



Pengelompokan Tindak Kejahatan Berdasarkan Tempat Kejadian Perkara di Kota Binjai Menggunakan Metode *Clustering* (Studi kasus: Polres Binjai)

Herdina Putri Ahmadi^{1*}, Magdalena Simanjuntak², Muammar Khadapi³

¹⁻³Sekolah Tinggi Manajemen Informatika dan Komputer Kaputama, Indonesia

herdinaputri03@gmail.com^{1*}, [magdalena.simanjuntak84@gmail.com](mailto:magdalenasimanjuntak84@gmail.com)², khadafi5@gmail.com³

Alamat Kampus: Jl. Veteran No.4A, Tangsi, Kec. Binjai Kota, Kota Binjai, Sumatera Utara

Korespondensi penulis: herdinaputri03@gmail.com

Abstract: *Crime is a social issue that continues to evolve alongside increasing community activity and regional development. This study aims to Cluster crime data in Binjai City based on the location of incidents using the K-Means algorithm and the Cross Industry Standard Process for Data Mining (CRISP-DM) approach. The data were obtained from the Binjai Police Department, with attributes including the type of crime, time of occurrence, and location, categorized by district. A comprehensive data preprocessing stage was carried out, involving the extraction of information from raw data, normalization of crime type labels, and conversion of categorical data into numerical form using label encoding. The optimal number of Clusters was determined using the Silhouette score method, which yielded the best result at $K = 10$. The Clustering results were further evaluated using the Davies-Bouldin Index (DBI) to ensure Cluster quality. The analysis revealed that Binjai Utara District has the highest number of crimes, particularly aggravated theft (*curat*), which frequently occurs from early morning to late morning. This Clustering is expected to provide valuable insights for authorities in formulating more targeted and data-driven regional security strategies.*

Keywords: *Crime, CRISP-DM, K-Means Clustering*

Abstrak: Kejahatan merupakan isu sosial yang terus berkembang seiring dengan meningkatnya aktivitas masyarakat dan pertumbuhan wilayah. Penelitian ini bertujuan untuk mengelompokkan data tindak kejahatan di Kota Binjai berdasarkan lokasi kejadian menggunakan metode K-Means dan pendekatan *Cross Industry Standard Process for Data Mining* (CRISP-DM). Data yang digunakan, diperoleh dari Polres Binjai dengan atribut berupa jenis kejahatan, waktu kejadian, dan lokasi kejadian yang diklasifikasikan berdasarkan kecamatan. Tahapan *preprocessing* data dilakukan secara menyeluruh, meliputi ekstraksi informasi dari data mentah, normalisasi format label jenis kejahatan, serta konversi data kategorikal menjadi numerik menggunakan *label encoding*. Untuk menentukan jumlah *Cluster* yang paling optimal, digunakan metode *Silhouette score* yang menunjukkan hasil terbaik pada nilai $K = 10$. Evaluasi hasil *Clustering* juga diperkuat dengan penggunaan *Davies-Bouldin Index* (DBI) untuk memastikan kualitas *Cluster*. Hasil analisis menunjukkan bahwa Kecamatan Binjai Utara merupakan wilayah dengan jumlah kejahatan tertinggi, terutama kasus pencurian dengan pemberatan (*curat*) yang sering terjadi pada waktu dini hari hingga pagi. Pengelompokan ini diharapkan dapat memberikan gambaran bagi pihak berwenang dalam menyusun strategi keamanan wilayah

Kata kunci: CRISP-DM, Kejahatan, K-Means Clustering

1. LATAR BELAKANG

Tindak kejahatan merupakan salah satu permasalahan sosial yang terus berkembang seiring meningkatnya aktivitas masyarakat dan pertumbuhan suatu wilayah. Kejahatan tidak hanya menimbulkan kerugian secara ekonomi, tetapi juga menciptakan rasa ketidaknyamanan dan ketidakamanan di tengah Masyarakat (Kharisma et al., 2024). Salah satu elemen penting dalam proses penyelidikan adalah Tempat Kejadian Perkara (TKP), yaitu lokasi di mana suatu

tindak pidana berlangsung, termasuk tempat ditemukannya barang bukti, korban, maupun pelaku (Sinaga & Simatupang, 2020). Informasi spasial dari TKP dapat memberikan kontribusi besar dalam proses investigasi dan perumusan strategi pencegahan kejahatan.

Kota Binjai merupakan salah satu kota berkembang di Provinsi Sumatera Utara yang dihadapkan pada tantangan dalam menjaga keamanan dan ketertiban masyarakat. Berdasarkan data dari BPS Kota Binjai (2023), angka kriminalitas di kota ini meningkat secara signifikan dari 1.189 kasus pada tahun 2021 menjadi 2.011 kasus pada tahun 2023. Hal ini menunjukkan perlunya pendekatan berbasis data dalam menganalisis wilayah rawan kejahatan guna meningkatkan efektivitas pengawasan dan penanggulangan. Salah satu metode yang relevan untuk menganalisis pola dan distribusi kejadian adalah data mining *Clustering*, khususnya algoritma K-Means yang mampu mengelompokkan data berdasarkan kemiripan karakteristik.

Penelitian ini bertujuan untuk mengelompokkan tindak kejahatan di Kota Binjai berdasarkan jenis kejahatan, waktu kejadian dan lokasi kejadian dengan menggunakan algoritma K-Means dan pendekatan CRISP-DM. Melalui visualisasi hasil *Clustering* yang diperoleh, diharapkan pihak kepolisian dapat menyusun strategi keamanan secara lebih efisien dan terfokus pada wilayah dengan tingkat kerawanan tinggi. Penelitian ini diharapkan dapat memperkuat pemanfaatan teknik data mining dalam studi kriminologi terapan, serta menjadi dasar bagi penelitian lanjutan di masa mendatang, khususnya yang berkaitan dengan analisis spasial kejahatan di wilayah lokal.

Penelitian oleh (Aswan et al., 2021) di Kepulauan Mentawai dan (Kharisma et al., 2024) di Kota Semarang, telah membuktikan bahwa metode ini efektif dalam mengelompokkan wilayah berdasarkan tingkat kerawanan kejahatan. Namun, sebagian besar studi sebelumnya masih berfokus pada satu atau dua atribut, seperti jenis kejahatan atau waktu kejadian secara terpisah. Belum banyak yang secara sistematis mengintegrasikan tiga elemen penting sekaligus jenis kejahatan, waktu kejadian, dan lokasi kejadian dalam satu model analisis, khususnya dengan pendekatan CRISP-DM yang terstruktur. Selain itu, belum ditemukan penelitian serupa yang fokus pada wilayah Kota Binjai sebagai konteks lokal. Oleh karena itu, penelitian ini hadir untuk mengisi kekosongan tersebut dengan menyajikan analisis yang lebih komprehensif dan berbasis data aktual dari wilayah Binjai.

2. KAJIAN TEORITIS

Tindak kejahatan merupakan salah satu bentuk pelanggaran yang sering disebut sebagai kriminalitas, merupakan pelanggaran terhadap hukum, baik dalam bentuk norma maupun peraturan yang telah ditetapkan (Yuliyanti & Martanto, 2024). Salah satu unsur penting dalam

proses penyidikan terhadap tindak kejahatan adalah lokasi peristiwanya, yang dikenal sebagai Tempat Kejadian Perkara (TKP). (Jasman et al., 2023) menjelaskan bahwa TKP tidak hanya mencakup lokasi di mana suatu tindak pidana terjadi atau dilakukan, tetapi juga mencakup area lain yang berkaitan dengan kejahatan tersebut, di mana barang bukti, tersangka, atau korban dapat ditemukan. Sehingga TKP memiliki peran penting dalam proses penyelidikan yang membantu dalam mengungkap serta memperjelas suatu perkara pidana.

Untuk menganalisis pola kejahatan secara lebih sistematis, pendekatan data mining banyak digunakan karena mampu menemukan pola tersembunyi dari data berukuran besar (Arhami & Nasir, 2020). Salah satu kerangka kerja yang cukup sering digunakan dalam proses ini adalah CRISP-DM (Cross-Industry Standard Process for Data Mining). Pendekatan ini membagi proses analisis menjadi enam tahap utama, mulai dari Business Understanding, Data Understanding, Data Preparation, Data Modeling, Evaluation serta Deployment (Sucahyo et al., 2022). Tahapan ini memberikan alur kerja yang runtut dan membantu peneliti tetap fokus pada tujuan analisis.

Di dalam proses modeling, metode *Clustering* menjadi salah satu pendekatan yang digunakan untuk mengelompokkan data berdasarkan kemiripan tertentu (Buaton et al., 2021). Metode *Clustering* ini membantu mengidentifikasi pola dan hubungan yang tidak terlihat pada pengamatan kasual. Hal ini menjadi sangat penting dalam analisis data berukuran besar, di mana volume dan kompleksitasnya melampaui kapasitas analisis manual (Rahayu et al., 2024). Salah satu algoritma yang paling umum dipakai adalah K-Means, yang bekerja dengan mengelompokkan data ke dalam sejumlah *Cluster* berdasarkan jarak ke titik pusat (centroid) (Juliawati et al., 2023). Algoritma ini dinilai efisien, namun tetap memiliki tantangan seperti kepekaan terhadap data ekstrem atau penentuan awal centroid yang kurang tepat. Karena itu, diperlukan evaluasi terhadap hasil *cluster* menggunakan metrik seperti *Silhouette score* dan *Davies-Bouldin Index* (DBI), yang membantu menilai seberapa baik data telah dikelompokkan.

Menurut (Syahkur & Hartama, 2024), nilai *silhouette score* menunjukkan seberapa baik data dikelompokkan dalam suatu *Cluster* dan seberapa terpisah dari *Cluster* lain. Skornya berkisar antara -1 hingga 1; semakin mendekati 1, semakin baik kualitas pengelompokan. Nilai yang rendah menunjukkan adanya tumpang tindih antar *Cluster*.

Silhouette score juga membantu menentukan jumlah *Cluster* optimal dengan membandingkan kesamaan data dalam dan antar kelompok, sehingga kerap digunakan untuk mengevaluasi efektivitas *Clustering* sebelum analisis lanjutan (Mulyawan et al., 2023).

Davies-Bouldin Index merupakan metrik evaluasi yang menilai kualitas *Clustering* berdasarkan rasio antara jarak antar *Cluster* dan dispersi dalam *Cluster*. Nilai DBI yang lebih

rendah menunjukkan bahwa *Cluster* yang terbentuk lebih kompak dan terpisah dengan baik dari *Cluster* lainnya, yang mengindikasikan hasil *Clustering* yang lebih baik (Fauzi et al., 2023).

Beberapa penelitian sebelumnya menunjukkan keberhasilan penggunaan algoritma K-Means dalam pengelompokan wilayah rawan kejahatan. (Aswan et al., 2021) mengklasifikasikan wilayah rawan pencurian kendaraan bermotor di Kepulauan Mentawai menjadi tiga kategori menggunakan K-Means. Sementara itu, (Kharisma et al., 2024) menerapkan teknik serupa pada data kejahatan di Kota Semarang dan berhasil membentuk tujuh *Cluster* berdasarkan atribut bulan, hari, waktu dan jenis kejahatan. Temuan dari kedua penelitian tersebut memperkuat relevansi metode ini dalam konteks analisis kejahatan berbasis lokasi.

3. METODE PENELITIAN

Penelitian ini bersifat eksploratif dengan pendekatan berbasis data numerik, menggunakan metode *Clustering* dan alur kerja CRISP-DM. Fokus utamanya adalah mengelompokkan kejadian tindak kejahatan berdasarkan pola yang terbentuk dari data historis. Model CRISP-DM digunakan karena menyediakan alur kerja yang terstruktur, mulai dari memahami tujuan analisis hingga mengevaluasi hasil, sehingga memudahkan proses dari pemahaman konteks hingga penilaian kualitas akhir (Khadapi & Pakpahan, 2024).

Data yang digunakan merupakan data sekunder yang diperoleh dari Kepolisian Resor (Polres) Binjai. Terdapat 876 entri data yang digunakan dalam penelitian ini, dengan atribut utama meliputi jenis kejahatan (seperti curat, curas, curanmor, dan cubis), waktu kejadian, serta lokasi kejadian berdasarkan kecamatan. Karena bentuk awal data bersifat kategorikal, maka dilakukan proses konversi ke dalam bentuk numerik menggunakan metode *label encoding*, agar dapat diproses lebih lanjut oleh algoritma *Clustering*.

Secara umum, penelitian ini mengikuti alur kerja CRISP-DM, dimulai dari:

1. Business Understanding: Merumuskan tujuan untuk mengetahui distribusi wilayah rawan kejahatan;
2. Data Understanding: Mengenal struktur dan karakteristik data yang tersedia;
3. Data Preparation: Melakukan pembersihan data, normalisasi label jenis kejahatan, dan transformasi nilai ke dalam bentuk yang dapat dianalisis;
4. Modeling: Menerapkan algoritma K-Means untuk mengelompokkan data kejahatan berdasarkan kemiripan;

5. Evaluation: Mengevaluasi kualitas *Cluster* yang terbentuk menggunakan *Silhouette score* untuk menentukan jumlah *cluster* terbaik, dan *Davies-Bouldin Index* (DBI) sebagai ukuran validitas *Cluster*.

<i>Business Understanding</i>	Identifikasi masalah dan menentukan tujuan
<i>Data Understanding</i>	Mengumpulkan data kejahatan dan menganalisis data
<i>Data Preparation</i>	Pembersihan data, melakukan kategorisasi waktu, standarisasi kecamatan, dan transformasi data
<i>Modelling</i>	Memilih algoritma dan mengevaluasi parameter model
<i>Evaluation</i>	Evaluasi kinerja model, validasi hasil <i>clustering</i> , interpretasi <i>cluster</i> , dan menilai kesesuaian model dengan tujuan analisis
<i>Deployment</i>	Menganalisis dan memvisualisasikan hasil <i>clustering</i>

Gambar 1. Alur Proses Penelitian Mengacu pada Model CRISP-DM

Gambar 1 menunjukkan tahapan yang digunakan dalam penelitian, mulai dari pemahaman konteks hingga proses evaluasi hasil *Clustering*.

Seluruh proses analisis dijalankan menggunakan bahasa pemrograman Python versi 3.11.11, dengan bantuan platform Google Colab sebagai lingkungan kerja. Selama proses berjalan, dilakukan pula visualisasi hasil *Clustering* untuk memudahkan interpretasi sebaran data dan karakteristik masing-masing *Cluster*.

Kualitas hasil pengelompokan kemudian dievaluasi menggunakan dua metrik, yaitu *Silhouette score* untuk menentukan jumlah *Cluster* yang paling sesuai, dan *Davies-Bouldin Index* (DBI) untuk mengukur validitas antar-*Cluster* yang terbentuk .

4. HASIL DAN PEMBAHASAN

1) Hasil Pengumpulan dan Persiapan Data

Penelitian ini menggunakan data tindak kejahatan yang diperoleh dari Polres Binjai. Data tersebut berisi 876 entri yang memuat informasi jenis kejahatan, waktu kejadian, dan lokasi kejadian dalam bentuk kecamatan. Data ini mencakup wilayah administrasi Kota Binjai yang terdiri atas lima kecamatan.

Proses awal meliputi ekstraksi informasi waktu kejadian dalam format jam serta identifikasi kelurahan dan kecamatan dari teks Tempat Kejadian Perkara (TKP). Data kemudian dinormalisasi agar variasi ejaan tidak menyebabkan kesalahan klasifikasi. Misalnya, variasi “curanmor”, “curan mor”, atau “curanmorr” diseragamkan menjadi “CURANMOR”.

2) Proses Analisis Data

a. Preprocessing dan Transformasi Data

Dalam penelitian ini, waktu kejadian dikategorikan ke dalam lima kelompok waktu, yaitu dini hari, pagi, siang, sore, dan malam. Pembagian ini dilakukan untuk mempermudah analisis pola temporal tindak kejahatan. Kategorisasi ini mengacu pada interval waktu yang digunakan

Tabel 1. Kategorisasi waktu kejadian ke dalam lima kelompok waktu

Waktu	WAKTU_ENCODED	WAKTU_LABEL
00.00 - 04.59	1	DINI HARI
05.00 - 10.59	2	PAGI
11.00 - 14.59	3	SIANG
15.00 - 17.59	4	SORE
18.00 - 23.59	5	MALAM

Sumber : (Ilham & Wibisono, 2023)

b. Menentukan Jumlah Cluster Optimal

Penentuan jumlah *Cluster* optimal dilakukan dengan metode *Silhouette score* pada rentang $K=2$ hingga $K=10$. Hasilnya menunjukkan nilai tertinggi diperoleh saat $K=10$ dengan skor 0,391.

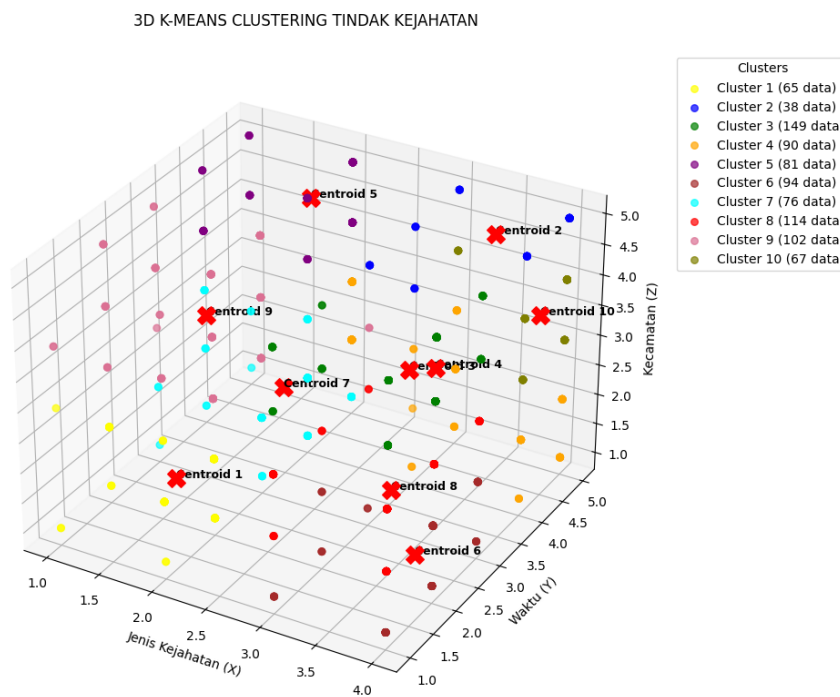


Gambar 2. Visualisasi *Silhouette score* pada rentang jumlah *Cluster* 2 – 10

3) Hasil *Clustering* dengan K-Means

a. Visualisasi Hasil *Clustering*

Hasil pengelompokan divisualisasikan dalam grafik 3D yang menggambarkan penyebaran data berdasarkan jenis kejahatan, waktu kejadian, dan kecamatan.



Gambar 3. Visualisasi 3D hasil *Clustering* K=10

Visualisasi pada gambar 3 memperlihatkan penyebaran data ke dalam 10 *Cluster*, masing-masing diwarnai berbeda, dengan centroid ditandai silang merah.

b. Interpretasi Setiap Cluster

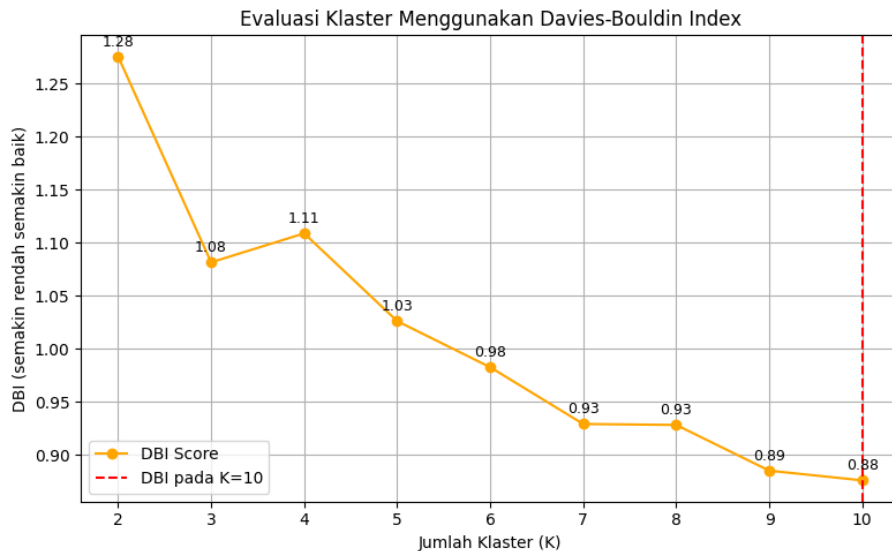
Hasil *Clustering* menunjukkan distribusi data ke dalam 10 *Cluster* dengan karakteristik centroid yang berbeda pada masing-masing variabel.

- *Cluster 1* (65 data): dominan CURANMOR pada dini hari hingga pagi hari, paling sering terjadi di Binjai Kota.
- *Cluster 2* (38 data): dominan CURAT pada sore hari di Binjai Utara.
- *Cluster 3* (149 data): dominan CURAT dini hari hingga pagi hari di Binjai Utara.
- *Cluster 4* (90 data): dominan CURANMOR sore hingga malam di Binjai Kota.
- *Cluster 5* (81 data): dominan CURANMOR sore hingga malam di Binjai Timur.
- *Cluster 6* (94 data): dominan CURAT pagi di Binjai Barat.
- *Cluster 7* (76 data): dominan CURANMOR sore di Binjai Kota.
- *Cluster 8* (114 data): dominan CURAT dini hari hingga pagi hari di Binjai Selatan.
- *Cluster 9* (102 data): dominan CURANMOR pagi di Binjai Timur.
- *Cluster 10* (67 data): dominan CURAT sore hingga malam di Binjai Timur.

Secara umum, Kecamatan Binjai Utara teridentifikasi sebagai wilayah dengan kasus kejahatan tertinggi, terutama curat pada waktu dini hari hingga pagi.

4) Evaluasi Model

Evaluasi kualitas *Clustering* dilakukan dengan *Davies-Bouldin Index* (DBI). Hasil pengujian menunjukkan nilai DBI terendah pada $K=10$ yakni 0,88 yang mengindikasikan pemisahan *Cluster* cukup baik.



Gambar 3. Grafik DBI pada jumlah *Cluster* 2 – 10

5) Implikasi Penelitian

Hasil *Clustering* ini memberi wawasan penting mengenai pola distribusi kejahatan di Kota Binjai. Temuan ini dapat digunakan oleh pihak kepolisian untuk menentukan prioritas wilayah patroli, menyusun strategi pencegahan pada waktu rawan, serta mengalokasikan sumber daya keamanan secara lebih efisien.

KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa algoritma K-Means efektif untuk mengelompokkan tindak kejahatan di Kota Binjai berdasarkan lokasi kejadian. Analisis diawali dengan *preprocessing* data, termasuk ekstraksi informasi, normalisasi label, dan label encoding pada atribut waktu, jenis kejahatan, dan kecamatan. Metode *silhouette score* menunjukkan hasil optimal pada K=10. Setiap *Cluster* memiliki karakteristik unik, dan ditemukan bahwa Kecamatan Binjai Utara memiliki kasus tertinggi, didominasi pencurian dengan pemberatan (curat) pada dini hari hingga pagi.

Evaluasi menggunakan *silhouette score* dan *Davies-Bouldin Index* menunjukkan hasil *Clustering* yang baik dan layak dijadikan dasar analisis wilayah rawan kejahatan. Temuan ini diharapkan dapat membantu kepolisian merancang strategi pencegahan yang lebih tepat. Namun, karena penelitian hanya menggunakan tiga variabel utama, disarankan untuk menambahkan faktor lain seperti data sosial-ekonomi atau kepadatan penduduk, serta mencoba metode analisis lain di penelitian selanjutnya.

DAFTAR REFERENSI

- Arhami, M., & Nasir, M.** (2020). *DATA MINING: Algoritma dan implementasi* (R. I. Utami, Ed.). Penerbit ANDI.
- Aswan, Y., Defit, S., & Nurcahyo, G. W.** (2021). Algoritma K-Means clustering dalam mengklasifikasi data daerah rawan tindak kriminalitas (Polres Kepulauan Mentawai). *Jurnal Sistem Informasi dan Teknologi*, 3, 245–250. <https://doi.org/10.37034/jsisfotek.v3i4.73>
- BPS Kota Binjai.** (2023). *Kota Binjai dalam angka 2023*.
- Buaton, R., Zarlis, M., & Yasin, V.** (2021). *Konsep data mining dalam implementasi (The concept of data mining in implementation)*. Mitra Wacana Media.
- Fauzi, I. F., Resmi, M. G., & Hermanto, T. I.** (2023). Penentuan jumlah cluster optimal menggunakan Davies–Bouldin index pada algoritma K-Means untuk menentukan kelompok penyakit. *[Nama Jurnal]*, 7(2), 2598–8069.
- Ilham, M., & Wibisono, B. H.** (2023). Pola spasial kejahatan pencurian berdasarkan aspek temporal di Kecamatan Kadia. *SPECTA Journal of Technology*, 7(3), 711–722. <https://doi.org/10.35718/specta.v7i3.1033>
- Jasman, M., Asba, P., Saputra, I. R., & Pinrang, K. R.** (2023). Urgensi olah tempat kejadian perkara dalam proses pembuktian. *[Nama Jurnal]*, 5(1), 101–112.
- Juliawati, F., Buaton, R., Saragih, R., & Kaputama, S.** (2023). Pengelompokan data mining penerimaan bantuan pangan non tunai (BPNT) menggunakan metode clustering (Studi kasus: Kantor Desa Payabakung Hamparan Perak). *Journal of Computer Science and Information Technology*, 3(2), 69–77.
- Khadapi, M., & Pakpahan, V. M.** (2024). Analisis sentimen berbasis jaringan LSTM dan BERT terhadap diskusi. *[Nama Jurnal]*, 6, 130–137.
- Kharisma, E. T., Jananto, A., & Teknologi, F. Stikubank, U.** (2024). Penerapan data mining clustering algoritma K-Means untuk menganalisa pola kejadian tindak kejahatan (Studi kasus Polrestabes Semarang). *[Nama Jurnal]*, 5(4), 1580–1587.
- Mulyawan, A. R., Gunawan, D., Basri, H., Alfarizi, S., & Ichsan, N.** (2023). Penerapan K-Medoids clustering dan silhouette method untuk strategi pemasaran program donasi pada lembaga amil zakat. *Information System for Educators and Professionals: Journal of Information System*, 8(1), 107. <https://doi.org/10.51211/isbi.v8i1.2468>
- Rahayu, P., Sudipa, I. G. I., Suryani, Surachman, A., Ridwan, A., Darmawiguna, I. G. M., Sutoyo, M., Slamet, I., Harlina, S., & May Sanjaya, I. M.** (2024). *Buku ajar data mining* (Vol. 1, Jan 2024).
- Sinaga, L. V., & Simatupang, M. Y. M.** (2020). Fungsi olah tempat kejadian perkara (TKP) guna mengungkapkan kasus penganiayaan berat ditinjau dari sudut hukum acara pidana. *Jurnal Rectum: Tinjauan Yuridis Penanganan Tindak Pidana*, 2(2), 129–137. <https://doi.org/10.46930/jurnalrectum.v2i2.645>
- Sucahyo, N., Kurniati, I., & Harvit, K.** (2022). *SWADHARMA (JRIS)*.

Syahkur, M. R., & Hartama, D. (2024). Evaluasi jumlah cluster pada algoritma K-Means++ menggunakan silhouette dan elbow dengan validasi nilai DBI dalam mengelompokkan gizi balita. [*Nama Jurnal*], 13(3), 487–496.

Yuliyanti, D., & Martanto, M. (2024). Clustering tingkat kejahatan kriminal menggunakan metode K-Means di wilayah Kabupaten Cirebon. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3509–3514. <https://doi.org/10.36040/jati.v7i6.8894>