



Terbit *online* pada laman web jurnal : <http://ejournal.upbatam.ac.id/index.php/comasiejournal>

## Jurnal Comasie

ISSN (Online) 2715-6265



# IMPLEMENTASI DATA MINING DENGAN METODE CLUSTERING ALGORITMA K-MEANS UNTUK PENGELOMPOKAN DATA TILANG DI POLDA KEPRI

Reni Alda Aristawaty Manihuruk<sup>1</sup>  
Rahmat Fauzi<sup>2</sup>

<sup>1</sup>Mahasiswa Program Studi Teknik Informatika , Universitas Putera Batam

<sup>2</sup>Dosen Program Studi Teknik Informatika, Universitas Putera Batam

email: [pb190210092@upbatam.ac.id](mailto:pb190210092@upbatam.ac.id)

## ABSTRACT

*Every community relies on transportation, but drivers must follow the rules to be safe. There are several variables that lead many individuals to get ticketed, including the public's lack of knowledge and awareness of excellent, accurate, and safe driving standards, and the community's failure to check vehicle conditions and paperwork before traveling to avoid tickets during special operations (raids). This research examined two-wheeled drivers' traffic offenses in Batam City, an issue investigated by the author due to the numerous traffic offences that have disrupted the regulatory system thus far. The K-Means method clusters. Implementing the K-Means Algorithm to group traffic violation data helps the Riau Regional Police discover the most traffic infractions and the ticketing service department locate Batam traffic violation data groupings. The author will utilize a K-Means algorithm data mining approach to segment infractions thus far. Data analysis from the cluster, segregated by mopabudget type, degree of violation, and fine amount, yielded the findings of Cluster 0: 11 items 13. Cluster 2: 11 items from 35 data treated as sample data with performance vector; best clustering is 0.499.*

*Keywords : Transport, Traffic Violation, Data Mining, K-Means Algorithm, cluster, RapidMiner*

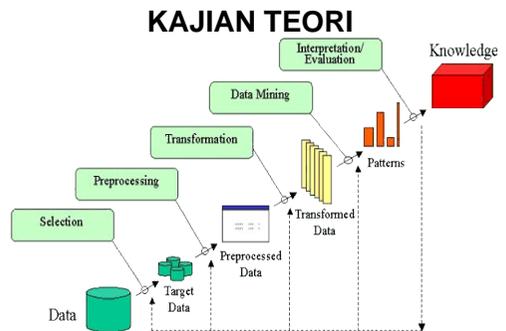
## PENDAHULUAN

Transportasi darat merupakan bagian integral dari masyarakat, baik secara individu maupun kelompok. Meningkatnya penggunaan kendaraan seperti kendaraan bermotor, mobil dan truk juga meningkatkan pelanggaran lalu lintas saat berkendara. Pelanggaran lalu lintas menyebabkan kecelakaan karena cara mengemudi kendaraan yang kurang baik, yaitu kurangnya pengetahuan pengemudi tentang disiplin lalu lintas dan pengguna jalan tidak memahami peraturan.

Perbuatan atau kelalaian yang melanggar hukum di jalan raya disebut pelanggaran lalu lintas. Mereka yang berada di belakang pengemudi dan melanggar hukum harus membayar mahal. Tiket lalu lintas adalah bukti pelanggaran dan dapat digunakan oleh penegak hukum untuk menyitakan properti pelanggar. Petugas polisi yang menghentikan pelanggar harus memberi mereka penjelasan terperinci tentang pelanggaran tersebut, salinan spesifik pelanggaran, dan tabel yang merinci hukuman terkait.

Di Indonesia, jumlah mobil berkembang pesat. Ditemukan bahwa pada tahun 2018 berdasarkan jenis kendaraan terdapat 15.423.968 mobil penumpang, 2.509.258 mobil penumpang, 7.289.910 truk, dan 111.988.683 sepeda motor. Khususnya mobil pribadi, masyarakat lebih memilih menggunakan mobil pribadi karena dianggap lebih nyaman untuk melakukan aktivitas sehari-hari. Demikian pula di Batam, jumlah kendaraan semakin meningkat setiap harinya karena kebutuhan masyarakat untuk beraktivitas sehari-hari seperti bekerja, pergi ke pasar, menyekolahkan anaknya dan kegiatan lain yang berkaitan dengan kebutuhannya sebagai makhluk sosial. Badan Pusat Statistik Kota Batam mencatat jumlah kendaraan bermotor di Kota Batam mencapai lebih dari 771.748 pada tahun 2021. Dari 2019 hingga 2021, jumlah mobil meningkat 9,7%. Jumlah bus bertambah 3,3%, dan jumlah sepeda motor berkurang 7,45%. Sehingga tentunya tidak semua pengendara tersebut mengikuti aturan berlalu lintas, sehingga pihak berwajib sering menindak pelanggaran seperti antara lain tidak membawa surat keterangan kendaraan yang lengkap, tidak menggunakan helm, tidak membawa surat keterangan pengemudi. SIM, menerobos lampu merah, memiliki penumpang berusia kurang dari 17 tahun di belakang kemudi, tidak mematuhi marka jalan, dll. Masyarakat juga perlu mengevaluasi kondisi mobil dan keaslian dokumen, yang keduanya telah berkontribusi ke sejumlah besar individu yang diberikan tiket. Untuk menghindari hukuman, Anda mungkin pergi berlibur sebelum melakukan operasi rahasia (razia).

Pelanggaran lalu lintas semakin meningkat, bervariasi setiap tahun dari pelanggaran ringan hingga pelanggaran yang membutuhkan denda terbesar. Kajian ini dilakukan oleh pihak Polda Kepri yang memiliki banyak data pelanggaran lalu lintas yang belum pernah digunakan sebelumnya untuk memberikan data baru untuk mengetahui sejauh mana pelanggaran lalu lintas khususnya sepeda motor roda dua meskipun digunakan itu memfasilitasi klasifikasi kategori pelanggaran lalu lintas. Oleh karena itu, penulis mengimplementasi teknik data mining menggunakan *k-means algorithm*.



Gambar 1 Proses Data Mining pada KDD

(Sumber: Data Penelitian, 2023)

Berikut Langkah proses data mining pada *knowledge discovery in database* (KDD) adalah sebagai berikut:

a. Penyeleksian Data

Tidak semua data dalam database dilakukan untuk proses data mining, sehingga diperlukan proses seleksi yang sesuai dengan analisis yang dilakukan dan hasil pemilihan kemudian dipakai cara selanjutnya.

b. *Preprocessing/Cleaning*

Saat sebelum proses data mining bisa dicoba, butuh dilakukan cara

pembersihkan kepada informasi yang jadi fokus KDD. Cara pembersihan mencakup penghapusan data duplikasi, pengecekan data yang tidak konsisten, serta pemeriksaan kekeliruan pada data tersebut, semacam kekeliruan cap (tipografi).

### c. Transformasi Data

Transformasi data dilakukan dengan memilih data yang cocok buat cara penambangan. Cara pengkodean KDD merupakan cara inovatif serta sungguh terkait pada tipe ataupun bentuk informasi yang didapat dari *database*.

### d. *Datamining*

*Datamining* merupakan sistem pencarian pola ataupun data yang menarik dari informasi tertentu dengan memakai metode ataupun tata cara khusus. Metode, tata cara, ataupun algoritma penambangan informasi bermacam-macam. Memilih tata cara ataupun algoritma yang pas sungguh tergantung pada tujuan serta totalitas sistem KDD.

### e. Interpretasi dan Evaluasi

Interpretasi dan evaluasi merupakan bentuk data yang diperoleh dari metode data mining, yang mesti disajikan dalam wujud yang gampang dimengerti oleh pihak yang bersangkutan. Tahap ini menggambarkan bagian dari sistem KDD yang disebut *rendering*.

Pada langkah ini kita memeriksa apakah pola ataupun data yang ditemui sebaliknya dengan kebenaran ataupun dugaan yang terdapat pada:

## 2.1. Pengertian Data Mining

Menurut (Nabila et al., 2021) *Data Mining* merupakan teknik bantu pengambilan keputusan pada penulis mencari pola dalam data. Ini dapat dilakukan secara manual oleh pengguna (melalui kueri, misalnya) atau secara otomatis oleh perangkat lunak yang

menanyakan basis data untuk mencari model data. Penemuan mengacu pada proses mencari informasi baru.

## 2.2. Pengertian *K-means Clustering*

*K-means clustering* merupakan satu atau lebih *cluster* data dapat dibuat tanpa menggunakan struktur hierarkis. Data dalam suatu kluster atau kelompok sangat mirip satu sama lain karena mereka memiliki kualitas yang sama, sedangkan data dalam kluster atau kelompok lain berbagi perbedaannya. Tentukan seberapa jauh masing-masing pusat massa objek dari pusat objek yang sebenarnya. Jarak dapat ditentukan dengan menggunakan Euclidean Distance, yaitu :

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2}$$

Keterangan:

De = Euclidean Distance

i = Banyaknya objek

(x, y) = Koordinat objek

(s, t) = Koordinat centroid

## 2.3. Pengertian *Rapidminer*



**Gambar 2** *Rapidminer*  
(Sumber: Data Penelitian, 2023)

RapidMiner merupakan perangkat lunak yang bersifat sumber terbuka (Open Source). RapidMiner adalah teknologi untuk memeriksa



Terbit online pada laman web jurnal : <http://ejournal.upbatam.ac.id/index.php/comasiejournal>

## Jurnal Comasie

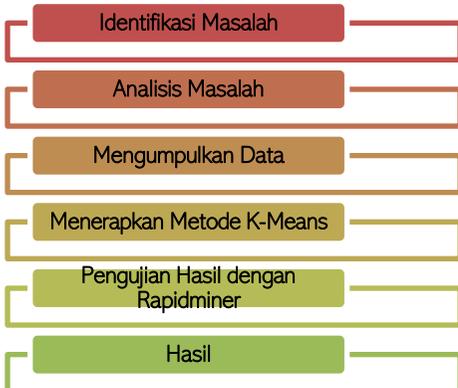
ISSN (Online) 2715-6265



penambahan data, penambahan teks, dan analitik prediktif. Sebagai alat analisis data dan mesin penambahan data, RapidMiner dapat digunakan sendiri atau dimasukkan ke dalam produk yang sudah ada. Karena dikembangkan di Java, RapidMiner dapat digunakan dengan sistem komputer apa pun. Pengguna bisa mendapatkan wawasan berharga dari berbagai metode deskriptif dan prediktif RapidMiner, yang memungkinkan mereka untuk membuat pilihan yang lebih tepat.

### METODE PENELITIAN

Desain penelitian secara teknis menggambarkan apa yang peneliti lakukan. Metodologi penelitian memiliki beberapa pedoman yang harus diikuti, dan urutan pedoman tersebut merupakan contoh prosedur harus diikuti untuk memastikan keberhasilan penelitian ini:



**Gambar 3** Desain Penelitian  
(Sumber: Data Penelitian, 2023)

Bagian ini menjelaskan metodologi yang digunakan dalam penelitian ini:

#### 1. Identifikasi Masalah

Dalam penelitian ini langkah utamanya adalah mengidentifikasi permasalahan yang muncul pada

objek penelitian, dimana penulis melakukan studi lapangan terhadap permasalahan yang muncul, pada tahap ini penulis melakukan wawancara dan observasi untuk mendapatkan informasi tentang pelanggaran lalu lintas. pengguna kendaraan

#### 2. Analisis Masalah

Dari masalah yang teridentifikasi pada tahap analisis, dilanjutkan dengan mengidentifikasi elemen situasi masalah dan mencari tahu komponen mana yang kritis. Disini penulis hanya menganalisis permasalahan terkait pelanggaran peraturan oleh pengguna roda dua.

#### 3. Mengumpulkan Data

Tahap selanjutnya adalah pengumpulan data, dimana penulis mengumpulkan informasi tentang pelanggaran lalu lintas yang dilakukan oleh kendaraan roda dua. Data yang terkumpul sebanyak 200 data dari tahun 2021-2022.

#### 4. Menerapkan Metode *K-Means*

Setelah tahap pengumpulan data, data dianalisis menggunakan teknik data mining yaitu algoritma *K-Means* untuk mengetahui tingkat segmental dari pelanggaran yang dilakukan oleh pengendara sepeda motor.

#### 5. Pengujian Hasil dengan *rapidminer*

Pada langkah selanjutnya, hasil analisis diuji kembali dengan software *rapidminer* dan hasil pengolahan manual dibandingkan dengan sistem.

#### 6. Hasil

Pada langkah terakhir, penulis merangkum hasil klasifikasi pelanggaran kendaraan lalu lintas dari informasi yang terkumpul selama ini dan menarik kesimpulan berupa informasi baru untuk menangani pelanggaran di masa yang akan datang.



Terbit online pada laman web jurnal : <http://ejournal.upbatam.ac.id/index.php/comasiejournal>

## Jurnal Comasie

ISSN (Online) 2715-6265



### HASIL DAN PEMBAHASAN

Berdasarkan analisis data, 35 data dibagi menjadi 3 cluster yaitu C0, C1 dan C2 dan setelah dilakukan ekstraksi data dan pemilihan atribut digunakan yaitu. jenis pelanggaran, tingkat pelanggaran, dan besaran denda, kemudian dilakukan

pengolahan data untuk menghindari duplikasi data. Perbaikan bug menyimpan data Excel baru. Diimplementasikan disini, data disanitasi atau disanitasi agar dapat diproses dan proses data mining dilakukan informasi tentangnya setelah langkah pra-pemrosesan disimpan dalam informasi berikut :

Tabel 1. Pre-Processing

No. Tilang	Jenis Pelanggaran	Tingkat Pelanggaran	Besar Denda
F8567177	3	2	3
E8566938	4	2	2
F8567606	1	1	4
F8567176	4	2	2
F8567361	1	2	3
F8567144	4	2	2
F8567587	3	2	3
F8567629	1	2	3
F8567667	1	2	4
F8567666	1	2	4
F8567670	1	2	4
F8567343	3	2	3
F8567346	1	2	3
F8567341	3	2	3
F8567348	2	2	3
E8566038	1	2	4
F8567669	4	2	2
F8567378	4	2	2
F8567422	4	2	2
F8567856	1	2	4
F8567857	4	2	2
F8567860	4	2	2
F8567804	1	2	4
F8567409	1	2	4



Terbit online pada laman web jurnal : <http://ejournal.upbatam.ac.id/index.php/comasiejournal>

## Jurnal Comasie

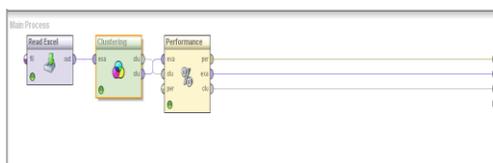
ISSN (Online) 2715-6265



No. Tilang	Jenis Pelanggaran	Tingkat Pelanggaran	Besar Denda
F8567374	3	1	3
F8567478	4	2	2
E8576582	3	2	3
F8567473	1	2	4
F8567489	2	2	3
F8567375	1	2	3
F8567412	4	2	2
F8567708	3	2	2
F8567143	4	2	2
E8566049	2	2	2
F8567353	3	2	3

(Sumber: Data Penelitian, 2023)

Kemudian dilakukan pengolahan data dengan software *rapidminer* dan didapatkan hasil sebagai berikut :



**Gambar 4** Proses *Rapidminer*  
(Sumber : Data Penelitian, 2022)

Dari gambar proses *Rapidminer* di atas, cluster khusus adalah 3 atau  $K=3$ , sedangkan ukuran yang dipilih adalah numerik menggunakan rumus jarak Euclidean, sehingga hasil perhitungan rata-rata  $K$  adalah sebagai berikut :



**Gambar 5** Text View  
(Sumber : Data Penelitian, 2022)

Terdapat total 35 data dari hasil text view, sehingga cluster 0 terdiri dari 11

elemen, cluster 1 terdiri dari 13 elemen, dan terakhir cluster 2 berisi 11 elemen. Jika  $C_0$  adalah karakter yang melanggar, tidak lengkap dan tidak patuh, maka  $C_1$  adalah jumlah pelanggaran yang dilakukan dan  $C_2$  adalah denda yang dibayarkan oleh vektor kinerja. Nilai cluster optimal adalah 0,499. Di gambar berikutnya.



Terbit online pada laman web jurnal : <http://ejournal.upbatam.ac.id/index.php/comasiejournal>

## Jurnal Comasie

ISSN (Online) 2715-6265



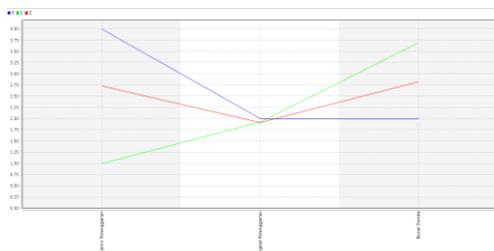
**Gambar 6 Performance Vector**  
(Sumber : Data Penelitian, 2022)

Angka-angka yang ditampilkan pada layar Tabel Centroid merupakan hasil akhir perhitungan menurut titik-titik centroid dan juga kelompok hasil yang sama seperti yang ditunjukkan pada gambar.

Attribute	cluster_0	cluster_1	cluster_2
Jenis Pelanggaran	4	1	2.727
Tingkat Pelanggaran Besar Denda	2	1.923	1.909
	2	3.692	2.818

**Gambar 7 Centroid Tabel**  
(Sumber : Data Penelitian, 2022)

*Centroid Plot View* menampilkan grafik berupa garis yang menggambarkan centroid akhir dari setiap cluster berikutnya.



**Gambar 8 Centroid Plot View**  
(Sumber : Data Penelitian, 2022)

Grafik di atas berupa garis memiliki tiga warna yaitu garis hijau untuk kelompok pelanggaran, garis merah untuk tingkat pelanggaran dan garis biru untuk denda.

## SIMPULAN

Berdasarkan penelitian yang telah dilakukan, diperoleh kesimpulan sebagai berikut :

1. Mengelompokkan data dengan algoritma K-Means dilakukan dengan cara menentukan jumlah cluster, hitung jarak terdekat dengan pusat cluster. Data dengan jarak terdekat menyatakan jenis pelanggaran dari cluster tersebut, dilakukan perhitungan kembali sampai data tidak berpindah pada cluster lain, untuk meminimalkan fungsi objektif.
2. Berdasarkan hasil pengujian 200 data menggunakan *rapidminer* pada penelitian data tilang di Provinsi Kepri pada tahun 2021-2022, menghasilkan pengelompokan yang berbeda dikarenakan jumlah pada jenis pelanggaran, tingkat pelanggaran dan besar denda. Untuk *cluster 0* adalah jenis pelanggaran terdapat 49 items, *cluster 1* adalah tingkat pelanggaran terdapat 101 items, *cluster 2* adalah besar denda terdapat 50 items.

## DAFTAR PUSTAKA

- Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 4(1), 51–58. <https://doi.org/10.30865/mib.v4i1.1784>
- Annisa Ekasetya, V., & Jananto, A. (2020). KLUSTERISASI OPTIMAL DENGAN ELBOW METHOD UNTUK PENGELOMPOKAN DATA KECELAKAAN LALU LINTAS DI KOTA SEMARANG. *Dinamika Informatika*, 12(1), 20–28.
- Daeli, F. F. W., Nofriansyah, D., & Yakub, S. (2022). Implementasi Data Mining Untuk Mengelompokkan



Terbit *online* pada laman web jurnal : <http://ejournal.upbatam.ac.id/index.php/comasiejournal>

## Jurnal Comasie

ISSN (Online) 2715-6265



- Lokasi Berdasarkan Tingkat Kejahatan Pada Kabupaten Nias Barat Menggunakan Metode K-Means Clustering. *Jurnal Cyber Tech*, 1(4).
- Dewi, E., Mulyani, S., Yusup, A. M., Tisna, A. K., Fauzi, F. A., Seta, I. B., Khairunas, R., & Ardiansyah, W. (2019). *Clustering Wilayah Dan Pelanggaran Berkendaraan Menggunakan Algoritma K-Means Pada Data Satlantas Polres Tasikmalaya Kota* (Vol. 8, Issue 1).
- Indraputra, R. A., Fitriana, R., & others. (2020). K-Means Clustering Data COVID-19. *Jurnal Teknik Industri*, 10(3), 275–282.

	<p>Reni Alda Aristawaty Manihuruk merupakan mahasiswa Prodi Teknik Informatika Universitas Putera Batam.</p>
	<p>Rahmat Fuzi merupakan Dosen Prodi Teknik Informatika Universitas Putera Batam.</p>