

# Pengelompokan Pelanggaran Lalu Lintas Menggunakan Algoritma K-Means pada Data CCTV

Oleh:

FITRIAH,

ADE EVIYANTI

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

Januari 2025



# Pendahuluan

- Peningkatan pelanggaran lalu lintas

- Efektivitas penggunaan CCTV dan tilang online

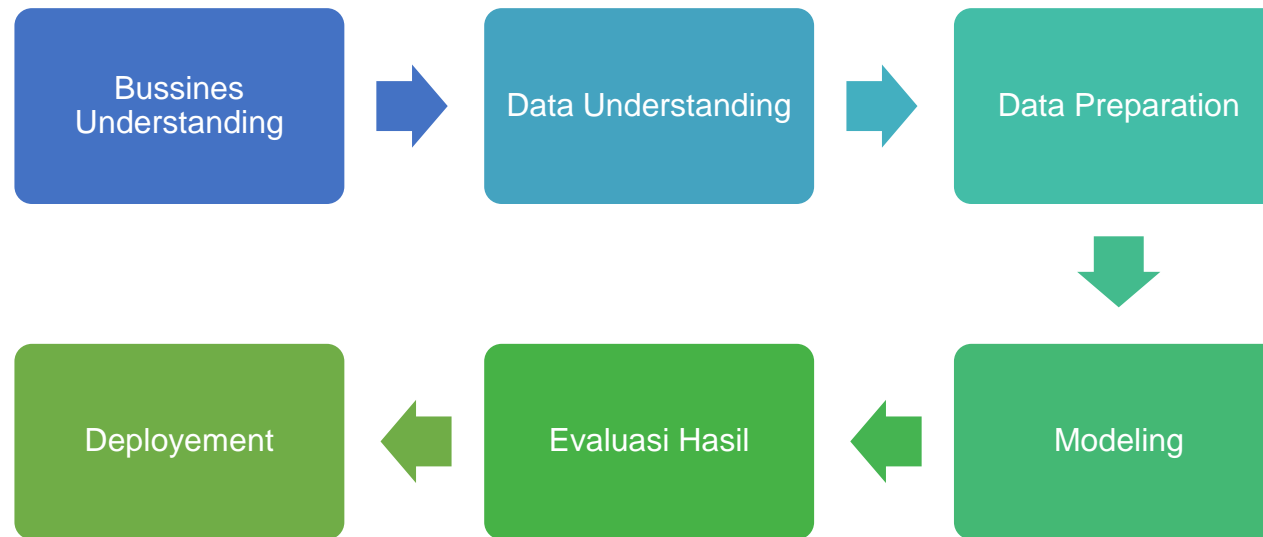
- Penerapan algoritma K-Means

# Pertanyaan Penelitian (Rumusan Masalah)

1. Bagaimana cara mengimplementasikan algoritma K-Means untuk mengelompokkan pelanggaran lalu lintas berdasarkan data CCTV?
2. Apa saja kelompok pelanggaran lalu lintas yang dapat diidentifikasi menggunakan metode ini?

# Metode Penelitian

Menggunakan metode penelitian CRISP-DM



# Metode Penelitian

## **Bussines Understanding**

Pada tahap awal, dilakukan pemahaman bisnis dan penjelasan mengenai masalah data mining agar tujuan penelitian dapat tercapai. Pemahaman masalah dalam penelitian ini fokus pada pelanggaran lalu lintas di Sidoarjo. Dikarenakan pelanggaran yang terjadi di Sidoarjo cukup tinggi, sehingga menjadi salah satu sektor penting untuk mendapatkan perhatian lebih atau perlakuan khusus, untuk dapat mengurangi angka pelanggaran lalu lintas yang terjadi dan demi keselamatan masyarakat sidoarjo.

## **Data Understanding**

Pada tahapan data understanding ini maka peneliti mencoba untuk memahami data yang akan digunakan. Data pelanggaran lalu lintas yang digunakan dalam penelitian ini berasal dari Januari 2023 hingga Juli 2024 di Kota Sidoarjo. Data ini mencakup berbagai jenis pelanggaran lalu lintas, seperti tidak memakai helm, melawan arah, dan lain-lain. Selain itu, data juga mencakup lokasi pelanggaran yang terdeteksi oleh CCTV di berbagai wilayah di Kota Sidoarjo.

## **Data Preparation**

Pada tahap persiapan data, langkah ini melibatkan mempersiapkan data yang telah dikumpulkan untuk proses data mining. Tahap ini mencakup pembersihan data, penyajian data, dan transformasi data. Dalam proses clustering ini menggunakan Bahasa pemrograman Python. Python merupakan sebuah bahasa pemrograman yang memiliki tingkat kesulitan yang tinggi, bersifat dinamis, interpretatif, serta memiliki kemampuan yang sangat bermanfaat dalam berbagai keperluan. Python juga populer digunakan untuk melakukan data science, machine learning, dan berbagai macam analisis data lainnya.

# Metode Penelitian

## Modeling

Modeling melibatkan penggunaan model pembelajaran mesin (machine learning) untuk memahami pola data dan memperoleh wawasan. Wawasan bisa berupa visualisasi pola atau prediksi nilai masa depan. Tahapan ini merupakan bagian paling menarik dari proyek ilmu data (data science) karena melibatkan pembelajaran mesin sebagai salah satu komponen terpentingnya. Pada tahap ini, kita akan mengeksekusi data dengan menggunakan algoritma clustering K-Means dengan bantuan software Google Colaboratory.

## Evaluasi Hasil

Setelah melakukan tahap pemodelan, tahap selanjutnya adalah evaluasi untuk mengukur model yang paling baik untuk digunakan. Evaluasi yang digunakan yaitu dengan teknik silhouette score .

## Deployment

Setelah tahap evaluasi dimana menilai secara detail hasil dari sebuah model maka dilakukan pengimplementasian dari keseluruhan model yang telah dibangun. Selain itu juga dilakukan penyesuaian terhadap model sehingga dapat menghasilkan suatu hasil yang sesuai dengan target awal tahap CRISP-DM ini . Dan tahap ini merupakan proses pembuatan laporan atau artikel jurnal menggunakan hasil penelitian.

# Hasil

Data yang diambil dari POLRES Sidoarjo berupa file excel. Proses untuk menarik data dengan cara mengambil data yang tersimpan pada cctv . Berikut pada Tabel merupakan datasheet yang diambil dari Polres Sidoarjo pada Januari 2023 hingga Juli 2024 :

No.	Lokasi	Plat	Pemilik	Warna Kendaraan	Tipe Kendaraan	Tanggal Pelanggaran	Jenis Pelanggaran
1	Jl. Pahlawan	S1072WU	KIKI NOVIANTO PT. PUSAKA	ABU ABU METALIK	Unknown	2023-07-28 04:02:21	Menerobos lampu merah
2	Jl. Teuku Umar	L1456PO	PRIMA TRANSPORT	HITAM METALIK	Unknown	2023-01-03 08:51:14	Menerobos lampu merah
3	Jl. Pahlawan	W1294QV	CHOIRUL WAHIB	PUTIH METALIK	Unknown	2023-01-03 08:50:47	Menerobos lampu merah
4	Jl. Teuku Umar	W9428NC	LUKINDARI PERMATA	HITAM (KANZAI)	Light-Duty Vehicle	2023-01-03 09:13:46	Melanggar rambu atau marka
5	Jl. Teuku Umar	W1438VH	SUBUR SANTUSO	ABU ABU METALIK	Unknown	2023-01-03 09:19:51	Melanggar rambu atau marka
6	Jl. Teuku Umar	L9115BA	KARYANI	PUTIH	Light-Duty Vehicle	2023-01-03 09:19:49	Melanggar rambu atau marka
7	Jl. Pahlawan	W5863WS	MUHAMAD ALI MASKUR	WHITE SILVER	Pedestrian	2023-01-03 06:33:04	Tidak mengenakan Helm
8	Jl. Pahlawan	W3685XW	LIVIA MEDI MARTHASARI	HITAM	Pedestrian	2023-01-03 08:41:39	Tidak mengenakan Helm
9	Jl. Pahlawan	L3796PI	AINUL ROFIQ	ABU-ABU GLP	Pedestrian	2023-01-03 08:08:13	Tidak mengenakan Helm
10	Jl. Pahlawan	W4019VT	HILDA NAFISAH	ABU-ABU	Pedestrian	2023-01-03 06:33:22	Tidak mengenakan Helm

# Hasil (Seleksi data )

Berdasarkan datasheet yang ada terdapat 8 atribut namun yang digunakan dalam penelitian ini hanya 3 atribut sehingga diperlukan seleksi data.

Lokasi	Tipe Kendaraan	Jenis Pelanggaran
Jl. Pahlawan	Unknown	Menerobos lampu merah
Jl. Teuku Umar	Unknown	Menerobos lampu merah
Jl. Pahlawan	Unknown	Menerobos lampu merah
Jl. Teuku Umar	Light-Duty Vehicle	Melanggar rambu atau marka
Jl. Teuku Umar	Unknown	Melanggar rambu atau marka
Jl. Teuku Umar	Light-Duty Vehicle	Melanggar rambu atau marka
Jl. Pahlawan	Pedestrian	Tidak mengenakan Helm
Jl. Pahlawan	Pedestrian	Tidak mengenakan Helm
Jl. Pahlawan	Pedestrian	Tidak mengenakan Helm
Jl. Pahlawan	Pedestrian	Tidak mengenakan Helm



# Hasil (pre-processing data)

Tahap processing ini dilakukan untuk membersihkan data yang tidak bisa untuk memasuki proses data mining, seperti data missing value. Berdasarkan dalam Tabel sebelumnya terdapat data yang tidak relevan seperti pada atribut tipe\_kendaraan = unknown atau pedestrian sehingga diharuskan penghapusan data.

Lokasi	Tipe Kendaraan	Jenis Pelanggaran
Simpang 4 krian 1	Mobil	Tidak menggunakan sabuk pengaman
Simpang 4 krian 1	Truk	Tidak menggunakan sabuk pengaman
Simpang 4 krian 1	Truk	Tidak menggunakan sabuk pengaman
Simpang 4 krian 1	Mobil	Tidak menggunakan sabuk pengaman
Simpang 4 krian 1	Mobil	Tidak menggunakan sabuk pengaman
Simpang 4 krian 1	Mobil	Tidak menggunakan sabuk pengaman
Simpang 4 krian 1	Truk	Tidak menggunakan sabuk pengaman
Simpang 4 krian 1	Truk	Tidak menggunakan sabuk pengaman
Simpang 4 krian 1	Truk	Tidak menggunakan sabuk pengaman
Simpang 4 krian 1	Truk	Tidak menggunakan sabuk pengaman
Simpang 4 krian 1	Mobil	Tidak menggunakan sabuk pengaman

# Hasil (Transformasi Data)

Transformasi data dilakukan untuk mengubah data atau dilakukan proses inisialisasi dimana data diubah kedalam bentuk angka atau numerik agar data dapat diolah secara baik menggunakan metode K-Means

Lokasi	Transformasi
Jalan Raya Kletek	0
Simpang 4 krian 1	1

Tipe Kendaraan	Transformasi
Bus	0
Mobil	1
Motor	2
Truk	3

Jenis Pelanggaran	Transformasi
Berbonceng 3 (Muatan)	0
Berbonceng 3 (Muatan), Tidak mengenakan Helm	1
Melanggar rambu atau marka	2
Menerobos lampu merah	3
Menggunakan Handphone / Mengemudi tidak wajar, Tidak menggunakan sabuk pengaman	4
Penumpang tidak menggunakan helm	5
Tidak mengenakan Helm	6
Tidak mengenakan Helm, Melanggar rambu atau marka	7
Tidak menggunakan sabuk pengaman	8

Lokasi	Tipe Kendaraan	Jenis Pelanggaran
1	1	8
1	3	8
1	3	8
1	1	8
1	1	8
1	1	8
1	3	8
1	3	8
1	3	8
1	3	8
1	3	8

# Hasil

Selanjutnya mengimplementasikan algoritma K-Means Clustering, data yang telah dikumpulkan dan ditransformasikan mulai dikelompokkan menggunakan algoritma K-means Clustering dengan menentukan titik pusat klaster (centroid) yang dipilih secara acak

```
[[0.86689715 0.98538683 8. ]  
 [0.90133897 2.08070472 5.88495419]  
 [0.85226188 3.      8. ]]
```

Berdasarkan data yang ditampilkan pada Gambar diatas, terdapat tiga centroid yang masing-masing memiliki koordinat tertentu. Centroid yang pertama [0.86698715, 0.98538683, 8.], cenroid kedua [9.96113987, 2.08074627, 5.84945194] dan centroid yang ke tiga [8.8522168, 3., 8.], Koordinat ini akan digunakan sebagai titik awal untuk proses iterasi selanjutnya dalam pengelompokan data.

# Hasil

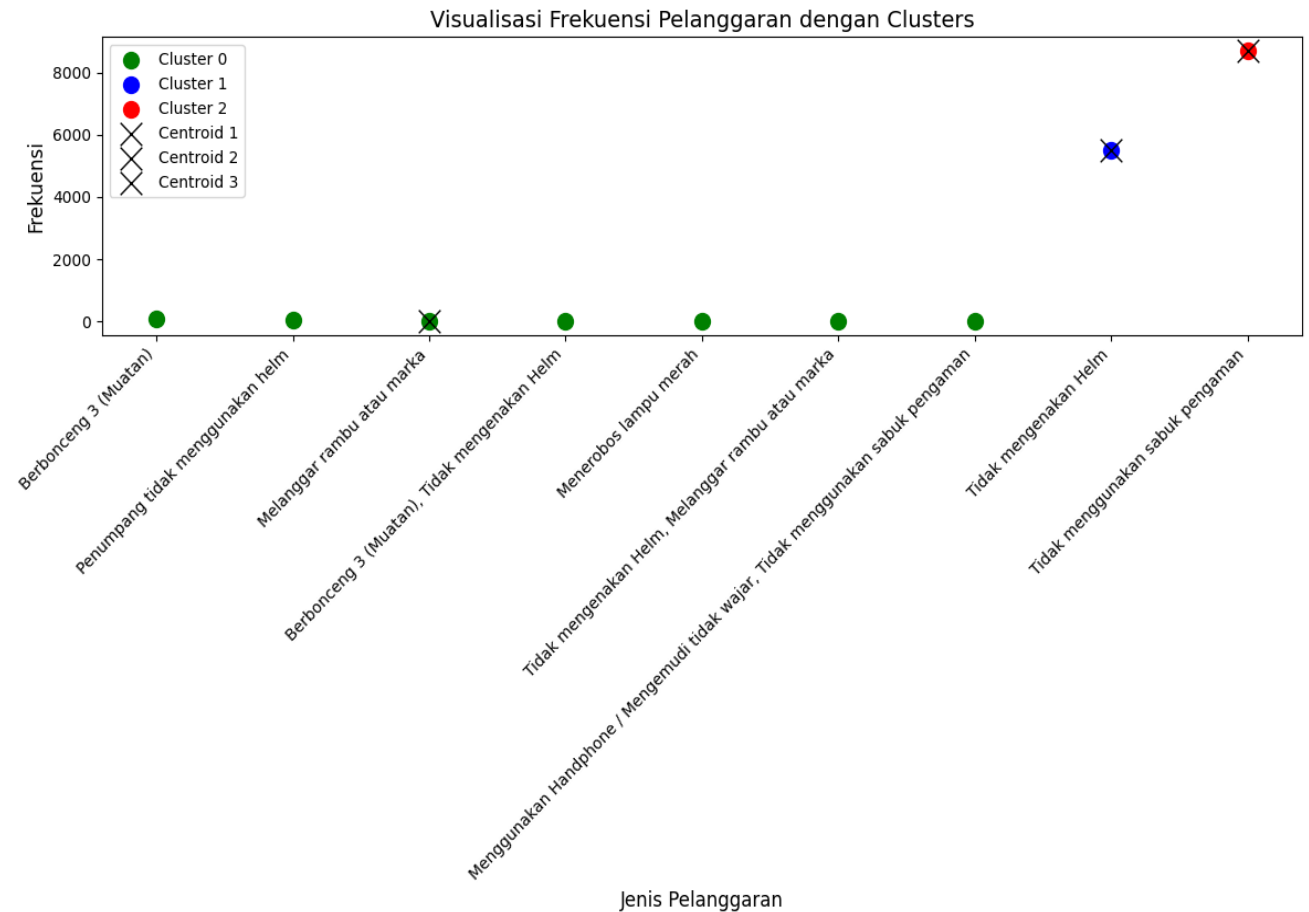
## Hasil clustering

Jenis Pelanggaran	Frequency	Cluster	Skala Pelanggaran
8	8710	2	Pelanggaran Tertinggi
6	5522	1	Pelanggaran Sedang
0	84	0	Pelanggaran Rendah
5	40	0	Pelanggaran Rendah
2	12	0	Pelanggaran Rendah
1	8	0	Pelanggaran Rendah
3	7	0	Pelanggaran Rendah
7	2	0	Pelanggaran Rendah
4	1	0	Pelanggaran Rendah

Pengelompokan Pelanggaran Lalu Lintas pada Data CCTV menghasilkan 3 cluster uji. Terlihat pada Tabel disamping skala pelanggaran tertinggi terdapat pada cluster 2 jenis pelanggaran yang terjadi yaitu tidak menggunakan sabuk pengaman memiliki 8710 data, pelanggaran sedang terdapat pada cluster 1 memiliki 5522 data jenis pelanggaran yang terjadi yaitu tidak mengenakan helm dan yang berada diposisi ke 3 yaitu cluster 0 memiliki 84 data jenis pelanggaran yang terjadi yaitu Berbonceng 3 (Muatan), diikuti dengan Penumpang tidak menggunakan helm dengan 40 data, Melanggar rambu atau marka 12 data, (Berbonceng 3 (Muatan), Tidak mengenakan Helm) 8 data, Menerobos lampu merah 7 data, (Tidak mengenakan Helm, Melanggar rambu atau marka) 2 data dan (Menggunakan Handphone / Mengemudi tidak wajar, tidak menggunakan sabuk pengaman) 1 data.

# Hasil

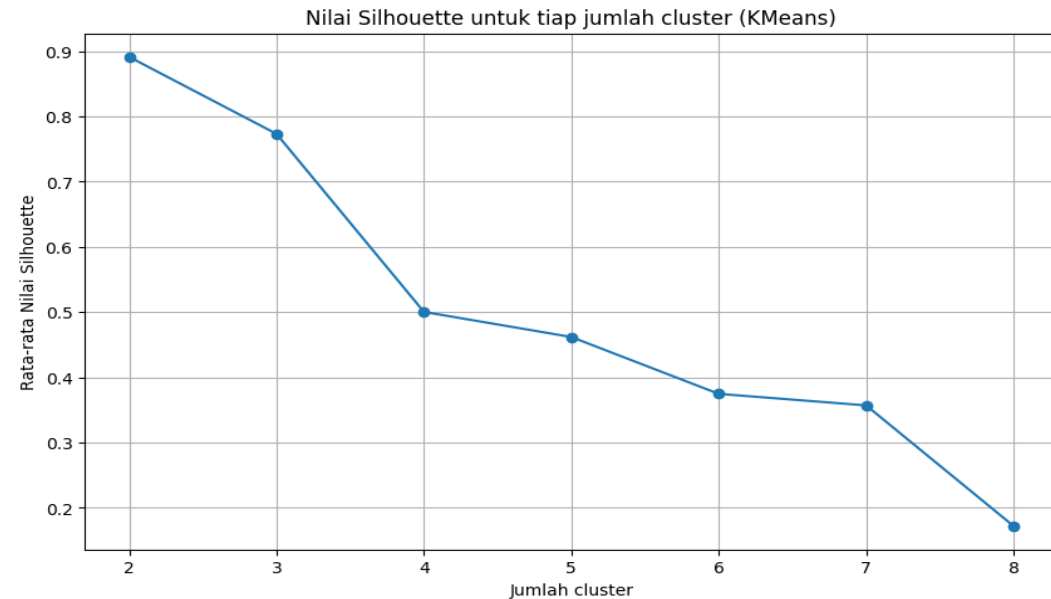
Pada visualisasi hasil clustering gambar 3, digambarkan dengan warna yang berbeda untuk setiap clusternya dan juga dapat diketahui untuk setiap letak titik centroid setiap clusternya. Untuk centroid 0 berwarna hijau termasuk dalam kategori pelanggaran rendah, cluster 1 berwarna biru termasuk kategori pelanggaran sedang dan cluster 2 berwarna merah termasuk dalam kategori pelanggaran tertinggi.



# Hasil

- Evaluasi hasil clustering

Data Uji	Nilai K	Nilai Rata-Rata Silhouette Score
14386	2	0.8913713235651775
	3	0.7733209565448501
	4	0.5004615786709502
	5	0.46197074733258936
	6	0.37483349983349984
	7	0.3571484071484072
	8	0.1722222222222222



Berdasarkan evaluasi menggunakan metrik Silhouette Score menggunakan python maka diperoleh 8 cluster yang dinilai dapat mengelompokan data dengan baik. Hasil evaluasi cluster terdapat k terbaik yaitu 0,8913 pada K=2 yang dapat dilihat pada Gambar grafik diatas.

# Hasil (perbandingan algoritma)

Algoritma	Sillhoutte Score
K-Means	0.8913713235651775
DBSCAN	0.37543770067855425

Hasil Pengujian validitas cluster terhadap algoritma DBSCAN dan K-Means menggunakan nilai Silhouette Index (SI) sudah dilakukan. Klasterisasi menggunakan K-Means, nilai SI terbaik diperoleh percobaan k=2 dengan nilai 0,8913, Sementara dari hasil klasterisasi menggunakan DBSCAN yang memiliki nilai SI terbaik yaitu 0.3754 dengan nilai Eps 5,00. Maka pada penelitian ini, algoritma K-Means memiliki nilai validitas cluster lebih baik dibandingkan algoritma DBSCAN

# Pembahasan

Penelitian ini membuktikan efektivitas *K-Means Clustering* dalam mengelompokkan data pelanggaran lalu lintas di Sidoarjo berdasarkan tingkat keparahan, dengan pelanggaran "tidak menggunakan sabuk pengaman" sebagai yang paling sering terjadi. Dibandingkan DBSCAN dan Apriori, K-Means unggul dalam menangani dataset besar tanpa label, meski sensitif terhadap ketidakseimbangan data. Data dari CCTV POLRES Sidoarjo menghasilkan rekomendasi strategis seperti pengawasan pelanggaran tinggi, patroli di lokasi pelanggaran rendah, dan kampanye edukasi. Penelitian ini memberikan panduan berbasis data untuk membantu Satlantas POLRES Sidoarjo meningkatkan keselamatan lalu lintas.



# Manfaat Penelitian

Sehingga tujuan dari penelitian ini dapat diketahui jenis pelanggaran seperti apa yang menyumbangkan penyebab kecelakaan lalu lintas juga dapat memberikan informasi dari pihak Satlantas POLRES Sidoarjo memberikan penanganan dalam kasus pelanggaran lalu lintas di Sidoarjo sehingga dapat disusun program yang sesuai guna mengurangi tingkat pelanggaran lalu lintas yang terjadi di Sidoarjo, dengan lebih memperhatikan atau memberi tindakan secara ketat atau perhatian khusus pada skala pelanggaran tertinggi seperti pada penelitian ini yaitu tidak menggunakan sabuk pengaman.

# Referensi

- [1] R. Manaek, Richardus Eko Indrajit, and Erick Dazki, "Arsitektur Perusahaan Untuk Infrastruktur Telekomunikasi Di Daerah Pedalaman Indonesia," *SATIN - Sains Dan Teknol. Inf.*, vol. 9, no. 2, pp. 01–11, Dec. 2023, doi: 10.33372/stn.v9i2.1000.
- [2] D. Wahyuni and A. Arianto, "PENERAPAN DATA MINING PADA DATA PELANGGARAN LALU LINTAS MENGGUNAKAN METODE K-MEANS CLUSTERING (STUDI KASUS : PENGADILAN NEGERI DUMAI)".
- [3] E. S. Wasih, S. Rahmatullah, and S. Mintoro, "Implementasi Data Mining Pada Data Pelanggaran Lalu Lintas di Lampung Utara Menggunakan Algoritma K-Means (Studi Kasus Kejaksaan Negeri Lampung Utara)," 2022.
- [4] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *J. Artif. Intell. Technol. Inf. JAITI*, vol. 1, no. 1, pp. 20–28, Feb. 2023, doi: 10.58602/jaiti.v1i1.22.
- [5] D. Sartika and J. Jumadi, "IMPLEMENTASI DATA MINING UNTUK PENGELOMPOKKAN WILAYAH PELANGGARAN LALU LINTAS MENGGUNAKAN METODE K-MEANS PADA POLRES BENGKULU".
- [6] R. Saragih, S. Kom, M. Kom, J. N. Sitompul, and M. Pd, "Analisis Perbandingan Data Mining Mengidentifikasi Pola Keterkaitan Variabel Kecelakaan Lalu Lintas Di Polresta Kota Medan," *Inf. Syst. Dev.*, vol. 4.
- [7] J. Rahmasari, "HALAMAN PERNYATAAN KEASLIAN SKRIPSI".
- [8] I. Budiman, T. Prahasto, and Y. Christyono, "Data Clustering Menggunakan Metodologi CRISP-DM Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma," *J. Sist. Inf. BISNIS*, vol. 1, no. 3, pp. 129–134, Jan. 2014, doi: 10.21456/vol1iss3pp129-134.
- [9] A. D. Adhi Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI J. Tek. Inform. Dan Sist. Inf.*, vol. 8, no. 2, pp. 636–646, Jun. 2021, doi: 10.35957/jatisi.v8i2.962.
- [10] M. Y. Nurzaman and B. N. Sari, "Implementasi K-Means Clustering Dalam Pengelompokkan Banyaknya Jumlah Petani Berdasarkan Kecamatan Di Provinsi Jawa Barat," vol. 10, no. 3, 2023.

# Referensi (Lanjutan)

- [11] A. Pambudi, "PENERAPAN CRISP-DM MENGGUNAKAN MLR K-FOLD PADA DATA SAHAM PT. TELKOM INDONESIA (PERSERO) TBK (TLKM) (STUDI KASUS: BURSA EFEK INDONESIA TAHUN 2015-2022)," *J. Data Min. Dan Sist. Inf.*, vol. 4, no. 1, p. 1, Mar. 2023, doi: 10.33365/jdmsi.v4i1.2462.
- [12] I. Fitrianti, A. Voutama, and Y. Umaidah, "Clustering Film Populer Pada Aplikasi Netflix Dengan Menggunakan Algoritma K-Means Dan Metode CRISP- DM," vol. 4, no. 2.
- [13] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *J. Teknol. Dan Inf.*, vol. 12, no. 1, pp. 64–77, Mar. 2022, doi: 10.34010/jati.v12i1.6674.
- [14] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *J. Appl. Inform. Comput.*, vol. 5, no. 2, pp. 103–108, Oct. 2021, doi: 10.30871/jaic.v5i2.3200.
- [15] A. S. Ritonga and I. Muhandhis, "Clustering Data Tweet E-Commerce Menggunakan Metode K-Means (Studi Kasus Akun Twitter Blibli Indonesia)," *SMATIKA J.*, vol. 12, no. 01, pp. 75–84, Jun. 2022, doi: 10.32664/smatika.v12i01.665.
- [16] N. Saniyah, N. Suarna, and W. Prihartono, "CLUSTERING PELANGGARAN LALU LINTAS PADA KENDARAAN BERMOTOR MENGGUNAKAN ALGORITMA K-PROTOTYPE (STUDI KASUS: PENGADILAN NEGERI CIREBON)," vol. 8, no. 1, 2024.
- [17] F. A. Nisa, A. Susanto, E. R. Pramudya, and U. W. Mulyono, "KLAUSTERISASI PERKARA PELANGGARAN LALU LINTAS MENGGUNAKAN ALGORITMA K-MEANS DAN DAVIES- BOULDIN INDEX," 2020.
- [18] M. R. Muttaqin, T. I. Hermanto, and M. A. Sunandar, "PENERAPAN K-MEANS CLUSTERING DAN CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM) UNTUK MENGELOMPOKAN PENJUALAN KUE," 2022.
- [19] I. F. Anshori and Y. Nuraini, "Pengelompokan Data Kecelakaan Lalu Lintas di Kota Tasikmalaya Menggunakan Algoritma K-Means," *J. Responsif Ris. Sains Dan Inform.*, vol. 2, no. 1, pp. 118–127, Mar. 2020, doi: 10.51977/jti.v2i1.198.
- [20] R. Adha, N. Nurhaliza, and U. Soleha, "Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia," vol. 18, no. 2, 2021.

