

Clustering Traffic Violations Using K-Means Algorithm on CCTV Data

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

Fitriah¹⁾, Ade Eviyanti²⁾, Hindarto³⁾, Novia Ariyanti⁴⁾

^{1,2,3,4)}Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi : adeeviyanti@umsida.ac.id

Abstract. *This study aims to cluster traffic violation data recorded by CCTV in the Sidoarjo area using the K-Means Clustering algorithm. The dataset used in this study was obtained from Sidoarjo Police, covering 43,055 traffic violation records in the period January 2023 to July 2024. The CRISP-DM approach is applied to ensure a systematic research flow, starting from problem understanding, data collection, to result evaluation. After the data selection and transformation stage, the dataset was processed into 14,386 data. Clustering was performed to divide violations into three categories based on severity, namely high, medium, and low. Evaluation of cluster quality using Silhouette Score showed the best result with a value of 0.8913 at $k=2$, indicating optimal cluster formation. The clustering results showed that the highest violation occurred in the category of "not using a seat belt" with 8,710 cases, while the moderate violation involved "not wearing a helmet" with 5,522 cases. This study confirms the effectiveness of the K-Means algorithm in clustering traffic violation data and provides valuable insights for the Sidoarjo Police Traffic Unit in designing more efficient traffic violation reduction programs.*

Keywords - Traffic Violations; CCTV, K-Means Clustering; CRISP-DM; Silhouette Score; Sidoarjo Police Station

Abstrak. *Penelitian ini bertujuan untuk mengelompokkan data pelanggaran lalu lintas yang terekam oleh CCTV di wilayah Sidoarjo menggunakan algoritma K-Means Clustering. Dataset yang digunakan dalam penelitian ini diperoleh dari Polresta Sidoarjo, mencakup 43.055 catatan pelanggaran lalu lintas pada periode Januari 2023 hingga Juli 2024. Pendekatan CRISP-DM diterapkan untuk memastikan alur penelitian yang sistematis, dimulai dari pemahaman masalah, pengumpulan data, hingga evaluasi hasil. Setelah tahap seleksi dan transformasi data, dataset diproses menjadi 14.386 data. Klusterisasi dilakukan untuk membagi pelanggaran ke dalam tiga kategori berdasarkan tingkat keparahan, yaitu tinggi, sedang, dan rendah. Evaluasi kualitas kluster menggunakan Silhouette Score menunjukkan hasil terbaik dengan nilai 0.8913 pada $k=2$, mengindikasikan pembentukan kluster yang optimal. Hasil klusterisasi menunjukkan bahwa pelanggaran tertinggi terjadi pada kategori "tidak menggunakan sabuk pengaman" dengan 8.710 kasus, sedangkan pelanggaran sedang melibatkan "tidak mengenakan helm" dengan 5.522 kasus. Penelitian ini menegaskan efektivitas algoritma K-Means dalam mengelompokkan data pelanggaran lalu lintas dan memberikan wawasan berharga bagi Satlantas POLRES Sidoarjo dalam merancang program pengurangan pelanggaran lalu lintas secara lebih efisien.*

Kata Kunci - Pelanggaran Lalu Lintas; CCTV; K-Means Clustering; CRISP-DM; Silhouette Score; Polresta Sidoarjo

1. Pendahuluan

Transportasi adalah sarana yang mempermudah aktivitas manusia, berfungsi untuk memindahkan barang atau orang dari satu lokasi ke lokasi lainnya menggunakan kendaraan yang digerakkan oleh manusia atau mesin. Peningkatan jumlah kendaraan yang signifikan berpotensi meningkatkan tingkat kemacetan. Semakin parah kemacetan di suatu wilayah, semakin tinggi pula kemungkinan terjadinya pelanggaran, seperti melanggar rambu lalu lintas, tidak memakai helm, kurangnya kelengkapan surat-surat berkendara, berkendara melawan arah, dan pelanggaran lainnya[1].

Penyebab pelanggaran lalu lintas sampai kecelakaan, dikarenakan masih rendahnya pengetahuan pengendalian akan kedisiplinan berlalu lintas dan sedikitnya pemahaman para pemakai jalan terhadap peraturan lalu lintas[2]. Undang undang Nomor 22 Tahun 2009 tentang Lalu Lintas dan Angkutan Jalan, mengungkapkan kecelakaan lalu lintas adalah suatu peristiwa di jalan yang tidak

Copyright © Universitas Muhammadiyah Sidoarjo. This preprint is protected by copyright held by Universitas Muhammadiyah Sidoarjo and is distributed under the Creative Commons Attribution License (CC BY). Users may share, distribute, or reproduce the work as long as the original author(s) and copyright holder are credited, and the preprint server is cited per academic standards.

Authors retain the right to publish their work in academic journals where copyright remains with them. Any use, distribution, or reproduction that does not comply with these terms is not permitted.

diduga dan tidak disengaja melibatkan kendaraan dengan atau tanpa pengguna jalan lain yang mengakibatkan korban manusia dan/atau kerugian harta benda [3].

Berdasarkan data yang peneliti peroleh dari POLRES Sidoarjo menunjukkan bahwa pada periode Januari hingga Desember 2023, terjadi 27.906 pelanggaran lalu lintas di wilayah tersebut. Pada periode Januari hingga Juli 2024, tercatat 15.149 pelanggaran lalu lintas yang semuanya terekam melalui CCTV. Menurut salah satu anggota Satlantas POLRES Sidoarjo, Bapak Akhmad Syamsul Rizal, tindakan terhadap pelanggaran ini dilakukan melalui surat tilang online yang dikirim sesuai dengan alamat yang terdeteksi pada data CCTV.

Meskipun demikian, masih banyak yang melanggar aturan lalu lintas. Di salah satu wilayah Kabupaten Sidoarjo dengan tingkat pelanggaran yang tinggi berdasarkan data CCTV yaitu di Jl. Simpang 4 Krian, berbagai upaya telah dilakukan oleh kepolisian dan dinas perhubungan untuk menertibkan lalu lintas, termasuk razia penilangan untuk memeriksa kelengkapan surat-surat kendaraan, kelayakan kendaraan, serta memberikan himbauan melalui spanduk. Penerapan aturan lalu lintas ini bertujuan untuk menciptakan, mendukung, dan menjaga keamanan, keselamatan, keteraturan, serta kelancaran lalu lintas[1].

Algoritma K-Means clustering adalah algoritma unsupervised learning yang berfungsi untuk mengelompokkan data dalam dataset yang tidak memiliki label menjadi beberapa cluster yang berbeda. Dengan algoritma ini, data dapat dikelompokkan ke dalam cluster berdasarkan variabel yang tersedia tanpa memerlukan proses pelatihan data sebelumnya. Tujuan utama K-Means clustering adalah mengurangi jarak antara setiap titik data dengan cluster yang sesuai.[4].

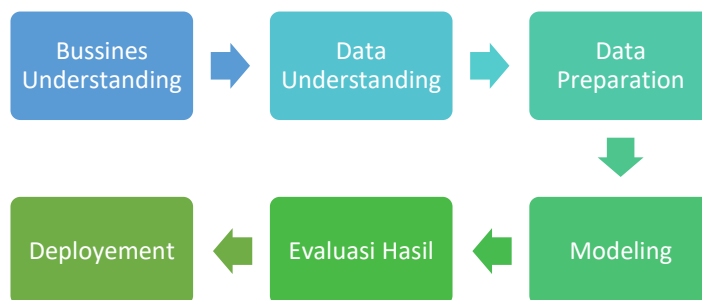
Menurut penelitian [5] Penelitian ini membahas penerapan data mining untuk pengelompokan wilayah pelanggaran lalu lintas menggunakan metode K-Means pada Polres Bengkulu. Dengan memanfaatkan algoritma K-Means Clustering, penelitian ini berhasil mengelompokkan wilayah berdasarkan tingkat pelanggaran menjadi tiga kategori: tinggi, sedang, dan rendah. Wilayah dengan tingkat pelanggaran tinggi, seperti Tanah Patah dan Betungan, diidentifikasi sebagai prioritas untuk penegakan hukum lebih lanjut. Penelitian ini menunjukkan bahwa metode K-Means efektif dalam mengelompokkan data dan dapat membantu pihak kepolisian dalam menganalisis pola pelanggaran serta merancang strategi penegakan hukum yang lebih tepat sasaran. Namun pada penelitian ini membuat aplikasi clustering sehingga tidak membahas evaluasi clusteringnya yang digunakan untuk mengetahui keakuratan hasilnya.

Beberapa studi tentang analisis pola kecelakaan lalu lintas menggunakan algoritma Apriori telah dilakukan, termasuk penelitian oleh Agus di Kota Yogyakarta. Penelitian-penelitian ini menerapkan berbagai pendekatan dengan itemset yang berbeda, seperti merek kendaraan, jenis kendaraan, umur pengemudi, dan faktor lainnya. Apriori adalah metode yang digunakan untuk mengidentifikasi frequent itemset dalam database besar. Dengan menerapkan Apriori, peneliti dapat menemukan pasangan item yang sering muncul, sehingga pola keterkaitan antar item dapat diidentifikasi. Dalam konteks data mining, Apriori merupakan algoritma klasik yang digunakan untuk mempelajari dan menganalisis aturan asosiasi[6].

Berdasarkan penelitian sebelumnya menunjukkan bahwa algoritma K-Means Clustering merupakan salah satu metode clustering yang populer, mudah digunakan, dan efisien dalam mengelompokkan data dalam jumlah besar yang dapat membantu Satlantas POLRES Sidoarjo dalam mengidentifikasi dan mengelompokkan jenis pelanggaran lalu lintas di Kota Sidoarjo berdasarkan dataset yang sudah ada. Maka dari itu penelitian ini bertujuan untuk mengelompokkan data pelanggaran lalu lintas yang terekam oleh CCTV di Sidoarjo menggunakan algoritma K-Means. Sehingga dari penelitian ini diharapkan hasil yang didapat bisa menjadi informasi bagi Satlantas POLRES Sidoarjo dalam menangani kasus pelanggaran lalu lintas di Sidoarjo sehingga dapat disusun program yang sesuai guna mengurangi tingkat pelanggaran lalu lintas yang terjadi di Sidoarjo [7].

2. Metode Penelitian

Dalam penelitian ini teknik data mining clustering digunakan dengan menerapkan algoritma K-means. Pada Gambar 1 merupakan pendekatan CRISP-DM yang digunakan sebagai alur penelitian yang terdiri dari beberapa langkah, termasuk 5 tahapan; tahap pemahaman bisnis (business understanding), tahap pemahaman data (data understanding), tahap persiapan data (data preparation), tahap pemodelan (modelling) dan tahap interpretasi hasil (evaluation dan deployment) [8]:



Gambar 1 Alur penelitian CRISP-DM

Berikut ini merupakan informasi mengenai alur penelitian pada Gambar 1 :

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 [9]. 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[10].

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[11]. Pada tahap ini, kita akan mengeksekusi data dengan menggunakan algoritma clustering K-Means dengan bantuan software Google Colaboratory[12].

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 [13].

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 [14]. Dan tahap ini merupakan proses pembuatan laporan atau artikel jurnal menggunakan hasil penelitian[13].

3. Hasil dan Pembahasan

Data yang diambil dari POLRES Sidoarjo berupa file excel. Proses untuk menarik data dengan cara mengambil data yang tersimpan pada cctv [15]. Berikut pada Tabel 1 merupakan datasheet yang diambil dari Polres Sidoarjo pada Januari 2023 hingga Juli 2024 :

Tabel 1 Dataset pelanggaran lalu lintas

No.	Lokasi	Plat	Pemilik	Warna Kendaraan	Tipe Kendaraan	Tanggal Pelanggaran	Jenis Pelanggaran
1	Jl. Pahlawan	S1072WU	KIKI NOVIANTO	ABU ABU METALIK	Unknown	2023-07-28 04:02:21	Menerobos lampu merah
2	Jl. Teuku Umar	L1456PO	PT. PUSAKA 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

Seleksi Data

Proses seleksi data adalah tahap pemilihan data yang sesuai dengan kebutuhan. Pada tahap ini, dipilih field yang relevan untuk keperluan penelitian[1].

Berdasarkan datasheet pada Tabel 1 terdapat 8 atribut nsamun yang digunakan dalam penelitian ini hanya 3 atribut sehingga diperlukan seleksi data. Berikut ini Tabel 2 hasil dari seleksi data:

Tabel 2 Hasil 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

Pre-processing Data

Setelah memilih atribut data yang diperlukan, langkah selanjutnya adalah tahap preprocessing. Tahap processing ini dilakukan untuk membersihkan data yang tidak bisa untuk memasuki proses data mining, seperti data missing value [16]. Tahap ini sangat penting karena merupakan langkah awal yang akan mempengaruhi hasil akhir dalam pembentukan pola pengelompokan. Selama proses ini, jumlah data akan berkurang karena data yang kurang relevan akan dihapus[1].

Berdasarkan dalam Tabel 2 terdapat data yang tidak relevan seperti pada atribut tipe_kendaraan = unknown atau pedestrian sehingga diharuskan penghapusan data. Berikut pada Tabel 3 merupakan hasil preprocessing data :

Tabel 3 Cleaning 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

Pada Tabel 3 sudah melewati tahap cleaning sehingga data yang tidak diperlukan sudah bersih untuk melanjutkan pada tahapan selanjutnya yaitu transformasi data.

Transformasi Data

Transformasi data dilakukan untuk mengubah data atau dilakukan proses inialisasi dimana data diubah kedalam bentuk angka atau numerik agar data dapat diolah secara baik menggunakan metode K-Means. Berikut merupakan tabel inialisasi dari atribut data [17]. Teknik transformasi yang digunakan adalah Label Encoder. Berikut pada Tabel 4, 5, 6 dan Tabel 7 merupakan hasil dari transformasinya :

Tabel 4 Transformasi Data

Lokasi	Transformasi
Jalan Raya Kletek	0
Simpang 4 krian 1	1

Pada data lokasi terdiri dua jenis data yaitu Jalan Raya Kletek dan Simpang 4 krian 1, di transformasi seperti Tabel 4.

Tabel 5 Transformasi Tipe Kendaraan

Tipe Kendaraan	Transformasi
Bus	0
Mobil	1
Motor	2
Truk	3

Tabel 5 diatas adalah transformasi dari tipe kendaraan, atribut yang akan digunakan proses clustering yang terdapat 4 data yaitu bus, mobil, motor dan truk.

Tabel 6 Transformasi Jenis Pelanggaran

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

Data jenis pelanggaran terdiri dari 9 jenis, ditransformasikan pada Tabel 6

Tabel 7 Transformasi Dataset

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

Pada Tabel 7 merupakan transformasi dari keseluruhan dataset untuk memudahkan proses clustering menggunakan k-means.

Penerapan K-Means Clustering

Selanjutnya, data yang telah dikumpulkan dan ditransformasikan mulai dikelompokkan menggunakan algoritma K-means Clustering dengan menentukan titik pusat kluster (centroid) yang dipilih secara acak yaitu pada Gambar 2[18].

```
[[0.86689715 0.98530683 8.
 [0.98133897 2.00870472 5.88495419]
 [0.85226188 3.      8. ]]]
```

Gambar 2 Menentukan Centroid

Berdasarkan data yang ditampilkan pada Gambar 2, terdapat tiga centroid yang masing-masing memiliki koordinat tertentu. Centroid yang pertama [0.86698715, 0.98538683, 8.], centroid 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.

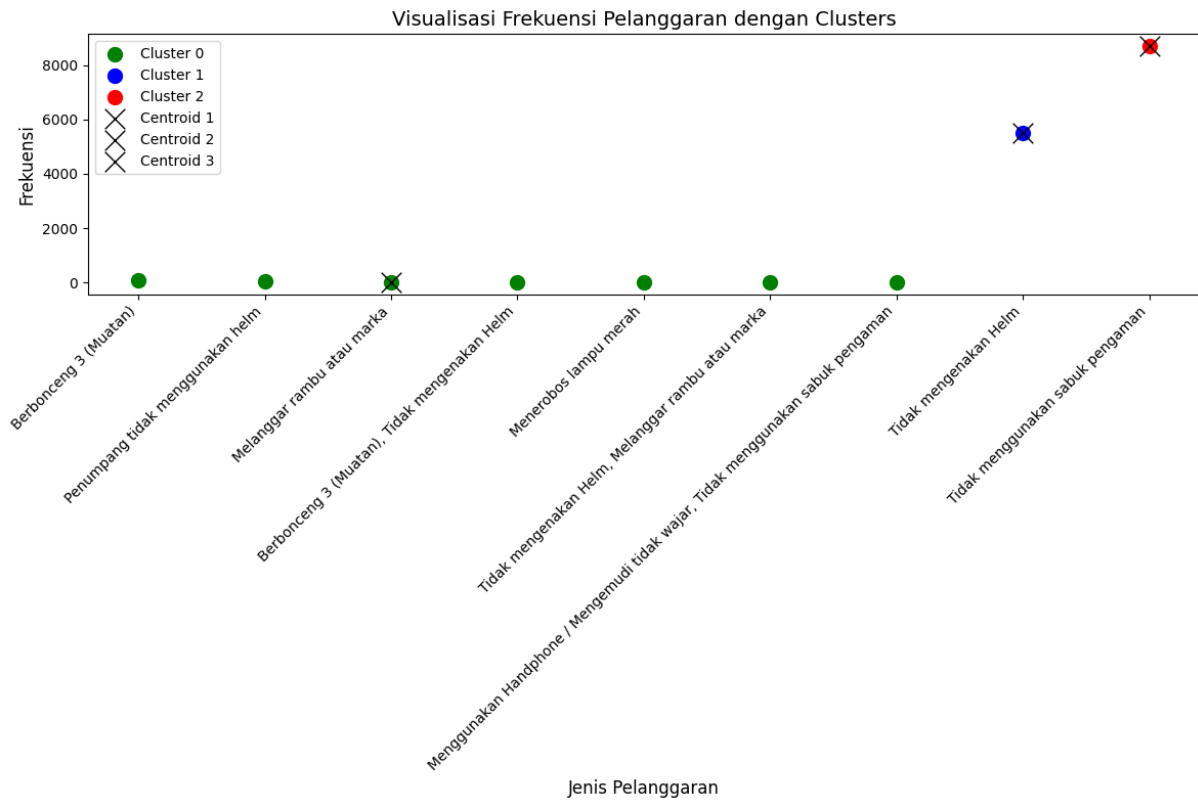
Hasil Clustering

Penelitian ini berhasil mengimplementasikan algoritma K-Means Clustering pada 14.386 data, yang dikelompokkan ke dalam tiga cluster: pelanggaran tertinggi, sedang, dan rendah. Pengelompokan ini memberikan gambaran yang jelas mengenai distribusi frekuensi pelanggaran. Hasil lengkapnya dapat dilihat pada Tabel 9 berikut :

Tabel 8 Hasil Akhir 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

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.



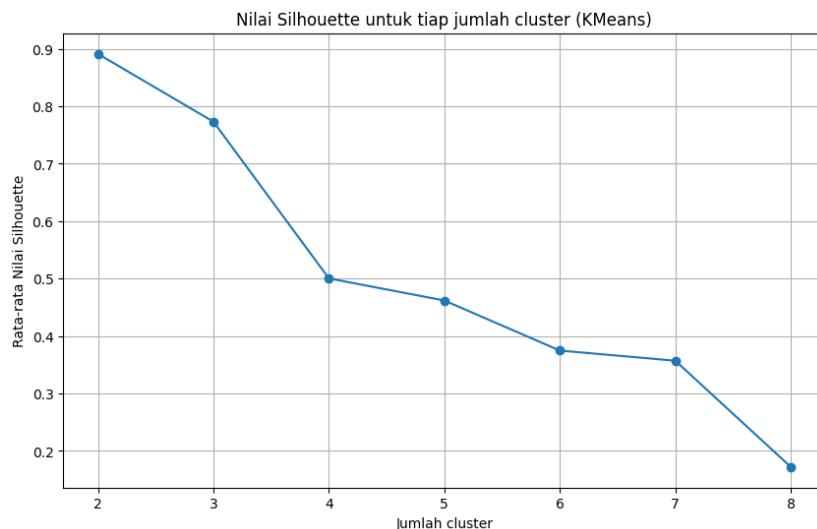
Gambar 3 Visualisasi Hasil Clustering

Evaluasi Hasil Clustering

Selanjutnya, dilakukan pengujian jumlah cluster untuk menentukan jumlah cluster yang optimal dengan melihat hasil dari nilai koefisien Silhouette yang dihasilkan dari pengujian yang telah dilakukan. Pengujian dilakukan dengan menggunakan jumlah nilai K mulai dari 2 hingga 8. Hasil pengujian ditampilkan dalam Tabel 10 dan visualisasi ditampilkan pada Tabel 11 berikut ini.

Tabel 9 Hasil Silhoutte Score

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



Gambar 4 Visualisasi Silhouette Score

Berdasarkan evaluasi menggunakan metrik Silhouette Score menggunakan python maka diperoleh 8 cluster yang dinilai dapat mengelompokkan data dengan baik. Hasil evaluasi cluster yang dibentuk sebanyak 8 cluster dengan nilai terbesar 0,8913 pada cluster 2 yang dapat dilihat pada Gambar 4 dan Tabel 10 [19].

Perbandingan Dengan Algoritma Lain

Sebagai bentuk pengembangan maka diperlukan perbandingan algoritma K-Means dengan algoritma DBSCAN dilakukan dengan menentukan jumlah cluster yang paling optimal dilihat dari nilai Silhouette Indexnya seperti pada Tabel 11 dibawah ini:

Tabel 10 Hasil Perbandingan Algoritma

Algoritma	Silhouette 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 [20].

Pembahasan

Penelitian ini menunjukkan bahwa algoritma K-Means Clustering sangat efektif dalam mengelompokkan data pelanggaran lalu lintas di Sidoarjo berdasarkan tingkat keparahan pelanggaran. Temuan utama menunjukkan bahwa pelanggaran "tidak menggunakan sabuk pengaman" menjadi kategori dengan frekuensi tertinggi, yang relevan dengan penelitian Sartika [5] yang juga menggunakan K-Means untuk mengelompokkan wilayah pelanggaran lalu lintas. Meskipun fokus mereka adalah pengelompokan wilayah, penelitian saya lebih mendetail karena

menggunakan evaluasi Silhouette Score untuk memastikan kualitas kluster yang terbentuk yang tidak dibahas dalam penelitian mereka. Penelitian ini juga membandingkan metode ini dengan algoritma DBSCAN, yang memiliki performa lebih rendah untuk dataset ini. Dengan demikian, penelitian ini tidak hanya mengonfirmasi efektivitas K-Means dalam menangani dataset besar tanpa label tetapi juga memberikan pedoman operasional berbasis data untuk membantu Satlantas POLRES Sidoarjo. Dan juga penelitian ini memiliki perbedaan dibandingkan menggunakan algoritma Apriori, seperti penelitian Agus di Yogyakarta [6], yang lebih menekankan pada analisis hubungan antar variabel. Kelebihan K-Means yang peneliti terapkan adalah kemampuannya menangani dataset besar tanpa memerlukan label, meskipun tantangannya adalah sensitivitas terhadap ketidakseimbangan data. Dalam konteks penelitian ini, data yang diambil dari CCTV POLRES Sidoarjo memberikan hasil yang lebih spesifik dan aplikatif, terutama dalam membantu menyusun kebijakan berbasis data untuk menurunkan tingkat pelanggaran lalu lintas.

Selain itu, analisis mendalam terhadap kluster pelanggaran rendah, seperti "berbonceng tiga" dan "melanggar rambu," menunjukkan bahwa meskipun jumlahnya kecil, kategori ini tetap relevan sebagai indikator ketidakdisiplinan yang mungkin bersifat situasional atau spesifik pada lokasi tertentu. Implikasi operasional yang dapat diambil meliputi pengawasan ketat pada pelanggaran tinggi, patroli periodik di lokasi dengan pelanggaran rendah, dan kampanye edukasi yang ditargetkan, seperti pentingnya mematuhi marka jalan atau larangan berbonceng tiga. Dengan langkah-langkah strategis ini, penelitian ini tidak hanya mengonfirmasi efektivitas K-Means Clustering tetapi juga memberikan panduan operasional yang dapat membantu Satlantas POLRES Sidoarjo menciptakan lalu lintas yang lebih aman dan tertib.

4. Penutup

Pada penelitian ini menghasilkan K Cluster terbaik yang mendapatkan nilai tertinggi pada silhouette score terlihat pada Tabel 10 sebesar 0.8913713235651775 pada $k = 2$ lebih besar dibanding cluster yang lain, dengan jumlah data uji sebanyak 14386 data pelanggaran lalu lintas. Pada pengujian DBSCAN hasil klasterisasi terbaiknya pada eps 5,00 yaitu 0.37543770067855425. Dengan demikian, diperoleh cluster paling optimal yaitu percobaan menggunakan algoritma K-Means dengan nilai $k=2$. Pengelompokan Pelanggaran Lalu Lintas pada Data CCTV menghasilkan 3 cluster uji. Terlihat pada Tabel 9 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. Berdasarkan penelitian yang telah peneliti lakukan dapat di simpulkan bahwa algoritma K-Means telah berhasil mengelompokkan atau mengklasterisasi pelanggaran lalu lintas berdasarkan Jenis Pelanggaran dengan mendapatkan anggota cluster yang optimal. Kelebihan pada penelitian ini adalah penggunaan silhouette score sebagai evaluasi nilai k terbaik dalam penelitian K-means clustering. 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.

Penelitian ini berhasil mengelompokkan data pelanggaran lalu lintas menggunakan algoritma K-Means, namun terdapat beberapa keterbatasan yang perlu diperhatikan. Salah satu tantangan utama adalah ketidakseimbangan data antar kategori pelanggaran, di mana kategori seperti "tidak menggunakan sabuk pengaman" memiliki jumlah kasus yang jauh lebih tinggi dibandingkan kategori lainnya. Ketidakseimbangan ini dapat memengaruhi hasil klusterisasi, membuat algoritma cenderung menghasilkan cluster yang didominasi oleh kategori tertentu. Selain itu, meskipun evaluasi dengan Silhouette Score menunjukkan hasil yang baik, nilai tersebut belum tentu mencerminkan representasi yang seimbang untuk semua kategori. Penelitian ini juga membandingkan algoritma DBSCAN, yang menunjukkan kinerja lebih rendah, kemungkinan disebabkan oleh parameter yang kurang optimal untuk data dengan distribusi tidak merata. Oleh karena itu, penelitian lanjutan disarankan untuk menerapkan teknik seperti oversampling atau undersampling untuk mengatasi ketidakseimbangan data dan mengeksplorasi kombinasi algoritma lain untuk meningkatkan hasil klusterisasi.

5. 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 Pengelompokan Banyaknya Jumlah Petani Berdasarkan Kecamatan Di Provinsi Jawa Barat," vol. 10, no. 3, 2023.
- [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.