

ANALISIS KONVERGENSI DAN EFEKTIVITAS ALGORITMA K-MEANS PADA KLASIFIKASI DATA KRIMINALITAS DAN PARTISIPASI KEAMANAN MASYARAKAT DI INDONESIA

Katrin Jenny Sirait¹, Cora Angelyna Hermanto²

Universitas Sumatera Utara, Medan

e-mail: ¹katrin@usu.ac.id, ²coraangelyna@gmail.com

Abstract: *This study aims to cluster provinces in Indonesia based on crime rates and community participation in security in 2024. The objective is to determine how effective the K-Means Clustering method is in analyzing this data. This study uses six crime indicators and five indicators of community participation at the provincial level. Before analysis, the data were cleaned and standardized using the Z-Score method. The clustering process employed the K-Means method with Euclidean distance. Cluster quality was evaluated using SSE and the Silhouette Score. The results show that the K-Means algorithm achieved convergence with a stable SSE value. However, determining the optimal number of clusters yielded different results between the Elbow method and the Silhouette Score. After considering stability and interpretability, three clusters were determined to be the optimal result. The clustering results indicate variations in characteristics across provinces regarding crime and community participation. The study also shows that cluster separation is better for crime data than for participation data. Thus, the K-Means Clustering method can be used as an initial mapping of regional security conditions in Indonesia. However, these results are also influenced by data characteristics and the sensitivity of centroid initialization.*

Keywords: *K-Means Clustering, Crime, Community Participation, Elbow Method, Silhouette Score.*

Abstrak: Penelitian ini bertujuan untuk mengelompokkan provinsi di Indonesia berdasarkan tingkat kriminalitas dan partisipasi masyarakat dalam keamanan pada tahun 2024. Tujuannya adalah untuk mengetahui seberapa efektif metode *K-Means Clustering* dalam menganalisis data ini. Penelitian ini menggunakan enam indikator kriminalitas dan lima indikator partisipasi masyarakat di tingkat provinsi. Sebelum dianalisis, data tersebut dibersihkan dan distandarisasi menggunakan metode Z-Score. Proses pengelompokan menggunakan metode *K-Means* dengan jarak Euclidean. Kualitas cluster dievaluasi menggunakan SSE dan *Silhouette Score*. Dari hasil penelitian, algoritma *K-Means* dapat mencapai konvergensi dengan nilai SSE yang stabil. Namun, penentuan jumlah cluster optimal menunjukkan hasil yang berbeda antara metode *Elbow* dan *Silhouette Score*. Setelah mempertimbangkan stabilitas dan interpretasi, ditetapkan bahwa tiga cluster adalah hasil yang optimal. Hasil klasifikasi menunjukkan adanya variasi karakteristik antarprovinsi pada aspek kriminalitas dan partisipasi masyarakat. Hasil penelitian juga menunjukkan bahwa pemisahan cluster pada data kriminalitas lebih baik dibandingkan data partisipasi. Dengan demikian, metode *K-Means Clustering* dapat digunakan sebagai pemetaan awal kondisi keamanan wilayah di Indonesia. Namun, hasil ini juga dipengaruhi oleh karakteristik data dan sensitivitas inisialisasi centroid.

Kata kunci: *K-Means Clustering, Kriminalitas, Partisipasi Masyarakat, Elbow Method, Silhouette Score.*

PENDAHULUAN

Perkembangan teknologi informasi

dan digitalisasi telah menghasilkan data dalam jumlah besar yang terus meningkat dari berbagai sektor. Ketersediaan data yang melimpah tersebut memerlukan metode analisis yang mampu mengidentifikasi pola dan karakteristik tersembunyi sehingga dapat dimanfaatkan dalam pengambilan keputusan yang lebih efektif. Salah satu teknik yang banyak digunakan dalam proses eksplorasi data adalah klasterisasi (*clustering*), yaitu metode pengelompokan objek berdasarkan tingkat kemiripan karakteristik yang dimilikinya. Tujuan utama klasterisasi adalah membentuk kelompok yang homogen di dalam klaster dan heterogen antar klaster sehingga pola yang terdapat dalam data dapat lebih mudah dipahami dan diinterpretasikan.

Salah satu metode klasterisasi yang banyak digunakan adalah K-Means Clustering. Cara kerja metode ini adalah dengan mengelompokkan sekumpulan objek ke dalam beberapa klaster berdasarkan kedekatan jaraknya terhadap titik pusat klaster (*centroid*), yang diperbarui secara berulang sampai mencapai kondisi konvergen. Tingkat popularitas K-Means tidak terlepas dari sifat algoritmanya yang sederhana, proses komputasi yang efisien, serta kemampuannya mengolah data berskala besar dengan hasil yang relatif mudah dipahami. Karena keunggulan tersebut, metode ini telah diaplikasikan secara luas di berbagai bidang, mulai dari pendidikan, kesehatan, ekonomi, pemasaran, hingga analisis kondisi keamanan suatu wilayah.

Meskipun memiliki keunggulan dalam kemudahan implementasi dan efisiensi komputasi, hasil klasterisasi menggunakan K-Means sangat dipengaruhi oleh pemilihan jumlah klaster dan inisialisasi *centroid*. Perbedaan jumlah klaster dapat menghasilkan struktur kelompok yang berbeda sehingga memengaruhi interpretasi hasil analisis. Selain itu, proses iteratif yang dilakukan oleh algoritma memerlukan evaluasi untuk memastikan bahwa solusi yang diperoleh telah mencapai kondisi stabil atau konvergen. Oleh karena itu, analisis

terhadap proses konvergensi dan efektivitas algoritma menjadi aspek penting dalam penerapan K-Means Clustering.

Beberapa penelitian sebelumnya menunjukkan bahwa metode K-Means mampu memberikan hasil yang baik dalam pengelompokan data kriminalitas. Maharani (2024) menggunakan kombinasi *Principal Component Analysis* (PCA) dan K-Means untuk mengelompokkan tingkat kriminalitas di Indonesia dan memperoleh struktur klaster yang optimal berdasarkan indeks *Davies–Bouldin*. Penelitian yang dilakukan oleh Sianturi (2025) menerapkan *Elbow Method* dan *Silhouette Score* untuk menentukan jumlah klaster optimal pada data kriminalitas di Provinsi Sumatera Utara. Sementara itu, Chairani (2024) berhasil mengidentifikasi wilayah dengan karakteristik kriminalitas yang berbeda menggunakan K-Means dan evaluasi *Silhouette Score*. Hasil serupa juga diperoleh Sulistina (2025) yang menunjukkan bahwa pemilihan jumlah klaster berpengaruh terhadap kualitas pengelompokan tingkat kriminalitas di Kota Jambi. Temuan-temuan tersebut menunjukkan bahwa K-Means merupakan metode yang potensial untuk mengidentifikasi pola dan karakteristik wilayah berdasarkan indikator keamanan.

Dalam konteks keamanan wilayah, tingkat kriminalitas tidak dapat dipisahkan dari peran masyarakat dalam menjaga keamanan lingkungan. Upaya keamanan berbasis masyarakat seperti pembangunan pos keamanan lingkungan (*poskamling*), pembentukan *regu kamling*, penambahan anggota perlindungan masyarakat (*Linmas*), pelaporan tamu menginap selama 1×24 jam, dan berbagai bentuk partisipasi sosial lainnya merupakan faktor yang dapat memengaruhi kondisi keamanan suatu daerah. Oleh karena itu, analisis keamanan wilayah yang hanya berfokus pada tingkat kriminalitas belum sepenuhnya menggambarkan kondisi keamanan secara menyeluruh. Integrasi antara indikator kriminalitas dan indikator

partisipasi masyarakat diperlukan untuk memperoleh pemetaan wilayah yang lebih komprehensif.

Penelitian ini menggunakan data kriminalitas dan partisipasi keamanan masyarakat tingkat provinsi di Indonesia tahun 2024 yang diperoleh dari Badan Pusat Statistik (BPS). Variabel kriminalitas terdiri atas enam indikator, yaitu intimidasi, pemerasan dan kekerasan terbuka, kejahatan kesusilaan dan kekerasan seksual, pencurian konvensional, kejahatan ekstrem terhadap nyawa dan harta benda, serta peredaran obat dan zat terlarang. Sementara itu, variabel partisipasi keamanan masyarakat terdiri atas lima indikator yang menggambarkan keterlibatan masyarakat dalam menjaga keamanan lingkungan.

Sebelum dilakukan proses klusterisasi, data perlu distandardisasi untuk menghilangkan pengaruh perbedaan satuan pengukuran antarvariabel. Standardisasi dilakukan menggunakan metode Z-Score dengan persamaan:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

dengan:

- z_{ij} = nilai hasil standardisasi observasi ke- i pada variabel ke- j ;
- x_{ij} = nilai asli observasi ke- i pada variabel ke- j ;
- $\bar{x}_j$ = rata-rata variabel ke- j ;
- s_j = simpangan baku variabel ke- j .

Standardisasi dilakukan agar seluruh variabel memiliki rata-rata 0 dan simpangan baku 1 sehingga tidak ada variabel yang mendominasi proses pengelompokan.

Proses pengelompokan menggunakan algoritma *K-Means* dilakukan berdasarkan jarak Euclidean yang dihitung menggunakan persamaan:

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^p (x_{ik} - c_{jk})^2}$$

dengan:

- $d(x_i, c_j)$ = jarak Euclidean antara objek ke- i dan centroid klaster ke- j ;
- x_{ik} = nilai objek ke- i pada variabel ke- k ;
- c_{jk} = nilai centroid klaster ke- j pada variabel ke- k ;
- p = jumlah variabel yang digunakan dalam analisis.

Setiap objek akan ditempatkan pada klaster yang memiliki jarak paling kecil terhadap centroid. Setelah seluruh objek teralokasi ke dalam klaster, centroid baru dihitung menggunakan rata-rata seluruh anggota klaster sebagai berikut:

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

dengan:

- c_j = centroid klaster ke- j ;
- n_j = jumlah anggota pada klaster ke- j ;
- x_i = data anggota klaster ke- j .

Proses alokasi objek dan pembaruan centroid dilakukan secara berulang hingga mencapai kondisi konvergen.

Konvergensi algoritma dievaluasi menggunakan nilai *Sum of Squared Error* (SSE), yaitu jumlah kuadrat jarak seluruh objek terhadap centroid klaster masing-masing yang dirumuskan sebagai:

$$SSE = \sum_{r=1}^k \sum_{x_i \in C_r} ||x_i - \mu_r||^2$$

dengan:

- SSE = jumlah kuadrat galat dalam klaster;
- k = jumlah klaster;
- C_r = klaster ke- r ;
- x_i = objek ke- i ;
- μ_r = centroid klaster ke- r .

Semakin kecil nilai SSE menunjukkan bahwa anggota dalam suatu klaster semakin dekat dengan centroid sehingga klaster yang terbentuk semakin homogen.

Penentuan jumlah kluster optimal dilakukan menggunakan *Elbow Method* dan *Silhouette Score*. Metode *Elbow* mengevaluasi perubahan nilai SSE pada berbagai jumlah kluster untuk menemukan titik penurunan yang mulai melambat. Selain itu, kualitas hasil klusterisasi juga dievaluasi menggunakan *Silhouette Score* yang dihitung dengan persamaan:

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

dengan:

- $s(i)$ = nilai silhouette untuk objek ke- i ;
- $a(i)$ = rata-rata jarak objek ke- i terhadap seluruh anggota dalam kluster yang sama;
- $b(i)$ = rata-rata jarak minimum objek ke- i terhadap anggota kluster lain.

Nilai *Silhouette Score* berada pada rentang -1 hingga 1 . Semakin mendekati 1 menunjukkan bahwa objek berada pada kluster yang tepat dan memiliki pemisahan yang baik terhadap kluster lainnya. Nilai rata-rata *Silhouette Score* dihitung menggunakan persamaan:

$$SC = \frac{1}{n} \sum_{i=1}^n s(i)$$

dengan:

- SC = rata-rata *Silhouette Score*;
- n = jumlah observasi;
- $s(i)$ = nilai silhouette observasi ke- i .

Semakin tinggi nilai rata-rata *Silhouette Score*, semakin baik kualitas kluster yang terbentuk.

Berdasarkan latar belakang tersebut, penelitian ini diarahkan untuk mengkaji proses konvergensi serta tingkat efektivitas algoritma K-Means Clustering dalam mengelompokkan provinsi-provinsi di Indonesia berdasarkan data kriminalitas dan partisipasi masyarakat dalam keamanan pada tahun 2024. Selain melakukan evaluasi terhadap proses konvergensi algoritma, penelitian ini juga membandingkan hasil penentuan jumlah

kluster yang diperoleh dari *Elbow Method* dan *Silhouette Score* untuk mendapatkan struktur kluster yang paling optimal. Hasil dari penelitian ini diharapkan dapat memberikan gambaran mengenai karakteristik kondisi keamanan di berbagai wilayah Indonesia, serta dapat dijadikan referensi dalam pengembangan metode klusterisasi maupun sebagai dasar pengambilan kebijakan keamanan yang berbasis data.

METODE

Penelitian ini menerapkan pendekatan kuantitatif eksploratif melalui metode unsupervised learning guna mengelompokkan provinsi-provinsi di Indonesia berdasarkan tingkat kriminalitas dan partisipasi masyarakat dalam bidang keamanan pada tahun 2024, dengan K-Means Clustering sebagai metode analisis utama yang digunakan. Proses klusterisasi dilakukan secara terpisah pada data kriminalitas dan data partisipasi keamanan masyarakat, kemudian hasilnya dibandingkan untuk memperoleh profil keamanan wilayah berdasarkan kategori rendah, sedang, dan tinggi.

Data penelitian merupakan data sekunder yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS), yaitu *Statistik Kriminal 2024/2025*. Unit analisis penelitian mencakup 34 provinsi di Indonesia yang memenuhi kriteria kelengkapan data. Variabel penelitian terdiri atas dua kelompok, yaitu variabel kriminalitas dan variabel partisipasi keamanan masyarakat.

Variabel Kriminalitas

- K1 = Intimidasi, pemerasan, dan kekerasan terbuka
- K2 = Kejahatan kesusilaan dan kekerasan seksual
- K3 = Pencurian konvensional
- K4 = Kejahatan ekstrem terhadap nyawa dan sabotase harta benda
- K5 = Peredaran obat dan zat terlarang

K6 = Pelanggaran ketertiban dan keamanan publik

Variabel Partisipasi Keamanan Masyarakat

P1 = Membangun poskamling

P2 = Membentuk regu kamling

P3 = Menambah anggota Hansip/Linmas

P4 = Pelaporan tamu menginap 1×24 jam

P5 = Pengaktifan sistem keamanan lingkungan dari inisiatif warga

Sebelum dilakukan analisis, data melalui tahap *pre-processing* yang meliputi pemeriksaan kelengkapan data (*missing value*), pembersihan data (*data cleaning*), dan standardisasi menggunakan metode *Z-Score*. Standardisasi dilakukan karena setiap indikator memiliki skala pengukuran yang berbeda sehingga diperlukan penyamaan skala sebelum proses klusterisasi.

Tahap analisis dimulai dengan statistik deskriptif untuk memetakan karakteristik data, mencakup nilai rata-rata, simpangan baku, nilai minimum dan maksimum, kuartil, serta median. Selanjutnya, dilakukan eksplorasi data melalui boxplot dan analisis korelasi Pearson untuk melihat pola distribusi data, mendeteksi kemungkinan outlier, serta mengidentifikasi keterkaitan antarvariabel.

Untuk menentukan jumlah kluster yang paling sesuai, digunakan dua metode, yaitu Elbow Method dan Silhouette Score. Elbow Method berfungsi untuk melihat titik di mana penurunan nilai Sum of Squared Error (SSE) mulai melandai secara signifikan, sementara Silhouette Score digunakan untuk menilai sejauh mana kluster yang terbentuk terpisah dengan baik dan memiliki tingkat kekompakan yang memadai.

Proses klusterisasi dilakukan menggunakan algoritma *K-Means* dengan tahapan menentukan jumlah kluster, menginisialisasi centroid awal, menghitung jarak antarobjek menggunakan jarak Euclidean,

mengelompokkan objek ke centroid terdekat, memperbarui centroid, dan mengulangi proses iterasi hingga tercapai kondisi konvergen. Analisis konvergensi dilakukan dengan mengamati perubahan nilai SSE pada setiap iterasi. Algoritma dinyatakan konvergen apabila nilai SSE telah stabil dan tidak mengalami perubahan yang berarti pada iterasi berikutnya.

Kualitas hasil klusterisasi dievaluasi menggunakan nilai SSE dan *Silhouette Score*. Nilai SSE digunakan untuk mengukur homogenitas anggota dalam suatu kluster, sedangkan *Silhouette Score* digunakan untuk menilai tingkat pemisahan antar kluster. Hasil klusterisasi kemudian divisualisasikan menggunakan *Principal Component Analysis* (PCA), grafik sebaran, dan peta tematik untuk mempermudah interpretasi hasil. Selanjutnya dilakukan analisis karakteristik setiap kluster berdasarkan nilai centroid dan distribusi anggotanya. Hasil kluster kriminalitas dan partisipasi masyarakat kemudian dikombinasikan melalui *cross tabulation* untuk memperoleh profil keamanan wilayah serta mengidentifikasi prioritas kewaspadaan pada masing-masing provinsi.

Seluruh tahapan pengolahan data, analisis statistik, proses klusterisasi, evaluasi hasil, hingga visualisasi dikerjakan dengan menggunakan bahasa pemrograman Python, didukung oleh sejumlah pustaka seperti Pandas, NumPy, Scikit-Learn, SciPy, Matplotlib, Seaborn, GeoPandas, dan Folium.

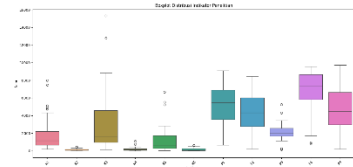
HASIL DAN PEMBAHASAN

Analisis Deskriptif Data

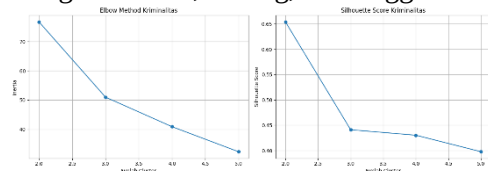
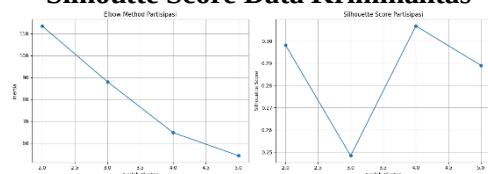
Hasil statistik deskriptif menunjukkan bahwa tingkat kriminalitas dan partisipasi keamanan masyarakat memiliki variasi yang cukup beragam antarprovinsi. Perbedaan karakteristik tersebut mengindikasikan perlunya pengelompokan wilayah untuk memperoleh pola yang lebih jelas.

Tabel 1 Statistik Deskriptif Variabel Penelitian

Variabel	Mean	Std	Min	25%	Median	75%	Max
K1	1967.24	2030.51	201.00	646.25	1168.00	2209.25	7970.00
K2	143.94	119.73	18.00	62.25	103.00	164.00	462.00
K3	3287.79	3784.97	155.00	969.00	1633.00	4577.50	15315.00
K4	198.32	232.43	16.00	54.00	113.50	232.50	1121.00
K5	1299.21	1603.87	39.00	298.50	650.00	1734.25	6680.00
K6	175.50	187.45	3.00	44.00	82.50	247.00	622.00
P1	5268.26	2287.55	669.00	3580.25	5475.00	6912.75	9101.00
P2	4264.91	2370.37	203.00	2771.50	4322.00	6033.25	8454.00
P3	2102.35	1042.28	147.00	1680.25	1982.00	2567.25	5274.00
P4	6708.06	2531.96	792.00	5821.25	7395.50	8644.00	9551.00
P5	4643.82	2941.64	199.00	2986.75	4481.00	6641.25	9726.00

**Gambar 1 Boxplot Distribusi Indikator Penelitian****Penentuan Jumlah Cluster**

Jumlah kluster ditentukan dengan menggunakan dua pendekatan, yaitu Elbow Method dan Silhouette Score. Keduanya mengarah pada kesimpulan yang sama bahwa penggunaan tiga kluster memberikan hasil paling optimal sekaligus mudah dimaknai, yakni sebagai kategori rendah, sedang, dan tinggi.

**Gambar 2 Hasil Elbow Method dan Silhouette Score Data Kriminalitas****Gambar 3 Hasil Elbow Method dan Silhouette Score Data Partisipasi Masyarakat****Hasil Klasterisasi K-Means**

Berdasarkan hasil K-Means Clustering, provinsi di Indonesia berhasil dikelompokkan ke dalam tiga cluster untuk data kriminalitas dan partisipasi keamanan masyarakat. Setiap cluster memiliki karakteristik yang berbeda sehingga dapat menggambarkan kondisi keamanan wilayah secara lebih terstruktur.

Tabel 2 Hasil Klasterisasi Provinsi Berdasarkan Kriminalitas dan Partisipasi Masyarakat

Provinsi	Cluster Kriminalitas	Kategori Kriminalitas	Cluster Partisipasi	Kategori Partisipasi
Aceh	C1	Sedang	C1	Sedang
Sumatera Utara	C2	Tinggi	C2	Rendah
Sumatera Barat	C1	Sedang	C1	Sedang
Riau	C1	Sedang	C1	Sedang
Jambi	C3	Rendah	C3	Tinggi
Sumatera Selatan	C1	Sedang	C1	Sedang
Bengkulu	C3	Rendah	C1	Sedang
Lampung	C1	Sedang	C3	Tinggi
Kep. Bangka Belitung	C3	Rendah	C1	Sedang
Kepulauan Riau	C3	Rendah	C1	Sedang
DKI Jakarta	C2	Tinggi	C3	Tinggi
Jawa Barat	C1	Sedang	C3	Tinggi
Jawa Tengah	C1	Sedang	C3	Tinggi
DI Yogyakarta	C3	Rendah	C3	Tinggi
Jawa Timur	C2	Tinggi	C3	Tinggi
Banten	C3	Rendah	C3	Tinggi
Bali	C3	Rendah	C3	Tinggi
Nusa Tenggara Barat	C3	Rendah	C3	Tinggi
Nusa Tenggara Timur	C1	Sedang	C2	Rendah
Kalimantan Barat	C3	Rendah	C1	Sedang
Kalimantan Tengah	C3	Rendah	C2	Rendah
Kalimantan Selatan	C3	Rendah	C1	Sedang
Kalimantan Timur	C3	Rendah	C3	Tinggi
Kalimantan Utara	C3	Rendah	C1	Sedang
Sulawesi Utara	C1	Sedang	C3	Tinggi
Sulawesi Tengah	C3	Rendah	C1	Sedang
Sulawesi Selatan	C2	Tinggi	C1	Sedang
Sulawesi Tenggara	C3	Rendah	C2	Rendah
Gorontalo	C3	Rendah	C1	Sedang
Sulawesi Barat	C3	Rendah	C2	Rendah
Maluku	C3	Rendah	C1	Sedang
Maluku Utara	C3	Rendah	C2	Rendah
Papua Barat	C3	Rendah	C2	Rendah
Papua	C1	Sedang	C2	Rendah

Tabel 3 Centroid Akhir Data Kriminalitas

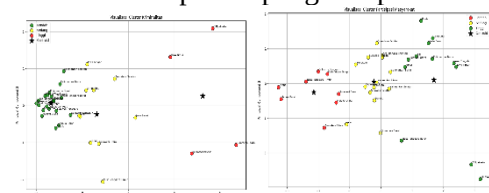
Cluster	K1	K2	K3	K4	K5	K6
C1	0.3308	0.3226	0.3123	0.3248	-0.0884	0.3130
C2	2.0778	2.2683	2.0625	2.0413	2.3840	2.0144
C3	-0.5810	-0.6150	-0.5686	-0.5707	-0.4326	-0.5594

Tabel 4 Centroid Akhir Data Partisipasi

Cluster	P1	P2	P3	P4	P5
C1	-0.1401	-0.0328	-0.0337	0.3408	-0.5304
C2	-1.2993	-0.6447	-0.7690	-1.3131	-0.7347
C3	1.0297	0.4680	0.5520	0.4777	1.1086

Visualisasi Sebaran Cluster

Sebaran cluster menunjukkan bahwa proses klasterisasi mampu memisahkan provinsi berdasarkan kemiripan karakteristik data. Visualisasi ini memperlihatkan perbedaan yang cukup jelas antarcluster sehingga mendukung efektivitas algoritma K-Means dalam proses pengelompokan.

**Gambar 4 Visualisasi Cluster Kriminalitas dan Partisipasi Masyarakat Menggunakan PCA**

Pemetaan Hasil Klasterisasi

Peta hasil klasterisasi menunjukkan persebaran geografis tingkat kriminalitas dan partisipasi keamanan masyarakat di Indonesia. Visualisasi spasial ini memudahkan identifikasi wilayah yang memerlukan perhatian lebih dalam upaya peningkatan keamanan dan partisipasi masyarakat.



Gambar 5 Peta Prioritas Kewaspadaan Wilayah Indonesia

Pembahasan

Temuan dalam penelitian ini mengindikasikan bahwa K-Means Clustering dapat diandalkan untuk mengklasifikasikan provinsi berdasarkan tingkat kriminalitas dan tingkat partisipasi masyarakat dalam menjaga keamanan. Pola pengelompokan yang dihasilkan dapat dimanfaatkan sebagai acuan dalam merumuskan kebijakan keamanan yang lebih sesuai dengan kondisi spesifik tiap wilayah.

SIMPULAN

Studi ini memperlihatkan bahwa metode K-Means Clustering efektif diterapkan untuk mengklasifikasikan provinsi-provinsi di Indonesia berdasarkan data tingkat kriminalitas dan keterlibatan masyarakat dalam keamanan. Hasil uji konvergensi mengindikasikan bahwa proses pengelompokan telah mencapai titik stabil, ditandai dengan posisi centroid yang relatif tetap pada iterasi-iterasi terakhir.

Dalam menentukan jumlah kluster optimal, penerapan Elbow Method dan Silhouette Score memberikan hasil yang tidak seragam. Meski demikian, dengan mempertimbangkan tren penurunan SSE serta kemudahan dalam menafsirkan hasil, tiga kluster dipilih sebagai jumlah

yang digunakan untuk kedua dataset. Pengelompokan yang dihasilkan memperlihatkan variasi karakteristik antarprovinsi, yang terbagi menjadi tiga kategori, yaitu rendah, sedang, dan tinggi. Dari sisi evaluasi, nilai Silhouette Score pada dataset kriminalitas menunjukkan kualitas kluster yang relatif memadai, sementara pada dataset partisipasi masyarakat masih ditemukan indikasi tumpang tindih antarkluster. Secara umum, dapat disimpulkan bahwa K-Means Clustering mampu menghasilkan gambaran yang jelas mengenai pola pengelompokan wilayah berdasarkan tingkat kriminalitas dan partisipasi keamanan masyarakat di Indonesia

DAFTAR PUSTAKA

- Badan Pusat Statistik. (2025). *Statistik kriminal 2024/2025*. Badan Pusat Statistik. Retrieved from <https://www.bps.go.id/assets/publication/2025/12/12/2edc8ea4c35b19ba912fc7e4/statistik-kriminal-2024-2025.html>
- Chairani, T. S., Listia, H., Wardaniah, S., Wulandari, S., Tampubolon, P. T. A., & Piliang, A. (2024). Clustering crime-prone areas in Indonesia using the K-Means method. *BIT (Fakultas Teknologi Informasi Universitas Budi Luhur)*, 21(2), 131–138. <https://doi.org/10.36080/bit.v21i2.3414>
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). John Wiley & Sons. <https://doi.org/10.1002/9780470977811>
- Jain, A. K. (2010). Data clustering: 50 years beyond K-Means. *Pattern Recognition Letters*, 31, 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical*

- Statistics and Probability* (Vol. 1, pp. 281–297).
- Maharrani, R. H., Abda'u, P. D., & Faiz, M. N. (2024). Clustering method for criminal crime acts using K-Means and PCA. *Indonesian Journal of Electrical Engineering and Computer Science*, 34(1), 224–232. <https://doi.org/10.11591/ijeecs.v34.i1.pp224-232>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Sianturi, M. D., Lubis, M. C., Payana, S. D., & Panjaitan, H. R. A. (2025). Analisis pengelompokan jenis kejahatan di Sumatera Utara berdasarkan pola kejadian tahunan menggunakan K-Means clustering. *AKSIOMA: Jurnal Sains Ekonomi dan Edukasi*, 2(6), 1308–1324. <https://doi.org/10.62335/aksioma.v2i6.1366>
- Sulistina, S., Utomo, P. E. P., & Hutabarat, B. F. (2025). Klasterisasi wilayah rawan kriminalitas di Kota Jambi (2022–2024) menggunakan algoritma K-Means. *Jurnal INSTEK (Informatika Sains dan Teknologi)*, 10(2), 546–560. <https://doi.org/10.24252/instek.v10i2.62051>