

Analisis Pengelompokan Jenis Kejahatan di Sumatera Utara Berdasarkan Pola Kejadian Tahunan Menggunakan Algoritma K-Means Clustering

Michael Dolly Sianturi^{a*}, Mery Christyn Lubis^a, Sandi Dwi Payana^a, Alya Nabilla Putri^a, Hotnauli Roni Arta Panjaitan^a

^a Program Studi Statistika, Universitas Negeri Medan, Indonesia

INFO ARTIKEL

Riwayat Artikel:

Received : 05-06-2025

Revised : 15-06-2025

Accepted : 18-06-2025

Keywords: Data Mining, K-Means Clustering, North Sumatra

Kata Kunci: Data Mining, K-Means Clustering, Sumatera Utara

Corresponding Author:
michaeldsianturi@gmail.com*

DOI: <https://doi.org/10.62335>

ABSTRACT

This study aims to cluster various types of crimes occurring in North Sumatra Province based on annual incident patterns using the K-Means clustering algorithm. The data utilized are secondary data obtained from the Central Bureau of Statistics (BPS), comprising 34 types of crimes recorded from 2007 to 2021. Prior to clustering, data were normalized using the Z-score standardization method to ensure uniform scaling across variables. The optimal number of clusters was determined using the Elbow Method and Silhouette Plot. The analysis results indicate that four clusters ($k = 4$) provide the best balance between model complexity and clustering quality. Each cluster reveals distinct crime patterns in terms of frequency and trend stability over the years. The clustering results offer a clearer understanding of crime characteristics in the region and can serve as a foundation for more targeted policy-making, such as resource allocation for law enforcement and data-driven crime prevention strategies. This study demonstrates that data mining approaches, particularly the K-Means algorithm, can significantly contribute to a systematic and comprehensive understanding of crime patterns.

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan jenis-jenis kejahatan yang terjadi di Provinsi Sumatera Utara berdasarkan pola kejadian tahunan menggunakan algoritma K-Means clustering. Data yang digunakan merupakan data sekunder dari Badan Pusat Statistik (BPS) yang mencakup 34 jenis kejahatan selama periode 2007 hingga 2021. Sebelum dilakukan proses klusterisasi, data terlebih dahulu dinormalisasi menggunakan metode standarisasi Z-score untuk menyamakan skala antar variabel. Selanjutnya, penentuan jumlah kluster optimal dilakukan dengan metode Elbow

dan Silhouette Plot. Hasil analisis menunjukkan bahwa jumlah kluster optimal adalah empat ($k = 4$), karena memberikan keseimbangan terbaik antara kompleksitas model dan kualitas pemisahan kluster. Masing-masing kluster menunjukkan pola kejahatan yang berbeda, baik dari segi frekuensi maupun kestabilan tren tahunan. Klasterisasi ini memberikan gambaran yang lebih jelas mengenai karakteristik kejahatan di wilayah tersebut dan dapat dijadikan dasar dalam pengambilan kebijakan yang lebih tepat sasaran, seperti alokasi sumber daya penegakan hukum dan strategi pencegahan kejahatan yang berbasis data. Penelitian ini menunjukkan bahwa pendekatan data mining dengan algoritma K-Means dapat memberikan kontribusi signifikan dalam pemahaman pola kejahatan secara komprehensif dan sistematis.

PENDAHULUAN

Tindak kejahatan merupakan salah satu permasalahan sosial yang berdampak luas terhadap rasa aman masyarakat. Kejahatan seperti pencurian, perampokan, pembunuhan, hingga kekerasan seksual tidak hanya merugikan korban secara langsung, tetapi juga menimbulkan keresahan publik. Untuk mengidentifikasi pola dan kecenderungan kejahatan, dibutuhkan pendekatan analisis data yang tepat guna membantu pengambilan kebijakan yang lebih akurat dan efektif.

Dalam beberapa tahun terakhir, metode data mining, khususnya clustering, telah banyak digunakan untuk menganalisis pola kejahatan. Salah satu algoritma clustering yang paling populer adalah K-Means Clustering, karena kemampuannya dalam mengelompokkan data dalam jumlah besar secara cepat dan efisien. Penelitian oleh Dewi et al. (2021) menunjukkan bahwa algoritma K-Means mampu mengelompokkan data kriminalitas berdasarkan wilayah secara akurat dan efisien, menghasilkan kluster yang dapat digunakan untuk analisis lebih lanjut.

Selain itu, studi oleh Prasetyo dan Nugroho (2020) mengaplikasikan K-Means dalam pengelompokan data kriminal berdasarkan waktu dan jenis kejahatan di Jakarta, yang mampu mengidentifikasi daerah rawan kejahatan secara spasial-temporal. Penelitian lainnya oleh Rahmadani et al. (2019) menggunakan metode ini dalam visualisasi tingkat kejahatan di kota Medan dan memberikan masukan berbasis data untuk aparat keamanan. Sementara itu, Putra dan Syahputra (2020) menggunakan kombinasi K-Means dengan GIS (Geographic Information System) untuk menganalisis sebaran kejahatan di wilayah tertentu, menunjukkan integrasi spasial yang memperkuat interpretasi hasil klustering.

Namun demikian, sebagian besar penelitian terdahulu hanya berfokus pada pendekatan spasial atau temporal secara terpisah, dan belum mengkaji secara khusus

bagaimana pola kejadian tahunan jenis-jenis kejahatan dapat digunakan untuk mengelompokkan kasus kejahatan secara holistik, terutama dalam konteks wilayah Sumatera Utara. Padahal, dinamika sosial, ekonomi, dan demografi di provinsi ini cukup kompleks dan dapat mempengaruhi persebaran dan jenis kejahatan yang terjadi. Hingga saat ini, belum banyak penelitian yang secara spesifik menyoroti pengelompokan kejahatan berdasarkan tren tahunan di Sumatera Utara dengan pendekatan K-Means Clustering.

Beberapa peneliti berfokus pada analisis spasial kejahatan atau pengelompokan berbasis waktu sesaat. Ada sedikit penelitian yang berkaitan dengan pengelompokan jenis kejahatan berdasarkan pola tahunan di wilayah Sumatera Utara. Oleh karena itu, penelitian ini bermaksud untuk mengeksplorasi dan menganalisis pengelompokan jenis-jenis kejahatan di Sumatera Utara menggunakan algoritma K-Means berdasarkan pola kejadian tahunan.

Tujuan penelitian ini adalah untuk mengidentifikasi tren jangka panjang, menemukan asosiasi antar kategori kejahatan, serta memberikan informasi berbasis data yang dapat digunakan untuk mendukung kebijakan keamanan yang lebih adaptif dan spesifik terhadap karakteristik wilayah Sumatera Utara.

Analisis pola kejahatan merupakan bagian penting dalam ilmu kriminologi modern yang berupaya memahami dinamika kriminalitas secara sistematis. Seiring berkembangnya teknologi informasi, pendekatan tradisional dalam memahami data kejahatan mulai digantikan oleh metode data mining, khususnya klasterisasi, yang mampu menyaring informasi tersembunyi dari kumpulan data dalam skala besar dan heterogen. Salah satu algoritma klaster yang paling banyak digunakan adalah K-Means Clustering karena kesederhanaan, kecepatan, dan efektivitasnya dalam mengelompokkan data numerik.

Beberapa penelitian sebelumnya telah menunjukkan efektivitas metode ini dalam konteks analisis kriminalitas. Penelitian oleh Dewi et al. (2019) menunjukkan bahwa algoritma K-Means dapat mengelompokkan data kriminalitas berdasarkan wilayah secara akurat dan efisien. Mereka menyoroti bahwa hasil klaster yang terbentuk dapat digunakan untuk menilai wilayah rawan kejahatan berdasarkan jumlah dan jenis kasus kriminal yang dilaporkan.

Lebih lanjut, Gultom et al. (2020) mengaplikasikan metode K-Means untuk mengklasifikasikan tingkat kejahatan di Kota Pematangsiantar. Studi ini menggabungkan data statistik kejahatan dengan atribut lokasi geografis, dan hasilnya memberikan wawasan baru bagi pemerintah lokal dalam mengidentifikasi kawasan prioritas untuk intervensi keamanan.

Ini menegaskan bahwa metode K-Means tidak hanya efisien secara komputasi,

tetapi juga relevan secara praktis dalam perumusan kebijakan berbasis bukti (evidence-based policy).

Selaras dengan itu, Septiani et al. (2020) menerapkan algoritma ini dalam konteks Kecamatan Medan Polonia dan menemukan bahwa pengelompokan jenis kejahatan berdasarkan waktu dan lokasi sangat membantu dalam menentukan pola temporal kriminalitas. Mereka juga menambahkan bahwa penggunaan metode ini dapat membantu kepolisian dalam perencanaan patroli berbasis data.

Sebagai pembaruan terhadap pendekatan spasial-temporal, Kharisma dan Jananto (2024) mengembangkan model analisis kejahatan yang mengintegrasikan data mining dengan sistem informasi geografis (GIS). Dalam studi kasus Polrestabes Semarang, mereka mampu memetakan dan mengelompokkan kejahatan berdasarkan lokasi dan waktu kejadian. Integrasi ini menunjukkan bahwa hasil klusterisasi tidak hanya membantu dalam klasifikasi, tetapi juga memberikan visualisasi yang mudah dipahami oleh pemangku kebijakan keamanan.

Sementara itu, Yuvrico (2023) melakukan penelitian dalam skala nasional yang mengelompokkan tingkat kejahatan antar provinsi menggunakan K-Means. Penelitian ini mengidentifikasi bahwa wilayah-wilayah tertentu memiliki karakteristik kejahatan yang unik dan memerlukan pendekatan kebijakan yang berbeda. Temuan ini memperkuat pentingnya pendekatan klusterisasi sebagai alat bantu dalam menyusun strategi keamanan nasional yang berbasis bukti.

Penelitian-penelitian tersebut menggarisbawahi kekuatan K-Means dalam menyederhanakan kompleksitas data kejahatan dan mengungkap pola tersembunyi yang tidak mudah terlihat melalui metode konvensional. Namun, keterbatasan metode ini juga perlu dicermati. Seperti dijelaskan oleh Hartanti (2020), penentuan jumlah kluster (k) sangat mempengaruhi hasil akhir, sehingga penggunaan metode Elbow atau Silhouette Plot sangat direkomendasikan untuk memastikan validitas pemisahan kelompok.

Elbow Plot digunakan untuk melihat titik di mana tambahan jumlah kluster tidak lagi memberikan pengurangan yang signifikan pada within-cluster sum of squares (WSS), sedangkan Silhouette Plot mengevaluasi seberapa baik objek dikelompokkan dalam kluster masing-masing.

Meskipun metode ini sangat populer, sebagian besar studi sebelumnya masih berfokus pada pendekatan spasial (berdasarkan wilayah) atau temporal sesaat (berdasarkan waktu bulanan atau mingguan). Sangat sedikit yang mengeksplorasi klusterisasi berdasarkan pola tahunan yang berlangsung dalam jangka panjang. Hal ini menunjukkan adanya celah penelitian yang penting, khususnya dalam konteks wilayah seperti Sumatera Utara yang memiliki dinamika sosial dan kriminalitas yang unik. Oleh karena itu, pengelompokan jenis kejahatan berdasarkan pola kejadian tahunan menjadi

pendekatan yang lebih strategis dan menyeluruh dalam memahami tren jangka panjang dan variasi antar jenis kejahatan.

Dalam konteks ini, penelitian yang menggunakan algoritma K-Means terhadap data tahunan berbagai jenis kejahatan dapat memberikan pemetaan risiko yang lebih akurat, pengidentifikasian kelompok kejahatan dengan karakteristik mirip, serta landasan untuk formulasi kebijakan keamanan yang lebih spesifik dan adaptif. Kajian ini secara akademik memperkuat kontribusi metode data mining sebagai pendekatan kuantitatif dalam analisis sosial, khususnya pada domain keamanan dan penanggulangan kriminalitas.

METODE

Penelitian ini menggunakan algoritma *K-Means Clustering*, untuk mengelompokkan jenis-jenis kejahatan berdasarkan pola kejadian tahunan di Provinsi Sumatera Utara selama periode 2007 hingga 2021. Tujuan dari pendekatan ini adalah untuk mengidentifikasi kelompok kejahatan yang menunjukkan pola temporal serupa, yang dapat memberikan wawasan dalam penyusunan strategi penanggulangan kriminalitas secara lebih terfokus.

Data dan Sumber

Data yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) melalui situs resmi (<https://sumut.bps.go.id/id>). Dataset terdiri dari total kejadian berbagai jenis kejahatan per tahun, termasuk di antaranya pencurian, penganiayaan, perkosaan, penipuan, dan lain-lain. Data ini disusun dalam format deret waktu tahunan, menghasilkan matriks berdimensi $n \times t$, dengan n mewakili jumlah jenis kejahatan dan $t = 15$ mewakili jumlah tahun dari 2007 hingga 2021.

Pra-pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk memastikan kualitas data yang layak dianalisis. Proses ini mencakup:

- Penghapusan nilai kosong, data duplikat, dan inkonsistensi penulisan.
- Standarisasi data menggunakan metode Z-Score Normalization, di mana setiap nilai dikurangi dengan rata-rata variabelnya, lalu dibagi dengan standar deviasi variabel tersebut.

Pendekatan ini digunakan agar setiap variabel memiliki skala yang setara (rata-rata 0 dan standar deviasi 1), sehingga algoritma K-Means tidak bias terhadap variabel yang memiliki skala lebih besar. Hal ini penting untuk menghasilkan klasterisasi yang lebih akurat dan adil antar variabel.

Rumus menghitung nilai Z-Score (Rahma Novita, Heru Pramanda, Kamalia, Muhajjir, Dedek Ariansyah), sebagai berikut:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Dimana:

X: nilai data asli

μ : rata-rata dari fitur

σ : standar deviasi dari fitur

z: nilai data setelah dinormalisasikan

Penentuan Jumlah Kluster Optimal

Penentuan jumlah kluster optimal dilakukan menggunakan metode Elbow. Dalam metode ini, algoritma K-Means dijalankan untuk berbagai nilai K (jumlah kluster), misalnya dari $K = 2$ hingga $K = 10$. Untuk setiap nilai K , dihitung nilai Sum of Squared Errors (SSE). Nilai-nilai SSE kemudian diplot terhadap K untuk membentuk Elbow Plot. Titik tekuk (elbow point) pada grafik menunjukkan nilai K optimal, yaitu ketika penurunan SSE mulai melambat secara signifikan. Rumus untuk menghitung nilai SEE (Ninik Tri Hartanti), sebagai berikut.

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} \|x_i - C_k\|_2^2 \quad (2)$$

Dimana:

K : jumlah kluster yang terbentuk

S_k : himpunan titik data yang termasuk dalam kluster ke- k

x_i : titik data ke- i yang berada di kluster S_k

C_k : centroid (titik pusat) dari kluster ke- k

Penerapan Algoritma K-Means

Setelah jumlah kluster optimal diperoleh, dilakukan proses klusterisasi menggunakan algoritma K-Means. Prosedur ini mencakup:

- Inisialisasi centroid secara acak sebanyak jumlah kluster yang ditentukan.
- Pengelompokan data berdasarkan jarak Euclidean antara masing-masing titik data dan centroid terdekat. Rumus jarak Euclidean (Berlian Juliartha Martin Putra, Dwi Ariani Finda Yuniarti) adalah sebagai berikut.

$$d(x_i, C_k) = \sqrt{\sum_{j=1}^m (x_{ij} - C_{kj})^2} \quad (3)$$

Dimana:

x_i : titik data ke-i

C_k : centroid kluster ke-k

- Pembaruan posisi centroid berdasarkan rata-rata anggota dalam tiap kluster. Rumus pembaruan centroid untuk kluster ke-k (Berlian Juliartha Martin Putra, Dwi Ariani Finda Yuniarti), sebagai berikut.

$$C_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i \quad (4)$$

Dimana:

C_k : himpunan titik data pada kluster ke-k

$|C_k|$: jumlah titik data dalam kluster ke-k

- Iterasi terus dilakukan hingga posisi centroid konvergen, yaitu perubahan menjadi sangat kecil atau tidak terjadi lagi.

Visualisasi Data

Untuk mendukung interpretasi hasil, dilakukan visualisasi data dalam bentuk:

- Elbow Plot: Menunjukkan proses identifikasi jumlah kluster optimal.
- Cluster Plot: Menggambarkan hasil klusterisasi dalam ruang dua dimensi hasil reduksi menggunakan PCA
- Scatter Plot: Memvisualisasikan pola tahunan masing-masing jenis kejahatan dalam kluster, guna memahami karakteristik temporal tiap kelompok.

ANALISIS DAN PEMBAHASAN

Dalam penelitian ini, data yang digunakan merupakan data kejadian tindak kejahatan yang terjadi di wilayah Provinsi Sumatera Utara. Data tersebut mencakup berbagai jenis kejahatan yang tercatat di masing-masing kabupaten/kota dan dihimpun berdasarkan kejadian tahunan. Rentang waktu data yang dianalisis meliputi dimulai dari tahun 2007 hingga tahun 2021. Data ini diperoleh dari Badan Pusat Statistik (BPS). Informasi tersebut disusun dalam bentuk tabel yang memuat jumlah kejadian dari masing-masing jenis kejahatan di setiap kabupaten/kota, sebagaimana ditampilkan pada Tabel 1 berikut.

Tabel 1. Kejahatan pada Tahun 2007-2021

Jenis Kejahatan	2007	2008	2009	2010	...	2021
Kejahatan Terhadap Kepala Negara	0	0	0	0	...	0
Kejahatan Terhadap Ketertiban Umum	9	0	5	0	...	0
Pembakaran	84	82	92	103	...	107
Kebakaran	201	224	229	252	...	56
Penyuapan	1	2	1	0	...	26
Kejahatan Mata Uang	32	21	24	0	...	8
Kejahatan Materai dan Merk	16	16	16	8	...	16
Melanggar Kesopanan, perzinahan	905	938	1023	1193	...	330
Perkosaan	233	207	214	207	...	203
Perjudian	936	1557	2908	4051	...	514
Penculikan	30	18	27	27	...	28
Pembunuhan	134	138	105	130	...	96
Penganiayaan Berat	4748	4304	3734	4217	...	2260
Penganiayaan Ringan	1682	1706	2476	2619	...	1176
Pencurian Ringan	1833	1738	1893	2030	...	28
Pencurian dengan Kekerasan	795	845	928	1001	...	532
Pencurian dengan Pemberatan	6138	6006	6219	7324	...	4738
Penghinaan	460	501	595	701	...	457
Pemerasan	691	493	407	379	...	715
Penggelapan	1592	1517	1697	2162	...	2531
Penipuan	1809	1392	1707	2317	...	2736
Pengrusakan	960	1080	1003	1217	...	723
Penadahan	0	13	19	9	...	4

Jenis Kejahatan	2007	2008	2009	2010	...	2021
Kejahatan Ekonomi	0	0	0	0	...	3
Pencurian Kendaraan Bermotor	2458	2342	3046	4999	...	2620
Melarikan Wanita dibawah Umur	460	208	200	253	...	0
Kejahatan Narkotik	0	2666	2802	2718	...	5950
Penyelundupan	1002	72	46	17	...	4
Korupsi	19	12	20	15	...	12
Penyalahgunaan Senjata Api	13	4	7	14	...	15
Kejahatan Surat-surat Sejenis	237	230	220	278	...	322
Sengketa Tanah	19	298	31	6	...	375
Ilegal Logging	208	148	172	169	...	18

Data di atas selanjutnya akan dilakukan pengolahan data dengan proses clustering menggunakan algoritma K-Means sehingga didapatkan hasil pengelompokan data yang diinginkan. Adapun langkah clustering dengan algoritma K-Means adalah sebagai berikut:

1) Standarisasi Z-score

2007	2008	2009	2010	2011	2012	2013	2014
-0.6112235	-0.6462279	-0.6692679	-0.6540909	-0.6540708	-0.6259916	-0.6369515	-0.6336970
-0.6046711	-0.6462279	-0.6658025	-0.6540909	-0.6520290	-0.6244632	-0.6369515	-0.6319240
-0.5500681	-0.5854629	-0.6055041	-0.5962178	-0.5887318	-0.5623087	-0.5556231	-0.5533207
-0.4648873	-0.4802359	-0.5105515	-0.5124984	-0.5203300	-0.5016825	-0.5279821	-0.5367727
-0.6104955	-0.6447458	-0.6685748	-0.6540909	-0.6540708	-0.6259916	-0.6369515	-0.6325150
-0.5879262	-0.6306661	-0.6526339	-0.6540909	-0.6367152	-0.6259916	-0.6305728	-0.6277869
2015	2016	2017	2018	2019	2020	2021	
-0.6841332	-0.6918335	-0.6918335	-0.6015972	-0.6262897	-0.6065499	-0.5619142	
-0.6841332	-0.6906104	-0.6906104	-0.6042036	-0.6269407	-0.6065499	-0.5619142	
-0.5993955	-0.5860376	-0.5860376	-0.5764012	-0.5553231	-0.5510140	-0.4873317	
-0.5816183	-0.5438416	-0.5438416	-0.5129768	-0.5247228	-0.4975870	-0.5228804	
-0.6835407	-0.6906104	-0.6906104	-0.6024660	-0.6262897	-0.6065499	-0.5437914	
-0.6758372	-0.6832720	-0.6832720	-0.5894336	-0.6139193	-0.6009260	-0.5563380	

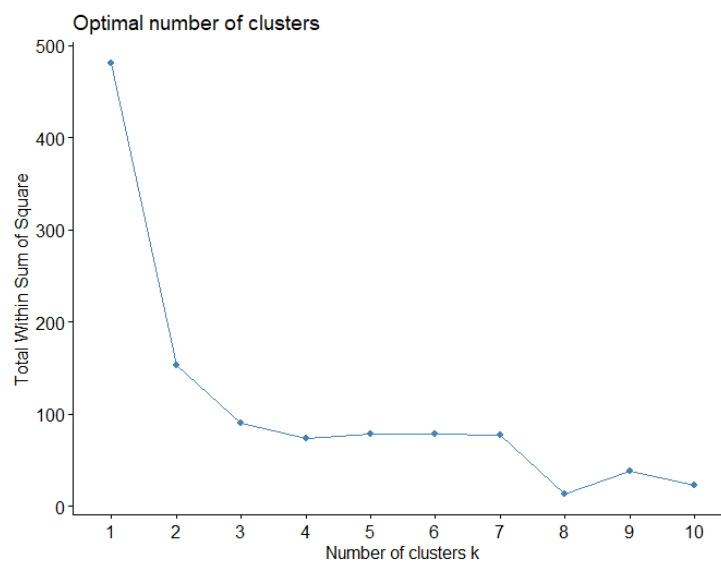
Gambar 1. Hasil dari Standarisasi Z-score

Berdasarkan hasil normalisasi data menggunakan metode standarisasi Z-score, seluruh nilai pada masing-masing variabel (tahun 2007 hingga 2021) telah ditransformasikan sedemikian rupa sehingga memiliki rerata (mean) sebesar nol dan simpangan baku (standar deviasi) sebesar satu. Proses ini menghasilkan distribusi nilai yang berskala seragam, memungkinkan perbandingan antar fitur tanpa terpengaruh oleh skala aslinya. Nilai-nilai Z-score yang negatif menunjukkan bahwa

nilai asli berada di bawah rata-rata tahun tersebut, sedangkan nilai positif berada di atas rata-rata. Misalnya, pada tahun 2021, entitas ke-3 memiliki Z-score tertinggi (0,4873), mengindikasikan bahwa nilainya lebih tinggi dibandingkan entitas lainnya relatif terhadap rerata populasi tahun tersebut. Standardisasi ini penting untuk menjaga konsistensi dan menghindari dominasi fitur tertentu dalam analisis lanjutan seperti klusterisasi.

2) Menentukan Jumlah Kluster

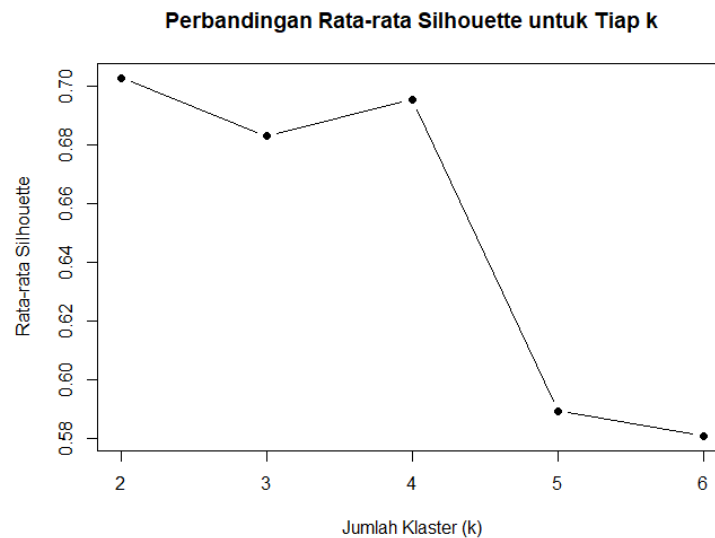
Langkah pertama dalam menentukan jumlah kluster untuk jenis-jenis kejahatan di Sumatera Utara dilakukan dengan menggunakan Metode Elbow, sebagaimana ditunjukkan pada Gambar 2.



Gambar 2. Hasil Grafik dengan Elbow-Method

Berdasarkan grafik Elbow yang menunjukkan hubungan antara jumlah kluster (k) dan total within-cluster sum of squares (WSS), penurunan WSS yang tajam terjadi dari $k = 1$ hingga $k = 3$, kemudian melandai setelahnya. Titik siku (elbow) terlihat jelas pada $k = 4$, yang menunjukkan bahwa penambahan kluster setelah titik ini tidak memberikan pengurangan WSS yang signifikan. Oleh karena itu, jumlah kluster yang optimal untuk segmentasi data ini adalah $k = 3$ atau $k = 4$, karena pada titik ini keseimbangan antara kompleksitas model dan kualitas pemisahan kluster tercapai.

Langkah kedua untuk mengevaluasi dan mengoptimalkan jumlah cluster, maka dilakukan dengan menggunakan Silhouette Plot, sebagaimana ditunjukkan pada Gambar 3.

**Gambar 3.** Hasil Grafik dengan Silhouette Plot**Tabel 2.** Nilai Rata-rata Silhouette

k	Nilai
k = 2	0.703
k = 3	0.683
k = 4	0.695
k = 5	0.589
k = 6	0.581

Berdasarkan hasil evaluasi menggunakan metode silhouette, nilai rata-rata silhouette tertinggi diperoleh pada jumlah kluster $k = 2$ sebesar 0.703, diikuti oleh $k = 4$ sebesar 0.695, dan $k = 3$ sebesar 0.683. Meskipun $k = 2$ memiliki nilai tertinggi, pemisahan hanya menjadi dua kelompok cenderung terlalu sederhana dan mungkin tidak cukup menggambarkan struktur kompleks data. Sebaliknya, $k = 4$ menawarkan kompromi yang lebih baik antara kualitas kluster dan detail segmentasi data. Penurunan tajam nilai silhouette setelah $k = 4$, yakni menjadi 0.589 pada $k = 5$ dan 0.581 pada $k = 6$, menunjukkan bahwa penambahan kluster selanjutnya justru memperburuk kualitas pemisahan. Oleh karena itu, jumlah kluster optimal disarankan pada $k = 2$ dan $k = 4$, karena memberikan nilai rata-rata yang cukup tinggi.

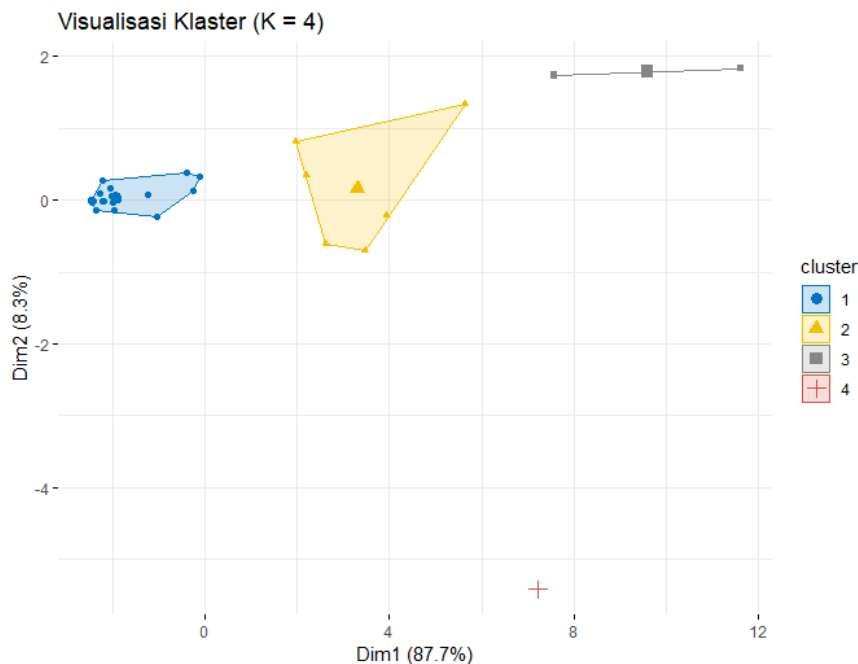
Tabel 3. Hasil dari Metode Elbow dan Silhouette Plot

Keterangan	Nilai k
Elbow-Method	$k = 3, k = 4$
Silhouette Plot	$k = 2, k = 4$

Dapat disimpulkan, berdasarkan Tabel 3, hasil dari metode Elbow dan Silhouette Plot menunjukkan adanya kesamaan pada nilai $k = 4$ sebagai jumlah kluster potensial yang optimal. Metode Elbow mengindikasikan bahwa nilai $k = 3$ dan $k = 4$ merupakan titik tekuk (elbow) yang menunjukkan penurunan signifikan pada total within-cluster sum of squares, sehingga menandakan bahwa penambahan jumlah kluster setelah titik tersebut memberikan manfaat yang semakin kecil. Sementara itu, metode Silhouette Plot mengindikasikan bahwa nilai $k = 2$ memberikan nilai rata-rata silhouette tertinggi (0,703), namun nilai $k = 4$ juga cukup tinggi (0,695) dan menawarkan segmentasi yang lebih rinci. Oleh karena itu, nilai $k = 4$ dapat dianggap sebagai pilihan paling optimal karena didukung oleh kedua metode, memberikan keseimbangan antara kualitas kluster dan kedalaman segmentasi.

3) Visualisasi Kluster

Proses klusterisasi dilakukan dengan algoritma K-Means menggunakan nilai $k = 4$, yang sebelumnya telah ditentukan sebagai jumlah kluster optimal berdasarkan hasil gabungan metode Elbow dan Silhouette Plot. Dalam proses ini, algoritma dijalankan berulang kali dengan variasi awal pusat kluster yang berbeda, untuk memastikan hasil klusterisasi yang paling optimal dan stabil. Pemilihan jumlah kluster sebanyak empat kelompok diharapkan mampu menggambarkan pola pengelompokan data secara lebih representatif dan bermakna, sehingga hasil analisis dapat memberikan wawasan yang lebih relevan terhadap karakteristik data yang dianalisis.



Gambar 4. Visualisasi Klaster

Berdasarkan dari Gambar 4. hasil analisis kluster K-Means dengan jumlah kluster optimal sebanyak empat ($k = 4$), diperoleh pemisahan yang jelas antar kelompok jenis kejahatan berdasarkan kemiripan pola kejadian dari tahun ke tahun.

- Klaster 1 (biru) terdiri dari mayoritas jenis kejahatan yang memiliki frekuensi dan pola kejadian yang relatif rendah dan seragam.
- Klaster 2 (kuning) mengelompokkan kejahatan yang memiliki frekuensi sedang dengan pola fluktuatif.
- Klaster 3 (abu-abu) mencakup jenis kejahatan dengan frekuensi tinggi namun cenderung stabil
- Klaster 4 (merah) merupakan outlier yang menunjukkan jenis kejahatan dengan tingkat kejadian sangat tinggi atau sangat unik dibanding klaster lain.

Pola pemisahan ini mengindikasikan bahwa pengelompokan dapat digunakan sebagai dasar klasifikasi risiko atau prioritas penanganan.

4) Menggabungkan Hasil Klaster dengan Nama Jenis Kejahatan

Langkah selanjutnya, penggabungan hasil klaster dengan nama jenis kejahatan merupakan tahap penting dalam interpretasi analisis K-Means, karena memungkinkan identifikasi langsung terhadap karakteristik masing-masing klaster berdasarkan jenis kejahatan yang tergolong di dalamnya, dapat dilihat pada Tabel 4.

Tabel 4. Distribusi Jenis Kejahatan Berdasarkan Hasil Klaster K-Means

Jenis Kejahatan/Pelanggaran	Total (Selama 15 Tahun)	Klaster
Kejahatan Terhadap Kepala Negara	6	1
Kejahatan Terhadap Ketertiban Umum	30	1
Pembakaran	1722	1
Kebakaran	2913	1
Penyuapan	42	1
Kejahatan Mata Uang	230	1
Kejahatan Materai dan Merk	144	1
Melanggar Kesopanan, perzinahan	13008	1
Perkosaan	3002	1
Perjudian	27991	2
Penculikan	415	1
Pembunuhan	1736	1
Penganiayaan Berat	49971	2
Penganiayaan Ringan	39873	2
Pencurian Ringan	28970	2
Pencurian dengan Kekerasan	13803	1
Pencurian dengan Pemberatan	87580	3
Penghinaan	7729	1
Pemerasan	8843	1
Penggelapan	36856	2
Penipuan	31550	2
Pengrusakan	14814	1
Penadahan	644	1

Jenis Kejahatan/Pelanggaran	Total (Selama 15 Tahun)	Klaster
Kejahatan Ekonomi	124	1
Pencurian Kendaraan Bermotor	64228	3
Melarikan Wanita dibawah Umur	2734	1
Kejahatan Narkotik	57954	4
Penyelundupan	1521	1
Korupsi	272	1
Penyalahgunaan Senjata Api	298	1
Kejahatan Surat-surat Sejenis	3662	1
Sengketa Tanah	3227	1
Ilegal Logging	1265	1

SIMPULAN, KETERBATASAN DAN SARAN

Berdasarkan hasil analisis klasterisasi menggunakan algoritma K-Means terhadap data jenis kejahatan di Provinsi Sumatera Utara selama tahun 2007 hingga 2021, diperoleh bahwa pengelompokan jenis kejahatan ke dalam empat klaster ($k = 4$) merupakan pilihan yang paling optimal. Keempat klaster ini menunjukkan karakteristik yang berbeda berdasarkan intensitas dan pola kejahatan dari waktu ke waktu, di mana terdapat kelompok kejahatan dengan frekuensi rendah dan stabil, sedang dengan fluktuasi, tinggi namun stabil, serta kelompok outlier yang mencerminkan kasus dengan tingkat kejadian sangat tinggi dan unik, seperti kejahatan narkoba. Hasil ini menunjukkan bahwa metode klasterisasi dapat memberikan gambaran yang lebih sistematis mengenai distribusi dan tren kejahatan, sehingga berpotensi menjadi dasar dalam penentuan prioritas kebijakan penegakan hukum maupun strategi pencegahan kriminalitas yang lebih terarah.

Namun demikian, penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Pertama, data yang digunakan hanya mencakup jumlah kejadian per jenis kejahatan tanpa mempertimbangkan variabel lain seperti faktor demografis, sosial ekonomi, atau spasial yang bisa berpengaruh terhadap tingkat kejahatan. Kedua, tingkat agregasi data tahunan membuat analisis musiman atau waktu-waktu spesifik dalam satu tahun tidak dapat dieksplorasi lebih lanjut. Ketiga, keterbatasan dari algoritma K-Means itu sendiri yang sensitif terhadap inisialisasi awal dan hanya efektif untuk klaster

berbentuk bulat (sferis) menjadikan hasilnya perlu diinterpretasikan secara hati-hati. Selain itu, karena data hanya sampai tahun 2021, maka hasil analisis ini mungkin belum sepenuhnya mencerminkan kondisi terkini yang terus berkembang.

Oleh karena itu, disarankan agar penelitian selanjutnya mempertimbangkan penggunaan metode klasterisasi alternatif seperti DBSCAN atau Hierarchical Clustering yang dapat menangani bentuk klaster yang tidak beraturan dan tidak memerlukan penentuan jumlah klaster di awal. Penambahan variabel pendukung seperti kondisi sosial-ekonomi, kepadatan penduduk, atau lokasi geografis juga akan sangat membantu dalam memperkaya analisis dan meningkatkan akurasi segmentasi. Selain itu, pembaruan data hingga tahun terkini perlu dilakukan agar hasil analisis tetap relevan dan dapat digunakan secara praktis dalam perumusan kebijakan. Hasil dari penelitian ini juga sebaiknya dimanfaatkan sebagai dasar penyusunan strategi pencegahan dan penanggulangan kejahatan yang berbasis data serta berbasis wilayah, dengan pendekatan yang menyesuaikan pada karakteristik masing-masing kelompok kejahatan.

DAFTAR PUSTAKA

- Amalina, T., Bima, D., Pramana, A., & Sari, B. N. (2022). Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food. *Jurnal Ilmiah Wahana Pendidikan*, 8(15), 574–583. <https://doi.org/10.5281/zenodo.7052276>
- Dewi Sinta, M., Windarto Agus, P., Damanik Irfan, S., & Satria Heru. (2019). *Seminar Nasional Sains & Teknologi Informasi (SENSASI) Analisa Metode K-Means pada Pengelompokan Kriminalitas Menurut Wilayah*. 620–625. <http://prosiding.seminar-id.com/index.php/sensasi/issue/archivePage|620>
- Galih Pradana, M., Nurcahyo, A. C., & Saputro, P. H. (2020). Penerapan Metode K-Means Klustering untuk Menentukan Kepuasan Pelanggan K-Means Clustering Method to Determine Customer Satisfaction. *Citec Journal*, 7(1).
- Gultom, D., Dame Tampubolon, H., Yolanda Hutabarat, L., Ilmi H Zer, F. R., Hartama, D., Sudirman Blok, J. A., Pematangsiantar, K., & Utara, S. (2020). PENERAPAN ALGORITMA K-MEANS UNTUK MENGETAHUI TINGKAT TINDAK KEJAHATAN DAERAH PEMATANGSIANTAR. *Jurnal Teknologi Informasi*, 4(1).
- Hartanti, N. T. (2020). Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 6(2), 82–89. <https://doi.org/10.25077/teknosi.v6i2.2020.82-89>
- Juliartha Martin Putra, B., Ariani Finda Yuniarti, D., Komputer dan Jaringan, P., & Komunitas Negeri Pacitan, A. (2022). Analisis Gaya Belajar terhadap Nilai Mahasiswa dengan Menggunakan Metode k-Means Analysis of Learning Styles on

- Student Grade using the k-Means Method. *Jurnal Techno COM*, 21(2), 343–354.
- Khalidin Basyir, M. (n.d.). *Klastering Penyakit Diabetes dengan Metode K-Means*.
- Kharisma, E. T., & Jananto, A. (2024). Penerapan Data Mining Clustering Algoritma K-Means Untuk Menganalisa Pola Kejadian Tindak Kejahatan (Studi Kasus Polrestabes Semarang). *Jurnal Penerapan Sistem Informasi*, 5(4), 1580–1587.
- Novita, R., Pramanda, H., & Ariansyah, D. (2024). *Analisis Pengaruh Pelanggaran (U-turn) pada Simpang Kota Banda Aceh (Studi Kasus : Sp. Surabaya, Sp. Jamboe Tape, dan Sp. PDAM) dengan Metode Rumus Z-Score*. 3(2), 433–441.
- Nur Afidah, N. (2023). Penerapan Metode Clustering dengan Algoritma K-means untuk Pengelompokkan Data Migrasi Penduduk Tiap Kecamatan di Kabupaten Rembang. *PRISMA, Prosiding Seminar Nasional Matematika*, 6, 729–738. <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Permata Sari, Y., & Basir, C. (2025). ANALISIS KLASER DENGAN METODE K-MEANS PADA PERSEBARAN KASUS COVID-19 BERDASARKAN DESA DI KECAMATAN KEMANG – BOGOR. *Pelita : Jurnal Penelitian Dan Karya Ilmiah*, 24(2), 33–45. <https://doi.org/10.33592/pelita.v24i2.5504>
- Prihati, Y., & Dharmawan, A. (2021). IMPLEMENTASI ALGORITMA K-MEANS UNTUK PEMETAAN PRESTASI AKADEMIK SISWA DISEKOLAH DASAR TERANG BAGI BANGSA PATI. *EJournal Universitas AKI*.
- Septiani, N., Erwansyah, K., Suryanata, M., Studi Sistem Informasi, P., & Triguna Dharma, S. (2020). ANALISIS DATA MINING PENGELOMPOKKAN KASUS TINDAK KEJAHATAN YANG TERJADI DI KECAMATAN MEDAN POLONIA DENGAN MENGGUNAKAN METODE K-MEANS CLUSTERING. *Jurnal Cyber Tech*, 3(2).
- Wenny. (2024). Normalisasi Data Kependudukan Dengan Model Min Max Dan Algoritma K-Means Untuk Pengelompokkan Tingkat Ekonomi Masyarakat. *Bulletin of Information System Research (BIOS)*, 2(2), 53–63. <https://journal.grahamitra.id/index.php/bios>
- Widad, R., & Fatah, Z. (2025). Penerapan Algoritma K-Means Clustering Pada Penjualan Produk Skincare. In *JAMASTIKA* (Vol. 4).
- Wisna, N., & Cahaya Rani, M. (2023). *ALGORITMA K-MEANS CLUSTERING ANALISIS RASIO AKTIVITAS MENGGUNAKAN PYTHON*. 10(2). <https://doi.org/10.36987/ecobi.v10i2>
- Yuvrico, B. F. (2023). *PENERAPAN METODE K-MEANS UNTUK CLUSTERING TINGKAT KEJAHATAN DI INDONESIA*. UIN SAMPEL AMPEL.