

IMPLEMENTASI ALGORITMA *K-MEANS* UNTUK *CLUSTERING* KASUS KRIMINAL DI WILAYAH KOTA DEPOK

Amelia Adhariani^{1*}, Triana Harmini², Aziz Musthafa³

^{1,2,3}Teknik Informatika, Universitas Darussalam Gontor, Ponorogo, Indonesia

*Penulis korespondensi: ameliaadhariani42066@mhs.unida.gontor.ac.id

ABSTRAK

Tindakan kriminal merupakan tindakan yang dilarang, ditindak tegas serta mendapatkan hukuman di setiap negara demi melindungi masyarakatnya. Pada portal Data Indonesia.id Kepolisian RI (Polri) mendata, ada 288.472 tindak kriminal yang terjadi di Indonesia sepanjang tahun 2023 jumlah kejahatan tersebut mengalami kenaikan sebesar 4,33% dibandingkan tahun sebelumnya sebanyak 276.507 kasus. Khususnya Kota Depok yang mengalami peningkatan sebanyak 32% pada tahun 2023 dibandingkan tahun 2022. Dengan adanya peningkatan angka kriminalitas, diperlukannya suatu metode yang efektif untuk memahami distribusi untuk meminimalisir kasus kriminal di masing- masing wilayah. Penelitian ini menggunakan data yang mencakup informasi tentang 5 jenis kejahatan yang memiliki nilai tertinggi pada tahun 2023 di masing- masing wilayah Kepolisian Sektor Kota Depok, yang kemudian dikelompokkan menggunakan algoritma *K-Means*. Sedangkan metode yang digunakan dalam penelitian ini adalah CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang terdiri dari beberapa fase yaitu, pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan *deployment*. Tujuan dari penelitian ini untuk menganalisis hasil *clustering* sehingga dapat mengetahui kategori wilayah berdasarkan kasus kriminal dengan kategori rawan, tidak rawan dan sedang. Berdasarkan hasil evaluasi yang telah dilakukan terhadap hasil *clustering* didapatkanlah cluster 3 dengan memperoleh hasil *Silhouette Score* sebesar 0.5470.

Kata kunci: Algoritma *K- Means*, *Clustering*, CRISP- DM, *Data Mining*, Kriminalitas

1 PENDAHULUAN

Dengan pertumbuhan populasi yang signifikan dalam beberapa tahun terakhir, Kota Depok telah mengalami peningkatan di berbagai macam aspek tidak hanya aspek positif namun juga negatif. Salah satu dari aspek negatif yang dapat dirasakan yaitu peningkatan angka kriminalitas. Berdasarkan Undang- Undang Dasar 1945 Pasal 28D ayat 1 mengatur bahwa setiap orang berhak atas pengakuan, jaminan, perlindungan, dan kepastian hukum yang adil serta perlakuan yang sama dihadapan hukum (Yudisial, 1945). Hal ini sebagai landasan hak rakyat Indonesia dalam penegakkan dan penagangan hukum dari pemerintahan.

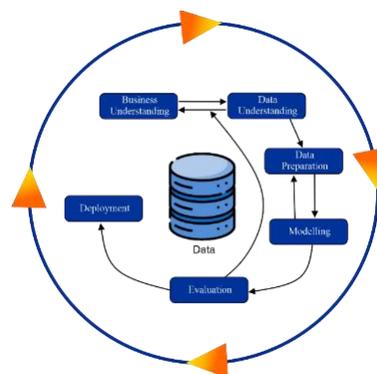
Menurut Data Mabes Polri yang telah diolah oleh Badan Pusat Statistika (BPS) Indonesia menunjukkan sejumlah daerah yang memiliki angka risiko waga terkena tindak kejahatan tertinggi pada tahun 2021. Wilayah hukum Polda Metro Jaya memiliki risiko 0.277% penduduk yang meliputi beberapa wilayah, salah satu diantaranya adalah Kota Depok. Hal ini dapat dilihat dari meningkatnya kasus kriminal yang digambarkan pada tahun 2020 sebanyak 1.246 kasus, Sedangkan 2021 sebanyak 1.792 kasus (Santika, 2023). Sebagaimana terlansir dalam *website* inilah Depok, Kota Depok mengalami peningkatan sebanyak 32% jumlah kejahatan meningkat

pada tahun 2023 dibandingkan tahun 2022 (Setyadi, 2023). Fenomena ini menimbulkan urgensi bagi pemerintah Kota Depok untuk menindak lanjuti hal ini agar keadaan normal kembali.

Ditengah masuknya era revolusi industri 5.0 yang mana hampir sebgayaan besar dari kegiatan yang dilakukan masyarakat terotomatisasi dan terdigitalisasi, salah satu teknologi yang sedang marak berkembang sekarang ialah analisis data menggunakan data mining memiliki beberapa metode yang diterapkan antara lain, estimasi, klasifikasi, klusterisasi, asosiasi, dan forecasting (Bahtiar, 2023). Salah satu metode yang populer dan sederhana ialah Clustering yang mana metode ini memiliki beberapa algoritma antara lain *K- Means*, *Mini Batch K-Means*, *Divisive Hierarchica*, *Gaussian Mixture Model (GMM)*, *DBSCAN (Density-Based Spatial Clustering of Applications with Noise)*, dll. Salah satu algoritma yang populer dan sederhana yaitu *K- Means* yang memiliki sifat *unsupervised learning* atau metode mandiri yang tanpa menggunakan label sebelumnya atau bekerja (Pudoli et al., 2024). Algoritma ini membagi data ke dalam kelompok K atau kluster berdasarkan kedekatan data terhadap *centroid* atau titik pusat, dan kelompok tersebut memiliki kesamaan karakteristik namun berbeda dengan data yang di kelompok lainnya. Pada penelitian ini *K- Means Clustering* melakukan pengelompokkan terhadap data 5 jenis kejahatan terbanyak pada tahun 2023 dalam 11 kecamatan yang berada di wilayah Kota Depok. Sehingga penelitian ini dapat menganalisis hasil *clustering* dan mengetahui kategori wilayah berdasarkan kasus kriminal dengan kategori rawan, tidak rawan dan sedang menggunakan *cluster* yang terbaik.

2 METODE

Untuk menjelaskan proses penelitian dari identifikasi masalah hingga penemuan solusinya, diperlukanlah metode penelitian sebagai langkah- langkah yang dapat diterapkan sampai ke tahap akhir. Pada penelitian ini peneliti memakai metode *Cross Industry Standard Process for Data Mining (CRISP-DM)* adalah suatu pendekatan sistematis yang digunakan dalam proses penambangan data atau data mining (Kusuma et al., 2023). Metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*) menyediakan kerangka kerja yang terstruktur yang biasa digunakan oleh peneliti yang terdiri dari beberapa tahapan dengan sistematis untuk pemecahan masalah dan analisis dalam *data mining*. Adapun tahapannya akan direpresentasikan dalam Gambar 1.



Gambar 1. Tahapan CRISP-DM

Penjelasan terhadap tahapan penelitian diatas antara lain:

1. Pemahaman Bisnis (*Business Understanding*)

Langkah awal yang sangat penting dalam penelitian yang menggunakan *Cross Industry Standard Process for Data Mining (CRISP- DM)* adalah pemahaman bisnis yang

dapat diambil dalam masalah yang ada pada studi kasus yang diambil. (Singgalen, 2024) Hal ini berfokus pada pemahaman isi dari data termasuk mendefinisikan masalah yang akan diselesaikan dalam penelitian.

2. Pemahaman Data (*Data Understanding*)

Setelah ditemukannya pemahaman terhadap penelitian, selanjutnya hal yang dilakukan ialah mempersiapkan dan mengumpulkan data yang relevan untuk dianalisis. Pada tahapan ini peneliti mengeksplorasi dan memahami data yang tersedia, termasuk mengevaluasi kualitas, distribusi, dan karakteristik dari data yang akan diteliti (Ruswanti et al., 2024). Data yang dikumpulkan pada tahap ini dapat berasal dari berbagai sumber yang terpercaya dan dapat dipertanggung jawabkan dengan berbagai format baik itu data berupa teks, gambar, suara, dll. Setelah pengumpulan data, data tersebut dianalisis untuk memahami struktur, kualitas, serta hubungan antar data yang diperoleh (Teknologi et al., 2021).

3. Persiapan Data (*Data Preparation*)

Tahap ini melibatkan persiapan data untuk dianalisis yang mana data mentah akan diolah menjadi dataset yang bersih dan siap untuk dianalisis. Hal ini melibatkan pembersihan data, transformasi, integrasi, dan pemilihan fitur (Pudoli et al., 2024). Setelah itu data diubah sesuai dengan format yang sesuai dengan algoritma yang akan digunakan. Pada tahap akhir dari dataset harus mencerminkan informasi yang sesuai dengan masalah yang diambil dan ditentukan sebelumnya, sehingga dataset dapat diolah dengan baik pada tahap *modelling*.

4. Pemodelan (*Modeling*)

Secara garis besar tahap ini melibatkan pengembangan model analisis yang sesuai dengan tujuan penelitian. Hal ini mencakup pemilihan teknik atau algoritma pemodelan yang tepat dan eksperimen dengan berbagai model untuk menemukan model yang paling efektif. Adapun beberapa *modelling* yang biasa dilakukan antara lain, *clustering*, *classification*, *scoring*, *ranking*, *finding relation*, dan *characterization* (Wahidah & Utari, 2022). Sedangkan algoritma yang dapat dipakai dalam *clustering* antara lain, *K-Means*, *Means Shift Clustering*, *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN), *Expectation-Maximization* (EM) *Clustering using Gaussian Mixture Models* (GMM) (Ashari et al., 2022)

5. Evaluasi (*Evaluation*)

Setelah model dibangun, tahap ini melibatkan evaluasi kinerja model yang dikembangkan untuk memastikan kelayakan dan kecocokannya dengan tujuan sebelumnya (Asyraf & Prasetya, 2024). Jika evaluasi model belum memenuhi kriteria maka perlu dilakukan revisi pada fase sebelumnya.

6. Penerapan (*Deployment*)

Tahapan ini adalah tahapan akhir yang mencakup implementasi model dalam lingkungan nyata (Basiri et al., 2024) Tahapan ini memberikan wawasan mengenai model atau solusi analisis yang dikembangkan ataupun diimplementasikan dalam penelitian berdasarkan data yang telah diteliti. Tahap ini dapat diakhiri dengan dokumentasi yang dapat dilakukan dengan proses pengembangan, hasil model, dan cara penggunaannya, supaya model dapat terus dikembangkan dan diperbaiki dimasa yang akan datang (Singgalen, 2024).

3 HASIL DAN PEMBAHASAN

3.1 Pemahaman Bisnis (*Business Understanding*)

Penelitian ini memiliki tujuan untuk mengetahui informasi dengan cara membentuk *cluster* tindakan kriminal berdasarkan wilayah di Kota Depok menggunakan algoritma *K- Means Clustering* yang mana hal ini menjadi salah satu usaha untuk memberikan dukungan kebijakan, bahan evaluasi berkala, serta keputusan kepada pihak kepolisian dan daerah setempat dalam mengambil keputusan baik pengamanan maupun pengawasan. Permasalahan utama yang dihadapi dalam penelitian ini yaitu kurangnya pemetaan wilayah criminal berdasarkan data yang tersedia, sehingga kebijakan yang diterapkan kurang efektif. Maka dari itu, penelitian ini bertujuan untuk mengelompokkan kasus criminal ke dalam beberapa cluster berdasarkan kategori kejahatannya, memberikan ambaran distribusi kasus 5 tindakan kriminal tertinggi di Kota Depok, dan menentukan area dengan kategori rawan, tidak rawan, dan sedang selama bulan Januari hingga Desember tahun 2023

3.2 Pemahaman Data (*Data Understanding*)

Data yang digunakan dalam penelitian ini diambil dari Polres (Polisi Resor) kota Depok dengan persetujuan pihak yang berwenang dimana dataset yang terkumpul terdiri dari 5 tindakan kriminal tertinggi selama 12 bulan yaitu dari bulan Januari hingga Desember pada tahun 2023. Untuk lokasi yang diambil adalah 11 wilayah hukum menurut kepolisian di kota Depok, dengan total baris 132 dan 7 kolom sebagaimana tertera di Tabel 1.

Tabel 1. Data Wilayah dan Tindakan Kriminal

Wilayah	Bulan	Anirat	Curat	Curas	Curanmor	Peras/ Anc
A	Januari	9	10	0	36	2
B	Januari	1	9	2	11	0
C	Januari	1	6	1	49	1
D	Januari	6	11	0	36	2

3.3 Persiapan Data (*Data Preparation*)

Tahapan ini adalah tahap dimana data diimpor dan dipersiapkan sehingga data menjadi data siap untuk diolah menggunakan algoritma *K- Means*. Setelah data diimpor, maka langkah selanjutnya yaitu ke tahap *preprocessing* dengan cara menstandarisasikan data menggunakan *standarscaler Scikit-Learn* sebagaimana Tabel 2.

Tabel 2. Hasil Standarisasi

Kolom 1	Kolom 2	Kolom 3	Kolom 4	Kolom 5	Kolom 6	Kolom 7
0	4	2.226201	1.476369	-0.417150	1.411135	1.088467
1	4	-0.606170	1.218930	3.253768	-0.091526	-0.611860
2	4	-0.606170	0.446616	1.418309	2.192519	0.238303
3	4	1.164062	1.733807	-0.417150	1.411135	1.088467
4	4	-0.252124	-0.068260	-0.417150	0.028687	0.238303

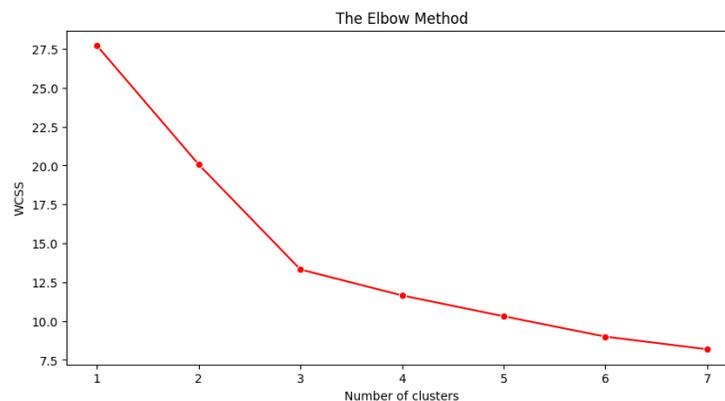
Tidak hanya standarisasi hal yang perlu dilakukan pada tahap ini yaitu melakukan normalisasi menggunakan pustaka *MinMaxScaler* yang mengubah semua nilai fitur ke dalam rentang [0, 1].

Proses ini penting untuk memastikan bahwa semua fitur memiliki skala yang sama, sehingga dapat meningkatkan kinerja model yang digunakan.

3.4 Pemodelan (Modeling)

Proses ini menggunakan metode klusterisasi dengan menggunakan algoritma *K-Means* yang mana proses tersebut mencakup, pemilihan jumlah cluster menggunakan elbow method untuk menentukan jumlah cluster yang optimal pada data ini.

Dalam analisis *clustering*, untuk tahap ini metode *Elbow* digunakan untuk menentukan jumlah *cluster* dan grafik perhitungan *WCSS* (*Within Cluster Sum of Squares*) yang ideal untuk algoritma *K-Means*. Metode ini membantu dalam menentukan nilai *K*, atau jumlah *cluster*, yang paling sesuai dengan data yang dianalisis. Sebagaimana tergambar pada Gambar 2.



Gambar 2. Hasil *Elbow Method*

Untuk menentukan jumlah *cluster* yang ideal untuk analisis *clustering*, metode *Elbow* sangat membantu untuk menentukan *centroid* yang sesuai dengan data yang diolah (Huang et al., 2024). Metode ini sangat penting dalam *aplikasi K-Means* dan membantu dalam memahami struktur data yang dianalisis. Sesuai dengan gambar dari rentang 1 hingga 7 maka dapat disimpulkan *cluster* 3 adalah *cluster* yang terbaik daripada *cluster* yang lainnya dikarenakan terjadi titik *elbow* pada nilai $K=3$

3.5 Evaluasi (Evaluation)

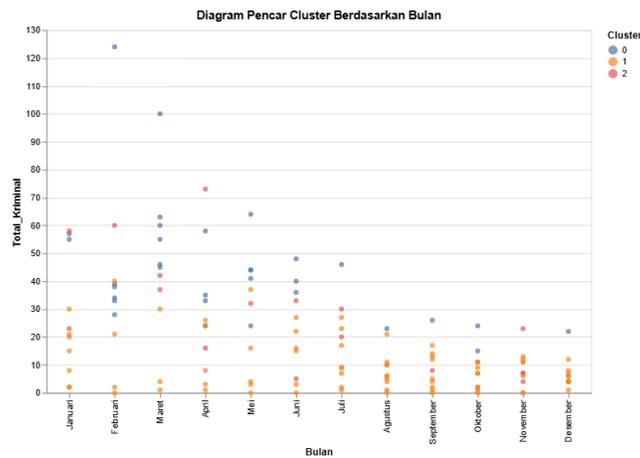
Salah satu tahapan penting yang harus dilakukan dalam data mining setelah pembuatan model ialah evaluasi, fungsi tahapan ini ialah menilai kualitas model dan memastikan bahwa itu memenuhi tujuan penelitian yang dilakukan. Evaluasi yang dilakukan dalam proses ini ialah dengan melihat hasil dari *Silhouette Score* atau biasa disebut menghitung rata-rata (*Silhouette Coefficient*) untuk semua sampel dalam data yang diolah. Dalam analisis data, skor *Silhouette* dapat disebut dengan metrik yang menunjukkan seberapa baik setiap objek dalam dataset dikelompokkan ke dalam *cluster* yang sesuai. Nilai *Silhouette Score* berkisar antara 0 dan 1 (Borlea et al., 2022). Untuk hasil dari *Silhouette Score* tertinggi untuk penelitian ialah 0.5470 untuk *cluster* 3, sebagaimana ditunjukkan dalam Tabel.3

Tabel 3. Hasil *Silhouette Score*

Cluster	Silhouette Score
2	0.50755
3	0.54706
4	0.53153
5	0.50652

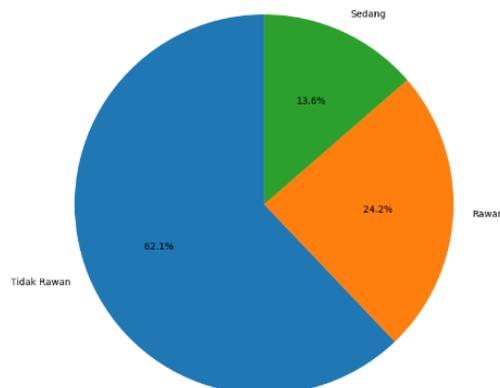
3.6 Penerapan (*Deployment*)

Hasil yang didapat dari analisis klusterisasi menunjukkan bahwa dari 11 wilayah di daerah Depok setiap bulan berhasil dikategorikan menjadi 3 kluster yang berbeda- beda, dengan masing- masing wilayah yang digambarkan dengan menggunakan diagram pencar sebagaimana Gambar 3.



Gambar 3. *Cluster* Berdasarkan Bulan

Adapun *presentase* nilai yang dapat dilihat dari hasil cluster berdasarkan kategori tindakan kriminal selama bulan Januari hingga Desember pada tahun 2023 yaitu , cluster 0: Rawan sebesar 24,2 % , cluster 1 : Tidak rawan sebesar 62,1% , dan cluster 2: Sedang sebesar 13,6 % , sebagaimana terlampir pada Gambar 4.



Gambar 4. *Presentase* kategori berdasarkan *cluster*

4 KESIMPULAN

Sebagaimana penelitian yang dilakukan maka dapat disimpulkan bahwa:

1. Analisis yang dihasilkan dalam klasterisasi berdasarkan kategori tindak kriminalitas di Kota Depok dengan menggunakan *K-Means* dari bulan Januari hingga Desember tahun 2023 dapat dikelompokkan menjadi 3 cluster yang berbeda, yaitu rawan, tidak rawan dan, sedang.
2. Penelitian ini menggunakan metode *elbow* untuk menentukan jumlah *cluster* yang tepat dengan melihat perbandingan *cluster* dari cluster 2 hingga 5 mendapatkan hasil evaluasi model dari masing – masing *cluster* ialah, *cluster* 2: 0.5075, *cluster* 3: 0.5470, *cluster* 4: 0.5315, *cluster* 5: 0.5065.
3. Evaluasi *cluster* tertinggi yang didapatkan pada penelitian ini menggunakan *Silhouette Score* sebesar 0.5470 pada *cluster* 3 yang mana hal ini menunjukkan nilai positif sehingga dapat digunakan untuk klasterisasi.
4. Sebagaimana hasil klasterisasi yang didapatkan pada penelitian ini hanya menggunakan 5 jenis tindakan kriminal dan lokasi kejadian selama tahun 2023, untuk penelitian kedepan dapat menambahkan faktor- faktor yang lain dan menggunakan metode yang lain sebagai pembanding dan menemukan metode yang lebih baik.

UCAPAN TERIMAKASIH

Penulis menyampaikan apresiasi yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, baik secara langsung maupun tidak langsung, dalam proses penelitian dan penulisan artikel ini. Di dalam artikel ini, penulis telah memperoleh banyak petunjuk, bantuan, dan bimbingan dari berbagai pihak dari mulai memilih data, mengolah data hingga artikel ini selesai. Selain itu penulis juga mendapatkan banyak dukungan dan nasihat yang membangun dari berbagai pihak ketika penulis mengalami kesulitan.

DAFTAR PUSTAKA

- Ashari, I. F., Banjarnahor, R., & Farida, D. R. (2022). *Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies*. 6(1), 7–15.
- Asyraf, H., & Prasetya, M. E. (2024). Implementasi Metode CRISP DM dan Algoritma Decision Tree Untuk Strategi Produksi Kerajinan Tangan pada UMKM A. *Jurnal Media Informatika Budidarma*, 8(1), 94. <https://doi.org/10.30865/mib.v8i1.7050>
- Bahtiar, R. (2023). *Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor*. 1(3), 203–214.
- Basiri, M. A., Parvin, M., & Sobhani, Z. (2024). *Implementing CRISP-DM and Logistic Regression for Predictive Analysis in Financial Transactions: A Case Study Annual Conference on Information Systems Implementing CRISP-DM and Logistic Regression*.
- Borlea, I., Precup, R., & Borlea, A. (2022). Improvement of K-means Cluster Quality by Post Processing Resulted Clusters. *Procedia Computer Science*, 199, 63–70. <https://doi.org/10.1016/j.procs.2022.01.009>
- Huang, Z., Zheng, H., Li, C., & Che, C. (2024). Application of Machine Learning - Based K - means Clustering for Financial Fraud Detection. *Academic Journal of Science and Technology*, 10(1), 33–39.
- Kusuma, D. A., Dwi, A., & Wicaksono, P. (2023). *ANALISIS KLASSTERING DAMPAK*

- LINGKUNGAN BERDASARKAN KONSUMSI ENERGI PERUSAHAAN BERBASIS INDUSTRI 4 . 0 MENGGUNAKAN METODE. 9(2).
<https://ejournal.poliban.ac.id/index.php/Positif/article/view/2050/1087>
- Pudoli, A., Nugroho, R., & Kusumaningsih, D. (2024). Implementasi Algoritma K-Means Clustering dalam Pengelompokan Produk Pigeon Pada PT . *Digital Niaga Indonesia Berdasarkan Analisis Recency , Frequency , Monetary (RFM)* . 6(1), 36–46.
- Ruswanti, D., Susilo, D., & Riani, R. (2024). Implementasi CRISP-DM pada Data Mining untuk Melakukan Prediksi Pendapatan dengan Algoritma C.45. *Go Infotech: Jurnal Ilmiah STMIK AUB*, 30(1), 111–121. <https://doi.org/10.36309/goi.v30i1.266>
- Santika, E. F. (2023). 10 Daerah dengan Risiko Penduduk Terkena Kejahatan Tertinggi, Jakarta Masuk Daftar. *Databoks.Katadata.Co.Id*, 2021, 1–9. <https://databoks.katadata.co.id/datapublish/2023/03/20/10-daerah-dengan-risiko-penduduk-terkena-kejahatan-tertinggi-jakarta-masuk-daftar>
- Setyadhi, I. I. (2023). *Angka Kriminalitas Tahun 2023 Meningkat 32 Persen*. Desember, 1–5. <https://depok.inilah.com/angka-kriminalitas-tahun-2023-meningkat-32-persen/>
- Singgalen, Y. A. (2024). Sentiment Classification of The Capsule Hotel Guest Reviews using Cross-Industry Standard Process for Data Mining (CRISP-DM). *Jurnal Media Informatika Budidarma*, 8(1), 632. <https://doi.org/10.30865/mib.v8i1.7329>
- Teknologi, J., Jtsi, I., Provinsi, D. I., Dengan, L., & Means, A. K.-. (2021). *ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19*. 2(2), 100–108.
- Wahidah, Z., & Utari, D. T. (2022). Implementation K-Means Algorithm to Group Provinces By Factors Influenced Criminal Act in Indonesia in 2019. *Enthusiastic : International Journal of Applied Statistics and Data Science*, 2(1), 37–46. <https://doi.org/10.20885/enthusiastic.vol2.iss1.art5>
- Yudisial, K. (1945). *PERUBAHAN KEDUA UNDANG-UNDANG DASAR NEGARA REPUBLIK INDONESIA TAHUN 1945 DENGAN*. 105(3), 129–133. <https://webcache.googleusercontent.com/search?q=cache:BDsuQOHoCi4J:https://media.neliti.com/media/publications/9138-ID-perlindungan-hukum-terhadap-anak-dari-konten-berbahaya-dalam-media-cetak-dan-ele.pdf+%&cd=3&hl=id&ct=clnk&gl=id>