

## **ANALISIS CLUSTER ANTARA METODE HIERARKI DAN NON-HIERARKI UNTUK KRIMINALITAS DI PROVINSI JAWA TIMUR**

**Melani Sinta Rahma Santosa**

Program Studi Matematika, FMIPA, Universitas Negeri Malang, Indonesia

e-mail: [melani.sinta.2203126@students.um.ac.id](mailto:melani.sinta.2203126@students.um.ac.id)\*

**Swasono Rahardjo**

Universitas Negeri Malang, Indonesia

e-mail: [swasono.rahardjo.fmipa@um.ac.id](mailto:swasono.rahardjo.fmipa@um.ac.id)

### **Abstrak**

Selama empat tahun terakhir, tindak kriminalitas di Jawa Timur mengalami fluktuasi. Kondisi tersebut dapat meresahkan dan mempengaruhi rasa aman masyarakat. Tindakan yang dapat diambil untuk menyelesaikan permasalahan tersebut adalah dengan mengelompokkan wilayah kabupaten/kota di Jawa Timur menurut indikator kriminalitas, yaitu dengan analisis *cluster*. Terdapat dua metode utama dalam analisis *cluster*, yaitu Hierarki dan Non-Hierarki. Dalam penelitian ini membahas analisis *cluster* menggunakan kedua metode tersebut untuk kriminalitas di Jawa Timur. Selanjutnya, pengelompokan wilayah kabupaten/kota di Jawa Timur menurut indikator kriminalitas dilakukan menggunakan hasil terbaik berdasarkan perhitungan indeks validitas *Davies-Bouldin (DB)*, *Silhouette*, *Calinski-Harabasz (CH)*, *Dunn*, dan *Connectivity*. Dengan demikian, penelitian ini bertujuan untuk mengetahui hasil pengelompokan daerah rawan kriminalitas di Provinsi Jawa Timur yang terbaik. Untuk mencapai tujuan dalam penelitian ini, analisis dilakukan dengan menggunakan alat bantu berupa *software Minitab 19* dan *RStudio*. Hasil analisis memperlihatkan bahwa metode terbaik dalam analisis *cluster* antara metode Hierarki dan Non-Hierarki untuk kriminalitas di Jawa Timur adalah *K-Medoids/PAM* dengan 2 *cluster*, dengan 37 kabupaten/kota pada *cluster 1* dan 1 kota pada *cluster 2*. Adapun karakteristik tiap *cluster*nya adalah *cluster 1 cluster 1* dapat dikategorikan sebagai kelompok wilayah dengan tingkat kriminalitas rendah. Sementara itu, pada *cluster 2* dapat dikategorikan sebagai kelompok wilayah dengan tingkat kriminalitas tinggi.

**Kata Kunci:** Kriminalitas, Analisis Cluster, Hierarki, Non-Hierarki

### **Abstract**

Over the past four years, the crime rate in East Java has fluctuated. This condition can cause concern and affect the public's sense of security. One action that can be taken to address this problem is to group districts/cities in East Java based on crime indicators, namely by using cluster analysis. There are two main methods in cluster analysis, namely Hierarchical and Non-Hierarchical. This study discusses cluster analysis using both methods for crime in East Java. Furthermore, the grouping of districts/cities in East Java based on crime indicators was carried out using the best results based on the calculation of the validity index *Davies-Bouldin (DB)*, *Silhouette*, *Calinski-Harabasz (CH)*, *Dunn*, and *Connectivity*. Thus, this study aims to determine the best results for grouping crime-prone areas in East Java Province. To achieve this research objective, the analysis was conducted using *Minitab 19* and *RStudio* software. The results of the analysis show that the best method in cluster analysis between Hierarchical and Non-Hierarchical methods for crime in East Java is *K-Medoids/PAM* with 2 clusters, with 37 districts/cities in *cluster 1* and 1 city in *cluster 2*. The characteristics of each cluster are *cluster 1* can be categorized as a group of areas with low crime rates. Meanwhile, *cluster 2* can be categorized as a group of areas with high crime rates.

**Keywords:** Crime, Cluster Analysis, Hierarchical, Non-Hierarchical

### **PENDAHULUAN**

Menurut Abdulsyani (1987) dalam Nahdliyah dkk. (2019), kriminalitas atau kejahatan diartikan sebagai suatu perbuatan yang dapat menimbulkan permasalahan dan keresahan dalam kehidupan

masyarakat. Secara umum, kriminalitas adalah perbuatan yang bertentangan dengan hukum dan beberapa normal sosial yang berlaku dalam suatu masyarakat. Perbuatan semacam ini berpotensi menimbulkan dampak negatif pada individu tertentu, masyarakat secara keseluruhan, dan

lingkungan sekitarnya (*Statistik Kriminal 2024/2025*, 2025). Dalam negara yang memegang dasar *rule of law* (negara hukum), kriminalitas didefinisikan sebagai perbuatan yang dilarang oleh aturan undang-undang dan dikenakan hukuman pidana terhadap pelakunya. (Dulkiah, 2020).

Selama empat tahun terakhir, kejadian tindak kriminalitas di Jawa Timur mengalami fluktuasi. Pada tahun 2024, pihak kepolisian kabupaten/kota di Jawa Timur menerima laporan sebanyak 28.456 perkara tindak pidana. Tidak semua tindak pidana tersebut dapat diselesaikan karena kurangnya barang bukti, keterbatasan sumber daya dan anggaran penyidik, serta sulitnya pelacakan pelaku. Selain itu, minimnya fasilitas CCTV, rendahnya partisipasi sebagai saksi, kemampuan teknologi yang terbatas, dan lemahnya koordinasi antar-instansi juga turut menghambat penyelesaian tindak pidana. Meskipun demikian, pihak kepolisian tetap terus berusaha meningkatkan kinerja penyelesaian tindak pidana. Hal ini tercermin dari jumlah tindak pidana yang diselesaikan pada tahun 2024 yang mencapai sekitar 91,46 persen, meningkat sebesar 20,1 persen dibandingkan tahun sebelumnya.

Melihat kondisi kriminalitas di Jawa Timur, menjadi suatu hal yang memprihatinkan karena dapat meresahkan dan mempengaruhi rasa aman dalam kehidupan masyarakat. Mengingat bahwa setiap individu harus memperoleh dan menikmati rasa aman sebagai salah satu hak asasinya. Dengan demikian, tindakan yang dapat diambil untuk menyelesaikan permasalahan tersebut adalah dengan mengelompokkan wilayah kabupaten/kota di Jawa Timur menurut indikator kriminalitas. Hal ini tidak hanya dapat membantu masyarakat untuk tetap waspada terhadap kriminalitas, tetapi juga dapat menjadi evaluasi bagi pemerintah dan berdampak pada kebijakan yang diambil oleh pemerintah nantinya.

Dalam bidang matematika, salah satu teknik pengelompokan yang mampu menangani permasalahan kriminalitas di Jawa Timur tersebut adalah analisis *cluster*. Analisis *cluster* memungkinkan pengelompokan data berdasarkan kesamaan pola data. Pengelompokan ini menyusun data berdasarkan kelompok yang sama tanpa adanya informasi awal mengenai kategori kelompok yang telah ada (Ardana Wijaya dkk., 2023). Dalam analisis cluster, terdapat dua metode

utama, yaitu Hierarki dan Non-Hierarki. Pada Hierarki, jumlah *cluster* yang dihasilkan belum ditentukan sejak awal, sementara dalam Non-Hierarki, jumlah *cluster* diasumsikan terlebih dahulu sebanyak k kelompok.

Penelitian mengenai analisis *cluster* telah dilakukan oleh Nahdliyah dkk. (2019) dengan metode *k-Medoids Clustering*. Dalam penelitian tersebut menekankan bahwa hasil analisis *cluster* dengan metode *K-Medoids* dipengaruhi oleh ukuran jarak yang digunakan serta menyarankan perlunya perbandingan antara *K-Means* dan *K-Medoids* untuk mengetahui metode mana yang lebih tahan terhadap penciran (*outlier*). Selain itu, Syafiyah dkk. (2022) juga sudah membuat penelitian tentang analisis *cluster* yang menggunakan metode Hierarki dan Non-Hierarki. Dalam penelitian tersebut, metode Hierarki yang digunakan hanya metode *Linkage* (*Single Linkage*, *Complete Linkage*, dan *Average Linkage*) dan metode Non-Hierarki yang digunakan hanya *K-Means*. Sedangkan, di buku *Methods of Multivariate Analysis*, metode Hierarki mencakup beberapa teknik, seperti *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Centroid Linkage*, *Median* dan *Ward's Method* (Rencher & Christensen, 2012). Sementara itu, menurut Raja dkk. (2024) dalam penelitian Sunandi dkk. (2025) metode Non-Hierarki mencakup beberapa teknik, seperti *K-Means* dan *K-Medoids*.

Pada tahun 2023, penelitian yang dilakukan Insiyah dkk. mengenai analisis *cluster* untuk pengelompokan daerah rawan kriminalitas di Jawa Timur menghasilkan 3 *cluster*, diantaranya 31 kabupaten/kota pada *cluster* 1, 5 kabupaten/kota pada *cluster* 2, dan = 2 kabupaten/kota pada *cluster* 3. Dalam penelitian tersebut menggunakan data indikator kriminalitas, yaitu pembunuhan, penganiayaan berat, KDRT, penculikan, pencurian bermotor, dan penipuan. Namun, dalam Statistik Kriminal 2024, terdapat sejumlah indikator yang umum digunakan untuk menilai tingkat kriminal dari perspektif makro. Indikator tersebut meliputi jumlah tindak pidana (*crime total*), tingkat risiko penduduk terjadi tindak pidana per 100.000 penduduk (*crime rate*), serta selang waktu terjadinya tindak pidana (*crime clock*). Selain itu, indikator lain yang digunakan adalah jumlah tindak pidana yang diselesaikan (*crime cleared*) dan persentase penyelesaian tindak pidana (*clearance rate*).

Berdasarkan paparan di atas, menunjukkan bahwa eksplorasi metode analisis *cluster* belum dilakukan. Dengan demikian, penelitian ini akan membahas analisis *cluster* menggunakan metode Hierarki (*Single Linkage*, *Complete Linkage*, *Average Linkage*, *Centroid Linkage*, *Centroid*, *Median* dan *Ward's Method*) dan Non-Hierarki (*K-Means* dan *K-Medoids*) dengan indikator kriminalitas yang lebih komprehensif (diambil indikator kriminalitas di Jawa Timur sebagai studi kasus). Selanjutnya, pengelompokan wilayah kabupaten/kota di Jawa Timur menurut indikator kriminalitas dilakukan menggunakan hasil terbaik berdasarkan perhitungan indeks validitas *Davies-Bouldin (DB)*, *Silhouette*, *Calinski-Harabasz (CH)*, *Dunn*, dan *Connectivity*.

## KAJIAN TEORI

### ANALISIS CLUSTER METODE HIERARKI

Analisis *cluster* metode Hierarki adalah teknik pengelompokan yang menyusun cluster sebanyak  $n-1$ , di mana  $n$  sama dengan jumlah pengamatan, dalam bentuk hierarki atau struktur berlapis (Hair JR. dkk., 2010). Metode Hierarki menghasilkan pengelompokan yang dapat direpresentasikan dalam diagram pohon, yang umumnya dikenal dengan dendrogram. Dendrogram ini memperjelas proses hierarki, termasuk jarak antara *cluster-cluster* yang dikelompokkan.

#### *Single Linkage*

*Single Linkage* atau yang juga disebut dengan metode *nearest neighbor* adalah metode yang mendefinisikan jarak antar *cluster* berdasarkan jarak minimum antar objek masing-masing *cluster*. Proses pengelompokan ini dilakukan dengan menemukan semua jarak antar objek di dua *cluster*, lalu memilih jarak terpendek sebagai ukuran kesamaan *cluster*. Kemudian, pasangan objek yang memiliki jarak terpendek tersebut digabungkan sehingga terbentuk satu kelompok keseluruhan. Metode *Single Linkage* ini mendefinisikan jarak antara dua *cluster* A dan B sebagai jarak minimum antara sebuah titik di A dan sebuah titik di B.

$$D(A, B) = \min \{d(y_i, y_j)\}, \text{ untuk } y_i \text{ di A dan } y_j \text{ di B} \quad (1)$$

dimana  $d(y_i, y_j)$  adalah jarak antara vektor  $y_i$  dan  $y_j$ .

#### *Complete Linkage*

*Complete Linkage* atau yang juga disebut metode *farthest neighbor* adalah metode yang mendefinisikan kesamaan antar *cluster* sebagai diameter minimum yang dapat melingkupi semua objek di kedua *cluster*. Diameter merujuk pada jarak maksimum antara objek di setiap *cluster*.

$$D(A, B) = \max \{d(y_i, y_j)\}, \text{ untuk } y_i \text{ di A dan } y_j \text{ di B,}$$

dimana  $d(y_i, y_j)$  adalah jarak antara vektor  $y_i$  dan  $y_j$ .

#### *Average Linkage*

*Average Linkage* adalah metode yang mendefinisikan jarak antara dua *cluster* A dan B sebagai rata-rata jarak  $n_A n_B$  antara titik  $n_A$  di A dan titik  $n_B$  di B.

$$D(A, B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(y_i, y_j) \quad (2)$$

dimana jumlahnya adalah untuk semua  $y_i$  di A dan  $y_j$  di B.

#### *Centroid Linkage*

Metode *Centroid* adalah metode yang mendefinisikan jarak antara dua *cluster* A dan B sebagai jarak antara vektor rata-rata (sering disebut *centroid*) dari kedua *cluster* tersebut.

$$D(A, B) = d(\bar{y}_A, \bar{y}_B) \quad (3)$$

dimana  $\bar{y}_A$  dan  $\bar{y}_B$  adalah vektor rata-rata untuk masing-masing objek pengamatan di A dan di B, dan  $\bar{y}_A = \sum_{i=1}^{n_A} y_i / n_A$

$$(4)$$

Pada tiap langkah, dua *cluster* dengan jarak terkecil antara *centroid* dikelompokkan. Kemudian, *centroid cluster* baru AB diberikan rata-rata tertimbang

$$\bar{y}_{AB} = \frac{n_A \bar{y}_A + n_B \bar{y}_B}{n_A + n_B} \quad (5)$$

#### *Median*

Metode *Median* adalah metode yang mendefinisikan median (titik tengah) dari garis yang menghubungkan antara dua titik A dan B sebagai titik untuk menghitung jarak baru ke *cluster* lain guna menghindari pembobotan vektor rata-rata yang bergantung pada ukuran *cluster*

$$m_{AB} = \frac{1}{2} (\bar{y}_A + \bar{y}_B) \quad (6)$$

Pada setiap langkahnya, dua *cluster* dengan jarak terkecil antar median dikelompokkan.

#### *Ward*

Metode *Ward* atau yang juga disebut metode *incremental sum of squares* adalah metode yang menggunakan jumlah jarak *within-cluster squared* dan jarak *between-cluster squared*. Jika AB adalah *cluster* yang didapat dari mengelompokkan *cluster* A dan B, maka jumlah jarak dalam *cluster* (dari item-item dari vektor rata-rata *cluster*) adalah

$$SSE_A = \sum_{i=1}^{n_A} (y_i - \bar{y}_A)' (y_i - \bar{y}_A) \quad (7)$$

$$SSE_B = \sum_{i=1}^{n_B} (y_i - \bar{y}_B)'(y_i - \bar{y}_B) \quad (8)$$

$$SSE_{AB} = \sum_{i=1}^{n_{AB}} (y_i - \bar{y}_{AB})'(y_i - \bar{y}_{AB}) \quad (9)$$

dimana  $\bar{y}_{AB} = (n_A \bar{y}_A + n_B \bar{y}_B) / (n_A + n_B)$  adalah rata-rata tertimbang dan  $n_A$  adalah jumlah titik di A,  $n_B$  adalah jumlah titik di B, dan  $n_{AB} = n_A + n_B$  adalah jumlah titik di AB. Karena jumlah jarak ini ekuivalen dengan *within-cluster sums of squares*, maka jumlah tersebut dilambangkan dengan  $SSE_A$ ,  $SSE_B$ , dan  $SSE_{AB}$ .

Metode *Ward* mengelompokkan dua *cluster* A dan B dengan memperkecil penaikan SSE, yang didefinisikan dengan

$$I_{AB} = SSE_{AB} - (SSE_A + SSE_B) \quad (10)$$

Hal ini menunjukkan bahwa peningkatan  $I_{AB}$  memiliki dua bentuk ekuivalen berikut:

$$I_{AB} = n_A(\bar{y}_A - \bar{y}_{AB})'(\bar{y}_A - \bar{y}_{AB}) + n_B(\bar{y}_B - \bar{y}_{AB})'(\bar{y}_B - \bar{y}_{AB}) \quad (11)$$

$$= \frac{n_A n_B}{n_A + n_B} (\bar{y}_A - \bar{y}_B)'(\bar{y}_A - \bar{y}_B) \quad (12)$$

Dengan demikian, memperkecil penaikan SSE sama dengan memperkecil jarak *between-cluster*. Jika A terdiri dari  $y_i$  saja dan B terdiri dari  $y_j$  saja, maka  $SSE_A$  dan  $SSE_B$  adalah nol, dan persamaan di atas direduksi menjadi

$$I_{ij} = SSE_{AB} = \frac{1}{2} (y_i - y_j)'(y_i - y_j) = \frac{1}{2} d^2(y_i, y_j) \quad (13)$$

## ANALISIS CLUSTER METODE NON-HIERARKI

Analisis *cluster* metode Non-Hierarki adalah teknik pengelompokan dimana jumlah *cluster* (k) ditentukan terlebih dahulu. Prosedur pengelompokan ini dilakukan berdasarkan jumlah *cluster* (k) yang telah ditentukan sebelumnya tersebut.

Untuk menentukan jumlah *cluster* (k), terdapat dua pendekatan, diantaranya metode *Elbow* dan metode *Silhouette*. Metode *Elbow* adalah metode yang menentukan nilai k optimal berdasarkan tingkat variasi data. Metode *Elbow* merupakan salah satu cara terbaik dalam menentukan nilai k optimal, terutama ketika nilai k relatif kecil (Cui, 2020). Metode ini menunjukkan jumlah *cluster* optimal dari grafik yang membentuk titik siku. Metode *Elbow* ditentukan dengan menghitung *Within-Cluster Sum of Squares* (WCSS). WCSS mengukur jarak kuadrat antara titik data dan *centroid* setiap *cluster*.

Adapun metode *Silhouette* adalah metode untuk menentukan jumlah *optimal* yang mengukur kualitas pengelompokan, seberapa baik penempatan setiap titik data dalam *cluster* tertentu dan seberapa jauh jaraknya dari *cluster* tetangga terdekatnya. Nilai *Silhouette* memiliki rentang antara -1 sampai 1, dimana nilai positif mengindikasikan bahwa titik data telah ditempatkan pada *cluster* yang tepat, nilai

negatif mengindikasikan kemungkinan titik data ditempatkan pada *cluster* yang salah, dan nilai yang mendekati 0 mengindikasikan bahwa titik data terletak di dekat batas antara dua *cluster*. Penentuan nilai k optimal dilakukan berdasarkan nilai koefisien *Silhouette* tertinggi (yang mendekati 1).

## K-Means

*K-Means* adalah metode analisis *cluster* partisional yang bekerja dengan cara membagi objek menjadi sejumlah *cluster* (k) yang sebelumnya telah ditentukan. Proses pengelompokan dilakukan secara iteratif dengan menetapkan ulang objek ke *cluster* tertentu sehingga beberapa kriteria numerik terpenuhi. Kriteria tersebut bertujuan untuk meminimalkan jarak *within-cluster* (objek satu sama lain dalam *cluster*) dan memaksimalkan *between-cluster* (jarak antar *cluster*). MacQueen mengemukakan dalam Johnson & Wichern (2002), bahwa *K-Means* digunakan untuk menempatkan setiap objek ke *cluster* yang memiliki *centroid* terdekat (*mean*). Secara sederhana, prosedur *K-Means* terdiri dari tiga langkah utama, yaitu:

1. Bagi objek menjadi k *cluster* awal.
2. Hitung *centroid* awal.
3. Kemudian, masukkan objek ke *cluster* yang *centroid*-nya (rata-rata) yang terdekat.
4. Hitung ulang *centroid cluster* tersebut.
5. Ulangi langkah 2 sampai 4 untuk jumlah *cluster* yang berbeda sampai semua posisi objek tiap *cluster* tidak berpindah kelompok lagi.

## K-Medoids/PAM

*K-Medoids* atau yang biasa disebut juga metode *Partitioning Around Medoid* (PAM) adalah metode analisis *cluster* partisional yang memiliki prosedur serupa dengan *K-Means*, yaitu membagi objek menjadi sejumlah *cluster* (k) yang sebelumnya telah ditentukan. Perbedaan mendasar antara *K-Medoids* dan *K-Means* terletak pada penentuan titik pusat *cluster* (*centroid*). Pada *K-Medoids* menggunakan *medoids* (objek dalam kumpulan objek yang merepresentasikan sebuah *cluster*), bukan menggunakan rata-rata seperti pada *K-Means*. Pengelompokan dilakukan dengan berdasarkan kedekatan antara *medoids* dan objek *non medoids* lainnya (Setiyawati, 2017). Leonard Kaufman dan Peter J. Rousseeuw merupakan yang pertama kali mengembangkan metode *K-Medoids* ini. Keunggulan *K-Medoids* terletak pada kemampuannya untuk menangani kelemahan *K-Means* yang sensitif terhadap *outlier*. Berikut adalah prosedur dalam *K-Medoids*:

1. Pilih k objek secara acak sebagai *medoid* awal.

2. Alokasikan setiap objek ke *medoid* terdekat sesuai dengan jarak yang telah ditentukan (jarak *Euclidean*).
3. Untuk tiap *cluster*, pilih objek secara acak sebagai *medoid* baru.
4. Ulangi langkah 2 dan 3 untuk jumlah *cluster* yang berbeda sampai *medoid* tidak ada perubahan lagi.

## PENENTUAN METODE TERBAIK ANTAR METODE

Penentuan metode terbaik dilakukan dengan melihat nilai indeks validitas *cluster* indeks, yaitu indeks *Davies-Bouldin (DB)*, *Silhouette*, *Calinski-Harabasz (CH)*, *Dunn*, dan *Connectivity*.

### 1. *Davies-Bouldin (DB)*

Indeks *Davies-Bouldin (DB)* mengukur kualitas hasil analisis *cluster* dengan membandingkan jarak antar *cluster* terhadap ukuran dalam *cluster*. Nilai indeks *DB* yang terendah menyatakan kualitas *cluster* yang baik.

### 2. *Silhouette*

Indeks *Silhouette* mengukur kualitas hasil analisis *cluster*, seberapa baik penempatan setiap titik data dalam *cluster* tertentu dan seberapa jauh jaraknya dari *cluster* tetangga terdekatnya. Nilai *Silhouette* memiliki rentang antara -1 sampai 1, dimana nilai positif mengindikasikan bahwa titik data telah ditempatkan pada *cluster* yang tepat, nilai negatif mengindikasikan kemungkinan titik data ditempatkan pada *cluster* yang salah, dan nilai yang mendekati 0 mengindikasikan bahwa titik data terletak di dekat batas antara dua *cluster*. Nilai indeks *Silhouette* yang tertinggi menyatakan kualitas *cluster* yang baik.

### 3. *Calinski-Harabasz (CH)*

Indeks *Calinski-Harabasz* mengukur kualitas hasil analisis *cluster* dengan melihat seberapa jauh *cluster* terpisah satu sama lain dan seberapa besar variasi data dalam setiap *cluster*. Prinsip dasar indeks *Calinski-Harabasz* adalah menghitung rasio antara *Within-Cluster Sum of Squares (WCSS)* dan *Between-Cluster Sum of Squares (BCSS)*. Nilai indeks *CH* yang tertinggi menyatakan kualitas *cluster* yang baik.

### 4. *Dunn*

Indeks *dunn* mengukur kualitas hasil analisis *cluster* dengan membandingkan jarak terdekat antar *cluster* dengan jarak terjauh pada masing-masing *cluster* data. Nilai indeks *Dunn* yang tertinggi menyatakan kualitas *cluster* yang baik.

### 5. *Connectivity*

Indeks *Connectivity* mengukur seberapa jauh objek ditempatkan dalam *cluster* yang sama dengan tetangga terdekatnya. Nilai indeks *Connectivity*

memiliki rentang 0 sampai tak hingga. Nilai indeks *Connectivity* yang terendah menyatakan kualitas *cluster* yang baik.

## METODE

Pada penelitian ini, menggunakan jenis data sekunder, yang didapat dari website resmi Badan Pusat Statistik (BPS), publikasi Statistik Politik dan Keamanan Provinsi Jawa Timur 2024. Data yang diambil merupakan indikator kriminalitas, diantaranya Data Jumlah Tindak Pidana yang Dilaporkan dan Jumlah Tindak Pidana yang Diselesaikan Menurut Kabupaten/Kota di Provinsi Jawa Timur pada tahun 2024 yang bersumber dari Polres/Polresta/Polrestabes dan Polda. Selain itu, juga mengambil data Jumlah Penduduk Menurut Kabupaten/Kota di Provinsi Jawa Timur, 2024 untuk mendukung perhitungan data Risiko Penduduk Terjadi Tindak Pidana per 100.000 Penduduk.

Untuk mencapai tujuan dalam penelitian ini, analisis dilakukan dengan menggunakan alat bantu berupa *software Minitab 19* dan *RStudio*. Alur penelitian ini mencakup hal-hal berikut:

1. Melakukan analisis statistik deskriptif untuk setiap variabel.
2. Melakukan analisis *cluster* metode Hierarki
  - a. Melakukan uji multikolinearitas dari data variabel berdasarkan nilai matriks korelasi.
  - b. Menentukan ukuran jarak.
  - c. Proses pengolahan data dengan metode Hierarki, yang diantaranya *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Centroid Linkage*, *Median* dan *Ward*.
  - d. Menentukan metode terbaik dari metode Hierarki berdasarkan nilai indeks validitas *Davies-Bouldin (DB)*, *Silhouette*, *Calinski-Harabasz (CH)*, *Dunn*, dan *Connectivity*.
3. Melakukan analisis *cluster* metode Non-Hierarki
  - a. Pendeteksian pencilan (*outlier*) dari data variabel dengan *boxplot*.
  - b. Menentukan jumlah *cluster* dengan metode *Elbow Plot* dan *Silhouette Plot*.
  - c. Proses pengolahan data dengan metode Non-Hierarki, yang

diantaranya *K-Means* dan *K-Medoids/PAM*.

- d. Menentukan metode terbaik dari metode Non-Hierarki berdasarkan nilai indeks validitas *Davies-Bouldin (DB)*, *Silhouette*, *Calinski-Harabasz (CH)*, *Dunn*, dan *Connectivity*.

4. Pemilihan metode terbaik antara metode Hierarki dan Non-Hierarki berdasarkan nilai indeks validitas *Davies-Bouldin (DB)*, *Silhouette*, *Calinski-Harabasz (CH)*, *Dunn*, dan *Connectivity*.

## HASIL DAN PEMBAHASAN

### ANALISIS STATISTIK DESKRIPTIF

Pada dasarnya, statistik deskriptif bermaksud untuk mengubah data penelitian menjadi bentuk yang lebih sederhana dan mudah dipahami. Para peneliti sering kali memanfaatkan statistik deskriptif untuk menggambarkan ciri-ciri variabel penelitian serta membantu analisis penelitian. Statistik deskriptif hanya menyajikan ringkasan dari data yang dikumpulkan, tanpa melakukan prediksi atau generalisasi. Ringkasan data yang disajikan meliputi nilai rata-rata, standar deviasi, minimum, dan maksimum. Hasil statistik deskriptif yang merepresentasikan indikator kriminalitas di Jawa Timur tahun 2024 disajikan pada Tabel 1.

Tabel 1. Statistik Deskriptif

| Variabel | Rata-rata | Standar Deviasi | Minimum | Maksimum |
|----------|-----------|-----------------|---------|----------|
| X1       | 749       | 967             | 105     | 5543     |
| X2       | 685       | 964             | 96      | 5731     |
| X3       | 10617     | 40050           | 2       | 189699   |
| X4       | 92,25     | 23,44           | 46,97   | 176,38   |
| X5       | 93750     | 71352           | 5689    | 248315   |

Dari Tabel 1 di atas, terlihat bahwa:

Variabel X1 (Jumlah Tindak Pidana yang Dilaporkan) memiliki nilai rata-rata sebesar 749

dengan standar deviasi 967, menunjukkan bahwa penyebaran data cukup besar dibandingkan dengan nilai rata-ratanya. Kabupaten/kota dengan jumlah tindak pidana yang dilaporkan terendah sebesar 105 adalah Kabupaten Pacitan. Sedangkan Kabupaten/kota dengan jumlah tindak pidana yang dilaporkan tertinggi mencapai 5.543 adalah Kota Surabaya.

Variabel X2 (Jumlah Tindak Pidana yang Diselesaikan) memiliki nilai rata-rata 685 dan standar deviasi 964, menunjukkan bahwa penyebaran data cukup besar dibandingkan dengan nilai rata-ratanya. Kabupaten/kota dengan jumlah tindak pidana yang diselesaikan terendah sebesar 96 adalah Kabupaten Pacitan. Sedangkan Kabupaten/kota dengan jumlah tindak pidana yang diselesaikan tertinggi mencapai 5.731 adalah Kota Surabaya.

Variabel X3 (Risiko Penduduk Terjadi Tindak Pidana per 100.000 Penduduk) menunjukkan rata-rata yang relatif besar, yaitu 10.617, dengan standar deviasi yang sangat tinggi sebesar 40.050. Hal ini menandakan bahwa data sangat bervariasi. Kabupaten/kota dengan risiko terendah sebesar 2 adalah Kabupaten Pacitan. Sedangkan Kabupaten/kota dengan risiko tertinggi mencapai 189.699 adalah Kota Surabaya. Rentang nilai yang lebar, dari minimum 2 hingga maksimum 189.699.

Variabel X4 (Persentase Penyelesaian Tindak Pidana) memiliki rata-rata 92,25 dengan standar deviasi 23,44. Dibandingkan dengan variabel lainnya, X4 memiliki variasi data yang relatif lebih kecil. Kabupaten/kota dengan persentase penyelesaian terendah sebesar 46,97 adalah Kabupaten Probolinggo. Sedangkan Kabupaten/kota dengan persentase penyelesaian tertinggi mencapai 176,38 adalah Kota Blitar.

Variabel X5 (Selang Waktu Terjadinya Tindak Pidana) memiliki nilai rata-rata 93.750 dengan standar deviasi 71.352. Kabupaten/kota dengan selang waktu terendah sebesar 5.689 adalah Kota Surabaya. Sedangkan Kabupaten/kota dengan selang waktu tertinggi mencapai 248.315 adalah Kota Blitar.

### ANALISIS CLUSTER METODE HIERARKI

Sebelum analisis *cluster* dilakukan, perlu dilakukan pendeteksian multikolinearitas. Multikolinieritas adalah kondisi dimana ditemui

hubungan yang erat antara dua variabel kelompok atau lebih. Kondisi ini menjadi hal yang menjadi perhatian dalam analisis multivariat, mengingat dampaknya yang signifikan terhadap solusi yang dihasilkan, sehingga dapat mengganggu proses analisis (Hair JR. dkk., 2010). Pendeteksian multikolinearitas penting dilakukan sebelum menentukan ukuran jarak dalam analisis *cluster*. Pendeteksian multikolinearitas ini sebagai dasar pertimbangan dalam penentuan ukuran jarak. Ukuran jarak yang dapat digunakan adalah jarak *Euclidean* dan jarak *Mahalanobis*. Jarak *Euclidean* diterapkan pada variabel-variabel yang tidak berkorelasi satu sama lain. Adapun jarak *Mahalanobis* memperhitungkan kovariansi atau korelasi yang ditemui di antara kedua variabel dalam mengukur jarak antara dua titik. Dengan demikian, jika tidak terdapat multikolinieritas antar variabel yang digunakan, maka menggunakan ukuran jarak *Euclidean*. Sebaliknya, jika terdapat multikolinearitas antar variabel yang digunakan, maka menggunakan ukuran jarak *Mahalanobis*.

Menurut Gujarati dalam Soemartini (2008), salah satu pendeteksian multikolinieritas dapat dilakukan dengan mengamati nilai korelasi antar variabel. Jika nilai korelasi antar variabel melebihi 0,85, maka menunjukkan terjadi multikolinieritas antar variabel. Berikut merupakan nilai korelasi antar variabel.

### Correlations

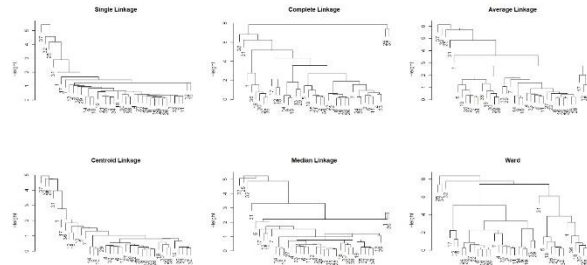
|    | X1     | X2     | X3     | X4    |
|----|--------|--------|--------|-------|
| X2 | 0,973  |        |        |       |
| X3 | 0,816  | 0,838  |        |       |
| X4 | -0,027 | 0,125  | 0,063  |       |
| X5 | -0,576 | -0,524 | -0,313 | 0,221 |

Gambar 1. Korelasi Antar Variabel

Dari Gambar 1 di atas, hasil pengolahan menggunakan software Minitab, terlihat bahwa terdapat nilai korelasi yang melebihi 0,85. Dengan demikian, kondisi tersebut menunjukkan terdapat multikolinieritas antar variabel. Variabel yang mengandung multikolinieritas meliputi X1 dengan X2, X1 dengan X3, dan X2 dengan X3. Karena terdapat multikolinearitas, maka analisis cluster metode Hierarki menggunakan ukuran jarak Mahalanobis. Proses pengolahan data dengan

ukuran jarak Mahalanobis tidak perlu dilakukan proses standarisasi.

Dalam penelitian ini, menggunakan beberapa pendekatan dalam metode Hierarki, diantaranya Single Linkage, Complete Linkage, Average Linkage, Centroid, Median, dan Ward. Berikut merupakan hasil analisis cluster dengan keenam pendekatan tersebut.



Gambar 2. Dendrogram Metode Hierarki

Berdasarkan Gambar 2 di atas, data indikator kriminalitas di Jawa Timur tahun 2024 dapat dikelompokkan menjadi 2, 3, dan 4 cluster. Penentuan pendekatan dan jumlah cluster optimal dilakukan dengan memperhatikan nilai indeks validitas cluster, yaitu indeks Davies-Bouldin (DB), Silhouette, Calinski-Harabasz (CH), Dunn, dan Connectivity. Nilai indeks validitas tersebut akan dibandingkan sehingga didapatkan jumlah cluster optimal berdasarkan kriteria masing-masing indeks validitas. Berikut merupakan daftar nilai indeks validitas cluster.

Tabel 2. Indeks Validitas Metode Hierarki

| Metode           | Jumlah Cluster | Indeks Validitas |            |            |            |              |
|------------------|----------------|------------------|------------|------------|------------|--------------|
|                  |                | DB               | Silhouette | CH         | Dunn       | Connectivity |
| Single Linkage   | 2              | 0.3645775        | 0.5949146  | 7.127533   | 0.69098759 | 0.125000     |
|                  | 3              | 0.7037426        | 0.5816039  | 4.205226   | 0.61810660 | 2.267.857    |
|                  | 4              | 0.5777691        | 0.5813012  | 5.280082   | 0.67013485 | 2.967.857    |
| Complete Linkage | 2              | 0.4506131        | 0.5881709  | 14.386.116 | 0.56812582 | 0.825000     |
|                  | 3              | 0.3753527        | 0.5795007  | 7.074066   | 0.61369701 | 0.825000     |

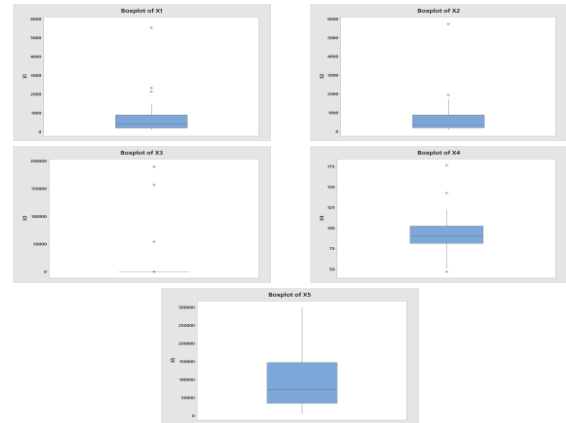
|                 |   |                             |                             |                              |                              |                            |
|-----------------|---|-----------------------------|-----------------------------|------------------------------|------------------------------|----------------------------|
|                 | 4 | 0.577<br>7691               | 0.581<br>3012               | 5.280.<br>082                | 0.670<br>13485               | 2.967.8<br>57              |
| Average Linkage | 2 | <b>0.364</b><br><b>5775</b> | <b>0.594</b><br><b>9146</b> | 7.127.<br>533                | <b>0.690</b><br><b>98759</b> | <b>0.12500</b><br><b>0</b> |
|                 | 3 | 0.703<br>7426               | 0.581<br>6039               | 4.205.<br>226                | 0.618<br>10660               | 2.267.8<br>57              |
|                 | 4 | 0.577<br>7691               | 0.581<br>3012               | 5.280.<br>082                | 0.670<br>13485               | 2.967.8<br>57              |
| Centroid        | 2 | <b>0.364</b><br><b>5775</b> | <b>0.594</b><br><b>9146</b> | 7.127.<br>533                | <b>0.690</b><br><b>98759</b> | <b>0.12500</b><br><b>0</b> |
|                 | 3 | 0.703<br>7426               | 0.581<br>6039               | 4.205.<br>226                | 0.618<br>10660               | 2.267.8<br>57              |
|                 | 4 | 0.577<br>7691               | 0.581<br>3012               | 5.280.<br>082                | 0.670<br>13485               | 2.967.8<br>57              |
| Median          | 2 | <b>0.364</b><br><b>5775</b> | <b>0.594</b><br><b>9146</b> | 7.127.<br>533                | <b>0.690</b><br><b>98759</b> | <b>0.12500</b><br><b>0</b> |
|                 | 3 | 0.375<br>3527               | 0.579<br>5007               | 7.074.<br>066                | 0.613<br>69701               | 0.82500<br>0               |
|                 | 4 | 0.577<br>7691               | 0.581<br>3012               | 5.280.<br>082                | 0.670<br>13485               | 2.967.8<br>57              |
| Ward            | 2 | 0.450<br>6131               | 0.588<br>1709               | 14.38<br>6.116               | 0.568<br>12582               | 0.82500<br>0               |
|                 | 3 | 0.678<br>7337               | 0.583<br>9911               | 8.065.<br>012                | 0.568<br>12582               | 2.967.8<br>57              |
|                 | 4 | 0.544<br>4337               | 0.292<br>3713               | <b>51.78</b><br><b>2.150</b> | 0.078<br>18089               | 10.020.<br>238             |

Berdasarkan Tabel 2 di atas, nilai DB dan Connectivity terendah ditunjukkan oleh *Single Linkage*, *Average Linkage*, *Centroid*, dan *Median* dengan 2 *cluster*. Nilai Silhouette dan Dunn terbesar juga ditunjukkan oleh *Single Linkage*, *Average Linkage*, *Centroid*, dan *Median* dengan 2 *cluster*. Sedangkan, nilai CH terbesar ditunjukkan oleh metode *Ward* dengan 4 *cluster*. Dengan demikian, dapat disimpulkan bahwa *Single Linkage*, *Average Linkage*, *Centroid*, dan *Median* dengan 2 *cluster* merupakan *cluster* dengan akurasi optimal.

## ANALISIS CLUSTER METODE NON-HIERARKI

Sebelum analisis *cluster* dilakukan, perlu dilakukan pendeteksian pencilan (*outlier*) dari setiap variabel. Pendeteksian pencilan (*outlier*) ini sebagai

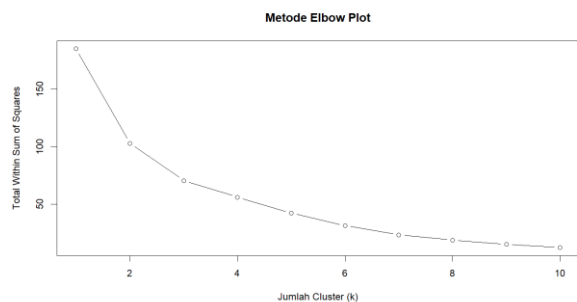
dasar pertimbangan dalam standarisasi data. Apabila variabel-variabel yang digunakan tidak menunjukkan adanya pencilan (*outlier*), maka tidak perlu standarisasi data. Sebaliknya, jika terdapat pencilan (*outlier*) antar variabel, maka perlu standarisasi data. Pendeteksian pencilan (*outlier*) dilakukan dengan mengamati *boxplot* yang disajikan pada Gambar 3 berikut.

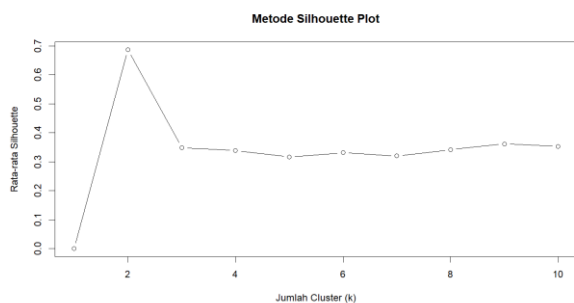


Gambar 3. Boxplot Setiap Variabel

Berdasarkan Gambar 3 di atas, masih terdapat pencilan (*outlier*) pada setiap variabel. Variabel yang mengandung pencilan (*outlier*) yaitu X1, X2, X3, dan X5. Karena terdapat pencilan (*outlier*), maka perlu dilakukan standarisasi data.

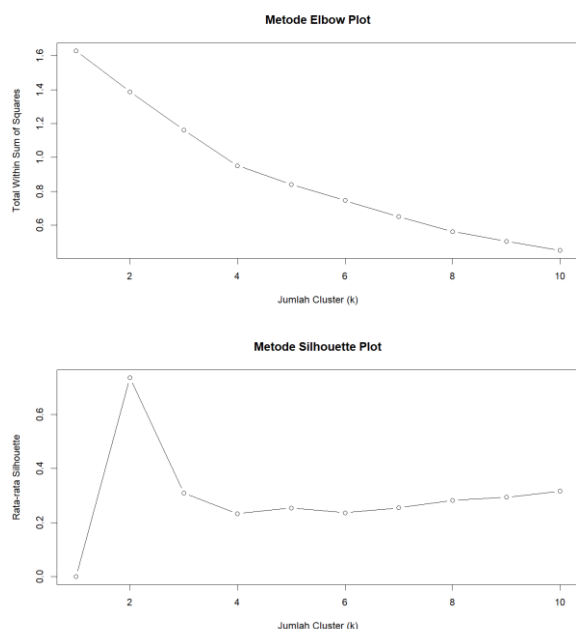
Dalam penelitian ini, metode Non-Hierarki menggunakan dua pendekatan, diantaranya *K-Means* dan *K-Medoids/PAM*. Tahapan sebelum analisis *cluster* metode Non-Hierarki adalah menentukan jumlah *cluster* terlebih dahulu. Metode Elbow Plot dan Silhouette Plot dapat digunakan dalam penentuan jumlah *cluster* optimal. Penentuan jumlah *cluster* menggunakan metode tersebut diperoleh hasil sebagai berikut:





Gambar 4. Metode Elbow Plot dan Silhouette Plot pada K-Means

Dari Gambar 4 di atas, pada Elbow Plot terlihat bahwa penurunan tajam pada nilai total *Within-Cluster Sum of Squares* (WCSS) terjadi sampai jumlah *cluster* ( $k$ ) = 2. Kemudian, cenderung melandai pada  $k > 2$ . Adapun pada Silhouette Plot, terlihat bahwa nilai rata-rata silhouette tertinggi terjadi pada jumlah *cluster* ( $k$ ) = 2. Maka, berdasarkan Elbow Plot dan Silhouette Plot dapat disimpulkan bahwa 2 *cluster* menjadi jumlah *cluster* optimal untuk K-Means.



Gambar 5 Metode Elbow Plot dan Silhouette Plot pada K-Medoids/PAM

Dari Gambar 5 di atas, pada Elbow Plot terlihat bahwa penurunan tajam pada nilai total *Within-Cluster Sum of Squares* (WCSS) terjadi sampai jumlah *cluster* ( $k$ ) = 4. Kemudian, cenderung melandai pada  $k > 4$ . Adapun pada Silhouette Plot, terlihat bahwa nilai rata-rata silhouette tertinggi terjadi pada jumlah *cluster* ( $k$ ) = 2. Dengan demikian, pengelompokan wilayah kabupaten/kota di Jawa

Timur berdasarkan data indikator kriminalitas dapat dilakukan dengan 2 dan 4 *cluster*.

Penentuan pendekatan dan jumlah *cluster* optimal dilakukan dengan memperhatikan nilai indeks validitas *cluster*, yaitu indeks *Davies-Bouldin* (DB), *Silhouette*, *Calinski-Harabasz* (CH), *Dunn*, dan *Connectivity*. Nilai indeks validitas tersebut akan dibandingkan sehingga didapatkan jumlah *cluster* optimal berdasarkan kriteria masing-masing indeks validitas. Berikut merupakan daftar nilai indeks validitas *cluster*.

Tabel 3. Indeks Validitas Metode Non-Hierarki

| Metode        | Jumlah Cluster | Indeks Validitas  |               |                   |                    |              |
|---------------|----------------|-------------------|---------------|-------------------|--------------------|--------------|
|               |                | DB                | Silhouette    | CH                | Dunn               | Connectivity |
| K-Means       | 2              | 0,64<br>6107<br>4 | 0,687<br>1536 | 2,86<br>8.37<br>8 | 0,404<br>9374      | 0,825        |
| K-Medoids/PAM | 2              | 0,19<br>4912<br>2 | 0,735<br>8207 | 24,9<br>368<br>1  | 0,782<br>279       | 0,1250<br>00 |
|               | 4              | 1,12<br>8484      | 0,232<br>7807 | 22,9<br>868       | 0,068<br>4167<br>9 | 10,728<br>57 |

Berdasarkan Tabel 3 di atas, nilai DB dan Connectivity terendah ditunjukkan oleh K-Medoids/PAM dengan 2 *cluster*. Nilai Silhouette, CH, dan Dunn terbesar juga ditunjukkan oleh K-Medoids/PAM dengan 2 *cluster*. Dengan demikian, dapat disimpulkan bahwa K-Medoids/PAM dengan 2 *cluster* merupakan *cluster* dengan akurasi optimal.

## PERBANDINGAN ANALISIS CLUSTER HIERARKI DAN NON-HIERARKI TERBAIK

Tabel 4. Perbandingan Analisis Cluster Hierarki dan Non-Hierarki Terbaik

| Metode | Jumlah Cluster | Indeks Validitas |            |    |      |              |
|--------|----------------|------------------|------------|----|------|--------------|
|        |                | DB               | Silhouette | CH | Dunn | Connectivity |
|        |                |                  |            |    |      |              |

|                                                                                                 |   |                            |                       |                      |                      |                      |
|-------------------------------------------------------------------------------------------------|---|----------------------------|-----------------------|----------------------|----------------------|----------------------|
| Sing<br>le<br>Link<br>age,<br>Ave<br>rage<br>Link<br>age,<br>Cent<br>roid,<br>dan<br>Med<br>ian | 2 | 0.36<br>4577<br>5          | 0.594<br>9146         | 7.12<br>7.53<br>3    | 0.690<br>9875<br>9   | <b>0.1250<br/>00</b> |
| K-<br>Med<br>oids                                                                               | 2 | <b>0,19<br/>4912<br/>2</b> | <b>0,735<br/>8207</b> | <b>24,9<br/>3681</b> | <b>0,782<br/>279</b> | <b>0,1250<br/>00</b> |

Dari Tabel 4 di atas, terlihat bahwa nilai DB dan Connectivity terendah ditunjukkan oleh metode *K-Medoids* dengan 2 *cluster*. Nilai Silhouette, CH, dan Dunn terbesar juga ditunjukkan oleh metode *K-Medoids* dengan 2 *cluster*. Dengan demikian, dapat disimpulkan bahwa *K-Medoids* dengan 2 *cluster* merupakan *cluster* dengan akurasi optimal.

Hasil analisis *cluster* menggunakan *K-Medoids/PAM* dengan 2 *cluster* adalah sebagai berikut:

1. *Cluster 1* terdiri dari 37 kabupaten/kota, hampir seluruh wilayah di Jawa Timur (selain Kota Surabaya)
2. *Cluster 2* hanya terdiri dari 1 kota, yaitu Kota Surabaya.

Untuk menentukan karakteristik *cluster* yang dihasilkan dari *K-Medoids/PAM*, akan dihitung rata-rata dari setiap variabel dalam setiap *cluster*. Perhitungan rata-rata dari setiap variabel diperoleh hasil sebagai berikut:

Tabel 5. Rata-Rata Setiap Variabel Metode *K-Medoids*

| Variabel | Rata-rata |           |
|----------|-----------|-----------|
|          | Cluster 1 | Cluster 2 |
| X1       | 619,2703  | 5543      |
| X2       | 548,4865  | 5731      |
| X3       | 5777,204  | 189698,8  |

|    |          |          |
|----|----------|----------|
| X4 | 91,95011 | 103,3917 |
| X5 | 96130,43 | 5689,338 |

Berdasarkan Tabel 5 di atas, menunjukkan bahwa:

#### 1. Cluster 1

Pada *cluster 1* nilai rata-rata jumlah tindak pidana yang dilaporkan (X1), jumlah tindak pidana yang diselesaikan (X2), tingkat risiko penduduk terkena tindak pidana (X3), dan persentase penyelesaian tindak pidana (X4) lebih rendah dibandingkan dengan *cluster 2*. Selanjutnya, selang waktu terjadinya tindak pidana (X5) pada *cluster* ini tergolong panjang, yang mengindikasikan bahwa kejadian kriminal relatif jarang terjadi. Oleh karena itu, *cluster 1* dapat dikategorikan sebagai kelompok daerah dengan tingkat kriminalitas rendah.

#### 2. Cluster 2

Sementara itu, pada *cluster 2* nilai rata-rata jumlah tindak pidana yang dilaporkan (X1), jumlah tindak pidana yang diselesaikan (X2), tingkat risiko penduduk terkena tindak pidana (X3), dan persentase penyelesaian tindak pidana (X4) lebih tinggi dibandingkan dengan *cluster 1*. Selanjutnya, selang waktu terjadinya tindak pidana (X5) pada *cluster* ini tergolong panjang, yang mengindikasikan bahwa kejadian kriminal relatif sering terjadi. Oleh karena itu, *cluster 2* dapat dikategorikan sebagai kelompok daerah dengan tingkat kriminalitas tinggi.

## PENUTUP

### SIMPULAN

Hasil analisis memperlihatkan bahwa metode terbaik dalam analisis *cluster* antara metode Hierarki dan Non-Hierarki untuk kriminalitas di Jawa Timur adalah *K-Medoids/PAM* dengan 2 *cluster*. Hal ini didapat dari perhitungan nilai indeks validitas *cluster* menggunakan indeks *Davies-Bouldin (DB)*, *Silhouette*, *Calinski-Harabasz (CH)*, *Dunn*, dan *Connectivity*. Hasil analisis *cluster* menggunakan *K-Medoids* dengan 2 *cluster* adalah 37 kabupaten/kota pada *cluster 1* dan 1 kota pada *cluster 2*. Adapun karakteristik tiap *clusternya* adalah *cluster 1 cluster 1* dapat dikategorikan sebagai kelompok wilayah dengan tingkat kriminalitas rendah. Sementara itu, pada *cluster 2* dapat dikategorikan sebagai kelompok wilayah dengan tingkat kriminalitas tinggi.

Metode *K-Medoids* terpilih menjadi metode terbaik karena mampu menghasilkan pengelompokan yang lebih stabil dan representatif. *K-Medoids* menggunakan objek pada kumpulan objek yang merepresentasikan sebuah *cluster* (disebut dengan *medoids*), sehingga lebih *robust* terhadap pencilaan (*outlier*). Temuan ini menjadi penting, mengingat data indikator kriminalitas memiliki ketimpangan yang cukup ekstrem antar wilayah, khususnya keberadaan Kota Surabaya dengan tingkat kriminalitas yang jauh lebih tinggi dibandingkan kabupaten/kota lain.

#### SARAN

Penelitian selanjutnya disarankan untuk menambahkan variabel lain yang relevan, seperti faktor ekonomi, kepadatan penduduk, tingkat pendidikan, atau tingkat pengangguran, dapat membantu menjelaskan faktor-faktor yang memengaruhi perbedaan karakteristik antar cluster. Penelitian lanjutan juga dapat menguji metode clustering lain, seperti DBSCAN atau model berbasis fuzzy clustering, untuk membandingkan kestabilan dan konsistensi hasil pengelompokan.

Pemerintah daerah dan aparat penegak hukum disarankan untuk memberikan prioritas penanganan dan pengawasan lebih intensif pada wilayah yang termasuk dalam *cluster* dengan tingkat kriminalitas tinggi. Hasil pengelompokan ini dapat dijadikan sebagai dasar perumusan kebijakan berbasis data dalam pengalokasian sumber daya keamanan agar lebih efektif dan tepat sasaran.

#### DAFTAR PUSTAKA

- Abdulsyani. (1987). *Sosiologi Kriminalitas*. Remadja Karya CV.
- Ardana Wijaya, M., Satria Prayoga, D., Karunia Rahman, A., & Puspita Sari, A. (2023). Perbandingan Algoritma K-Means dan DBSCAN dalam Metode Clustering dengan PCA untuk Analisis Data Statistik Negara Dunia. *Seminar Nasional Informatika Bela Negara (SANTIKA)*, 3, 63–70. <https://ourworldindata.org/>
- Cui, M. (2020). *Introduction to the K-Means Clustering Algorithm Based on the Elbow Method*. 1, 5–8. <https://doi.org/10.23977/accaf.2020.010102>
- Dulkiah, M. (2020). *Sosiologi Kriminal*. LP2M UIN SGD Bandung.
- Gujarati, D. (1978). *Ekonometrika Dasar*. Erlangga.
- Hair JR., J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (Seventh Edition). Pearson.
- Insiyah, I., Khasanah, M., & Hendarsyah, T. P. (2023). Penerapan Metode Ward Clustering Untuk Pengelompokan Daerah Rawan Kriminalitas Di Jawa Timur Tahun 2021. *Jurnal Statistika dan Komputasi*, 2(1), 44–54. <https://doi.org/10.32665/statkom.v2i1.1664>
- Johnson, R. A., & Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis* (Fifth Edition). Prentice-Hall, Inc. .
- Nahdliyah, M. A., Widiharini, T., & Prahutama, A. (2019). Metode K-Medoids Clustering dengan Validasi Silhouette Index dan C-Index. *Jurnal Gaussian*, 8(2), 161–170. <http://ejournal3.undip.ac.id/index.php/gaussian>
- Raja, N. A., Tinungki, G. M., & Sirajang, N. (2024). Implementasi Algoritma Centroid Linkage dan K-Medoids dalam Mengelompokkan Kabupaten/Kota di Sulawesi Selatan Berdasarkan Indikator Pendidikan. *ESTIMASI: Journal of Statistics and Its Application*, 61–74. <https://doi.org/10.20956/ejsa.v5i1.13605>
- Rencher, A. C. ., & Christensen, W. F. . (2012). *Methods of Multivariate Analysis* (Third Edition). John Wiley & Sons, Inc., Hoboken, New Jersey .
- Setiyawati, A. W. (2017). *Implementasi Algoritma Partitioning Around Medoids (PAM) untuk Pengelompokan Sekolah Menengah Atas Berdasarkan Nilai Daya Serap Ujian Nasional*.
- Soemartini. (2008). *Principal Component Analysis (PCA) sebagai Salah Satu Metode untuk Mengatasi Masalah Multikolinearitas*.
- Statistik Kriminal 2024 (Vol. 15). (2024). Badan Pusat Statistik.
- Statistik Kriminal 2024/2025 (Vol. 16). (2025). Badan Pusat Statistik.
- Sunandi, E., Simanullang, A. R., Elvira, V., Syah, S., & Afni, D. N. (2025). Analisis Perbandingan Metode K-Medoid dan Ward dalam Klasterisasi Nasabah Bank Churners. *MATHunesa: Jurnal Ilmiah Matematika*, 13(02), 118–123. <https://www.kaggle.com/datasets/imanema/g/bankchurner>
- Syafiyah, U., Asrafi, I., Wicaksono, B., Puspitasari, D. P., & Sirait, M. (2022). Analisis Perbandingan Hierarchical dan Non-Hierarchical Clustering Pada Data Indikator Ketenagakerjaan di Jawa Barat Tahun 2020. *Seminar Nasional Official Statistics*, 803–813.