

## Klasterisasi Wilayah Rawan Kriminalitas Di Kota Jambi (2022-2024) Menggunakan Algoritma K-Means

Sri Sulistina<sup>\*1</sup>, Pradita Eko Prasetyo Utomo<sup>2</sup>, Benedika Ferdian Hutabarat<sup>3</sup>

<sup>1,2,3</sup>Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Jambi, Indonesia  
Email: <sup>1</sup>srisulistina06@gmail.com, <sup>2</sup>pradita.eko@unja.ac.id, <sup>3</sup>benedika@unja.ac.id

### Abstrak

Penelitian ini bertujuan mengelompokkan kerawanan kriminalitas di 11 kecamatan Kota Jambi pada periode 2022–2024 menggunakan algoritma K-Means. Data kriminalitas diperoleh dari Polresta Kota Jambi, sedangkan data demografis (jumlah penduduk, luas wilayah, dan kepadatan penduduk) diperoleh dari BPS Kota Jambi. Seluruh variabel numerik dinormalisasi menggunakan metode min–max. Penentuan jumlah klaster dievaluasi pada  $K = 2$ – $10$  menggunakan Elbow Method, Silhouette Coefficient, dan Davies Bouldin Index (DBI). Pada tahun 2022 dan 2023, Elbow menunjukkan titik siku pada  $K = 3$  (masing-masing bernilai 0,872 dan 1,124), dengan nilai Silhouette maksimum pada  $K = 3$  (0,476 dan 0,466), sedangkan DBI mencapai nilai minimum pada  $K = 9$  (0,197 dan 0,3963). Pada tahun 2024, Elbow kembali mengarah ke  $K = 3$  (1,182), Silhouette tertinggi diperoleh pada  $K = 5$  (0,536) dengan nilai yang masih kompetitif pada  $K = 3$  (0,492), dan DBI kembali minimum pada  $K = 9$  (0,2458). Dengan mempertimbangkan konsistensi Elbow dan Silhouette serta kemudahan interpretasi, dipilih  $K = 3$  sebagai jumlah klaster optimal. Evaluasi konsistensi menggunakan Rand Index menunjukkan nilai 1 sepanjang periode, sehingga menguatkan bahwa tiga klaster merupakan struktur pengelompokan yang stabil.

**Kata Kunci:** Kriminalitas, Clustering, Elbow, Silhouette Coefficient, Davies–Bouldin Index, K-Means

### Abstract

*This study categorizes crime vulnerability in 11 districts of Jambi City for the 2022–2024 period using the K-Means Clustering algorithm. Criminal case data were obtained from the Jambi City Police, while demographic data (population, land area, density) were collected from Statistics Indonesia (BPS) Jambi City. All numerical variables were normalized using the min–max method. The number of Klasters was evaluated for  $K = 2$ – $10$  using the Elbow Method, Silhouette Coefficient, and Davies–Bouldin Index (DBI). In 2022 and 2023, the Elbow Method indicated an Elbow at  $K = 3$  (0.872 and 1.124), with the maximum Silhouette value also at  $K = 3$  (0.476 and 0.466), while DBI reached its minimum at  $K = 9$  (0.197 and 0.3963). In 2024, the Elbow Method again pointed to  $K = 3$  (1.182); the highest Silhouette value was obtained at  $K = 5$  (0.536) with a competitive value at  $K = 3$  (0.492), and DBI again reached its minimum at  $K = 9$  (0.2458). Considering the consistency of Elbow and Silhouette and the need for interpretability,  $K = 3$  was selected as the optimal number of Klasters. Evaluation using the Rand Index yielded a value of 1, confirming that three Klasters form a stable structure.*

**Keywords:** Crime, Clustering, Elbow, Silhouette Coefficient, Davies–Bouldin Index, K-Means

*This work is an open access article and licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0)*



## 1. PENDAHULUAN

Kriminalitas di Indonesia menunjukkan peningkatan yang signifikan dalam beberapa tahun terakhir. Berdasarkan data Kepolisian Republik Indonesia (Polri), pada tahun 2023 jumlah kriminalitas mencapai 288.472 kasus. Kriminalitas cenderung menaik sekitar 4,33% dibandingkan tahun sebelumnya di tahun 2022 yang mencatat 276.507 kasus [1]. Hal ini mencerminkan kenaikan sebesar 11.965 kasus hanya dalam kurun waktu satu tahun dan menjadikannya angka tertinggi dalam lima tahun terakhir, setelah sempat mengalami penurunan kriminalitas di tahun 2019. Tingginya angka kriminalitas di Indonesia juga tercermin di berbagai daerah, termasuk di Kota Jambi yang angka kasus kejahatannya turut mengalami peningkatan. Kota Jambi mencatat 2.419 kasus kriminalitas di tahun 2023, dibandingkan dengan kasus kriminalitas pada tahun 2022 sebanyak 1.671 kasus. Terjadinya 748 kasus baru dalam kurun waktu satu tahun ini menunjukkan bahwa upaya penanggulangan kejahatan masih

menghadapi tantangan besar dan memerlukan informasi berbasis data mengenai sebaran wilayah rawan kriminalitas sebagai bahan pertimbangan kebijakan keamanan [2].

Kriminal atau kejahatan mengacu pada segala jenis tindakan atau kegiatan yang merugikan secara ekonomi dan psikologis, serta melanggar hukum dan norma sosial dan agama yang berlaku di Indonesia. Dapat diartikan bahwa tindak kriminalitas adalah segala sesuatu perbuatan yang melanggar hukum dan mengabaikan norma-norma sosial yang berlaku di masyarakat [3]. Kriminalitas menjadi satu isu yang sering muncul dalam masyarakat yang harus diperhatikan karena merugikan banyak kepentingan dan menimbulkan efek negatif bagi masyarakat, seperti kecemasan, perasaan tidak aman, kepanikan, dan ketakutan [4]. Sejalan dengan itu, negara bertanggung jawab untuk menjaga keselamatan dan keamanan rakyatnya dari tindakan kriminal, yang merupakan bagian dari usaha untuk menjaga perlindungan negara. Tindak kejahatan merupakan bahaya bagi keamanan, pengelolaan, dan hukum negara, yang bisa mengganggu stabilitas [5]. Untuk mendukung upaya tersebut, penelitian ini menerapkan teknik data mining Klastering guna mengidentifikasi wilayah rawan kriminalitas di Kota Jambi.

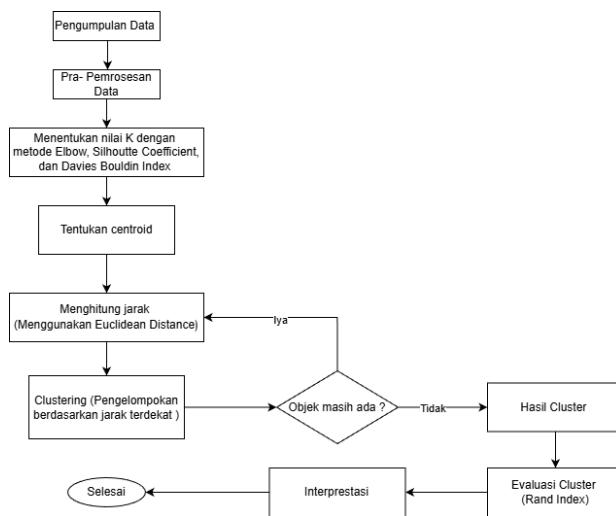
Salah satu pendekatan yang dapat dimanfaatkan untuk menganalisis pola kerawanan kriminalitas adalah teknik data mining, terutama melalui proses Klastering, yang mengelompokkan data berdasarkan kesamaan karakteristik sehingga pola-pola tertentu dapat dikenali [6]. Salah satu algoritma Klastering non-hierarki yang sering digunakan adalah K-Means Klastering [7]. Algoritma ini dipilih karena kesederhanaan dan efisiensinya dalam mempartisi data menjadi beberapa klaster, termasuk pada dataset dengan volume besar. K-Means mampu mengidentifikasi pola yang relevan untuk pemahaman tingkat kerawanan kriminalitas [8]. Meski membutuhkan penentuan jumlah klaster di awal dan sensitif terhadap outlier [9], keunggulan efisiensi dan kemudahannya menjadikan K-Means metode yang tepat untuk penelitian ini. K-Means Klastering dipilih karena algoritma kesederhanaan dan efisiensinya, dengan mempartisi data ke dalam kelompok-kelompok yang berbeda, juga pada dataset bervolume tinggi.

Penelitian terkait pemetaan wilayah rawan kriminalitas menggunakan K-Means telah dilakukan di beberapa daerah di Indonesia. Misalnya, Rahmat Hidayat (2021) mengelompokkan daerah rawan kejahatan di Kabupaten Solok berdasarkan jumlah kasus per kecamatan, dan menemukan variasi tingkat kerawanan antar kecamatan [10]. Penelitian lain, berjudul *Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta* [11], membahas penentuan jumlah klaster optimal (K). Dalam studi tersebut, K optimal ditentukan berdasarkan konsensus mayoritas: K = 3 direkomendasikan oleh metode Elbow dan Calinski-Harabasz Index (K = 3 dan K = 6), serta didukung oleh Silhouette Score (K = 2, 3, dan 6), meskipun Davies-Bouldin Index menunjukkan kecenderungan ke K = 6.

Meskipun demikian, sebagian besar penelitian terdahulu terbatas pada analisis total kasus atau tidak mempertimbangkan faktor-faktor lokal seperti kepadatan penduduk, luas wilayah, dan variabel struktural lainnya secara bersamaan. Penelitian ini bertujuan untuk mengisi kekosongan tersebut dengan melakukan klasterisasi wilayah kriminalitas di Kota Jambi secara lebih mendalam, dengan mempertimbangkan berbagai variabel sosio-ekonomi dan geografis, serta menentukan jumlah klaster optimal menggunakan kombinasi beberapa metode evaluasi. Hasil penelitian diharapkan memberikan gambaran pengelompokan wilayah rawan kriminalitas yang lebih komprehensif sebagai bentuk interpretasi dari penerapan algoritma K-Means dalam analisis data kriminalitas. Berdasarkan latar belakang yang telah diuraikan, penelitian ini akan melakukan pengelompokan wilayah rawan kriminalitas di Kota Jambi menggunakan algoritma K-Means.

## 2. METODE PENELITIAN

### 2.1 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

### 2.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari dua sumber utama, yaitu Polresta Jambi dan Badan Pusat Statistik (BPS) Kota Jambi. Data dari Polresta Jambi mencakup jumlah kasus kejahatan yang terjadi dalam wilayah kerja masing-masing Polsek jajaran. Sementara itu, data dari BPS Kota Jambi memberikan informasi tambahan berupa data demografis dan geografis, yaitu jumlah penduduk, luas wilayah, dan kepadatan penduduk di setiap kecamatan. Data yang digunakan berasal dari rentang tahun 2022 hingga 2024.

Tabel 1. Variabel Data

No	Variabel	Tipe Data
1.	Kecamatan	Kategorikal
2.	Wilayah (Polsek Jajaran)	Kategorikal
3.	Bulan	Numerik
4.	Tahun	Numerik
5.	Total Kejahatan	Numerik
6.	Jumlah Penduduk	Numerik
7.	Luas Wilayah	Numerik
8.	Kepadatan Penduduk	Numerik

Penelitian ini menggunakan delapan variabel awal sebagaimana tercantum pada Tabel 1. Hanya empat variabel numerik yang digunakan dalam analisis *K-Means* Klastering, yaitu Total Kejahatan, Jumlah Penduduk, Luas Wilayah, dan Kepadatan Penduduk. Data dianalisis per tahun oleh karena itu variabel Bulan tidak dimasukkan sebagai fitur. Sementara itu, variabel kategorikal berupa Kecamatan dan Wilayah Polsek Jajaran tidak dimasukkan dalam proses perhitungan jarak, tetapi tetap digunakan pada tahap interpretasi hasil klaster untuk memberikan identitas wilayah.

### 2.3 Pre Processing

Data *preprocessing*, atau pra-pemrosesan data, merujuk pada serangkaian proses yang diterapkan pada data mentah sebelum digunakan untuk analisis lebih lanjut atau pembangunan model. Tujuan utamanya adalah untuk meningkatkan kualitas data, menjamin ketepatan hasil analisis, serta mengatasi potensi masalah atau kekurangan yang ada pada data awal [12].

Berikut adalah beberapa tahapan *preprocessing* yang dilakukan dalam penelitian ini :

1. Seleksi Fitur (*Data Selection*)

Pemilihan subset fitur mempertimbangkan variable yang paling relevant dan signifikan untuk analisis atau pemodelan [12]. Variabel yang dipilih dalam analisis *K-Means* Klastering untuk penelitian ini adalah Total Kejahatan, Jumlah Penduduk, Luas Wilayah, dan Kepadatan Penduduk.

2. Pembersihan Data (*Data Cleaning*)

Proses ini dilakukan untuk meminimalkan data yang tidak konsisten sebelum digunakan pada analisis lebih lanjut [13]. Pembersihan data dilakukan dalam penelitian ini dengan memeriksa dan mengatasi duplikasi data dan mengatasi nilai yang hilang, serta mengoreksi inkonsistensi atau kesalahan input yang berpotensi memengaruhi hasil analisis *K-Means* Klastering

3. Transformasi Data

Transformasi data merupakan langkah, modifikasi, dan pemodelan data agar sesuai untuk berbagai tujuan. Tujuan dari proses ini adalah untuk mengonversi data menjadi format yang berbeda. Dengan transformasi data, informasi dari berbagai sumber dan berbagai format dapat digabungkan, dimodelkan, dan dianalisis secara menyeluruh [14]. Untuk mengatasi perbedaan skala dan inkonsistensi antar variabel, peneliti ini melakukan manipulasi atau transformasi data dengan menerapkan metode *normalisasi min-max* [15].

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Pada persamaan (1)  $x_i$  adalah nilai yang akan dinormalisasi, sedangkan  $x'$  adalah hasil normalisasi. Nilai  $\min(x)$  merupakan nilai minimum atribut, dan  $\max(x)$  adalah nilai maksimumnya.

## 2.4 Evaluasi Jumlah Klaster Optimal

Nilai K optimal ditentukan dengan memanfaatkan tiga metrik evaluasi, yaitu metode Elbow, Silhouette Score, dan Davies–Bouldin Index (DBI). Elbow Method biasanya ditampilkan dalam bentuk grafik untuk mengetahui lebih jelas siku yang terbentuk. Tujuannya untuk memilih nilai k yang kecil dan masih memiliki nilai *withinss* yang rendah sehingga hasil Klaster yang di berikan jelas dan bermakna tanpa terlalu banyak detail yang berlebihan [16]

Jika nilai Klaster pertama dibandingkan dengan nilai Klaster kedua menunjukkan sudut dalam grafik atau nilainya mengalami penurunan paling besar, maka jumlah nilai Klaster tersebut adalah yang tepat. Untuk memperoleh perbandingannya, kita perlu menghitung *Sum of Square Error* (SSE) dari setiap nilai Klaster. Karena semakin tinggi jumlah nilai Klaster K, maka nilai SSE akan semakin menurun [17]. Adapun persamaan yang digunakan dalam metode Elbow yakni nilai total WSS (*Within Klaster Sum of Squares*) yang juga dikenal sebagai SSE (*Sum Square Error*) dengan rumus [16] :

$$SSE = \sum_{k=1}^K \sum_{x_i \in C_k} (X_i - \varphi_k)^2 \quad (2)$$

Pada persamaan di atas (2),  $x_i$  adalah data ke-i dalam klaster dan  $\varphi_k$  adalah rata-rata (*centroid*) klaster ke-k. Nilai SSE dihitung untuk beberapa nilai K, lalu titik “siku” pada grafik SSE dipilih sebagai jumlah klaster optimal.

Adapun Koefisien Silhouette menggabungkan konsep kohesi dan pemisahan dalam proses klasterisasi data. Nilai *koefisien Silhouette* bervariasi antara -1 sampai 1, dan sebuah sistem pengelompokan dinilai efektif jika nilai *koefisien Silhouette* mendekati 1 [18]. Adapun cara untuk menghitung Silhouette Coefficient dalam persamaan berikut [19] :

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

Pada persamaan (3),  $a(o)$  merepresentasikan rata-rata jarak suatu objek terhadap anggota lain di klaster yang sama, sedangkan  $b(o)$  menunjukkan rata-rata jarak objek tersebut terhadap objek-objek pada klaster terdekat.

Sementara itu, Davies-Bouldin Index (DBI) menilai seberapa baik kluster terpisah satu sama lain dan seberapa seragam data yang ada di dalam kluster. Nilai DBI yang paling rendah menunjukkan jumlah kluster yang paling optimal. Tahapan dari perhitungan Davies-Bouldin Index adalah sebagai berikut [17]. Persamaan (4) merupakan rumus DBI yang akan digunakan untuk melakukan evaluasi jumlah klaster optimal:

$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (4)$$

Dengan,

$$R_{i,j} = \frac{SSW_I + SSW_J}{SSB_{i,j}} \quad (5)$$

Dimana pada persamaan (5) SSW merepresentasikan tingkat kekompakan dalam klaster, sedangkan SSB menggambarkan tingkat pemisahan antar klaster.

## 2.5 K-Means

Pada titik ini, teknik pengelompokan dilaksanakan dengan memanfaatkan metode K-Means Klustering. K-Means merupakan salah satu algoritma Klustering yang masuk ke dalam kelompok *unsupervised learning* yang dapat di pakai untuk membagi data menjadi beberapa kelompok dengan sistem partisi [20]. Algoritma K-Means digunakan untuk mengelompokkan data, observasi, atau kasus berdasarkan kemiripan dari apa yang sedang dipelajari. K-means Klustering adalah teknik analisis Klaster *non-hierarki* yang mencoba membagi objek-objek yang ada ke dalam satu atau lebih Klaster atau kelompok berdasarkan sifat-sifat yang dimilikinya. Langkah-langkah algoritma K-Means dapat dijelaskan sebagai berikut [21] :

1. Menentukan jumlah Klaster (k) pada dataset.
2. Menentukan titik pusat (*Centroid*) secara random

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (6)$$

Pada persamaan (6),  $V_{ij}$  menyatakan *centroid* klaster ke-i pada variabel ke-j,  $N_i$  menunjukkan banyaknya anggota klaster ke-i, sedangkan  $X_{kj}$  adalah nilai data ke-k pada variabel ke-j di klaster tersebut. Keterangan:

3. Menghitung jarak terdekat titik pusat atau *centroid* dapat menggunakan rumus :

$$d = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (7)$$

Pada persamaan (7),  $d$  merupakan jarak *euclidean* antara suatu objek dan titik *centroid*, dengan  $(x_i, y_i)$  sebagai koordinat objek ke-i dan  $(s_i, t_i)$  sebagai koordinat *centroid* yang bersesuaian.

4. Golongkan objek menurut jarak ke *centroid* terdekat
5. Lakukan Kembali langkah 3 dan 4 sehingga menghasilkan *centroid* secara optimal

## 2.6 Evaluasi Rand Indek

Rand Index merupakan ukuran yang digunakan untuk mengevaluasi seberapa baik hasil suatu proses Klustering. Indeks ini menilai proporsi pasangan data yang dikelompokkan secara benar, baik berada dalam klaster yang sama maupun berbeda sesuai dengan pengelompokan ideal. Nilainya berada pada rentang 0 hingga 1, di mana angka yang semakin mendekati 1 menandakan hasil klasterisasi yang semakin akurat karena menunjukkan lebih banyak pasangan data yang terklasifikasi dengan tepat [22]. Rumus untuk menghitung Rand Index score [23] :

$$R = \frac{(a+b)}{(nC_2)} \quad (8)$$

Dalam evaluasi Rand Index, a menunjukkan jumlah pasangan elemen yang ditempatkan dalam Klaster yang sama, sedangkan b menunjukkan jumlah pasangan elemen yang ditempatkan pada Klaster yang berbeda. Adapun  $nC_2$  menggambarkan banyaknya pasangan tak berurutan yang dapat dibentuk dari suatu himpunan beranggotakan n elemen, sehingga mencerminkan total seluruh pasangan elemen yang menjadi objek perbandingan.

## 3. HASIL DAN PEMBAHASAN

Setelah data dikumpulkan maka dapat dilakukan tahapan data preprocessing di mana tahapan ini merupakan langkah awal dalam pengolahan data. Di tahap ini, data yang akan diproses ditujukan untuk menghindari adanya gangguan dari data yang tidak relevan atau data yang tidak konsisten [13]. Dataset tersebut, dilakukan proses *pre-processing* dimulai dengan proses seleksi data yaitu memilih variabel yang digunakan dalam proses Klaster, variabel yang dipilih berupa data bertipe numerik yaitu variabel Total Kejahatan, Jumlah Penduduk, Luas Wilayah dan Kepadatan Penduduk. Untuk memastikan kembali kebersihan data fungsi “*is.na()*” digunakan untuk membersihkan data yang hilang dan

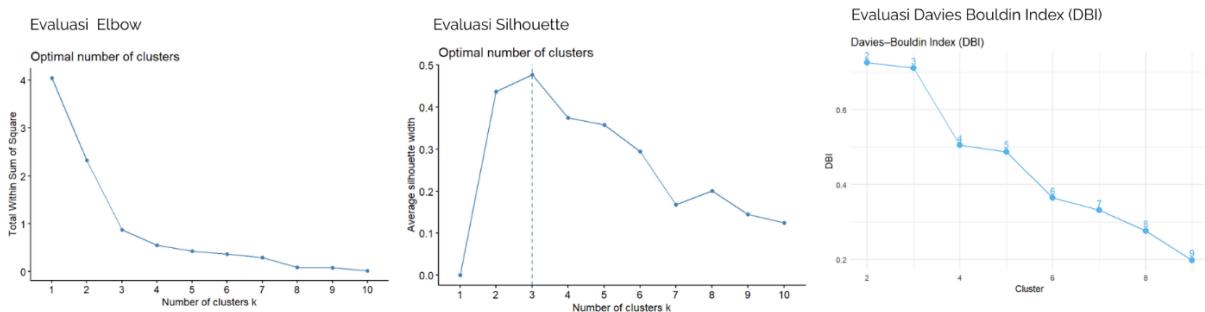
“duplicated()” untuk memeriksa duplikasi data. Berdasarkan hasil pengecekan tersebut, tidak ditemukan adanya *missing value* maupun data duplikat pada dataset, sehingga tidak diperlukan tindakan imputasi atau penghapusan data. Sementara itu untuk menangani perbedaan antar variabel, penelitian ini menerapkan metode normalisasi min-max dengan menggunakan fungsi  $(x - \min(x)) / (\max(x) - \min(x))$  yang dilakukan menggunakan Rstudio.

Setelah proses *pre processing* selesai di lakukan maka proses k means Klastering bisa dilakukan numun sebelum itu langkah pertama yang harus di lakukan adalah menentukan jumlah Klaster optimal menggunakan metode Elbow

### 3.1. Evaluasi Jumlah Klaster Optimal

Tahap awal sebelum menerapkan analisis klaster dengan algoritma K-Means adalah menentukan jumlah klaster. Dalam penelitian ini, jumlah klaster optimal ditentukan menggunakan Elbow *Method*, Silhouette *Method*, dan Davies Bouldin Index (DBI) untuk memastikan validitas klaster. Grafik Elbow dan Silhouette dibuat dengan fungsi “fviz\_nbclust()” dari paket “factoextra”, sedangkan grafik DBI dibuat manual menggunakan “ggplot2” karena metode ini tidak didukung “fviz\_nbclust()” di RStudio.

#### 1. Jumlah Klaster Optimal Tahun 2022

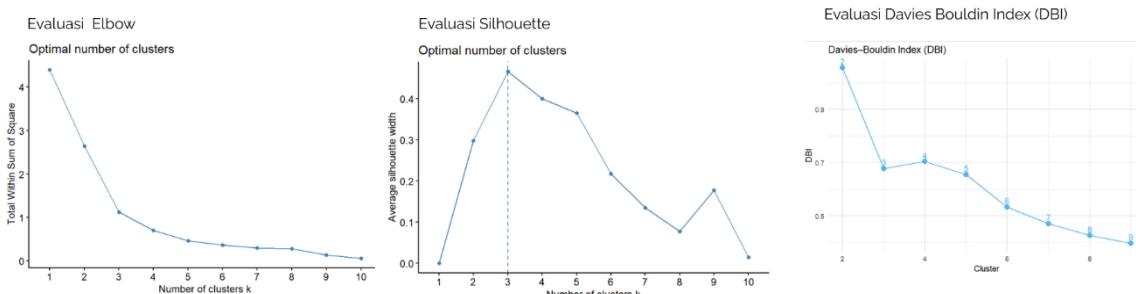


Gambar 2. Grafik Klaster Optimal Berdasarkan Elbow, Silhouette dan DBI Tahun 2022

Tabel 1. Nilai K Optimal Elbow, Silhouette, dan Davies Bouldin Index Tahun 2022

Klaster	Elbow	Silhouette	Davies Bouldin Index
2	1.874	0.436	0.726
3	<b>0.872</b>	<b>0.476</b>	0.711
4	0.559	0.374	0.505
5	0.432	0.310	0.487
6	0.306	0.238	0.364
7	0.181	0.230	0.331
8	0.091	0.211	0.276
9	0.028	0.181	<b>0.197</b>

#### 2. Jumlah Klaster Optimal Tahun 2023

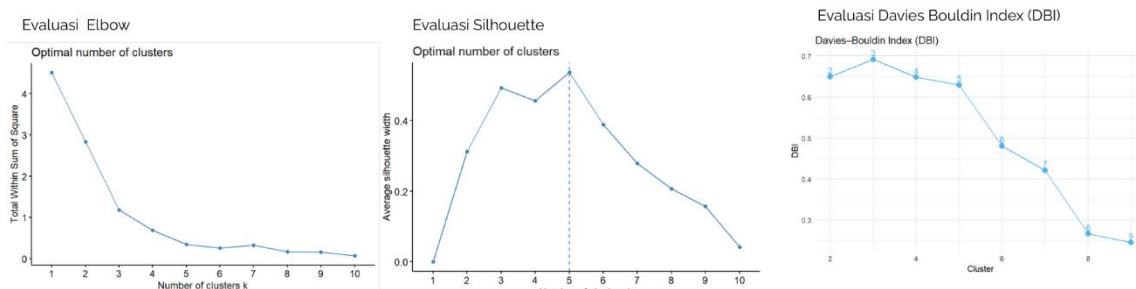


Gambar 3. Grafik Klaster Optimal Berdasarkan Elbow, Silhouette dan DBI Tahun 2023

Tabel 2. Nilai K Optimal Elbow, Silhouette, dan Davies Bouldin Index Tahun 2023

Klaster	Elbow	Silhouette	Davies Bouldin Index
2	2.614	0.298	1.0571
3	<b>1.124</b>	<b>0.466</b>	0.6777
4	0.704	0.399	0.7041
5	0.457	0.365	0.6549
6	0.315	0.329	0.5331
7	0.213	0.250	0.4707
8	0.134	0.192	0.4260
9	0.072	0.160	<b>0.3963</b>

### 3. Elbow Method Tahun 2024



Gambar 4. Grafik Klaster Optimal Berdasarkan Elbow, Silhouette dan DBI Tahun 2024

Tabel 3. Nilai K Optimal Elbow, Silhouette, dan Davies Bouldin Index Tahun 2024

Klaster	Elbow	Silhouette	Davies Bouldin Index
2	2.688	0.312	0.6492
3	<b>1.182</b>	0.492	0.6919
4	0.693	0.456	0.6483
5	0.349	<b>0.536</b>	0.6297
6	0.255	0.443	0.4802
7	0.167	0.352	0.4209
8	0.092	0.341	0.2667
9	0.024	0.294	<b>0.2458</b>

### 4. Hasil Evaluasi K Optimal

Tabel 4. Hasil Evaluasi K Optimal Tahun 2022-2024

Tahun	Metode	Jumlah K Optimal
2022	1. Metode Elbow : 3 Klaster 2. Silhouette Coefficient : 3 Klaster 3. DBI : 9 Klaster	3 Klaster
2023	1. Metode Elbow : 3 Klaster 2. Silhouette Coefficient : 3 Klaster 3. DBI : 9 Klaster	3 Klaster
2024	1. Metode Elbow : 3 Klaster 2. Silhouette Coefficient : 5 Klaster 3. DBI : 9 Klaster	3 Klaster

Penentuan jumlah klaster optimal dalam penelitian ini dilakukan menggunakan tiga metode, yaitu Elbow, Silhouette, dan Davies–Bouldin Index (DBI), sebelum penerapan algoritma K-Means. Metode

Elbow secara konsisten menunjukkan titik siku pada  $K = 3$ . Metode Silhouette juga memberikan nilai koefisien tertinggi pada  $K = 3$  untuk sebagian besar tahun pengamatan, meskipun pada tahun 2024 nilai maksimum berpindah ke  $K = 5$  dengan peningkatan yang relatif tidak terlalu jauh dibandingkan  $K = 3$ . Sementara itu, DBI menghasilkan nilai indeks terkecil pada  $K = 9$ .

Pemilihan jumlah klaster dalam penelitian ini tidak hanya didasarkan pada satu indeks validasi, tetapi juga mempertimbangkan karakteristik data dan tujuan analisis klaster. Pada data dengan jumlah kecamatan yang terbatas, penggunaan  $K$  yang terlalu besar seperti  $K = 9$  berpotensi menghasilkan klaster dengan anggota sangat sedikit sehingga kurang informatif untuk analisis kelompok wilayah. Dengan mempertimbangkan hasil metode Elbow dan Silhouette yang sama-sama mendukung penggunaan jumlah klaster yang relatif sedikit serta kemudahan interpretasi hasil klaster, penelitian ini menetapkan  $K = 3$  sebagai jumlah klaster optimal yang digunakan dalam proses klasterisasi dengan algoritma K-Means.

### 3.2. K-Means Klastering

Proses klasterisasi data dengan algoritma K-Means dilakukan menggunakan fungsi `kmeans()` dari paket `stats` di RStudio. Sebelum menjalankan klasterisasi, jumlah klaster ditetapkan sebanyak tiga, merujuk pada hasil penentuan  $K$  optimal melalui hasil evaluasi penentuan nilai  $K$ . Untuk memastikan hasil yang lebih stabil dan dapat direplikasi, digunakan beberapa parameter penting, yaitu `set.seed(123)` untuk menjamin *reproducibility*, `nstart = 25` sebagai jumlah inisialisasi awal pusat klaster sehingga algoritma memilih solusi dengan nilai within-klaster sum of squares terbaik, serta `iter.max = 100` sebagai batas maksimum iterasi dalam proses konvergensi K-Means. Nilai-nilai parameter tersebut kemudian digunakan dalam pemanggilan fungsi `kmeans` (`x, centers = 3, nstart = 25, iter.max = 100`) untuk memperoleh hasil klasterisasi yang selanjutnya dianalisis pada bagian berikutnya.

Tabel 5. K-Means Klastering Tahun 2022

Klaster	Total Anggota	Anggota Klaster
1	3	Kota Baru, Alam Barajo, Paal Merah
2	5	Jambi Selatan, Jambi Timur, Jelutung, Danau Sipin, Telanaipura
3	3	Pasar Jambi, Pelayangan, Danau Teluk

Tabel 6. K-Means Klastering Tahun 2023

Klaster	Total Anggota	Anggota Klaster
1	5	Kota Baru, Alam Barajo, Paal Merah, Telanaipura, Jambi Timur
2	4	Jambi Selatan, Jelutung, Pasar Jambi, Danau Sipin
3	2	Pelayangan, Danau Teluk

Tabel 7. K-Means Klastering 2024

Klaster	Total Anggota	Anggota Klaster
1	5	Kota Baru, Alam Barajo, Paal Merah, Telanaipura, Jambi Timur
2	4	Jambi Selatan, Jelutung, Pasar Jambi, Danau Sipin
3	2	Pelayangan, Danau Teluk

### 3.3 Evaluasi Hasil Klaster

Setelah proses klasterisasi K-Means dilakukan, kualitas klaster dinilai menggunakan Rand Index yang dihitung dengan fungsi `"randIndex()"` pada paket `flexclust` di RStudio. Penilaian ini dilakukan dengan membandingkan hasil klasterisasi pada berbagai jumlah klaster, yaitu  $K = 1$  hingga  $K = 10$ . Nilai Rand Index tertinggi kemudian dijadikan acuan untuk menentukan jumlah klaster dengan tingkat kesesuaian terbaik, yang digunakan sebagai dasar dalam melakukan interpretasi terhadap hasil klaster.

Tabel 8. Evaluasi Rand Index Tahun 2022

Nilai K	Rand Index
1	0.00000000

Nilai K	Rand Index
2	0.48210923
<b>3</b>	<b>1.00000000</b>
4	0.80968858
5	0.70270270
6	0.52449568
7	0.39195980
8	0.32098765
9	0.16846652
10	0.08637874

Tabel 9. Evaluasi Rand Index Tahun 2023

Nilai K	Rand Index
1	0.00000000
2	0.69863014
<b>3</b>	<b>1.00000000</b>
4	0.71698113
5	0.35176791
6	0.32610745
7	0.19038273
8	0.22845691
9	0.15557830
10	0.07949791

Tabel 10. Evaluasi Rand Index Tahun 2024

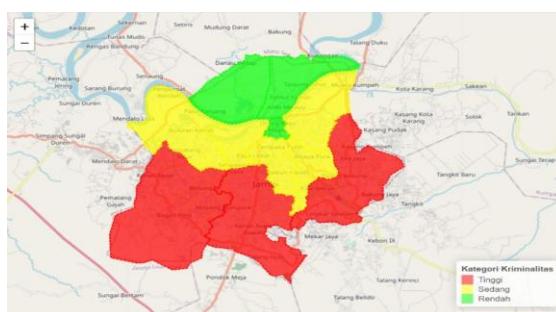
Nilai K	Rand Index
1	0.00000000
2	0.69863014
<b>3</b>	<b>1.00000000</b>
4	0.71698113
5	0.49168207
6	0.42978322
7	0.29833170
8	0.22845691
9	0.15557830
10	0.07949791

Hasil evaluasi menggunakan Rand Index pada data tahun 2022-2024 menunjukkan bahwa nilai tertinggi, yaitu 1.0000, diperoleh ketika jumlah klaster ditetapkan pada K = 3. Nilai Rand Index yang hampir mencapai angka 1 menandakan bahwa proses klasterisasi menghasilkan pengelompokan data dengan tingkat akurasi dan konsistensi yang sangat tinggi. Oleh karena itu, jumlah klaster K = 3 dianggap paling tepat untuk merepresentasikan pola distribusi wilayah rawan kriminalitas di Kota Jambi pada tahun 2022-2024.

### 3.4 Interpretasi Hasil K-Means Klustering

Untuk memahami distribusi klaster secara lebih jelas, hasil klasterisasi divisualisasikan dalam bentuk peta. Warna merah menunjukkan Klaster 1 dengan tingkat kriminalitas tinggi, warna kuning merepresentasikan Klaster 2 dengan tingkat kriminalitas sedang, dan warna hijau menunjukkan Klaster 3 dengan tingkat kriminalitas rendah. Pola warna ini diterapkan secara konsisten pada peta tahun 2022, 2023, dan 2024 untuk memudahkan analisis perubahan tingkat kejahatan antar wilayah dari waktu ke

waktu. Berikut disajikan interpretasi hasil klasterisasi kriminalitas di Kota Jambi menggunakan algoritma K-Means Klustering :



Gambar 5. Visualisasi Peta Klaster 2022

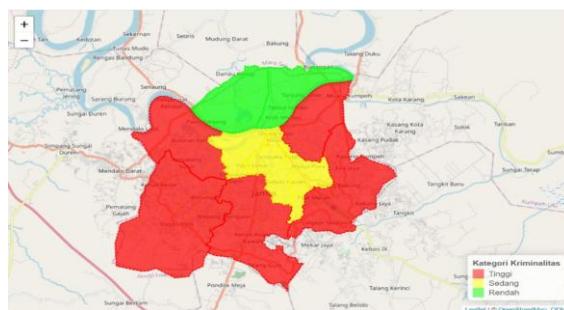
Tahun 2022, wilayah yang ditandai dengan warna merah yang menunjukkan tingkat kriminalitas tinggi dan berada pada Klaster 1 meliputi Kota Baru, Alam Barajo, dan Paal Merah. Area berwarna kuning yang merepresentasikan tingkat kriminalitas sedang dan termasuk dalam Klaster 2 mencakup Jambi Selatan, Jambi Timur, Jelutung, Danau Sipin, dan Telanaipura. Sementara itu, daerah yang berwarna hijau, yang mencerminkan tingkat kriminalitas rendah dan tergolong dalam Klaster 3, terdiri atas Pasar Jambi, Pelayangan, dan Danau Teluk.

Karakteristik setiap klaster dianalisis melalui nilai rata-rata variabel numerik yang digunakan dalam pemodelan, yaitu Total Kejahatan, Jumlah Penduduk, Luas Wilayah, dan Kepadatan Penduduk.

Tabel 11. Hasil *Centroid* K-Means Klustering Tahun 2022

Klaster	Total Kejahatan	Jumlah Penduduk	Luas Wilayah	Kepadatan Penduduk
1	0.8042929	0.88650542	0.8234772	0.32432144
2	0.4106061	0.43510819	0.2427278	0.60756683
3	0.1035354	0.01230404	0.2037826	0.09775328

- Klaster 1 menunjukkan tingkat kriminalitas tertinggi. Nilai *centroid* terendah berada pada variabel Kepadatan Penduduk 0.32432144, sedangkan nilai tertinggi terdapat pada Jumlah Penduduk 0.88650542. Klaster ini menjadi prioritas utama dengan ditandai oleh populasi besar dan area yang luas, sehingga aktivitas tinggi mendorong munculnya kasus kriminalitas.
- Klaster 2 berada pada tingkat kriminalitas menengah. Nilai *centroid* terendah terdapat pada Luas Wilayah 0.2427278, sedangkan nilai tertinggi berada pada Kepadatan Penduduk 0.60756683. Klaster ini termasuk prioritas menengah dengan kondisi wilayah area yang lebih kecil tetapi dihuni oleh jumlah penduduk yang cukup besar, sehingga menghasilkan kepadatan tinggi. Situasi tersebut menjadi faktor kerentanan sosial yang dapat memicu potensi kejahatan.
- Klaster 3 menunjukkan tingkat kriminalitas terendah. Nilai *centroid* terendah berada pada Jumlah Penduduk 0.01230404, sedangkan nilai tertinggi terdapat pada Luas Wilayah 0.2037826. Klaster ini menjadi prioritas rendah dengan wilayah dalam klaster ini memiliki area yang luas namun dihuni oleh sedikit penduduk, sehingga aktivitas sosial dan potensi kriminal sangat rendah.



Gambar 6. Visualisasi Peta Klaster 2023

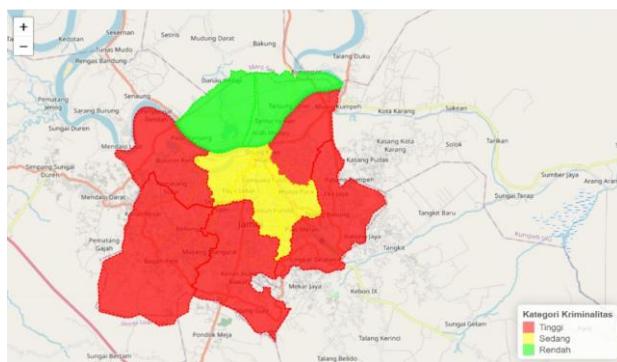
Tahun 2023, wilayah yang ditandai dengan warna merah yang menunjukkan tingkat kriminalitas tinggi dan termasuk dalam Klaster 1 cukup luas, mencakup Kota Baru, Alam Barajo, Paal Merah, Telanaipura, dan Jambi Timur. Area berwarna kuning yang merepresentasikan tingkat kriminalitas sedang dan berada pada Klaster 2 meliputi Jambi Selatan, Jelutung, Pasar Jambi, dan Danau Sipin. Sementara itu, wilayah berwarna hijau yang menunjukkan tingkat kriminalitas rendah dan tergolong dalam Klaster 3 mencakup Pelayangan dan Danau Teluk.

Karakteristik setiap klaster dianalisis melalui nilai rata-rata variabel numerik yang digunakan dalam pemodelan, yaitu Total Kejahatan, Jumlah Penduduk, Luas Wilayah, dan Kepadatan Penduduk.

Tabel 12. Hasil *Centroid* K-Means Klustering Tahun 2023

Klaster	Total Kejahatan	Jumlah Penduduk	Luas Wilayah	Kepadatan Penduduk
1	0.77936508	0.71067502	0.6585635	0.38208661
2	0.58134921	0.30613832	0.1271445	0.88593228
3	0.07936508	0.01906116	0.3022681	0.01869362

- Klaster 1 menunjukkan tingkat kriminalitas tertinggi. Nilai *centroid* terendah terdapat pada Kepadatan Penduduk (0.38208661), sedangkan indikator tertinggi berada pada Total Kejahatan (0.77936508). Klaster ini termasuk prioritas utama dengan kondisi wilayah dengan jumlah penduduk yang besar dan cakupan area yang luas, sehingga aktivitas masyarakat lebih tinggi. Meskipun tingkat kepadatan tidak sebesar klaster lainnya, kombinasi wilayah yang luas dan populasi besar tetap mendorong munculnya angka kriminalitas yang tinggi.
- Klaster 2 berada pada tingkat kriminalitas menengah. Nilai *centroid* terendah terdapat pada Luas Wilayah (0.1271445), sedangkan nilai tertinggi berada pada Kepadatan Penduduk (0.88593228). Klaster ini tergolong prioritas menengah dengan kondisi wilayah cakupan wilayah yang sempit tetapi dihuni oleh penduduk dalam jumlah besar. Konsentrasi penduduk yang tinggi menciptakan potensi kerawanan sosial dan risiko kejahatan meskipun jumlah kasusnya tidak setinggi Klaster 1.
- Klaster 3 menunjukkan tingkat kriminalitas terendah. Nilai *centroid* terendah terdapat pada Kepadatan Penduduk (0.01869362), sedangkan nilai tertinggi berada pada Luas Wilayah (0.3022681). Klaster ini tergolong prioritas rendah dicerminkan dengan wilayah yang memiliki area relatif luas tetapi dihuni oleh sedikit penduduk. Rendahnya intensitas interaksi sosial dan sebaran penduduk yang tidak padat berkontribusi pada minimnya kejadian kriminalitas yang terbatas



Gambar 7. Visualisasi Peta Klaster 2024

Tahun 2024, wilayah yang ditandai dengan warna merah yang menunjukkan tingkat kriminalitas tinggi dan termasuk dalam Klaster 1 cukup luas, mencakup Kota Baru, Alam Barajo, Paal Merah, Telanaipura, dan Jambi Timur. Area berwarna kuning yang merepresentasikan tingkat kriminalitas sedang dan berada pada Klaster 2 meliputi Jambi Selatan, Jelutung, Pasar Jambi, dan Danau Sipin. Sementara itu, wilayah berwarna hijau yang menunjukkan tingkat kriminalitas rendah dan tergolong dalam Klaster 3 mencakup Pelayangan dan Danau Teluk.

Karakteristik setiap klaster dianalisis melalui nilai rata-rata variabel numerik yang digunakan dalam pemodelan, yaitu Total Kejahatan, Jumlah Penduduk, Luas Wilayah, dan Kepadatan Penduduk.

Tabel 13. Hasil *Centroid* K-Means Klustering Tahun 2024

Klaster	Total Kejahatan	Jumlah Penduduk	Luas Wilayah	Kepadatan Penduduk
1	0.80810811	0.71696205	0.6585635	0.38817918
2	0.59459459	0.31054810	0.1271445	0.88762259
3	0.08783784	0.01649274	0.3022681	0.01861485

- Klaster 1 menunjukkan tingkat kriminalitas tertinggi. Nilai *centroid* terendah terdapat pada Kepadatan Penduduk (0.38817918), sedangkan nilai tertinggi berada pada Total Kejahatan (0.80810811). Klaster ini termasuk prioritas utama dengan kondisi dicirikan oleh jumlah penduduk yang besar dan cakupan area yang luas. Meskipun tingkat kepadatan penduduk tidak setinggi klaster lain, kombinasi populasi besar dan aktivitas wilayah yang tinggi berkontribusi pada tingginya angka kriminalitas.
- Klaster 2 berada pada tingkat kriminalitas menengah. Nilai *centroid* terendah terdapat pada Luas Wilayah (0.1271445), sedangkan nilai tertinggi berada pada Kepadatan Penduduk (0.88762259). Klaster ini digolongkan sebagai prioritas menengah dengan area yang lebih sempit namun dihuni oleh penduduk dalam jumlah besar. Konsentrasi penduduk yang tinggi menciptakan tekanan sosial dan meningkatkan potensi kerawanan.
- Klaster 3 menunjukkan tingkat kriminalitas terendah. Nilai *centroid* terendah terdapat pada Jumlah Penduduk (0.01649274), sedangkan nilai tertinggi berada pada Luas Wilayah (0.3022681). Klaster ini termasuk prioritas rendah cakupan wilayah yang relatif luas dengan jumlah penduduk yang sangat sedikit. Rendahnya aktivitas sosial dan interaksi masyarakat sejalan dengan minimnya kejadian kriminalitas.

#### 4. DISKUSI

Penelitian ini berfokus pada wilayah administratif Kota Jambi dengan menggunakan data kriminalitas dan indikator struktural yang bersumber dari instansi resmi. Dataset kriminalitas diperoleh dari Polresta Jambi berdasarkan wilayah Polsek jajaran Kota Jambi untuk periode Januari 2022 hingga Oktober 2024, sedangkan data demografis dan geospasial, seperti jumlah penduduk, luas wilayah, dan kepadatan penduduk, diperoleh dari Badan Pusat Statistik (BPS) Kota Jambi untuk rentang tahun 2022–2024. Fokus utama penelitian ini adalah menerapkan metode Klustering menggunakan algoritma K-

Means untuk mengelompokkan wilayah rawan kriminalitas di Kota Jambi berdasarkan kombinasi variabel kriminalitas, demografi, dan karakteristik wilayah.

Berdasarkan hasil klasterisasi dengan algoritma K-Means, penelitian ini mengidentifikasi tiga kelompok kecamatan di Kota Jambi dengan karakteristik yang berbeda. Analisis *centroid* menunjukkan bahwa Klaster 1 secara konsisten memiliki nilai tertinggi pada variabel Total Kejahatan, Jumlah Penduduk, dan Luas Wilayah, sedangkan nilai tertinggi untuk Kepadatan Penduduk berada pada Klaster 2. Sementara itu, Klaster 3 cenderung memiliki nilai terendah untuk Total Kejahatan, Jumlah Penduduk, dan Kepadatan Penduduk, tetapi Luas Wilayahnya relatif lebih besar dibandingkan Klaster 2. Temuan ini mengindikasikan bahwa kategori wilayah rawan kriminalitas tidak hanya terkait dengan kepadatan penduduk, tetapi juga berkaitan dengan karakteristik struktur wilayah tertentu di Kota Jambi.

Dibandingkan dari penelitian sebelumnya yang hanya mengandalkan total kasus kriminalitas, studi ini fokus pada Kota Jambi dan menambahkan indikator struktural seperti jumlah penduduk, luas wilayah, dan kepadatan penduduk dalam pembentukan klaster. Hasil klaster diinterpretasikan secara visual dalam peta untuk mempermudah pemahaman pola wilayah berdasarkan kategori rawan kriminalitas dan karakteristik struktural. Penelitian ini berkontribusi pada analisis distribusi wilayah rawan kriminalitas menggunakan K-Means Klustering dan mengisi celah literatur terkait penerapan teknik data mining untuk analisis kriminalitas di tingkat kecamatan di Indonesia.

Adapun keterbatasan yang perlu diperhatikan, ruang lingkup penelitian ini hanya mencakup wilayah Kota Jambi dengan periode data 2022-2024, serta bergantung pada ketersediaan data kriminalitas dari Polresta Jambi dan data demografis serta geospasial dari BPS Kota Jambi. Selain itu, variabel yang digunakan masih terbatas pada indikator kriminalitas dan variabel struktural dasar, sehingga belum memasukkan faktor sosial, ekonomi, maupun variabel lainnya yang berpotensi memengaruhi pola klaster kriminalitas. Oleh karena itu, penelitian selanjutnya diharapkan dapat memasukkan indikator tambahan, seperti variabel sosial ekonomi dan lingkungan, serta membandingkan hasil K-Means dengan teknik Klustering lainnya, seperti DBSCAN, Fuzzy C-Means, atau metode Klustering lain untuk memberikan perspektif yang lebih kaya dalam analisis pengelompokan wilayah kriminalitas di Kota Jambi.

## 5. KESIMPULAN

Berdasarkan penelitian yang dilakukan untuk mengelompokkan wilayah rawan kriminalitas di Kota Jambi menggunakan algoritma K-Means, diperoleh bahwa jumlah klaster optimal adalah  $K = 3$ . Penentuan ini didasarkan pada evaluasi jumlah klaster menggunakan Elbow *Method*, Silhouette Coefficient, dan Davies Bouldin Index (DBI) pada rentang  $K = 2-10$ , di mana Elbow dan Silhouette secara dominan mengarah pada  $K = 3$ , serta diperkuat oleh evaluasi konsistensi menggunakan Rand Index yang menunjukkan struktur tiga klaster yang stabil pada periode 2022–2024. Hasil klasterisasi diinterpretasikan ke dalam tiga kategori wilayah, yaitu Klaster 1 yang merepresentasikan wilayah dengan kategori kriminalitas tinggi, Klaster 2 untuk wilayah dengan kategori kriminalitas sedang, dan Klaster 3 untuk wilayah dengan kategori kriminalitas rendah. Dari tahun 2022 hingga 2024, kecamatan yang konsisten berada pada kategori kriminalitas tinggi adalah Kota Baru, Alam Barajo, dan Paal Merah. Kecamatan Jambi Selatan, Jelutung, dan Danau Sipin secara konsisten berada pada kategori kriminalitas sedang, sedangkan Pelayangan dan Danau Teluk tetap berada pada kategori kriminalitas rendah. Perubahan komposisi klaster terjadi pada Telanaipura dan Jambi Timur yang beralih dari kategori sedang menjadi tinggi sejak 2023 dan bertahan hingga 2024, serta Pasar Jambi yang naik dari kategori rendah pada 2022 menjadi kategori sedang sejak 2023 hingga akhir periode penelitian. Temuan ini memberikan pengetahuan dan wawasan mengenai pengelompokan wilayah rawan kriminalitas di Kota Jambi.

## DAFTAR PUSTAKA

- [1] Data Indonesia, “Data jumlah kejahatan di Indonesia pada 2023,” *Data Indonesia*, 2023. [Online]. Available: <https://dataindonesia.id/varia/detail/data-jumlah-kejahatan-di-indonesia-pada-2023> (accessed Sep. 5, 2025)
- [2] Badan Pusat Statistik Provinsi Jambi, *Provinsi Jambi dalam Angka 2024*, Jambi: Badan Pusat Statistik Provinsi Jambi, 2024. [Online]. Available: <https://jambi.bps.go.id>

- [3] L. Suriani, "Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Klastering," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 1, no. 2, p. 151, Jan. 2020, doi: 10.30865/json.v1i2.1955.
- [4] R. N. Fahmi, M. Jajuli, N. Sulistiyowati, and U. S. Karawang, "Analisis pemetaan tingkat kriminalitas di Kabupaten Karawang menggunakan algoritma K-Means / Mapping analysis of criminality level in Karawang using K-Means algorithm," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 4, no. 1, pp. 67–79, 2021. doi: 10.31539/intecom.v4i1.2413
- [5] Badan Pusat Statistik, *Statistik Kriminal 2024*, Jakarta: Badan Pusat Statistik, 2024. [Online]. Available: <https://www.bps.go.id/id/statistics-table/2/MTY2MSMy/statistik-kriminal.html>
- [6] L. Wulandari and B. O. Yogantara, "Algorithm Analysis of K-Means and Fuzzy C-Means for Klastering Countries Based on Economy and Health," *Faktor Exacta*, vol. 15, no. 2, Aug. 2022, doi: 10.30998/faktorexacta.v15i2.12106.
- [7] D. Yulyanti, "Klastering tingkat kejahatan kriminal menggunakan metode K-Means di wilayah Kabupaten Cirebon," *JATI: Jurnal Administrasi dan Teknologi Informasi*, vol. 7, no. 6, 2023, doi: 10.36040/jati.v7i6.8894.
- [8] Kolavennu, M. S. T. Saravanan, B. Nishanth, K. Nithin, D. Vishal, and U. G. Student, "Crime Rate Prediction and Analysis Using K-Means Algorithm," *International Journal of Research Publication and Reviews*, vol. 5, no. 5, pp. 8487–8490, May 2024. [Online]. Available: <https://www.ijrpr.com>
- [9] I. Kamila and U. Khairunnisa, "Perbandingan algoritma K-Means dan K-Medoids untuk pengelompokan transaksi bongkar muat di Provinsi Riau," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, pp. 119–125, 2019. doi: 10.24014/rmsi.v5i1.7381.
- [10] R. Hidayat, "Klastering menggunakan algoritma K-Means untuk mengelompokan wilayah rawan kejahatan di wilayah Kabupaten Solok," *Jurnal Ilmiah Mahasiswa Manajemen, Bisnis dan Akuntansi (JIMMBA)*, vol. 4, no. 5, pp. 646–654, 2022. doi: 10.32639/jimmbs.v4i5.169.
- [11] I. F. Ashari, E. D. Nugroho, R. Baraku, I. N. Yanda, and R. Liwardana, "Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta," *Jurnal Analisis dan Informatika Computasional (JAIC)*, vol. 7, no. 1, Jul. 2023. doi: 10.30871/jaic.v7i1.4947.
- [12] P.W.Rahayu, I.G.I.Sudipa, Suryani, A. Surachman, A. Ridwan, I. G. M. Darmawiguna, M. N. Sutoyo, I. Slamet, S. Harlina, and I. M. D. Maysanjaya, *Buku Ajar Data Mining*. Indonesia: PT Sonpedia Publishing Indonesia, 2024.
- [13] F. Alghifari and D. Juardi, "Penerapan Data Mining pada Penjualan Makanan dan Minuman Menggunakan Metode Algoritma Naïve Bayes: Studi Kasus: Makan Barbeque Sepuasnya," *Jurnal Ilmiah Informatika*, vol. 9, no. 2, pp. 75–81, 2021. doi: 10.33884/jif.v9i02.3755.
- [14] E. Rossalina Fitria, F. Rozci, and E. R. Fitria, "Penerapan Metode Regresi Least Absolute Shrinkage and Selection Operator (LASSO) dan Regresi Linier untuk Memprediksi Tingkat Kemiskinan di Indonesia," *Jurnal Informatika dan Statistik Aplikasi (JISA)*, vol. 22, no. 2, 2022, doi: 10.30742/jisa22220222620.
- [15] I. Permana and F. N. Salisah, "Pengaruh normalisasi data terhadap performa hasil klasifikasi algoritma Backpropagation / The Effect of Data Normalization on the Performance of the Classification Results of the Backpropagation Algorithm," *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, vol. 2, no. 1, pp. 67–72, Mar. 2022. doi: 10.57152/ijirse.v2i1.311.
- [16] M. Qusyairi, Zul Hidayatullah, and Arnila Sandi, "Penerapan K-Means Klastering Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode Elbow," *Infotek: Jurnal Informatika dan Teknologi*, vol. 7, no. 2, pp. 500–510, Jul. 2024, doi: 10.29408/jit.v7i2.26375.
- [17] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Klastering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix : Jurnal Manajemen Teknologi dan Informatika*, vol. 9, no. 3, pp. 102–109, Nov. 2019, doi: 10.31940/matrix.v9i3.1662.

- [18] A. Tira and B. Nurina Sari, "Penerapan Silhouette Coefficient, Elbow Method dan Gap Statistics untuk Penentuan Klaster Optimum dalam Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Kebahagiaan," *Jurnal Ilmiah Wahana Pendidikan*, vol. 9, no. 17, pp. 76–86, 2023. doi: 10.5281/zenodo.8282638.
- [19] N. Nugroho and F. D. Adhinata, "Penggunaan Metode K-Means dan K-Means++ Sebagai Klastering Data Covid-19 di Pulau Jawa," *Teknika*, vol. 11, no. 3, pp. 170–179, 2022. doi: 10.34148/teknika.v11i3.502.
- [20] P. Abela, R. Buaton, M. Simanjuntak, S. Kaputama, and I. Binjai, "Penerapan Metode Klastering Pada Kasus Kecelakaan Kerja," vol. 2, no. 5, pp. 3031–8904, 2024, doi: 10.61132/merkurius.v2i4.
- [21] D. Gultom *et al.*, "Penerapan Algoritma K-Means untuk Mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar," *Jurnal Teknologi Informasi*, vol. 4, no. 1, pp. 146–151, Jun. 2020. doi: 10.36294/jurti.v4i1.1263.
- [22] M. N. Zhafar, K. Usman, and F. Akhyar, "Penerapan Metode Klastering dengan Algoritma K-Means untuk Analisa Persebaran Varian COVID-19 (Studi Kasus Kelurahan Antapani Kidul)," *e-Proceeding of Engineering*, vol. 10, no. 5, Okt. 2023. [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/21215>
- [23] M. Mutmainah and W. Yustanti, "Studi Komparasi Local Outlier Factor (LOF) dan Isolation Forest (IF) pada Analisis Anomali Kinerja Dosen," *JINACS: Journal of Informatics and Computer Science*, vol. 6, no. 2, pp. 532–540, 2024. doi: 10.26740/jinacs.v6n02.p532-540.