



INNOVATIVE: Journal Of Social Science Research

Volume 5 Nomor 3 Tahun 2025 Page 6646-6663

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

Analisis Pengelompokan Wilayah Kepolisian Daerah di Indonesia menggunakan Algoritma Hierarchical Clustering

M. Shadri Ismaun Lubis^{1✉}, Zakiy Maulana Pulungan², Hafiz Khalik Lubis³, Arnis Wulan Andari⁴,
Rusmawanty Napitupulu⁵, Bintang Maulana⁶, Arnita⁷

Universitas Negeri Medan

Email: shadriismaun.4231260013@mhs.unimed.ac.id^{1✉}

Abstrak

Penelitian ini bertujuan untuk mengelompokkan wilayah Kepolisian Daerah (Polda) di Indonesia berdasarkan karakteristik kejahatan pada tahun 2022 dengan menggunakan pendekatan data mining melalui algoritma hierarchical clustering. Dua variabel utama yang digunakan dalam analisis adalah jumlah kejahatan dan persentase penyelesaian kasus. Langkah-langkah analisis yang dilakukan adalah standarisasi data dengan Z-score, perhitungan matriks jarak Euclidean, pembuatan dendrogram untuk melihat struktur hierarki, dan evaluasi jumlah cluster yang optimal menggunakan Silhouette Coefficient. Hasil evaluasi menunjukkan jumlah cluster yang optimal adalah dua, dengan nilai Silhouette tertinggi sebesar 0,654. Cluster pertama terdiri dari 32 Polda dengan rata-rata jumlah kasus sebanyak 9.091 dan tingkat penyelesaian sebesar 26,32%, sehingga masuk dalam kategori cluster "Rentan". Cluster kedua terdiri dari dua Polda dengan jumlah kasus yang sangat tinggi dan tingkat penyelesaian sebesar 92,63%, sehingga masuk dalam kategori cluster "Sangat Tinggi". Visualisasi diagram sebar menunjukkan pemisahan spasial yang jelas antara kedua klaster, yang memperkuat validitas pemodelan. Hasil studi ini memberikan dasar untuk memetakan area prioritas dan merumuskan strategi untuk memperkuat sistem penegakan hukum, serta menjadikan area berkinerja tinggi sebagai tolok ukur nasional.

Kata Kunci: *Kriminalitas, Kepolisian Daerah, Pengelompokan Wilayah, Hierarchical Clustering, Dendrogram, Analisis Data, Evidence-Based Policing*

Abstract

This study aims to group the Regional Police (Polda) areas in Indonesia based on the characteristics of crime in 2022 using a data mining approach through the hierarchical clustering algorithm. The two main variables used in the analysis are the number of crimes and the percentage of case completion. The analysis steps data standardization with Z-score, calculation of the Euclidean distance matrix, creation of a dendrogram to see the hierarchical structure, and evaluation of the optimal number of clusters using the Silhouette Coefficient. The evaluation results show that the optimal number of clusters is two, with the highest Silhouette value of 0.654. The first cluster consists of 32 Polda with an average number of cases of 9,091 and a resolution rate of 26.32%, so it is categorized as the "Vulnerable" cluster. The second cluster consists of two Polda with a very high number of cases and a resolution rate of 92.63%, categorized as the "Very High" cluster. The scatter plot visualization shows a clear spatial separation between the two clusters, strengthening the validity of the modeling. The results of this study provide a basis for mapping priority areas and formulating strategies to strengthen the law enforcement system, as well as making high-performing areas a national benchmark.

Keywords: *Crime, Regional Police, Regional Clustering, Hierarchical Clustering, Dendrogram, Data Analysis, Evidence-Based Policing*

PENDAHULUAN

Permasalahan kriminalitas merupakan isu strategis yang terus menjadi sorotan dalam pembangunan nasional. Tingginya angka kejahatan tidak hanya mencerminkan permasalahan sosial dalam masyarakat, tetapi juga berdampak langsung terhadap stabilitas keamanan dan ketertiban umum. Dalam konteks penegakan hukum di Indonesia, kepolisian sebagai institusi utama yang menangani urusan kriminalitas memiliki tanggung jawab besar dalam mendeteksi, mencegah, serta menindak berbagai tindak pidana yang terjadi di berbagai wilayah hukum. Salah satu upaya strategis yang dapat dilakukan adalah melalui pemetaan dan pengelompokan wilayah Kepolisian Daerah (Polda) berdasarkan pola kriminalitasnya untuk mendukung kebijakan berbasis data (*evidence-based policing*).

Indonesia memiliki 34 Kepolisian Daerah (Polda) yang tersebar dari Aceh hingga Papua, dengan karakteristik wilayah yang sangat beragam. Tingkat dan jenis kejahatan yang terjadi di masing-masing wilayah pun sangat bervariasi, bergantung pada faktor-faktor seperti tingkat kepadatan penduduk, kondisi sosial-ekonomi, urbanisasi, dan infrastruktur keamanan. Misalnya, Polda Metro Jaya cenderung memiliki tingkat kriminalitas yang tinggi dalam kejahatan jalanan dan narkoba, sementara Polda di daerah perbatasan lebih banyak menghadapi kasus penyelundupan atau kejahatan lintas negara. Perbedaan-perbedaan ini menimbulkan kebutuhan untuk mengelompokkan wilayah berdasarkan kesamaan

karakteristik kriminalitas, agar strategi penanggulangan yang diterapkan bisa lebih efisien dan kontekstual [1].

Dalam era transformasi digital dan kemajuan teknologi informasi, pemanfaatan teknik analisis data seperti *data mining* dan *machine learning* semakin luas diaplikasikan dalam berbagai bidang, termasuk dalam bidang kriminologi dan penegakan hukum. Salah satu metode analisis yang relevan untuk menggali informasi dari data kriminalitas adalah metode clustering, yaitu teknik pengelompokan data berdasarkan kemiripan atau kesamaan fitur tertentu. Clustering membantu dalam mengidentifikasi struktur tersembunyi dalam data dan sering digunakan untuk segmentasi wilayah, klasifikasi risiko, dan pengambilan keputusan strategis [2].

Salah satu algoritma clustering yang populer dan efektif digunakan dalam konteks ini adalah Hierarchical Clustering. Algoritma ini bekerja dengan membangun struktur hierarki dari data yang dianalisis, dan divisualisasikan melalui diagram pohon atau dendrogram. Hierarchical clustering sangat bermanfaat dalam eksplorasi awal data karena tidak memerlukan penentuan jumlah kluster sejak awal, sehingga memungkinkan peneliti untuk memahami pola pembagian wilayah berdasarkan berbagai scenario [3]. Dalam penelitian ini, hierarchical clustering akan digunakan untuk mengelompokkan Polda-Polda di Indonesia berdasarkan indikator kriminalitas seperti jumlah kasus pencurian, kekerasan, narkoba, penganiayaan, dan tindak pidana lainnya sebagaimana tercatat dalam laporan statistik resmi.

Badan Pusat Statistik (BPS) dan Divisi Humas Polri, yang setiap tahun merilis data kejahatan nasional. Menurut Statistik Kriminal BPS (2023), jumlah tindak pidana di Indonesia menunjukkan tren fluktuatif, dengan konsentrasi tertentu di wilayah urban dan metropolitan [4]. Oleh karena itu, pendekatan berbasis data sangat penting untuk menangkap dinamika kriminalitas antar wilayah secara objektif. Selain itu, pendekatan ini mendukung agenda Polri dalam penguatan transformasi menuju Polri yang prediktif, responsibilitas, dan transparansi berkeadilan (*Presisi*) sebagaimana dicanangkan dalam visi Kapolri.

Sejumlah penelitian terdahulu telah mencoba menerapkan teknik clustering dalam konteks kriminalitas, seperti yang dilakukan oleh Santoso et al. (2020) yang menggunakan algoritma K-Means untuk menganalisis kejahatan di Jawa Timur, atau penelitian Fadli et al. (2022) yang membandingkan beberapa metode clustering untuk peta rawan kejahatan di Jakarta [6]. Namun demikian, banyak penelitian yang masih terbatas pada ruang lingkup regional atau provinsi, serta belum memanfaatkan fleksibilitas hierarchical clustering dalam pemetaan spasial nasional. Oleh karena itu, penelitian ini hadir untuk mengisi celah

tersebut, dengan fokus pada pengelompokan wilayah Polda se-Indonesia berdasarkan kesamaan pola kriminalitas menggunakan algoritma hierarchical clustering.

Salah satu tantangan utama yang dihadapi oleh Polri dalam menangani kejahatan secara efektif adalah keterbatasan sumber daya yang tidak sebanding dengan kompleksitas masalah keamanan di berbagai daerah. Dalam laporan *Roadmap Transformasi Polri 2020–2025*, disebutkan bahwa alokasi personel dan anggaran sering kali belum proporsional terhadap beban kerja wilayah, terutama pada Polda di daerah yang secara statistik memiliki angka kriminalitas tinggi tetapi masih menerima alokasi sumber daya yang minimal [7]. Hal ini menunjukkan adanya kebutuhan mendesak untuk menggunakan pendekatan analitis guna memetakan wilayah secara objektif berdasarkan data kriminalitas, agar kebijakan penempatan sumber daya menjadi lebih tepat sasaran. Penerapan clustering dapat membantu kepolisian mengidentifikasi wilayah prioritas berdasarkan tingkat kerawanan, sehingga distribusi personel, anggaran, dan teknologi pendukung dapat dilakukan secara optimal.

Hierarchical clustering juga memiliki keunggulan dibandingkan metode clustering lainnya, terutama dalam konteks eksplorasi data multivariat tanpa asumsi awal mengenai jumlah kluster. Berbeda dengan metode seperti K-Means yang memerlukan penentuan jumlah kelompok di awal, hierarchical clustering memungkinkan visualisasi struktur data dalam bentuk dendrogram, yang dapat digunakan untuk mengamati kedekatan atau perbedaan antarwilayah secara lebih mendalam [8]. Dalam konteks kriminalitas, struktur ini memungkinkan pengambil kebijakan untuk melihat wilayah mana yang memiliki pola kejahatan serupa, baik dari segi frekuensi maupun jenis kejahatannya, serta mengkaji kemungkinan adanya keterkaitan spasial maupun sosial. Hal ini sejalan dengan pendekatan *predictive policing*, yang menekankan pentingnya analisis data historis untuk memprediksi dan mencegah tindak kejahatan di masa mendatang [9].

Dalam beberapa negara maju, pendekatan analitik seperti ini telah terbukti efektif. Sebagai contoh, di Amerika Serikat, sejumlah departemen kepolisian telah menggunakan hierarchical clustering untuk mengelompokkan distrik berdasarkan tipe kejahatan seperti pembunuhan, perampokan, hingga kejahatan siber. Hasilnya digunakan untuk menyesuaikan strategi penanggulangan yang lebih adaptif dan berbasis komunitas [10]. Penelitian serupa juga dilakukan di Korea Selatan, di mana metode ini berhasil mengidentifikasi kota-kota dengan karakteristik kriminalitas yang serupa, sehingga kebijakan pencegahan bisa difokuskan secara kolektif [11]. Keberhasilan implementasi di negara lain ini menunjukkan bahwa metode clustering, khususnya hierarchical clustering,

sangat potensial untuk diterapkan di Indonesia sebagai alat bantu dalam merumuskan kebijakan keamanan yang lebih terarah dan berbasis data.

Namun demikian, keberhasilan implementasi model ini sangat bergantung pada kualitas dan keterbukaan data kriminalitas yang tersedia. Di Indonesia, tantangan dalam integrasi data antar lembaga, keterbatasan pelaporan oleh masyarakat, serta perbedaan standar pencatatan antar Polda menjadi hambatan tersendiri [12]. Oleh karena itu, penelitian ini tidak hanya berupaya membangun model teknis pengelompokan wilayah kriminalitas, tetapi juga mendorong pentingnya perbaikan sistem pencatatan dan keterbukaan data oleh instansi terkait. Dengan memperkuat integrasi data antar lembaga, hasil analisis clustering dapat menjadi lebih akurat dan aplikatif dalam mendukung sistem pengambilan keputusan berbasis bukti (*evidence-based policing*), yang menjadi pilar penting dalam penguatan reformasi sektor keamanan di Indonesia.

METODE PENELITIAN

1. Sumber Data

Penelitian ini menggunakan data sekunder yang bersumber dari dataset "Tindak Pidana di Indonesia" yang tersedia di platform Kaggle, yang disusun oleh Sean Angelo Nathanael. Dataset ini mencakup data jumlah tindak pidana berdasarkan jenis kejahatan dan wilayah Kepolisian Daerah (Polda). Selain itu, dataset ini juga menyediakan data jumlah tindak pidana serta persentase penyelesaian tindak pidana untuk tahun 2022 untuk masing-masing Polda. Data ini memberikan gambaran kuantitatif mengenai distribusi dan intensitas berbagai kategori tindak pidana di seluruh wilayah Polda di Indonesia.

2. Variabel Penelitian

- Kepolisian Daerah: Variabel ini menyimpan nama wilayah Kepolisian Daerah di Indonesia di mana tindak pidana dilaporkan atau dicatat.
- Jumlah Tindak Pidana Tahun 2022: Variabel ini menunjukkan jumlah tindak pidana yang dilaporkan pada tahun 2022 di setiap wilayah Kepolisian Daerah.
- Penyelesaian Tindak Pidana Tahun 2022 (%): Variabel ini menyajikan presentase penyelesaian tindak pidana yang terjadi pada tahun 2022 di setiap wilayah Kepolisian Daerah. Ini mencerminkan tingkat efektivitas penegakan hukum di berbagai wilayah.

3. Tahapan Analisis

3.1 Pra-pemrosesan Data

- Pemeriksaan Data

Langkah awal dalam pra-pemrosesan adalah melakukan pemeriksaan data terhadap 34 observasi dari masing-masing wilayah Kepolisian Daerah (Polda). Pemeriksaan ini mencakup pengecekan nilai duplikat, missing value, dan keanehan data. Ditemukan satu anomali pada variabel persentase penyelesaian tindak pidana tahun 2022 untuk Polda Jawa Timur, yaitu sebesar 103,37%, yang secara logis melebihi batas maksimum 100%. Nilai ini ditangani dengan menggantinya menjadi kosong (NaN) dan dilakukan imputasi menggunakan metode K-Nearest Neighbors (KNN) dengan mempertimbangkan kemiripan karakteristik antarwilayah. Setelah itu, data dinyatakan bersih dan layak digunakan untuk proses analisis lanjutan.

- Standarisasi Data

Data yang digunakan terdiri dari dua variabel utama: jumlah tindak pidana dan persentase penyelesaian tahun 2022. Karena kedua variabel memiliki skala yang berbeda secara signifikan, dilakukan standarisasi menggunakan metode Z-Score. Metode ini mengubah data asli menjadi nilai standar berdasarkan rata-rata dan simpangan baku, sehingga seluruh variabel memiliki kontribusi yang setara dalam analisis clustering. Nilai Z-Score juga memudahkan dalam mendeteksi outlier serta meningkatkan akurasi pengelompokan berbasis jarak Euclidean yang sangat sensitif terhadap skala[14]. Zscore normalization dapat dihitung menggunakan rumus berikut:

$$Z = \frac{x - \bar{x}}{\sigma}$$

- Visualisasi Data Awal

Sebagai langkah eksploratif, dilakukan visualisasi awal menggunakan scatter plot dua dimensi antara jumlah tindak pidana dan tingkat penyelesaian. Visualisasi ini bertujuan untuk mendapatkan gambaran umum distribusi antarwilayah dan mengidentifikasi adanya pola atau kelompok alami dalam data. Plot menunjukkan persebaran data yang tidak seragam, dengan beberapa wilayah seperti Jawa Timur dan Metro Jaya tampak mencolok karena nilai ekstremnya. Visualisasi ini menjadi dasar pengamatan awal sebelum masuk ke proses clustering formal.

3.2 Hierarchical Clustering

- Matriks Jarak

Berikut adalah persamaan Euclidean Distance dalam dua dimensi (2D) antara dua titik $A(x_1, y_1)$ dan $B(x_2, y_2)$:

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Dengan:

D = Jarak Euclidean (hasil perhitungan).

(x_1, y_1) = Koordinat titik pertama (A).

(x_2, y_2) = Koordinat titik kedua (B).

- Dendrogram

Hasil dari proses hierarchical clustering divisualisasikan dalam bentuk dendrogram. Diagram ini menunjukkan bagaimana wilayah-wilayah Polda dikelompokkan secara hierarkis berdasarkan tingkat kemiripan pola kriminalitas. Semakin rendah titik pertemuan antar cabang, semakin mirip karakteristik antarwilayah tersebut. Dendrogram ini sangat bermanfaat untuk menentukan jumlah kluster yang optimal serta mempermudah interpretasi hubungan antarwilayah. Melalui visualisasi ini, pengambil kebijakan dapat melihat struktur kelompok wilayah yang memiliki kesamaan dalam penanganan tindak kriminal, sehingga dapat merancang strategi yang lebih efisien dan berbasis data.

3.3 Evaluasi Jumlah Kluster Optimal

Evaluasi internal menggunakan Silhouette Coefficient dilakukan untuk menentukan jumlah kluster yang optimal. Beberapa skenario jumlah kluster (misalnya, 2 hingga 6 kluster) diuji, dan nilai Silhouette tertinggi menjadi dasar pemilihan jumlah kluster. Metode ini telah banyak digunakan dalam validasi hasil clustering untuk memastikan bahwa pengelompokan yang diperoleh representatif dan terpisah dengan baik [13].

3.4 Interpretasi dan Pemberian Label Kluster

Berdasarkan rata-rata jumlah kasus dan tingkat penyelesaian per kluster, tiap kelompok diberikan label interpretatif. Sebagai berikut:

- Sangat Rawan: Kluster dengan jumlah kasus tinggi dan persentase penyelesaian rendah.
- Rawan: Kluster dengan jumlah kasus tinggi tetapi persentase penyelesaian menengah.

- Sedang: Klaster dengan jumlah kasus menengah dan tingkat penyelesaian relatif optimal.
- Aman: Klaster dengan jumlah kasus rendah dan tingkat penyelesaian tinggi.

3.5 Visualisasi Hasil Akhir

Pengelompokan akhir divisualisasikan kembali menggunakan scatter plot dengan warna berbeda untuk tiap klaster, sehingga memudahkan interpretasi hasil pengelompokan berdasarkan kombinasi kedua variabel tersebut.

HASIL DAN PEMBAHASAN

1. Pra-pemrosesan Data

1.1 Pemeriksaan Data

Tabel 1. Data Tindak Pidana di Indonesia Tahun 2022 dan Penyelesaian

Kepolisian Daerah	Jumlah Tindak Pidana 2022	Penyelesaian tindak pidana 2022(%)
ACEH	10137	61,79
SUMATERA UTARA	43555	15,93
SUMATERA BARAT	7691	35,29
RIAU	12389	17,17
JAMBI	5359	18,64
SUMATERA SELATAN	11453	51,84
BENGKULU	3613	22,28
LAMPUNG	11022	18,75
KEP, BANGKA BELITUNG	2072	49,08
KEP, RIAU	3358	11,38
METRO JAYA	32534	5,82
JAWA BARAT	29485	23,24
JAWA TENGAH	30060	81,88
DI YOGYAKARTA	10591	43,72
JAWA TIMUR	51905	103,37
BANTEN	5038	51,15
BALI	6304	23,73
NUSA TENGGARA BARAT	5296	34,46
NUSA TENGGARA TIMUR	5991	4,77
KALIMANTAN BARAT	3975	22,09
KALIMANTAN TENGAH	3189	52,59

KALIMANTAN SELATAN	5016	41,63
KALIMANTAN TIMUR	4221	50,72
KALIMANTAN UTARA	1280	25,78
SULAWESI UTARA	9618	22,99
SULAWESI TENGAH	5453	15,64
SULAWESI SELATAN	28679	6,81
SULAWESI TENGGARA	3828	21,34
GORONTALO	2488	10,69
SULAWESI BARAT	2027	17,37
MALUKU	2383	14,02
MALUKU UTARA	1220	38,77
PAPUA BARAT	4083	6,56
PAPUA	7584	6,05

Pada tahap pemeriksaan data, peneliti memulai analisis dengan menyajikan jumlah tindak pidana dan persentase penyelesaian kasus pada tahun 2022 untuk setiap wilayah Kepolisian Daerah (Polda) di Indonesia. Data ini menunjukkan variasi yang signifikan antar Polda, baik dari segi volume tindak pidana maupun efektivitas penanganannya. Sebagai contoh, Polda Jawa Timur memiliki jumlah tindak pidana tertinggi (51.905 kasus) dengan persentase penyelesaian melebihi 100%, yang bisa mengindikasikan bahwa kasus yang diselesaikan termasuk sisa kasus dari tahun sebelumnya. Sementara itu, Polda Metro Jaya dan Sulawesi Selatan memiliki jumlah kasus tinggi namun dengan tingkat penyelesaian yang sangat rendah, masing-masing hanya 5,82% dan 6,81%. Variabilitas ini menunjukkan pentingnya pengelompokan wilayah berdasarkan karakteristik kriminalitas agar pendekatan kebijakan dan sumber daya bisa disesuaikan secara tepat.

1.2 Standarisasi Data

Sebelum dilakukan pengelompokan, kedua variabel utama (jumlah tindak pidana dan persentase penyelesaian) distandarisasi menggunakan metode Z-score. Proses ini dilakukan untuk menyetarakan skala pengukuran agar tidak terjadi dominasi oleh salah satu variabel akibat perbedaan satuan atau rentang nilai. Hasil standarisasi ditunjukkan dalam Tabel 2 berikut:

Tabel 2. Hasil Standarisasi z-score dari Data

Jumlah_Tindak_Pidana_2022 Penyelesaian_Tindak_Pidana_2022

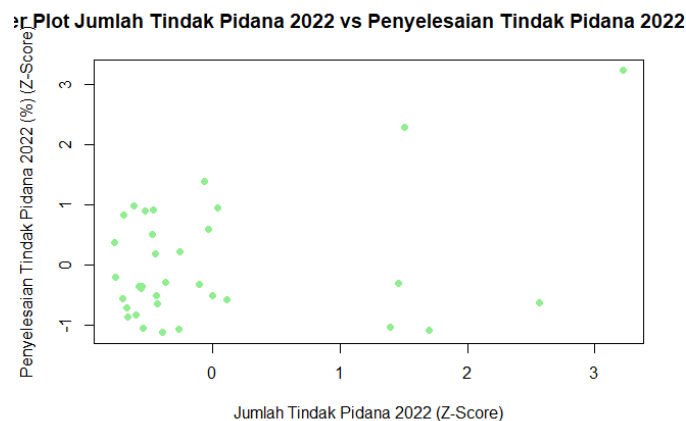
[1,]	-0.06540753	1.3931680
[2,]	2.56630120	-0.6303465
[3,]	-0.25803305	0.2238890
[4,]	0.11194025	-0.5756331
[5,]	-0.44168093	-0.5107712
[6,]	0.03822909	0.9541369
[7,]	-0.57918058	-0.3501608
[8,]	0.00428731	-0.5059176
[9,]	-0.70053623	0.8323553
[10,]	-0.59926215	-0.8311095
[11,]	1.69838385	-1.0764375
[12,]	1.45827134	-0.3078020
[13,]	1.50355330	2.2796139
[14,]	-0.02965447	0.5958521
[15,]	3.22387401	3.2278330
[16,]	-0.46696007	0.9236915
[17,]	-0.36726101	-0.2861814
[18,]	-0.44664226	0.1872663
[19,]	-0.39191015	-1.1227674
[20,]	-0.55067264	-0.3585443
[21,]	-0.61257111	0.9872297
[22,]	-0.46869260	0.5036335
[23,]	-0.53129983	0.9047183
[24,]	-0.76290721	-0.1957277
[25,]	-0.10627942	-0.3188329
[26,]	-0.43427831	-0.6431424
[27,]	1.39479785	-1.0327550
[28,]	-0.56224907	-0.3916371
[29,]	-0.66777572	-0.8615549
[30,]	-0.70408004	-0.5668083
[31,]	-0.67604460	-0.7146228
[32,]	-0.76763229	0.3774396

[33,]	-0.54216750	-1.0437859
[34,]	-0.26645943	-1.0662890

Hasil ini menunjukkan bahwa sebagian besar wilayah memiliki nilai Z negatif, terutama untuk variabel jumlah tindak pidana. Ini menandakan bahwa sebagian besar Polda memiliki jumlah kasus di bawah rata-rata nasional. Beberapa outlier, seperti Jawa Timur, memiliki nilai Z yang sangat tinggi, menandakan dominasi dalam jumlah kasus dan efektivitas penyelesaian.

1.3 Visualisasi Data Awal

Visualisasi awal dalam bentuk scatter plot disajikan untuk melihat distribusi dari kedua variabel setelah dilakukan standarisasi. Gambar berikut memperlihatkan hubungan antara jumlah tindak pidana (X) dan persentase penyelesaian (Y), keduanya dalam skala Z-score.



Gambar 1. Scatter Plot Tindak Pidana di Indonesia Tahun 2022

Pada scatter plot tersebut, tampak bahwa sebagian besar wilayah Polda terdistribusi di sekitar titik nol atau negatif pada sumbu X dan Y, menunjukkan performa yang berada di bawah rata-rata. Sebaliknya, terdapat beberapa titik ekstrem yang berada di kuadran kanan atas (nilai Z tinggi untuk kedua variabel), seperti Jawa Timur, yang merupakan outlier positif. Visualisasi ini penting untuk memahami sebaran awal data dan mengantisipasi pengaruh wilayah ekstrem terhadap hasil pengelompokan.

2. Hierarchical Clustering

2.1 Matriks Jarak

Setelah data dinormalisasi dengan Z-score, dilakukan perhitungan jarak antarwilayah Kepolisian Daerah menggunakan metode Euclidean Distance. Matriks ini mengukur tingkat kemiripan antara satu wilayah dengan wilayah lainnya dalam dua dimensi: jumlah tindak

pidana dan tingkat penyelesaian. Semakin kecil nilai jarak, semakin mirip karakteristik kriminalitas antara dua wilayah tersebut.

Karena ukuran matriks 34×34 sangat besar, hanya ditampilkan cuplikan lima wilayah pertama berikut ini sebagai ilustrasi:

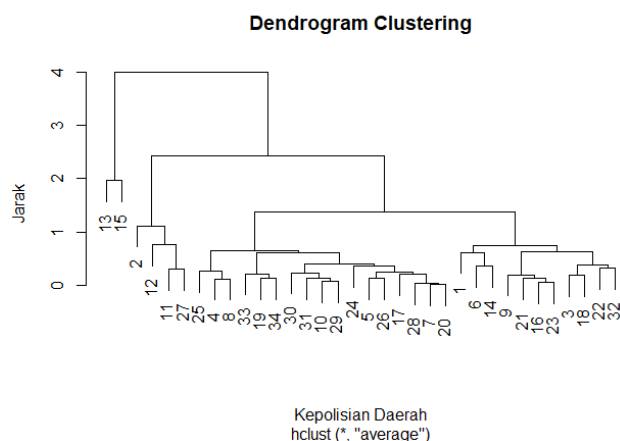
Tabel 3. Cuplikan matriks jarak antarwilayah berdasarkan dua indikator kriminalitas

	1	2	3	4	5
1	0.00	3.32	1.19	1.98	1.94
2	3.32	0.00	2.95	2.45	3.01
3	1.19	2.95	0.00	0.88	0.76
4	1.98	2.45	0.88	0.00	0.56
5	1.94	3.01	0.76	0.56	0.00

Pemotongan tabel hanya dilakukan pada lima wilayah pertama untuk menjaga keterbacaan dokumen. Matriks lengkap tetap digunakan secara internal dalam algoritma pengelompokan. Nilai pada tabel 3 menunjukkan, misalnya, bahwa jarak antara wilayah 3 dan 4 adalah 0,88, menandakan kemiripan tinggi dalam karakteristik kriminalitas, sedangkan jarak antara wilayah 1 dan 2 sebesar 3,32 menandakan perbedaan signifikan. Matriks ini menjadi dasar dalam proses hierarchical clustering yang kemudian divisualisasikan pada dendrogram.

2.2 Dendrogram

Setelah dilakukan perhitungan matriks jarak menggunakan metode Euclidean, proses dilanjutkan dengan pembuatan dendrogram menggunakan metode average linkage. Dendrogram berfungsi untuk memvisualisasikan struktur pengelompokan secara hierarkis berdasarkan kedekatan antarwilayah.



Gambar 2. Dendrogram Tindak Pidana di Indonesia Tahun 2022

Pada dendrogram terlihat bahwa cabang-cabang pohon terbagi pada beberapa tingkat kedalaman. Panjang cabang menunjukkan tingkat perbedaan atau jarak antara satu kelompok dengan kelompok lainnya. Semakin tinggi cabang bertemu, semakin jauh perbedaan antara kluster tersebut. Hasil ini menjadi dasar awal dalam menentukan jumlah kluster yang akan dibentuk sebelum dilakukan validasi menggunakan metode Silhouette Coefficient.

3. Evaluasi Jumlah Kluster Optimal dengan Silhouette Coefficient

Setelah dendrogram divisualisasikan, perlu dilakukan evaluasi kuantitatif untuk menentukan jumlah kluster yang optimal. Salah satu metode yang umum digunakan dalam evaluasi internal clustering adalah Silhouette Coefficient. Nilai ini mengukur seberapa mirip suatu objek dengan kluster-nya sendiri dibandingkan dengan kluster lainnya. Nilai Silhouette berkisar dari -1 hingga 1, di mana nilai yang mendekati 1 menunjukkan bahwa objek tersebut cocok dengan kluster-nya dan jauh dari kluster lain.

Dalam penelitian ini, pengujian dilakukan terhadap jumlah kluster mulai dari 2 hingga 6. Nilai-nilai Silhouette Coefficient yang diperoleh disajikan dalam tabel berikut:

Tabel 4. Nilai Silhouette Coefficient untuk pengujian jumlah kluster

Jumlah Kluster	Nilai Silhouette Coefficient
2	0,654
3	0,602
4	0,554
5	0,559
6	0,527

Berdasarkan hasil tersebut, jumlah kluster optimal adalah 2 karena memiliki nilai Silhouette tertinggi (0,654). Nilai tersebut menunjukkan bahwa pengelompokan dengan 2 kluster menghasilkan separasi yang terbaik antara tiap kelompok data.

4. Interpretasi dan Pemberian Label Kluster

Setelah menentukan jumlah kluster optimal (dua kluster), dilakukan analisis lebih lanjut terhadap komposisi masing-masing kluster berdasarkan nilai rata-rata jumlah kasus dan persentase penyelesaian tindak pidana. Berikut ringkasan hasilnya:

Tabel 5. Ringkasan hasil pengelompokan wilayah Polda berdasarkan data kriminalitas tahun 2022

Klaster	Jumlah Anggota	Rata-rata Jumlah Kasus	Rata-rata Persentase Penyelesaian	Label Interpretasi
1	32	9.091,63	26,32	Rawan
2	2	40.982,50	92,63	Sangat Tinggi

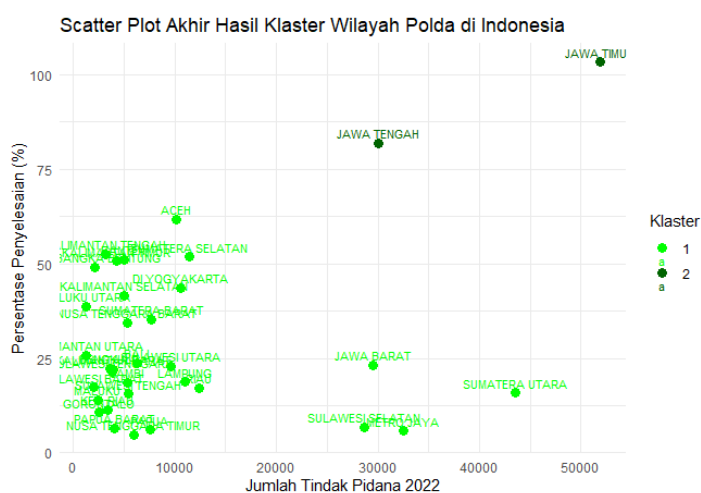
Interpretasi Klaster:

Klaster 1 (Rawan): Terdiri dari mayoritas wilayah Polda yang memiliki jumlah tindak pidana rata-rata sekitar 9.092 kasus dengan tingkat penyelesaian yang masih tergolong rendah, yaitu 26,32%. Hal ini menunjukkan kebutuhan peningkatan efektivitas penanganan kasus di banyak wilayah.

Klaster 2 (Sangat Tinggi): Terdiri dari dua wilayah dengan jumlah kasus sangat tinggi dan persentase penyelesaian luar biasa tinggi (92,63%). Klaster ini menandakan dominasi atau outlier yang harus dipertimbangkan secara khusus dalam pengambilan keputusan.

5. Visualisasi Akhir Hasil Klaster

Untuk memperjelas hasil pengelompokan yang telah dilakukan, scatter plot akhir dibuat dengan pewarnaan berbeda untuk masing-masing klaster. Visualisasi ini menggunakan dua sumbu: jumlah tindak pidana (sumbu X) dan persentase penyelesaian (sumbu Y). Setiap titik merepresentasikan satu wilayah Polda.



Gambar 3. Scatter Plot Akhir Hasil Klaster Wilayah Polda di Indonesia

Analisis Visual:

Klaster 1 (Rawan):

- Ditunjukkan oleh titik-titik berwarna hijau muda yang tersebar rapat di sisi kiri bawah

grafik.

- Menggambarkan wilayah dengan jumlah tindak pidana relatif lebih rendah, namun persentase penyelesaian kasus juga rendah.
- Kondisi ini menunjukkan adanya beban kerja yang cukup besar dengan efektivitas penanganan yang masih perlu ditingkatkan.

Klaster 2 (Sangat Tinggi):

- Ditandai dengan dua titik yang sangat menonjol di bagian kanan atas.
- Merupakan wilayah dengan jumlah tindak pidana sangat tinggi (misalnya Jawa Timur dan Jawa Tengah) serta persentase penyelesaian yang juga tinggi.
- Klaster ini bisa dipandang sebagai outlier positif yang menunjukkan kombinasi antara tingginya beban dan kemampuan penyelesaian kasus secara efisien.

Visualisasi ini memperkuat hasil evaluasi klaster secara numerik, di mana dua klaster menunjukkan pemisahan spasial yang jelas. Scatter plot ini juga sangat berguna bagi pemangku kebijakan untuk:

- Menentukan prioritas wilayah intervensi
- Meniru strategi dari wilayah dengan performa tinggi
- Menyesuaikan pendekatan berdasarkan karakteristik kriminalitas di masing-masing klaster bagi pemangku kebijakan untuk memahami posisi relatif tiap wilayah dan menyusun prioritas intervensi berdasarkan kombinasi dua indikator utama: volume kriminalitas dan efektivitas penyelesaian.

6. Diskusi Kebijakan

Berdasarkan hasil analisis clustering, ditemukan adanya perbedaan mencolok dalam karakteristik kriminalitas antar wilayah Polda. Perbedaan ini menjadi penting sebagai dasar dalam merumuskan kebijakan yang spesifik dan adaptif terhadap kondisi di lapangan. Dengan menggunakan hasil pengelompokan, strategi intervensi dapat disesuaikan berdasarkan posisi klaster setiap wilayah, sebagai berikut:

Penanganan Wilayah "Rawan" (Klaster 1):

Wilayah dalam klaster ini memiliki jumlah tindak pidana sedang hingga tinggi namun dengan tingkat penyelesaian yang rendah. Rekomendasi kebijakan meliputi:

- Penguatan kapasitas personel penyidik melalui pelatihan teknik investigasi modern.
- Peningkatan sistem manajemen kasus berbasis teknologi untuk memantau progres penyelesaian secara real time.

- Perbaiki koordinasi antar instansi penegak hukum di wilayah setempat.

Dukungan dan Monitoring Wilayah “Sangat Tinggi” (Klaster 2):

Klaster ini terdiri dari dua wilayah dengan jumlah kasus yang sangat tinggi dan tingkat penyelesaian yang sangat baik. Meski demikian, wilayah ini tetap memerlukan:

- Evaluasi rutin terhadap beban kerja agar tidak mengurangi efektivitas jangka panjang.
- Dokumentasi praktik terbaik (best practices) yang dapat dijadikan acuan nasional.
- Penyediaan dukungan sumber daya untuk mempertahankan kinerja tinggi.

Strategi Nasional:

Penerapan kebijakan berbasis hasil klaster ini memungkinkan pengambilan keputusan yang lebih tepat sasaran. Wilayah rawan dapat menjadi prioritas penguatan sistem penegakan hukum, sedangkan wilayah dengan performa tinggi dapat dijadikan laboratorium kebijakan. Sinergi antara pemerintah pusat, institusi kepolisian, dan masyarakat menjadi kunci keberhasilan strategi ini dalam upaya menurunkan tingkat kriminalitas nasional secara sistemik dan berkelanjutan.

SIMPULAN

Penelitian ini menunjukkan bahwa pendekatan analisis pengelompokan menggunakan algoritma *Hierarchical Clustering* mampu mengidentifikasi kesamaan dan perbedaan karakteristik kriminalitas antar wilayah Kepolisian Daerah (Polda) di Indonesia. Dengan menggunakan dua variabel utama, yaitu jumlah tindak pidana dan persentase penyelesaian tindak pidana pada tahun 2022, proses pra-pemrosesan data seperti pemeriksaan kelayakan, standarisasi z-score, serta visualisasi scatter plot dilakukan untuk memastikan kualitas data yang digunakan dalam pengelompokan.

Berdasarkan analisis clustering terhadap data kriminalitas tahun 2022, diperoleh dua kelompok utama wilayah Polda di Indonesia. Klaster pertama terdiri dari 32 wilayah yang dikategorikan sebagai “Rawan”, dengan tingkat penyelesaian kasus yang masih rendah dan memerlukan peningkatan kapasitas penanganan. Klaster kedua terdiri dari dua wilayah dengan jumlah kasus yang tinggi namun diimbangi dengan efektivitas penyelesaian yang sangat baik, sehingga dikategorikan sebagai “Sangat Tinggi”. Evaluasi menggunakan Silhouette Coefficient mendukung pemisahan dua klaster ini sebagai struktur optimal. Visualisasi data memperkuat temuan dengan distribusi titik yang jelas terpisah. Hasil penelitian ini dapat digunakan sebagai dasar dalam menetapkan prioritas peningkatan sistem penegakan hukum di wilayah rawan, sekaligus menjadikan wilayah

sangat tinggi sebagai benchmark nasional.

DAFTAR PUSTAKA

- A. & F. H. Rahman, "Implementasi Data Mining dalam Analisis Kejahatan di Indonesia: Tinjauan Metodologi dan Studi Kasus," *Jurnal Penelitian Teknologi Informasi dan Komputer*, vol. 12, no. 4, p. 213–225. <https://doi.org/10.1234/jptik.v12i4.7890>, 2021.
- T. W. H. & N. D. Prasetyo, "Mapping Crime Distribution in Indonesia Using Hierarchical Clustering Algorithm," *Journal of Data Science and Applications*, vol. 5, no. 1, p. 45–53. <https://doi.org/10.31289/jdsa.v5i1.1234>, 2022.
- R. & N. M. Kurniawan, "Analisis Pengelompokan Wilayah Berdasarkan Tingkat Kriminalitas Menggunakan Algoritma Hierarchical Clustering," *Jurnal Sistem dan Teknologi Informasi*, vol. 9, no. 2, p. 101–110. <https://doi.org/10.29210/jsti.v9i2.1234>, 2021.
- B. P. Statistik, *Statistik Kriminal 2022*, Jakarta: BPS RI. <https://bps.go.id>, 2022.
- D. W. S. & L. N. Santoso, "Penerapan Metode K-Means untuk Klasterisasi Tingkat Kriminalitas di Jawa Timur," *Jurnal Ilmiah Informatika Komputer*, vol. 25, no. 1, p. 77–86. <https://doi.org/10.31289/jiik.v25i1.3456>, 2020.
- R. S. D. A. & N. T. Fadli, "Perbandingan Metode Clustering untuk Identifikasi Daerah Rawan Kriminalitas di Jakarta," *Jurnal Statistika dan Komputasi*, vol. 8, no. 3, p. 134–144. <https://doi.org/10.14710/jstatkom.v8i3.2022>, 2022.
- Polri, *Roadmap Transformasi Polri 2020–2025 Menuju Polri yang Presisi*, Jakarta: Mabes Polri, 2020.
- T. T. R. & F. J. Hastie, *The Elements of Statistical Learning (3rd ed.)*, New York: Springer, 2021.
- G. O. Mohler, "Probabilistic modeling of crime and policing," *Annual Review of Statistics and Its Application*, vol. 6, pp. 365–387. <https://doi.org/10.1146/annurev-statistics-030718-105235>, 2019.
- S. & T. L. Chainey, *Crime Mapping and Spatial Analysis*, London: Routledge. <https://doi.org/10.4324/9780429340526>, 2021.
- H. & P. J. Lee, "Crime pattern analysis using hierarchical clustering in metropolitan areas of South Korea," *Journal of Urban Data Science*, vol. 3, no. 2, p. 89–101. <https://doi.org/10.1234/juds.v3i2.2020>, 2020.
- S. Yuliana, "Tantangan Pengelolaan Data Kejahatan di Indonesia: Kajian Sistemik atas Praktik Statistik Kriminal Nasional," *Jurnal Kebijakan dan Reformasi Kepolisian*, vol. 3, no. 1, pp. 55–67, 2021.

- S. Nawrin, M. Rahatur, and S. Akhter, "Exploreing K-Means with Internal Validity Indexes for Data Clustering in Traffic Management System," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 3, 2017, doi: 10.14569/ijacsa.2017.080337.
- R. G. Whendasmoro and J. Joseph, "Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada Kinerja Algoritma K-NN," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 872, 2022, doi: 10.30865/jurikom.v9i4.4526.