

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

SKRIPSI



Oleh :

Reni Alda Aristawaty Manihuruk

190210092

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2023**

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

SKRIPSI

**Untuk memenuhi salah satu syarat
memperoleh gelar sarjana**



Oleh :

**Reni Alda Aristawaty Manihuruk
190210092**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KOMPUTER
UNIVERSITAS PUTERA BATAM
TAHUN 2023**

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini saya:

Nama : Reni Alda Aristawaty Manihuruk
NPM : 190210092
Fakultas : Teknik dan Komputer
Program Studi : Teknik Informatika

Menyatakan bahwa “skripsi” yang saya buat dengan judul:

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

Adalah hasil karya sendiri dan bukan “duplikasi” dari karya orang lain. Sepengetahuan saya, didalam naskah skripsi ini tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip didalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka. Apabila ternyata di dalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia naskah skripsi ini digugurkan dan hasil penilaian yang saya peroleh dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada paksaan dari siapapun.

Batam, 27 Januari 2023



Reni Alda Aristawaty Manihuruk
190210092

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

SKRIPSI

**Untuk memenuhi salah satu syarat
memperoleh gelar sarjana**

Oleh:

Reni Alda Aristawaty Manihuruk

190210092

**Telah disetujui oleh Pembimbing pada tanggal
seperti yang tertera di bawah ini**

Batam, 27 Januari 2023



Rahmat Fauzi, S.Kom., M.Kom.

Pembimbing

ABSTRAK

Setiap komunitas mengandalkan transportasi, tetapi pengemudi harus mengikuti aturan agar aman. Ada beberapa variabel yang menyebabkan banyak individu ditilang antara lain kurangnya pengetahuan dan kesadaran masyarakat akan standar berkendara yang baik, akurat dan aman, serta kegagalan masyarakat untuk memeriksa kondisi kendaraan dan dokumen sebelum bepergian untuk menghindari tilang selama operasi khusus (penggerebekan). Penelitian ini mengkaji pelanggaran lalu lintas pengendara roda dua di Kota Batam yang menjadi permasalahan yang diteliti oleh penulis karena banyaknya pelanggaran lalu lintas yang mengganggu sistem regulasi selama ini. Kluster metode *K-Means*, penerapan Algoritma *K-Means* untuk mengelompokkan data pelanggaran lalu lintas membantu Polda Kepri menemukan pelanggaran lalu lintas terbanyak dan dinas tilang menemukan pengelompokan data pelanggaran lalu lintas di Kota Batam. Penulis akan memanfaatkan pendekatan *data mining* algoritma *K-Means* untuk melakukan segmentasi pelanggaran selama ini. Analisis data dari kluster dipisahkan berdasarkan jenis pelanggaran, tingkat pelanggaran dan besaran denda, menghasilkan temuan kluster 0: 11 item, kluster 1: 13 item, kluster 2: 11 item, dari 35 data yang diperlakukan sebagai data sampel dengan vektor kinerja pengelompokan terbaik adalah 0,499.

Kata kunci : Transportasi, Pelanggaran Lalu Lintas, Pengelompokan Data, Algoritma *K-Means*, *cluster*, *RapidMiner*.

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 offences 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 Island 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, Cluster1: 13 items, 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.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan yang Maha Esa, karena atas segala berkat dan rahmat-Nya sehingga tugas akhir ini dapat diselesaikan sesuai dengan waktu yang telah ditentukan. Penyusunan tugas akhir ini merupakan syarat untuk mencapai kelulusan dan meraih gelar Strata Satu (S1) Program Studi Akademik Teknik Informatika pada Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati. Dengan segala keterbatasan penulis, penulis menyadari pula bahwa skripsi ini dapat diselesaikan dengan tepat waktu karena berbagai dukungan, dorongan bimbingan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Ibu Dr. Nur Elfi Husda, S.Kom., M.SI. selaku Rektor Universitas Putera Batam;
2. Bapak Welly Sugianto, S.T., M.M. selaku Dekan Fakultas Teknik dan Komputer Universitas Putera Batam;
3. Bapak Andi Maslan, S.T., M.SI. selaku Ketua Program Studi Teknik Informatika Universitas Putera Batam;
4. Bapak Rahmat Fauzi, S.Kom., M.Kom. selaku pembimbing Skripsi pada Program Studi Teknik Informatika Fakultas Teknik dan Komputer Universitas Putera Batam;
5. Ibu Alfannisa Annurrullah Fajrin, S.Kom., M.Kom. selaku pembimbing Akademik Program Studi Teknik Informatika Fakultas Teknik dan Komputer Universitas Putera Batam;
6. Dosen dan Staff Universitas Putera Batam;
7. Pihak Subdit Gakkum Ditlantasi Polda Kepulauan Riau;
8. Pihak Bagdalpers Ro SDM Polda Kepulauan Riau;
9. Kepada kedua orang tua dan keluarga tercinta atas nasihat, doa dan dukungannya;
10. Teman-teman dan sahabat yang selalu memberikan dukungan dan

semangat;

11. Pihak lain yang turut memberikan dorongan dalam dukungan penyusunan skripsi ini;

Semoga Tuhan Yang Maha Esa membalas kebaikan dan selalu mencurahkan hidayat serta taufik-Nya, Amin.

Batam, 27 Januari 2022



Reni Alda Aristawaty Manihuruk

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
SURAT PERNYATAAN ORISINALITAS	iii
HALAMAN PENGESAHAN	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR RUMUS	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Identifikasi Masalah.....	4
1.3. Batasan Masalah	4
1.4. Rumusan Masalah.....	4
1.5. Tujuan Penelitian	5
1.6. Manfaat Penelitian	5
1.6.1. Manfaat Teoritis.....	5
1.6.2. Manfaat Praktis	6
BAB II TINJAUAN PUSTAKA	7
2.1. Teori Umum.....	7
2.1.1. <i>Knowledge Discovery in Database (KDD)</i>	7
2.1.2. Data Mining	9
2.1.3. <i>Clustering</i>	9
2.1.4. <i>K-Means</i>	11
2.1.5. <i>RapidMiner</i>	12
2.2. Teori Khusus.....	14
2.2.1. Pengertian Lalu Lintas	14
2.2.2. Sepeda Motor.....	16
2.2.3. Pelanggaran Lalu Lintas	17
2.2.4. Faktor Pelanggaran Lalu Lintas.....	17
2.2.5. Dampak Pelanggaran Lalu Lintas.....	19
2.3. Penelitian Terdahulu	20
2.4. Kerangka Pemikiran	23
BAB III METODE PENELITIAN	26
3.1. Desain Penelitian	26
3.2. Teknik Pengumpulan Data	28
3.3. Operasional Variabel	28
3.4. Algoritma.....	29
3.5. Lokasi dan Jadwal Penelitian.....	29
3.5.1. Lokasi	29
3.5.2. Jadwal Penelitian	30

BAB IV HASIL DAN PEMBAHASAN	31
4.1. Hasil Analisa dengan data mining Algoritma <i>clustering K-Means</i>	31
4.2. Melakukan Pra Proses Data Penelitian	31
4.3. Analisa Data.....	37
4.3.1. Pembersihan Data	38
4.3.2. Seleksi Data	38
4.3.3. Transformasi Data	38
4.3.4. Perhitungan Algoritma <i>Clustering K-Means</i>	41
4.4. Pengujian Menggunakan Software RapidMiner 5.3.....	116
BAB V SIMPULAN DAN SARAN	127
5.1. Simpulan.....	127
5.2. Saran	127
DAFTAR PUSTAKA	129
LAMPIRAN	
Lampiran 1. Dokumen Pendukung	
Lampiran 2. Daftar Riwayat Hidup	
Lampiran 3. Surat izin penelitian	
Lampiran 4. Balasan Surat Izin Penelitian	
Lampiran 5. Turnitin Skripsi	
Lampiran 6. Turnitin Jurnal	
Lampiran 7. Data Tilang	
Lampiran 8. Data Cleaning	
Lampiran 9. Transformasi Data	

DAFTAR GAMBAR

Gambar 2.1 <i>Knowledge Discovery in Database</i>	7
Gambar 2.2 Rapidminer	12
Gambar 2.3 Sistem Komponen Lalu Lintas	15
Gambar 2.4 Kerangka Pemikiran	24
Gambar 3.1 Desain Penelitian	26
Gambar 3.2 <i>Flowchart</i> Algoritma <i>K-Means</i>	29
Gambar 3.3 Lokasi Penelitian	30
Gambar 4.1 Data Tilang Skripsi.xlsx	117
Gambar 4.2 Tampilan Awal Rapidminer 5.3	117
Gambar 4.3 Menu <i>New Process</i>	118
Gambar 4.4 Menu <i>New Process</i>	118
Gambar 4.5 Mencari <i>Read Excel</i>	119
Gambar 4.6 <i>Read Excel</i>	119
Gambar 4.7 <i>Import Configuration Wizard</i>	120
Gambar 4.8 <i>Data Import Wizard Step 1</i>	120
Gambar 4.9 <i>Data Import Wizard Step 2</i>	121
Gambar 4.10 <i>Data Import Wizard Step 3</i>	121
Gambar 4.11 <i>Data Import Wizard Step Terakhir</i>	122
Gambar 4.12 <i>Clustering K-Means</i>	122
Gambar 4.13 Penentuan Titik Pusat	123
Gambar 4.14 Hubungkan <i>Read Excel</i> ke <i>Clustering K-Means</i>	123
Gambar 4.15 <i>Cluster Distance Performance</i>	124
Gambar 4.16 Proses <i>Run</i>	124
Gambar 4.17 <i>Example Set (Read Excel)</i>	125
Gambar 4.18 <i>Cluster Model</i>	125
Gambar 4.19 <i>Centroid Table</i>	126
Gambar 4.20 <i>Graph View</i>	126
Gambar 4.21 <i>Centroid Plot View</i>	126
Gambar 4.22 <i>Performance Vector</i>	127

DAFTAR TABEL

Tabel 3.1 Jadwal Penelitian.....	30
Tabel 4.1 Data Sampel	32
Tabel 4.2 Transformasi Data	38
Tabel 4.3 Perhitungan Iterasi Pertama.....	43
Tabel 4.4 Perhitungan Iterasi Kedua	53
Tabel 4.5 Perhitungan Iterasi Ketiga	63
Tabel 4.6 Perhitungan Iterasi Keempat	69
Tabel 4.7 Perhitungan Iterasi kelima.....	75
Tabel 4.8 Perhitungan iterasi keenam	81
Tabel 4.9 Perhitungan iterasi ketujuh.....	87
Tabel 4.10 Perhitungan iterasi kedelapan.....	93
Tabel 4.11 Perhitungan Iterasi kesembilan	99
Tabel 4.12 Perhitungan Iterasi Kesepuluh.....	105
Tabel 4.13 Hasil Akhir Perhitungan Algoritma <i>Clustering</i> K-Means.....	111

DAFTAR RUMUS

Rumus 2.1 Perhitungan Euclidean	11
--	----

BAB I

PENDAHULUAN

1.1. Latar Belakang

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 menyita 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, 2.509.258 mobil penumpang, 7.289.910 truk dan 111.988.683 sepeda motor. Khususnya sepeda motor, masyarakat lebih memilih menggunakan sepeda motor karena dianggap lebih nyaman untuk melakukan aktivitas sehari-hari. Demikian pula di Kota 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 terberat. 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*.

Seperti yang digunakan dalam penelitian sebelumnya, lalu lintas yang teratur di jalan sangat penting bagi pengendara di jalan raya, ketidaktahuan pengguna

kendaraan bermotor dan peraturan lalu lintas yang buruk bertanggung jawab atas tingkat pelanggaran lalu lintas saat mengemudi. Jalan tol terus berkembang, sehingga banyak informasi bagus di Pengadilan Negeri Dumai, karakteristik yang dianalisis dalam penelitian ini adalah jenis pelanggaran dan kendaraan yang berjumlah 616 pelanggaran. Data pusat klaster terdiri dari tiga klaster yaitu C1 = banyak, C2 = sedang dan C3 = sedikit pelanggar. Hasil data demikian diperoleh, dimana C1 menghasilkan 1 data, C2 menerima hingga 4 data dan C3 menerima hingga 7 data. Sedangkan jenis pelanggaran yang sering dilanggar adalah memakai helm dan jenis kendaraannya adalah sepeda motor. Hasil penelitian ini dapat digunakan atau ditelusuri dengan sosialisasi untuk mengurangi pelanggaran lalu lintas sebuah pada penelitian lain dilakukan analisis data mining menggunakan metode *clustering* data pelanggaran lalu lintas Polres Tasikmalaya, menggunakan *k-means clustering* sebagai algoritma yang mengelompokkan beberapa data atau objek menjadi satu *cluster* atau kelompok sehingga setiap data *cluster* dibuat semirip mungkin dan berbeda satu sama lain ke objek *cluster* lain. Informasi pelanggaran lalu lintas dari Polres Tasikmalaya di kota ini diolah menggunakan *Knowledge Data Discovery* (KDD) untuk menemukan tes *RapidMiner* yang menghasilkan kelompok pelanggaran lalu lintas. Sampel yang digunakan adalah dari tabel data pelanggaran lalu lintas yang dimodifikasi. Karakteristik yang ditentukan ada 6 karakteristik yaitu area, helm, zona aman, pelanggaran rambu lalu lintas, pembekuan SIM dan STNK. Dimana disajikan klaster dan jenis pelanggaran lalu lintas tiap kelompok wilayah (Dewi et al., 2019).

Berdasarkan permasalahan yang ada, maka peneliti membuat penelitian yang

berjudul “IMPLEMENTASI DATA MINING DENGAN METODE *CLUSTERING* ALGORITMA *K-MEANS* UNTUK PENGELOMPOKAN DATA TILANG DI POLDA KEPRI”.

1.2. Identifikasi Masalah

Sebagai pedoman peneliti dalam menjelaskan bagian pembahasan skripsi ini, berdasarkan latar belakang permasalahan yang telah dipaparkan sebelumnya, maka peneliti mengidentifikasi beberapa permasalahan yaitu:

1. Semakin bertambahnya kendaraan yang ada di Kota Batam sehingga pengendara juga banyak melakukan pelanggaran.
2. Belum dimanfaatkannya data tilang atau pelanggaran lalu lintas selama ini untuk di analisis guna menemukan informasi baru yang bermanfaat untuk pengelompokan jenis tilang.

1.3. Batasan Masalah

Batasan masalah dibuat karena adanya berbagai keterbatasan yang dialami oleh peneliti dan bertujuan agar penelitian yang dilakukan oleh peneliti menjadi lebih terpusat. Berikut adalah batasan masalah yang ditentukan oleh peneliti :

1. Peneliti ini dilaksanakan di Kota Batam pada Polda Kepri terkhusus untuk pelanggaran kendaraan roda dua.
2. Metode data mining yang digunakan adalah metode *K-Means* dan pengujian dengan *software RapidMiner*.

1.4. Rumusan Masalah

Berdasarkan Latar Belakang Masalah diatas maka dapat dirumuskan masalah

dalam penelitian ini :

1. Bagaimana teknik analisis data mining dengan menggunakan algoritma *K-Means* menganalisis pengelompokan tingkat pelanggaran lalu lintas kendaraan roda dua?
2. Bagaimana hasil pengujian analisis data mining dengan algoritma *K-Means* dengan *software RapidMiner*?

1.5. Tujuan Penelitian

Untuk Tujuan dari penelitian ini berdasarkan rumusan diatas adalah sebagai berikut :

1. Untuk menganalisis pengelompokan tingkat pelanggaran lalu lintas kendaraan roda dua dengan algoritma *K-Means*.
2. Untuk menguji hasil analisis segmentasi pelanggaran lalu lintas dengan *software RapidMiner*.

1.6. Manfaat Penelitian

Manfaat dari hasil penelitian mengacu pada dua hal, yaitu manfaat teoritis dan manfaat praktis:

1.6.1. Manfaat Teoritis

Manfaat teoritis dari penelitiannya ini berdasarkan manfaat yang berkenaan dengan ilmu pengetahuan, dalam hal ini ilmu linguistik atau kebahasaan:

1. Dapat menambah wawasan mengenai data mining.
2. Dapat menambah wawasan mengenai cara kerja data mining.
3. Dapat digunakan sebagai referensi dalam membuat karya lainnya, baik karya

ilmiah maupun penelitian yang relevan.

1.6.2. Manfaat Praktis

Manfaat Praktis Penelitian dilakukan karena ada masalah yang ingin diselesaikan atau dipecahkan. Manfaat praktis menjelaskan manfaat yang berguna untuk memecahkan masalah secara tersebut secara praktis:

1. Penelitian ini memberikan manfaat bagi penulis sebagai referensi dalam pembelajaran dan pemahaman yang lebih interaktif.
2. Penelitian ini bisa dijadikan sebagai bahan referensi dan saran bagi peneliti selanjutnya bagi yang mengambil tema tentang data mining pada fakultas Teknik dan komputer Program Studi Teknik Informatika.
3. Dengan adanya penelitian ini, diharapkan bisa menyadari masyarakat bahwasanya peran masyarakat dalam pelayanan publik sangat penting sehingga akan terus ikut berpartisipasi dalam program yang diselenggarakan oleh pemerintah melalui masukan dan keluhan dari masyarakat, guna memperbaiki sistem pelayanan publik agar berjalan dengan efektif dan efisien.

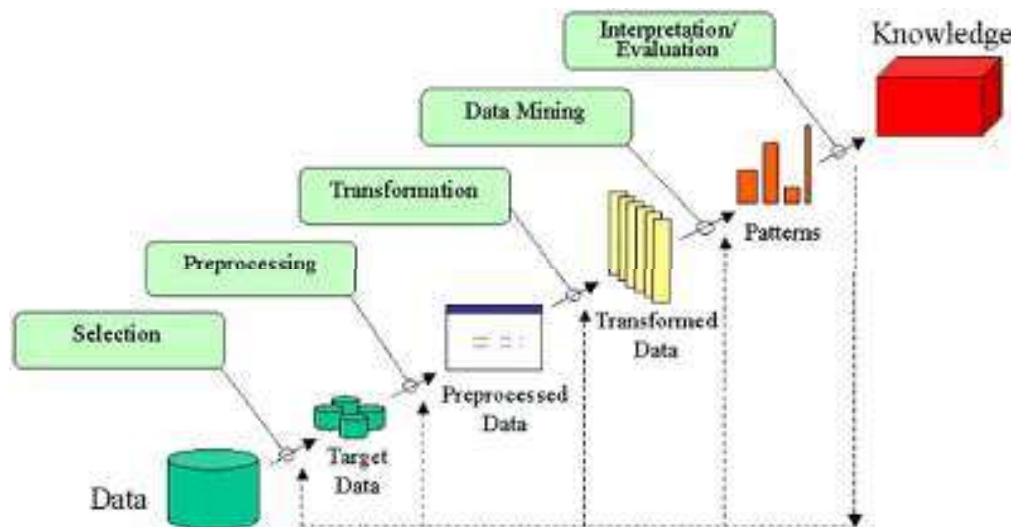
BAB II

TINJAUAN PUSTAKA

2.1. Teori Umum

2.1.1. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) merupakan proses yang terdiri dari pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar *Knowledge Discovery in Database* (KDD) didefinisikan sebagai ekstraksi informasi potensial, implisit dan tidak dikenal dari sekumpulan data. Proses *Knowlegde Discovery in Database* melibatkan hasil proses data mining (proses pengekstrak kecenderungan suatu pola data) kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami P. P. Putra & Chan, 2018).



Gambar 2.1 Knowledge Discovery in Database
Sumber: (Data Penelitian, 2023)

Berikut Langkah proses data mining pada *knowledge discovery in database* (KDD) adalah sebagai berikut (R. R. Putra & Wadisman, 2018):

1. 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.

2. *Preprocessing/Cleaning*

Saat sebelum proses data mining bisa dicoba, butuh dilakukan cara membersihkan kepada informasi yang jadi fokus KDD. Cara membersihkan mencakup penghapusan data duplikasi, pengecekan data yang tidak konsisten, serta pemeriksaan kekeliruan pada data tersebut, semacam kekeliruan cap (tipografi).

3. Transformasi Data

Transformasi data adalah 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.

4. Data mining

Data mining 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.

5. 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 fakta.

2.1.2. 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. Pemilihan tugas data mining, pemilihan goal dari proses KDD misalnya klasifikasi, regresi, *clustering*, dll.

Pemilihan algoritma data mining untuk pengelompokkan (*clustering*). Proses data mining yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan (P. P. Putra & Chan, 2018).

2.1.3. Clustering

Menurut (Aditya et al., 2020) *Clustering* data mining adalah nama lain untuk

metode *segmentation*. Tujuan dari segmentasi pada metode data mining adalah mengelompokkan suatu class ke dalam beberapa segmen berdasarkan atribut yang ditentukan. Penentuan atribut harus sesuai kesamaan yang dimiliki beberapa *class* tadi. *K-means* merupakan algoritma *clustering*. *K-means Clustering* adalah salah satu “*unsupervised machine learning algorithms*” yang paling sederhana dan populer. *K-Means Clustering* adalah suatu metode penganalisaan data atau metode data mining yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistemn partisi. *Cluster* adalah suatu kumpulan dari entitas yang hampir sama

Clustering adalah metode penganalisaan data yang sering dimasukkan sebagai salah satu metode data mining yang tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama ke suatu wilayah yang sama dan data dengan karakteristik yang berbeda ke wilayah yang lain. Ada beberapa pendekatan yang digunakan dalam mengembangkan metode *clustering*. Dua pendekatan utama adalah *clustering* dengan pendekatan partisi dan *clustering* dengan pendekatan hirarki. *Clustering* dengan pendekatan partisi atau sering disebut dengan *partition-based clustering* mengelompokkan data dengan memilah-milah data yang dianalisa ke dalam *cluster-cluster* yang ada (Nabila et al., 2021).

Clustering dengan pendekatan hirarki atau sering disebut dengan *hierarchical clustering* mengelompokkan data dengan membuat suatu hirarki berupa dendogram dimana data yang mirip akan ditempatkan pada hirarki yang berdekatan dan yang tidak pada hirarki yang berjauhan. Di samping kedua pendekatan tersebut, ada juga *clustering* dengan pendekatan *automatic mapping*.

2.1.4. *K-Means*

Menurut (Nabila et al., 2021) metode *K-means* adalah metode pengelompokan data dengan mengambil parameter sejumlah k *cluster* dan mempartisi data kedalam *cluster* tersebut, dengan berpatokan pada kemiripan antara data dalam suatu *cluster* dan ketidakmiripan di antara *cluster* berbeda, pusat dari *cluster* adalah rata-rata dari nilai anggota *cluster* yang disebut dengan *centroid* atau *center of gravity*. Selain itu *K-means* melakukan pengelompokan dengan meminimalkan jumlah kuadrat dari jarak (*distance*) antara data dengan *centroid cluster* yang cocok. Dengan kata lain, metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu *cluster* dan memaksimalkan variasi dengan data yang ada di *cluster* lainnya.

Adapun dalam melakukan perhitungan untuk mengetahui jarak data terhadap pusat data adalah dengan menggunakan rumus *Euclidean* yaitu:

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

Rumus 2.1 Perhitungan *Euclidean*

Sumber: (Data Penelitian, 2023)

Keterangan:

De = *Euclidean Distance*

I = Banyaknya objek

(x, y) = Koordinat objek

(s, t) = Koordinat centroid

Menurut (Aditya et al., 2020) Algoritma *k-means* dimulai dengan pembentukan prototipe *cluster* di awal kemudian secara iteratif prototipe *cluster* ini diperbaiki hingga konvergen (tidak terjadi perubahan yang signifikan pada prototipe *cluster*). Perubahan ini diukur menggunakan fungsi objektif yang

umumnya didefinisikan sebagai jumlah atau rata-rata jarak tiap item data dengan pusat massa kelompoknya.

2.1.5. RapidMiner



Gambar 2.2 Rapidminer
Sumber: (Data Penelitian, 2023)

Penerapan teknik data mining dengan algoritma *K-Means* pada penelitian ini yaitu untuk menentukan pola pelaksanaan tilang dalam proses razia yang membutuhkan software pendukung data mining yang bersifat *open source* yaitu *RapidMiner*. *RapidMiner* merupakan perangkat lunak yang bersifat terbuka (*open source*). *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap data mining, *text mining* dan analisis prediksi. *RapidMiner* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *RapidMiner* memiliki kurang lebih 500 operator data mining termasuk operator untuk input, output, data preprocessing dan visualisasi (Nabila et al., 2021).

RapidMiner merupakan perangkat lunak yang bersifat sumber terbuka (*Open Source*). *RapidMiner* adalah teknologi untuk memeriksa penambangan data, penambangan teks dan analitik prediktif. Sebagai alat analisis data dan mesin

penambangan 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. *RapidMiner* sebelumnya bernama YALE (*Yet Another Learning Environment*) dimana versi awalnya mulai di kembangkan pada tahun 2001 oleh RalfKlinkenberg, Ingo Mierswa, dan Simon Fischer di *Artificial Intelligence Unit* dari *University of Dortmund*. *RapidMiner* distribusikan di bawah lisensi AGPL (*GNU Affero General Public License*) versi tiga. Hingga saat ini telah ribuan aplikasi yang dikembangkan menggunakan *RapidMiner* di lebih dari 40 negara.

Menurut (Zaki Muhammad et al., 2018) *RapidMiner* adalah platform perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, yang menyediakan lingkungan terpadu untuk pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*text mining*) dan analisis prediktif (*predictive analytics*).

Rapidminer memiliki beberapa sifat sebagai berikut :

1. Ditulis dengan bahasa pemrograman Java sehingga dapat dijalankan di berbagai sistem operasi.
2. Proses penemuan pengetahuan dimodelkan sebagai operator *trees*.
3. Representasi XML internal untuk memastikan format standar pertukaran data.

4. Bahasa *scripting* memungkinkan untuk eksperimen skala besar dan otomatisasi eksperimen.
5. Konsep *multi-layer* untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
6. Memiliki GUI, *command line mode* dan Java API yang dapat dipanggil dari program lain.

Beberapa Fitur dari Rapidminer, antara lain:

1. Banyaknya algoritma data mining, seperti *decision tree* dan *self-organization map*.
2. Bentuk grafis yang canggih, seperti tumpang tindih diagram histogram, *tree chart* dan *3D Scatter plots*.
3. Banyaknya variasi *plugin*, seperti *text plugin* untuk melakukan analisis teks.
4. Menyediakan prosedur data mining dan *machine learning* termasuk ETL (*extraction, transformation, loading*), data *preprocessing*, visualisasi, *modelling* dan evaluasi.
5. Proses data mining tersusun atas operator-operator yang *nestable*, dideskripsikan dengan XML dan dibuat dengan GUI.
6. Mengintegrasikan proyek data mining Weka dan statistika R.

2.2. Teori Khusus

2.2.1. Pengertian Lalu Lintas

Lalu lintas diartikan sebagai gerak bolak-balik manusia atau barang dari suatu tempat ke tempat lainnya dengan menggunakan sarana jalan (Rahayu Nurfauziah & Hetty Krisnan, 2021). Menurut Pasal 1 Undang-Undang Nomor 22 Tahun 2009

Tentang Lalu Lintas dan Angkutan Jalan mendefinisikan bahwa lalu lintas adalah gerak kendaraan dan orang di ruang lalu lintas jalan. Lalu lintas menurut Poerwodarminto (1993:55) yaitu:

1. Perjalanan bolak-balik
2. Perihal perjalanan di jalan dan sebagainya
3. Berhubungan antara sebuah tempat.

Definisi-definisi tersebut dapat diartikan bahwa lalu lintas adalah segala sesuatu hal yang berhubungan langsung dengan sarana jalan yang menjadi sarana utamanya untuk dapat mencapai satu tujuan yang dituju baik disertai maupun tidak disertai oleh alat angkut. Jadi, di dalam lalu lintas ada tiga komponen penyusunnya yaitu manusia, kendaraan, dan jalan yang saling berinteraksi dalam proses pergerakan.



Gambar 2.3 Sistem Komponen Lalu Lintas
Sumber: (Data Penelitian, 2023)

a. Manusia

Dalam komponen lalu lintas manusia berperan sebagai pengendara atau penumpang atau pejalan kaki dan mempunyai keadaan yang berbeda beda.

b. Kendaraan

Dalam komponen lalu lintas kendaraan merupakan suatu sarana angkut penumpang maupun barang yang terdiri atas kendaraan bermotor dan kendaraan tidak bermotor. Dalam Undang - Undang Nomor 22 Tahun 2009 tentang Lalu Lintas dan Angkutan Jalan, jenis kendaraan bermotor dibagi menjadi beberapa jenis yaitu sepeda motor, mobil penumpang, mobil bus mobil barang dan mobil khusus.

c. Jalan

Dalam komponen lalu lintas jalan merupakan lintasan yang direncanakan dan digunakan kepada pengguna kendaraan bermotor maupun tidak bermotor, jalan juga digunakan untuk mengalirkan aliran lalu lintas dengan lancar, mendukung beban muatan kendaraan.

2.2.2. Sepeda Motor

Sepeda Motor atau angkutan bermotor menurut Nasution (1996) adalah roda dua transportasi yang menggunakan kendaraan bermotor sebagai fasilitasnya yang bergerak di jalan raya. Salah satu angkutan bermotor itu adalah sepeda motor, sepeda motor pertama kali dirancang oleh *Ernest Michaud* pada Tahun 1868 dengan menggunakan mesin berjenis uap. Tapi pada saat itu proyek sepeda motor dengan mesin uap gagal, lalu pada Tahun 1885 *Edward Butler* memperbaiki kendaraan tersebut dengan menggunakan mesin berjenis mesin pembakaran dalam dan pada tahun tersebut juga *Gottlieb Daimler* dan *Wilhelm Maybach* sebagai ahli mesin di Jerman menjadi seorang perakit motor pertama kali di dunia. Sejak saat itu lah banyak penemuan lainnya dalam perkembangan jenis sepeda motor ini.

Selain itu sepeda motor juga dapat membantu masyarakat untuk melakukan aktivitas dan dapat pula menghemat perjalanan untuk sampai tujuan. Selain menghemat perjalanan sepeda motor juga dapat menghindari macet, mempermudah aktivitas, menghemat waktu dan dapat meminimalisir pengeluaran bahan bakar minyak (Vivin Tannia & Nyoman Yulianthini, 2021).

2.2.3. Pelanggaran Lalu Lintas

Pelanggaran Lalu Lintas merupakan perbuatan melanggar sesuatu dan berhubungan dengan hukum. Pelanggaran lalu lintas itu adalah satu perbuatan atau tindakan pengendara yang bertentangan dengan ketentuan dan peraturan Undang-Undang Lalu Lintas. Jika ketentuan tersebut dilanggar oleh pengendara maka kecelakaan dalam berkendara kemungkinan dapat terjadi. Jadi dari definisi pelanggaran lalu lintas diatas dapat diketahui bahwa pelanggaran lalu lintas itu adalah suatu perbuatan yang bertentangan dengan hukum dan akan menimbulkan akibat dari perbuatan itu (Nurfauziah, 2021).

2.2.4. Faktor Pelanggaran Lalu Lintas

Menurut (Suryandari et al., 2022) faktor Pelanggaran lalu lintas saat ini terdapat banyak faktor yang mempengaruhi terjadinya pelanggaran lalu lintas. Faktor-faktor tersebut antara lain yaitu:

a. Faktor manusia

Menurut Suwardjoko mengatakan bahwa hampir semua bentuk pelanggaran lalu lintas dan kecelakaan lalu lintas yang ada disebabkan oleh pengendara.

Hal ini dipertegas juga oleh Hoobs mengatakan bahwa manusia adalah

penyebab paling banyak dalam pelanggaran dan kecelakaan yang ada. Faktor manusia ini mencakup psikologi dan sistem indra seperti penglihatan dan pendengaran dan pengetahuan akan tata cara lalu lintas. Ada beberapa indikator yang membentuk sikap dan perilaku manusia dalam berkendara, diantaranya adalah:

1) Pengetahuan

Pemerintah telah membuat peraturan lalu lintas yang ditujukan kepada setiap pengguna jalan demi menciptakan keamanan, keselamatan, ketertiban dan kelancaran lalu lintas dalam bentuk Undang-Undang, Perpu, Perda, dan aturan lainnya. Oleh karena itu setiap pengguna jalan wajib memahami dan menjalani setiap aturan yang telah dibuat sehingga terbentuk satu persepsi dalam pola pikir dan tindakan dalam berinteraksi di jalan raya. Pola pikir yang terbentuk dapat terjadinya perbedaan tingkat pemahaman dan pengetahuan antara pengguna jalannya terhadap peraturan yang ada sehingga berpotensi munculnya masalah dalam berlalu lintas, baik antara pengguna jalan sendiri maupun antara pengguna jalan dengan aparat yang bertugas. Selain pemahaman terhadap peraturan yang ada, pengendara juga harus memiliki pemahaman tentang karakteristik kendaraanya. Setiap kendaraan memiliki karakteristik yang berbeda-beda. Pemahaman terhadap karakteristik kendaraan sangat berpengaruh terhadap operasional kendaraan di jalan raya yang berpengaruh juga terhadap situasi lalu lintas.

2) Mental

Salah satu yang menjadi faktor utama terhadap situasi lalu lintas adalah

mental pengendara tersebut. Untuk menciptakan sebuah interaksi dengan hasil seperti keamanan, keselamatan, kelancaran lalu lintas pengendara harus bisa menjaga etika, sopan-santun, toleransi antar pengguna jalan dan mengendalikan emosi. Jika pengendara tidak bisa menjaga hal itu maka dampak negatif yang diperoleh seperti menimbulkan kemacetan, pelanggaran lalu lintas dan bahkan dapat menyebabkan kecelakaan lalu lintas.

3) Keterampilan

Demi menciptakan keamanan, keselamatan, ketertiban, dan kelancaran lalu lintas pengendara harus memiliki keterampilan dalam mengendalikan kendaraannya, karena hal ini akan berpengaruh besar terhadap situasi lalu lintas yang ada. Setiap pengendara harus memiliki lisensi terhadap kemampuan dalam mengendalikan kendaraan yang diwujudkan secara formal melalui Surat Izin Mengemudi (SIM). Di antara faktor-faktor yang ada, faktor manusia menjadi penyebab yang paling tinggi dalam pelanggaran lalu lintas karena untuk faktor manusia berkaitan erat dengan tingkah laku, etika dan tata cara berkendara di jalan.

2.2.5. Dampak Pelanggaran Lalu Lintas

Dampak yang akan terjadi akibat pelanggaran-pelanggaran lalu lintas yang ada antara lain seperti:

1. Terjadinya peningkatan angka kecelakaan di jalan baik pada perempatan maupun tidak.
2. Rawan terjadi kecelakaan tunggal maupun beruntun.
3. Dapat merugikan diri sendiri maupun orang lain.

4. Tindakan melanggar rambu lalu lintas dapat membuat meningkatnya angka kecelakaan.
5. Kurangnya kesadaran pengendara dalam mematuhi peraturan yang ada dapat membuat kemacetan semakin parah.
6. Tindakan melanggar peraturan lalu lintas akan menciptakan suatu kebiasaan melanggar lalu lintas yang terjadi secara terus menerus.

2.3. Penelitian Terdahulu

Berikut ini adalah beberapa penelitian terdahulu sebagai acuan dalam penelitian ini:

1. Berdasarkan jurnal penelitian (Joseph & Thanakumar, 2019) dengan judul *“Survey Of Data mining Algorithms For Intelligent Computing System”*. *The Intelligent computing system, described to be a collection of the connected device working in mutual understanding to attain a particular purpose, is an incorporation of artificial intelligence and the computational intelligence, and are employed in variety of applications. The paper presents the survey on the data mining algorithms and the techniques that could be employed with the intelligent computing system, presenting a basic conception of the data mining along with the prominent algorithms of the data mining and the classification of its techniques, further the survey concludes with the challenges included in the overview of the survey done along with the future enhancement in the research that analyses the data mining techniques in the intelligent computing applications.*
2. Berdasarkan jurnal penelitian (Jalal & Ali, 2021) dengan judul *“Text*

documents clustering using data mining techniques”. Increasing progress in numerous research fields and information technologies, led to an increase in the publication of research papers. Therefore, researchers take a lot of time to find interesting research papers that are close to their field of specialization. Consequently, in this paper we have proposed documents classification approach that can cluster the text documents of research papers into the meaningful categories in which contain a similar scientific field. Our presented approach based on essential focus and scopes of the target categories, where each of these categories includes many topics. Accordingly, we extract word tokens from these topics that relate to a specific category, separately. The frequency of word tokens in documents impacts on weight of document that calculated by using a numerical statistic of term frequency-inverse document frequency (TF-IDF). The proposed approach uses title, abstract, and keywords of the paper, in addition to the categories topics to perform the classification process. Subsequently, documents are classified and clustered into the primary categories based on the highest measure of cosine similarity between category weight and documents weights.

3. Berdasarkan jurnal penelitian (Simanjuntak et al., 2021) dengan judul “*Data mining Rekomendasi Pemakaian Skincare*”. Pada proses pengolahan data dengan menggunakan algoritma *naïve bayes* yang merupakan salah satu metode klasifikasi *data mining* untuk menentukan pola pembelian konsumen terhadap produk yang ditawarkan oleh Toko *Skincare* dengan menggunakan data *training* dari transaksi selama ini dengan mengambil 4 variabel yang

sangat mempengaruhi di mana untuk membuat penjualan mereka laris di pasaran harus memperhatikan produk, jenis kulit, manfaat dan harga. Penerapan teknik data ini sangat efisien dan efektif untuk memprediksi bisnis ke depannya hal ini dapat dilihat dari hasil probabilitas dari penjualan mereka selama ini terhadap pembelian memiliki 72 % prediksi yang benar dan 28% prediksi yang salah.

4. Berdasarkan jurnal penelitian (Indraputra et al., 2020) dengan judul “*K-Means Clustering Data COVID-19*”. Tujuan dari penelitian ini adalah untuk memperoleh dan mengolah data *COVID-19* yang terdapat pada *Kaggle* menggunakan metode *Data mining* yaitu *K-Means Clustering* Untuk *K-Means Clustering* pada penelitian ini akan digunakan tiga metode untuk mengolah data yaitu pengolahan menggunakan *software. Microsoft Excel* dan *software Data mining* yaitu *Weka* dan *KNIME*. Dari hasil pengolahan data, diperoleh dua *cluster* data, dimana *cluster 2* memiliki jumlah terjangkit dan meninggal yang lebih tinggi dibandingkan dengan *cluster 1*, maka daerah - daerah *cluster* tersebut perlu diprioritaskan penanganannya. Sampel data ini dapat dibagi menjadi 2 *cluster*, dimana pada *cluster 1* terdapat 29 data dan pada *cluster 2* terdapat 19 data. Dengan informasi ini, dapat diketahui *cluster* daerah-daerah yang memerlukan penanganan COVID-19 yang lebih darurat, di mana pada sampel ini merupakan *cluster 2* dengan jumlah terjangkit dan meninggal yang lebih tinggi dibandingkan dengan *cluster 1*.
5. Berdasarkan jurnal penelitian (Daeli et al., 2022) dengan judul “Implementasi *Data mining* Untuk Mengelompokkan Lokasi Berdasarkan Tingkat

Kejahatan Pada Kabupaten Nias Barat Menggunakan Metode *K-Means Clustering*". Kabupaten Nias Barat terdiri dari 115 lokasi atau desa yang memiliki tingkat kejahatan berbeda-beda. Pihak kepolisian pada Kabupaten Nias Barat akan membagi tugas untuk menjaga setiap lokasi di mana jumlah petugas pada setiap desa disesuaikan dengan tingkat kejahatan pada lokasi tersebut. Sulitnya untuk mengetahui atau mengelompokkan lokasi berdasarkan tingkat kejahatan pada Kabupaten Nias Barat menimbulkan permasalahan yang mempersulit pembagian tugas oleh pihak kepolisian untuk menjaga keamanan setiap lokasi. Atas dasar masalah tersebut, maka dibutuhkan sebuah aplikasi *data mining* yang mampu mengelompokkan lokasi berdasarkan tingkat kejahatan pada Kabupaten Nias Barat. Di mana setiap data yang diperoleh dari kepolisian Kabupaten Nias Barat akan dihitung menggunakan salah satu metode *data mining* yang mampu dalam mengelompokkan lokasi kejahatan yaitu metode *K-means Clustering*. Hasil dari penelitian adalah sebuah aplikasi *data mining* yang mengadopsi metode *K-Means Clustering* yang mampu menjawab permasalahan yang ada pada Kabupaten Nias Barat terkait pengelompokan lokasi berdasarkan tingkat kejahatan.

2.4. Kerangka Pemikiran

Kerangka pemikiran adalah alur utama penelitian, dalam urutan masalah, peluang, pendekatan, identifikasi dan pemetaan, pemodelan, evaluasi dan hasil. Pada dasarnya, penelitian adalah kerangka hubungan antar konsep yang ingin diamati atau diukur melalui penelitian yang ingin dilakukan. Karena standar

kualitas pendaftar berada dalam kisaran yang cukup luas, studi ini tidak konsisten berdasarkan jurusan yang dapat diterima (Annisa Ekasetya & Jananto, 2020).



Gambar 2.4 Kerangka Pemikiran
Sumber: (Data Penelitian, 2023)

Kerangka Pemikiran merupakan tahapan sebelum pelaksanaan penelitian.

Berikut langkah-langkah kerangka pemikiran yang akan dilaksanakan, yaitu :

1. Melaksanakan studi literatur dengan cara membaca jurnal untuk memahami mengenai data mining dan algoritma *K-Means*.
2. Melakukan observasi, wawancara dan pengumpulan data yang akan diperlukan.
3. Mengemukakan masalah serta Batasan masalah yang akan diteliti.
4. Melakukan pengolahan data tilang kendaraan roda dua untuk mempermudah

clustering.

5. Menerapkan metode *K-Means* untuk mendapatkan hasil *cluster* dan melakukan pengujian dengan software *RapidMiner*.
6. Menyimpulkan hasil akhir dari pengolahan data yang telah dilakukan.

BAB III

METODE PENELITIAN

3.1. Desain 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 3.1 Desain Penelitian
Sumber: (Data Penelitian, 2023)

Berdasarkan gambar 3.1 diatas maka tahapan yang akan dilakukan, 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 yaitu 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 *system*.

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.

3.2. Teknik Pengumpulan Data

Pengumpulan data merupakan salah satu kegiatan yang sangat penting pada penelitian ini, untuk itu maka penulis melakukan teknik pengumpulan data berikut:

1. Observasi

Teknik ini merupakan teknik pengumpulan data berdasarkan observasi lingkungan. Pada titik ini, penulis melakukan observasi terhadap data tilang yaitu objek penelitian untuk mendapatkan informasi berdasarkan fakta.

2. Wawancara

Teknik ini merupakan teknik langsung dengan menghubungi pihak yang berkaitan dengan objek dan permasalahan yang diteliti kemudian akan melakukan wawancara kepada pihak Subdit Gakkum Ditlantas Polda Kepri..

3. Teknik Studi Literatur

Teknik ini merupakan perdebatan dalam bidang studi tertentu. Kami juga mendapatkan gambaran singkat tentang apa yang bisa dipelajari, penalaran dan pemecahan masalah melalui penelitian masalah. Oleh karena itu dalam penelitian ini penulis mencari berbagai informasi pendukung dari jurnal. Penulis mengkajinya dan menarik kesimpulan yang dapat membantu dalam penelitian.

3.3. Operasional Variabel

Adapun operasional variabel yang akan digunakan dalam penelitian ini adalah:

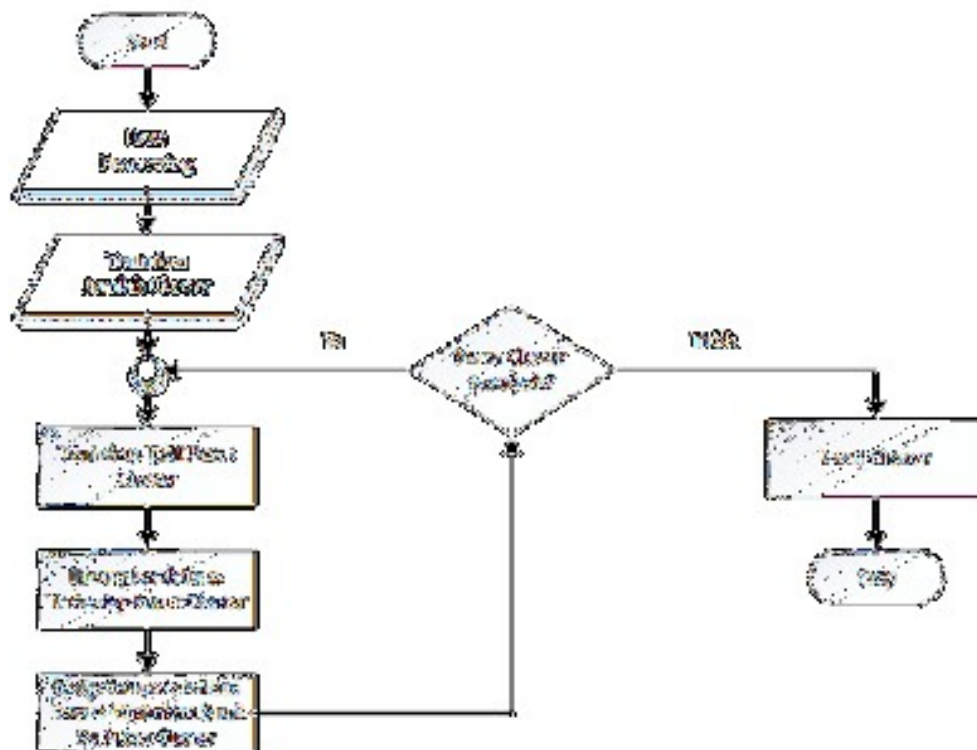
1. Kategori Pelanggaran

2. Jenis Pelanggaran

3. Denda

3.4. Algoritma

Algoritma yang di manfaat penulis adalah algoritma *K-means* dimana memiliki langkah kerja sebagai berikut :



Gambar 3.2 *Flowchart* Algoritma *K-Means*
 Sumber: (Data Penelitian, 2023)

3.5. Lokasi dan Jadwal Penelitian

3.5.1. Lokasi

Lokasi penelitian akan dilakukan pada kantor yang berlokasi di Polda Kepri Jalan Hang Jebat No.81 Batu besar, Nongsa. Untuk lebih jelasnya dapat dilihat dari

map berikut.



Gambar 3.3 Lokasi Penelitian
Sumber: Google Map (2023)

3.5.2. Jadwal Penelitian

Dalam melakukan penelitian ini tentunya penulis memiliki jadwal agar dapat terselesaikan tepat waktu adapun jadwalnya dapat dilihat pada tabel berikut :

Tabel 3.1 Jadwal Penelitian

Kegiatan	Waktu Kegiatan (Tahun)																			
	2022																			
	Sept				Okt				Nov				Des				Jan			
	Minggu Ke-				Minggu Ke-				Minggu Ke-				Minggu Ke-				Minggu Ke-			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Pengajuan Judul																				
Penyusunan BAB I																				
Penyusunan BAB II																				
Penyusunan BAB III																				
Penyusunan BAB IV																				

Sumber: (Data Penelitian, 2023)