapan Hidup (AHH), dan Rata-rata Lama Sekolah (RLS). Tahapan penelitian meliputi preprocessing data, normalisasi menggunakan StandardScaler, penentuan jumlah cluster optimal dengan Elbow Method dan Silhouette Score, serta visualisasi hasil clustering menggunakan Principal Component Analysis (PCA). Hasil penelitian menunjukkan bahwa wilayah di Pulau Jawa terbagi ke dalam empat cluster dengan karakteristik sosial ekonomi yang berbeda, yaitu sangat baik, menengah, menengah-bawah, dan tertinggal. Hasil ini diharapkan dapat menjadi dasar pertimbangan dalam perumusan kebijakan pembangunan wilayah yang lebih efektif dan berbasis data.

Kata kunci : *K-Means, Sosial Ekonomi, Clustering, Pulau Jawa.*

1. PENDAHULUAN

Sebagai wilayah dengan konsentrasi penduduk tertinggi di Indonesia, Pulau Jawa memiliki peran penting dalam mendorong pertumbuhan dan pemerataan pembangunan nasional. Perbedaan kondisi sosial ekonomi antar wilayah, seperti tingkat pendidikan, kemiskinan, dan pengangguran, menyebabkan terjadinya kesenjangan pembangunan yang perlu dianalisis secara sistematis. Analisis berbasis data diperlukan agar kebijakan pembangunan dapat lebih tepat sasaran dan sesuai dengan karakteristik wilayah masing-masing [1], [2]. Permasalahan utama dalam perencanaan pembangunan wilayah adalah belum adanya pengelompokan kabupaten/kota yang mampu merepresentasikan kesamaan karakteristik sosial ekonomi secara komprehensif. Oleh karena itu, teknik *clustering* digunakan untuk mengelompokkan wilayah berdasarkan kemiripan indikator sosial ekonomi agar pola pembangunan dapat diidentifikasi dengan lebih jelas [3], [4]. Berdasarkan permasalahan tersebut, tujuan penelitian ini adalah mengelompokkan kabupaten/kota di Pulau Jawa berdasarkan indikator sosial ekonomi guna memperoleh gambaran struktur dan karakteristik pembangunan wilayah yang lebih terukur. Untuk mencapai tujuan tersebut, penelitian ini menerapkan algoritma *K-Means* karena kemampuannya dalam mengelompokkan data numerik secara efisien dengan mekanisme perhitungan yang relatif sederhana serta menghasilkan kelompok yang mudah diinterpretasikan. Dengan demikian, hasil penelitian ini diharapkan dapat menjadi dasar pertimbangan bagi pemangku kebijakan dalam merumuskan strategi pembangunan wilayah yang lebih tepat sasaran dan berbasis data.

2. TINJAUAN PUSTAKA

2.1. Data Mining dan *Clustering*

Data mining merupakan proses analisis data untuk menemukan pola atau informasi yang bermakna dari kumpulan data berukuran besar. Salah satu teknik yang digunakan dalam data mining adalah *clustering*, yaitu proses pengelompokan data ke dalam beberapa kelompok berdasarkan tingkat kemiripan karakteristik tertentu [3],[8].

2.2. *K-Means Clustering*

K-Means adalah metode *clustering* berbasis partisi yang mengelompokkan data ke dalam sejumlah *k cluster* berdasarkan jarak terdekat terhadap pusat *cluster* (*centroid*). Metode ini banyak digunakan dalam penelitian sosial ekonomi di Indonesia karena mampu menangani data numerik secara efisien dan menghasilkan pengelompokan yang mudah diinterpretasikan [5], [6].

2.3. Normalisasi Data

Normalisasi data dilakukan untuk menyamakan skala antar variabel agar tidak terjadi dominasi *variabel* tertentu dalam proses *clustering*. Tahap ini sangat penting dalam penerapan *algoritma K- Means* yang berbasis perhitungan jarak [6],[7].

2.4. Evaluasi Kualitas *Cluster*

Evaluasi kualitas cluster diperlukan untuk memastikan bahwa hasil pengelompokan yang dihasilkan memiliki tingkat pemisahan dan homogenitas yang baik. *Elbow Method* digunakan untuk menentukan jumlah cluster optimal dengan menganalisis perubahan nilai *Within-Cluster Sum of Squares* (WCSS) terhadap jumlah cluster [10], [14]. Selain itu, *Silhouette Score* digunakan untuk mengukur tingkat kedekatan data dalam satu cluster dibandingkan dengan cluster lainnya, sehingga dapat menilai kualitas struktur cluster secara kuantitatif [12], [14]. Penggunaan kedua metode evaluasi ini membantu menghasilkan cluster yang lebih representatif dan dapat diinterpretasikan secara ilmiah.

2.5. Penelitian Terdahulu

Sejumlah studi di Indonesia telah mengembangkan analisis pengelompokan wilayah berbasis indikator sosial ekonomi diantaranya Fauzan dan Dini [3], [4] melakukan *clustering* provinsi di Indonesia berdasarkan indikator kesejahteraan masyarakat dan menunjukkan bahwa teknik *clustering* mampu mengungkap pola ketimpangan antar wilayah secara jelas. Agustine et al. [5] menerapkan algoritma *K-Means* untuk mengelompokkan wilayah berdasarkan aspek sosial ekonomi dan memperoleh hasil cluster dengan karakteristik kesejahteraan yang berbeda-beda. Penelitian oleh Saputra dan Kurniawan [6] juga membuktikan bahwa *K-Means* efektif digunakan untuk pengelompokan data sosial ekonomi dan mendukung analisis pembangunan wilayah. Selain itu, Nugroho dan Setiawan [10] menegaskan bahwa kombinasi preprocessing, standarisasi, dan *K-Means* menghasilkan pengelompokan wilayah yang lebih akurat dan mudah diinterpretasikan. Berdasarkan penelitian terdahulu tersebut, metode *K-Means* dinilai relevan dan layak diterapkan dalam pengelompokan kabupaten/kota di Pulau Jawa berdasarkan indikator sosial ekonomi.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan teknik data mining untuk mengelompokkan wilayah berdasarkan indikator sosial ekonomi. Data yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) tahun 2023 yang mencakup 119 kabupaten/kota di Pulau Jawa [1], [2]. Variabel yang dianalisis meliputi Tingkat Pengangguran Terbuka (TPT), Indeks Kedalaman Kemiskinan (P1), Indeks Pembangunan Manusia (IPM), Angka Harapan Hidup (AHH), dan Rata-rata Lama Sekolah (RLS). Tahapan penelitian diawali dengan pengumpulan data, kemudian dilanjutkan dengan preprocessing yang meliputi penanganan missing value, deteksi outlier, serta pemilihan variabel numerik untuk memastikan kualitas data sebelum dianalisis. Selanjutnya, dilakukan normalisasi dan standarisasi data menggunakan *StandardScaler* untuk menyamakan skala antar variabel agar tidak terjadi dominasi variabel tertentu dalam perhitungan jarak. Penentuan jumlah cluster optimal dilakukan menggunakan *Elbow Method* dan *Silhouette Score* untuk memperoleh jumlah cluster yang paling representatif. Setelah itu, algoritma K-Means diterapkan untuk mengelompokkan kabupaten/kota di Pulau Jawa berdasarkan tingkat kemiripan indikator sosial ekonomi, dan hasil *clustering* dianalisis melalui nilai rata-rata setiap variabel pada masing-masing cluster guna menginterpretasikan karakteristik pembangunan wilayah secara lebih terukur [6], [10], [14]. Tabel berikut menunjukkan contoh sebagian data asli sebelum dilakukan preprocessing:

Tabel 1. Data Awal

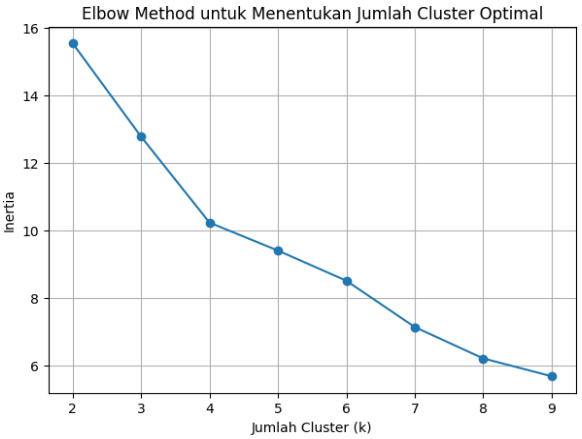
No	Kabupaten /Kota	TPT	P1	IPM	AHH	RLS
1	Kabupaten A	8,11	1,52	75,91	74,89	9,03
2	Kabupaten B	5,37	0,37	86,71	76,02	11,75
3	Kabupaten C	7,24	0,76	84,26	75,37	11,88
4	Kabupaten D	6,76	0,71	83,99	74,75	11

Data mentah selanjutnya diproses melalui tahap preprocessing yang meliputi penanganan nilai kosong, deteksi outlier, serta standarisasi data sebelum diterapkan ke dalam model *K-Means*. Tahap ini bertujuan untuk memastikan kualitas dan konsistensi data sehingga dapat menjadi dasar yang valid dalam analisis pengelompokan wilayah [6], [10]. Setelah itu, dilakukan normalisasi terhadap seluruh variabel indikator sosial ekonomi karena perbedaan satuan, skala, dan rentang nilai antar variabel berpotensi memengaruhi hasil perhitungan jarak dalam proses *clustering.*

4. HASIL DAN PEMBAHASAN

Setelah data awal diperoleh, tahap selanjutnya adalah melakukan normalisasi terhadap seluruh variabel indikator sosial ekonomi. Normalisasi diperlukan karena setiap variabel memiliki satuan, skala, dan rentang nilai yang berbeda, sehingga perbedaan skala tersebut berpotensi menimbulkan bias dalam proses *clustering*. Tanpa normalisasi, variabel dengan rentang nilai yang lebih besar dapat mendominasi perhitungan jarak dan memengaruhi hasil pengelompokan wilayah [10],[14]. Dalam penelitian ini, proses normalisasi dilakukan dengan menyesuaikan karakteristik masing-masing indikator sosial ekonomi. Indikator yang bersifat positif, yaitu indikator dengan makna semakin besar nilainya menunjukkan kondisi yang semakin baik. Sementara itu, indikator yang bersifat negatif, seperti Tingkat Pengangguran Terbuka (TPT) dan Indeks Kedalaman Kemiskinan (P1), dinormalisasi secara terbalik agar interpretasi seluruh variabel menjadi konsisten, yaitu semakin tinggi nilai hasil normalisasi menunjukkan kondisi sosial ekonomi yang semakin baik [6]. Hasil normalisasi data menunjukkan bahwa seluruh variabel indikator sosial ekonomi telah berada pada skala yang sebanding. Proses standarisasi ini dilakukan

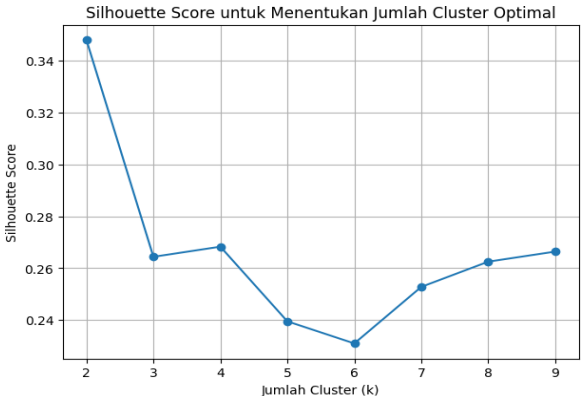
menggunakan *StandardScaler* sehingga setiap variabel memiliki kontribusi yang seimbang dalam perhitungan jarak pada algoritma *K-Means*. Dengan demikian, hasil *clustering* yang diperoleh mencerminkan kemiripan karakteristik wilayah secara objektif dan tidak dipengaruhi oleh perbedaan skala antar variable[10]. Penentuan jumlah *cluster* optimal dalam penelitian ini dilakukan menggunakan *Elbow Method* dan *Silhouette Score* untuk memperoleh jumlah *cluster* yang paling representatif. *Elbow Method* digunakan dengan menganalisis perubahan nilai *Within-Cluster Sum of Squares* (WCSS) terhadap variasi jumlah *cluster* [12],[14].



Gambar 1. *Elbow Method*

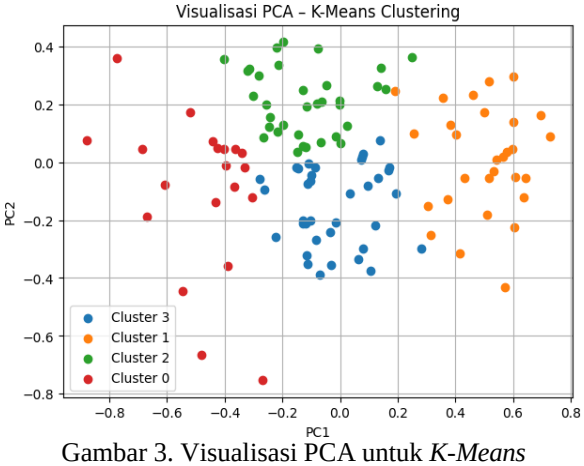
Berdasarkan visualisasi grafik *Elbow* pada Gambar 1, terlihat bahwa penurunan nilai WCSS terjadi secara signifikan pada rentang k = 2 hingga k = 4. Setelah k = 4, penurunan nilai WCSS cenderung melandai, yang menunjukkan bahwa penambahan jumlah *cluster* selanjutnya tidak memberikan peningkatan kualitas pengelompokan yang signifikan. Oleh karena itu, k = 4 diidentifikasi sebagai titik *elbow* yang merepresentasikan keseimbangan antara kompleksitas model dan kualitas *clustering*. Dengan demikian, jumlah *cluster* sebanyak empat dipilih sebagai jumlah *cluster* optimal dalam penelitian ini sesuai dengan pendekatan *Elbow Method* [10], [14].

Sebagai evaluasi tambahan, dilakukan pengukuran kualitas pemisahan *cluster* menggunakan *Silhouette Score***.** Berdasarkan visualisasi pada Gambar 2, nilai *Silhouette Score* pada beberapa jumlah *cluster* menunjukkan struktur *cluster* yang cukup stabil. Meskipun nilai *Silhouette Score* tertinggi diperoleh pada k = 2, nilai pada k = 4 tetap menunjukkan tingkat pemisahan *cluster* yang memadai dan konsisten [14].



Gambar 2. *Silhouette Score.*

Dengan mempertimbangkan *Elbow Method* sebagai indikator utama serta *Silhouette Score* sebagai evaluasi pendukung, jumlah *cluster* yang digunakan dalam penelitian ini ditetapkan sebanyak empat *cluster* (k = 4) [10],[14]**.** Untuk memudahkan interpretasi hasil *clustering*, dilakukan visualisasi menggunakan *Principal Component Analysis* (PCA). PCA mereduksi dimensi data dari 5 variabel menjadi 2 komponen utama sehingga hasil *cluster* dapat divisualisasikan dalam bentuk plot 2D. Visualisasi ini membantu Berdasarkan hasil analisis menggunakan metode *K-Means*, 119 wilayah kabupaten/kota di Pulau Jawa terbagi ke dalam empat *cluster*, yaitu *Cluster* 1, *Cluster* 2, Cluster 3, dan *Cluster* 0. Interpretasi *cluster* dilakukan dengan menganalisis nilai rata-rata masing-masing indikator pada setiap *cluster* [6],[10].



Gambar 3. Visualisasi PCA untuk *K-Means*

Tabel 2. Hasil Clustering K-Means

Cluster *K-Means*	**Kabupaten/Kota**
0	['Pandeglang', 'Lebak', 'Serang', '3303 Kabupaten Purbalingga', '3304 Kabupaten Banjarnegara', '3307 Kabupaten Wonosobo', '3327 Kabupaten Pemalang', '3329 Kabupaten Brebes', 'Kabupaten Lumajang', 'Kabupaten Jember', 'Kabupaten Bondowoso', 'Kabupaten Situbondo', 'Kabupaten Probolinggo', 'Kabupaten Pasuruan', 'Kabupaten Bojonegoro', 'Kabupaten Tuban', 'Kabupaten Bangkalan', 'Kabupaten Sampang', 'Kabupaten Pamekasan', 'Kabupaten Sumenep']

Cluster ***K-Means***	**Kabupaten/Kota**
1	['Jakarta Selatan', 'Jakarta Timur', 'Jakarta Pusat', 'Jakarta Barat', 'Jakarta Utara', 'Kota Bogor', 'Kota Sukabumi', 'Kota Bandung', 'Kota Cirebon', 'Kota Bekasi', 'Kota Depok', 'Kota Cimahi', 'Kota Tangerang', 'Kota Tangerang Selatan', '3371 Kota Magelang', '3372 Kota Surakarta', '3373 Kota Salatiga', '3374 Kota Semarang', 'Bantul', 'Sleman', 'Kota Yogyakarta', 'Kabupaten Sidoarjo', 'Kota Kediri', 'Kota Blitar', 'Kota Malang', 'Kota Mojokerto', 'Kota Madiun', 'Kota Surabaya','Kota Batu']
2	['Tasikmalaya', 'Ciamis', 'Majalengka', 'Pangandaran', '3305 Kabupaten Kebumen', '3306 Kabupaten Purworejo', '3308 Kabupaten Magelang', '3309 Kabupaten Boyolali', '3310 Kabupaten Klaten', '3311 Kabupaten Sukoharjo', '3312 Kabupaten Wonogiri', '3313 Kabupaten Karanganyar', '3314 Kabupaten Sragen', '3315 Kabupaten Grobogan', '3316 Kabupaten Blora', '3317 Kabupaten Rembang', '3318 Kabupaten Pati', '3319 Kabupaten Kudus', '3320 Kabupaten Jepara', '3321 Kabupaten Demak', '3322 Kabupaten Semarang', '3323 Kabupaten Temanggung', '3326 Kabupaten Pekalongan', 'Kulonprogo', 'Gunungkidul', 'Kabupaten Pacitan', 'Kabupaten Ponorogo', 'Kabupaten Trenggalek', 'Kabupaten Blitar', 'Kabupaten Mojokerto', 'Kabupaten Jombang', 'Kabupaten Nganjuk', 'Kabupaten Magetan', 'Kabupaten Ngawi']
3	['Kepulauan Seribu', 'Bogor', 'Sukabumi', 'Cianjur', 'Bandung', 'Garut', 'Kuningan', 'Cirebon', 'Sumedang', 'Indramayu', 'Subang', 'Purwakarta', 'Karawang', 'Bekasi', 'Bandung Barat', 'Kota Tasikmalaya', 'Kota Banjar', 'Tangerang', 'Kota Cilegon', 'Kota Serang', '3301 Kabupaten Cilacap', '3302 Kabupaten Banyumas', '3324 Kabupaten Kendal', '3325 Kabupaten Batang', '3328 Kabupaten Tegal', '3375 Kota Pekalongan', '3376 Kota Tegal', 'Kabupaten Tulungagung', 'Kabupaten Kediri', 'Kabupaten Malang', 'Kabupaten Banyuwangi', 'Kabupaten Madiun', 'Kabupaten Lamongan', 'Kabupaten Gresik', 'Kota Probolinggo', 'Kota Pasuruan']

Berdasarkan hasil analisis menggunakan metode *K-Means*, wilayah kabupaten/kota di Pulau Jawa terbagi ke dalam empat cluster, yaitu Cluster 1, Cluster 2, Cluster 3, dan Cluster 0. Perbedaan antar cluster ditentukan berdasarkan kedekatan nilai indikator Tingkat Pengangguran Terbuka (TPT), Indeks Kedalaman Kemiskinan (P1), Indeks Pembangunan Manusia (IPM), Angka Harapan Hidup (AHH), dan Rata-rata Lama Sekolah (RLS). Interpretasi cluster dilakukan dengan menganalisis nilai rata-rata masing-masing indikator pada setiap cluster [6],[7],[10].

- *Cluster* 0
  Cluster 0 memiliki nilai indikator yang berada pada tingkat menengah-bawah**,** dengan nilai IPM, RLS, dan AHH lebih rendah dibandingkan *Cluster* 2, namun masih lebih tinggi dibandingkan *Cluster* 3. Tingkat kemiskinan pada *cluster* ini berada di atas *Cluster* 1 dan *Cluster* 2, tetapi tidak setinggi *Cluster* 3. Pola tersebut menunjukkan bahwa *Cluster* 0 merupakan kelompok wilayah dengan karakteristik transisi antara *cluster* menengah dan *cluster* dengan nilai indikator rendah
- *Cluster 1*
  Cluster 1 memiliki nilai rata-rata IPM, RLS, dan AHH yang relatif paling tinggi**, serta** nilai P1 yang relatif rendah dibandingkan *cluster* lainnya. Pola ini menunjukkan bahwa wilayah dalam *cluster* ini memiliki capaian pendidikan dan kesehatan yang lebih baik, serta tingkat kemiskinan yang lebih rendah. Kesamaan nilai indikator membentuk kelompok wilayah dengan karakteristik sosial ekonomi yang relatif homogen.
- *Cluster* 2
  Cluster 2 ditandai oleh nilai IPM dan RLS pada tingkat menengah hingga cukup tinggi**,** dengan nilai P1 dan TPT berada pada kisaran sedang**.** Karakteristik tersebut menunjukkan bahwa wilayah dalam *cluster* ini memiliki kondisi sosial ekonomi yang cukup stabil, namun belum mencapai nilai indikator setinggi *Cluster* 1. Kedekatan nilai indikator mencerminkan pola perkembangan wilayah yang relatif serupa.
- *Cluster* 3
  Cluster 3 menunjukkan nilai IPM, RLS, dan AHH yang relatif lebih rendah**,** serta nilai P1 yang lebih tinggi dibandingkan *Cluster* 1 dan *Cluster* 2. Kondisi ini menggambarkan kelompok wilayah dengan keterbatasan pada aspek pendidikan, kesehatan, dan kesejahteraan ekonomi. Kesamaan karakteristik indikator antar wilayah membentuk *cluster* dengan tingkat homogenitas yang cukup tinggi.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa metode *K-Means* mampu mengelompokkan 119 wilayah kabupaten/kota di Pulau Jawa ke dalam empat *cluster* berdasarkan indikator sosial ekonomi, yaitu Tingkat Pengangguran Terbuka (TPT), Indeks Kedalaman Kemiskinan (P1), Indeks Pembangunan Manusia (IPM), Angka Harapan Hidup (AHH), dan Rata-rata Lama Sekolah (RLS). Hasil *clustering* menunjukkan adanya perbedaan karakteristik sosial ekonomi antar wilayah, sehingga pengelompokan ini dapat memberikan gambaran kondisi pembangunan yang lebih terstruktur dan objektif. Penentuan jumlah *cluster* optimal menggunakan *Elbow Method* yang didukung oleh *Silhouette Score* menghasilkan empat *cluster* yang representatif, sementara visualisasi PCA membantu dalam interpretasi pola pemisahan antar *cluster*. Oleh karena itu, hasil penelitian ini dapat dimanfaatkan sebagai bahan pertimbangan bagi pemangku kebijakan dalam perumusan strategi pembangunan wilayah yang lebih tepat sasaran. Sebagai saran, penelitian selanjutnya dapat mengembangkan analisis dengan menambahkan variabel sosial ekonomi lainnya, membandingkan metode *clustering* yang berbeda, serta menggunakan data lintas waktu untuk memperoleh gambaran

dinamika perubahan kondisi wilayah secara lebih komprehensif.

DAFTAR PUSTAKA

[1] Badan Pusat Statistik, Statistik Indonesia 2023. Jakarta: Badan Pusat Statistik, 2023. https://www.bps.go.id

[2] Badan Pusat Statistik, Indikator Kesejahteraan Rakyat 2023. Jakarta: Badan Pusat Statistik, 2023. https://www.bps.go.id

[3] A. Fauzan dan S. K. Dini, "Clustering Provinsi di Indonesia Berdasarkan Indikator Kesejahteraan Masyarakat," EKSAKTA: Journal of Sciences and Data Analysis, vol. 1, no. 1, 2020.

[4] A. Fauzan dan S. K. Dini, "Clustering Provinsi di Indonesia Berdasarkan Indikator Kesejahteraan Masyarakat," EKSAKTA: Journal of Sciences and Data Analysis, vol. 1, no. 1, 2020. https: //journal.uii.ac.id/Eksakta/article/view/14436

[5] V. Agustine et al., "Clustering Wilayah Berdasarkan Aspek Sosial Ekonomi Menggunakan Algoritma K-Means," Jurnal Ekonomi dan Studi Pembangunan, vol. 26, no. 2, 2025.https://journal. umy.ac.id/index.php/esp/article/view/26212

[6] R. A. Saputra dan D. Kurniawan, "Penerapan Algoritma K-Means untuk Pengelompokan Data Sosial Ekonomi," Jurnal RESTI, vol. 4, no. 2, 2020.https://jurnal.iaii.or.id/index.php/RESTI/article/view/1847

[7] R. Y. Farifah and D. Pramesti, "Cluster Analysis of Inclusive Economic Development Using K-Means Algorithm," Jurnal Varian, vol. 5, no. 2, 2022. https://doi.org/10.30812/varian.v5i2.1894

[8] "Clustering of Countries Through UMAP and K-Means," Logistics (MDPI), vol. 9, no. 3, 2025. https://www.mdpi.com/2305- 6290/9/3/108

[9] F. Khudin et al., "Analisis Komparatif Penerapan K-Means Clustering pada Lima Dataset Nyata," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 2, 2025.

[10] A. A. Nugroho dan R. Setiawan, "Implementasi Algoritma K-Means untuk Pengelompokan Data Sosial Ekonomi," *Jurnal RESTI*, vol. 6, no. 2, 2022.

[11] Parjito dan Permata, "Penerapan Data Mining untuk Clustering Data Penduduk Miskin Menggunakan Metode K-Means," *Ainet: Jurnal Informatika*, vol. 3, no. 1, 2025..

[12] F. Khudin et al., "Analisis Komparatif Penerapan K-Means Clustering pada Lima Dataset Nyata," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 2, 2025.

[13] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining, 2nd ed. Parjito dan Permata, "Penerapan Data Mining untuk Clustering Data Penduduk Miskin Menggunakan Metode K-Means," Ainet: Jurnal Informatika, vol. 3, no. 1, 2025. https://journal.unismuh.ac.id/ index.php/ainet/article/view/5878

[14] F. Khudin et al., "Analisis Komparatif Penerapan K-Means Clustering pada Lima Dataset Nyata," RIGGS: Journal of Artificial Intelligence and Digital Business, vol. 4, no. 2, 2025.https://journal.ilmudata.co.id/index.php/RIGGS/article/view/1361