

**IMPLEMENTASI ALGORITMA *K-PROTOTYPE* PADA
PENGELOMPOKAN DATA PELANGGARAN LALU LINTAS
DI KOTA MALANG**

SKRIPSI

**OLEH:
IVAN BINTANG SABRANI
NIM. 210601110076**



**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2024**

**IMPLEMENTASI ALGORITMA *K-PROTOTYPE* PADA
PENGELOMPOKAN DATA PELANGGARAN LALU LINTAS
DI KOTA MALANG**

SKRIPSI

**Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh
Ivan Bintang Sabrani
NIM. 210601110076**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2024**

**IMPLEMENTASI ALGORITMA *K-PROTOTYPE* PADA
PENGELOMPOKAN DATA PELANGGARAN LALU LINTAS
DI KOTA MALANG**

SKRIPSI

**Oleh
Ivan Bintang Sabrani
NIM. 210601110076**

Telah Disetujui Untuk Diuji

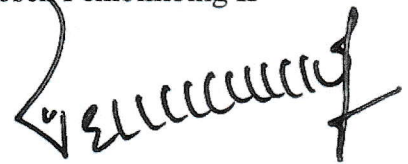
Malang, 5 Desember 2024

Dosen Pembimbing I



Hisyam Fahmi, M.Kom
NIP. 19890727 201903 1 18

Dosen Pembimbing II



Evawati Alisah, M.Pd
NIP. 19720604 199903 2 001

Mengetahui,
Ketua Program Studi Matematika



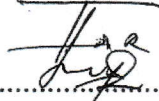
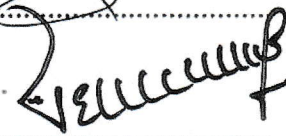

Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005

**IMPLEMENTASI ALGORITMA *K-PROTOTYPE* PADA
PENGELOMPOKAN DATA PELANGGARAN LALU LINTAS
DI KOTA MALANG**

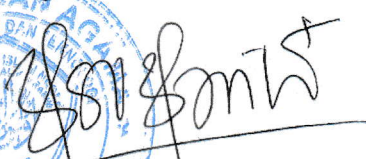
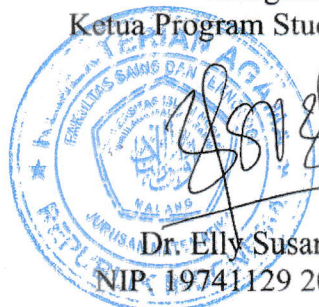
SKRIPSI

Oleh
Ivan Bintang Sabrani
NIM. 210601110076

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S. Mat)
Tanggal 20 Desember 2024

Ketua Penguji	: Dr. Fachrur Rozi, M.Si	
Anggota Penguji 1	: M. Nafie Jauhari, M.Si	
Anggota Penguji 2	: Hisyam Fahmi, M.Kom	
Anggota Penguji 3	: Evawati Alisah, M.Pd	

Mengetahui,
Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Ivan Bintang Sabrani
NIM : 210601110076
Program Studi : Matematika
Fakultas : Sains dan Teknologi
Judul Skripsi : Implementasi Algoritma *K-Prototype* Pada
Pengelompokan Data Pelanggaran Lalu Lintas Di
Kota Malang

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan dan pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 20 Desember 2024

Yang membuat pernyataan,



Ivan Bintang Sabrani
NIM. 210601110076

MOTO

“Apa yang tidak dimulai hari ini, tidak akan pernah selesai besok”

(Johann Wolfgang von Goeth)

PERSEMBAHAN

Dengan penuh rasa syukur, kupersembahkan skripsi ini untuk Ayahku tersayang, Alm. Bambang Budi Wahono. Meskipun ragamu telah tiada, kenangan dan kasih sayangmu terus hidup dalam setiap langkahku. Engkau adalah pelita yang tak pernah padam, mengajarkan keteguhan hati dan semangat juang yang selalu menjadi teladan bagiku.

Untuk Ibuku tercinta, Widiyawati, wanita luar biasa yang tak pernah lelah memberikan cinta, dukungan, dan doa. Dari kelembutanmu aku belajar ketulusan, dari kekuatanmu aku memahami makna harapan, dan dari kasihmu aku menemukan alasan untuk terus berjuang.

Kepada Kakakku, Nando, yang selalu hadir dengan nasihat bijak, canda tawa, dan dukungan yang menguatkan, terima kasih atas segala inspirasi yang telah kau berikan. Juga

untuk Adikku, Andhika, sumber semangat yang mengingatkanku untuk terus bermimpi dan berjuang. Skripsi ini adalah wujud kecil dari rasa cinta, doa, dan terima kasihku kepada kalian. Semoga setiap usaha ini membawa keberkahan dan kebahagiaan untuk kita semua.

.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan proposal skripsi ini. Skripsi ini disusun sebagai bagian dari pemenuhan beban studi untuk mata kuliah Skripsi Matematika di Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Penyusunan skripsi ini tentu tidak terlepas dari dukungan dan bantuan berbagai pihak, oleh karena itu penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan dan motivasi selama proses ini. Ucapan terima kasih kepada: Ucapan terima kasih dituliskan sebagai berikut:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim.
2. Prof. Dr. Sri Harini, M.Si, selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim.
3. Dr. Elly Susanti, S.Pd., M.Sc, selaku ketua Program Studi Matematika, Universitas Islam Negeri Maulana Malik Ibrahim.
4. Hisyam Fahmi, M.Kom selaku dosen pembimbing I
5. Evawati Alisah, M.Pd selaku dosen pembimbing II
6. Seluruh dosen Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim.
7. Orang tua, serta keluarga besar penulis atas doa, dukungan, dan cinta kasih yang selalu menyertai. Segala bentuk dukungan menjadi kekuatan dalam menyelesaikan skripsi ini. Semoga segala kebaikan yang telah diberikan senantiasa mendapatkan balasan dari Allah SWT.

8. Seluruh mahasiswa angkatan 2021 dan teman penulis atas kebersamaan, dukungan, serta motivasi yang telah diberikan selama masa studi ini. Kehadiran dan semangat kalian menjadi bagian penting dalam perjalanan ini. Semoga kebersamaan ini terus terjalin dan membawa manfaat bagi kita semua.

Malang, 20 Desember 2024



Ivan Bintang Sabrani
NIM. 210601110076

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTO	vi
PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR SIMBOL	xiv
DAFTAR LAMPIRAN	xv
ABSTRAK	xvi
ABSTRACK	xvii
مستخلص البحث	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah	6
1.6 Definisi Istilah	6
BAB II KAJIAN TEORI	8
2.1 Teori Pendukung	8
2.1.1 <i>Data mining</i>	8
2.1.2 <i>Clustering</i>	10
2.1.3 <i>Algoritma K- Prototype</i>	12
2.1.4 <i>Normalisasi Data</i>	16
2.1.5 <i>Metode Elbow</i>	17
2.1.6 <i>Pelanggaran Lalu Lintas</i>	18
2.2 Kajian Topik dalam Al-Qur'an/Hadits	23
2.3 Kajian Topik Dengan Teori Pendukung	25
BAB III METODE PENELITIAN	27
3.1 Jenis Penelitian	27
3.2 Data dan Sumber Data	27
3.3 Teknik Analisis Data	28
BAB IV HASIL DAN PEMBAHASAN	31
4.1 Seleksi Data (<i>Data selection</i>)	31
4.2 Transformasi Data	31
4.3 Preprocessing Data	34
4.4 Implementasi Algoritma <i>K-Prototype</i>	36
4.4.1 Implementasi Data Sampel Perhitungan Manual	
Algoritma <i>K-Prototype</i>	36

4.4.2 Implementasi Algoritma <i>K-Prototype</i> Menggunakan Python..	55
4.5 Evaluasi <i>Clustering</i>	57
4.6 Hasil <i>Clustering K-Prototype</i>	58
4.7 Kajian Keislaman Dengan Hasil Penelitian.....	70
BAB V PENUTUP	75
5.1 Kesimpulan.....	75
5.2 Saran	76
DAFTAR PUSTAKA	77
LAMPIRAN	79
RIWAYAT HIDUP	82

DAFTAR TABEL

Tabel 4.1	Seleksi Data	31
Tabel 4.2	Inisialisasi Atribut Pasal	32
Tabel 4.3	Inisialisasi Atribut Formulir	33
Tabel 4.4	Inisialisasi Atribut Jenis Kendaraan	33
Tabel 4.5	Normalisasi Atribut Denda	34
Tabel 4.6	Data Sampel Pelanggaran Lalu Lintas	37
Tabel 4.7	<i>Centroid</i> Sampel Data	37
Tabel 4.8	Data Denda dan Hasil Kuadrat Selisih terhadap Mean	38
Tabel 4.9	Hasil Klaster Iterasi 1	41
Tabel 4.10	Pengelompokan Untuk Klaster 0 Iterasi 1	41
Tabel 4.11	Pengelompokan Untuk Klaster 1 Iterasi 1	41
Tabel 4.12	<i>Centroid</i> Baru dari Iterasi 1	42
Tabel 4.13	Hasil Klaster Iterasi ke-2	42
Tabel 4.14	Pengelompokan Untuk Klaster 0 Iterasi ke-2	43
Tabel 4.15	Pengelompokan Untuk Klaster 1 Iterasi ke-2	43
Tabel 4.16	<i>Centroid</i> Baru Dari Iterasi ke-2	43
Tabel 4.17	Hasil Klaster Iterasi ke-3	43
Tabel 4.18	<i>Centroid</i> Baru Dari Iterasi ke-3	44
Tabel 4.19	WCSS Untuk $K = 2$	45
Tabel 4.20	Hasil Jarak Untuk $K = 1$	46
Tabel 4.21	WCSS Untuk $K = 1$	46
Tabel 4.22	Hasil Jarak Untuk $K = 3$	47
Tabel 4.23	WCSS Untuk $K = 3$	47
Tabel 4.24	Hasil Jarak Untuk $K = 4$	48
Tabel 4.25	WCSS Untuk $K = 4$	49
Tabel 4.26	Hasil Jarak Untuk $K = 5$	50
Tabel 4.27	WCSS Untuk $K = 5$	50
Tabel 4.28	Hasil Jarak Untuk $K = 6$	51
Tabel 4.29	WCSS Untuk $K = 6$	51
Tabel 4.30	Hasil Jarak Untuk $K = 7$	52
Tabel 4.31	WCSS Untuk $K = 7$	52
Tabel 4.32	WCSS Setiap Klaster Pada Data Sampel	53
Tabel 4.33	WCSS Setiap Klaster Pada Data Asli	55
Tabel 4.34	Hasil <i>Clustering</i> Jenis Kendaraan	59
Tabel 4.35	Hasil <i>Clustering</i> Pasal	59
Tabel 4.36	Hasil <i>Clustering</i> Barang Bukti	59
Tabel 4.37	Hasil <i>Clustering</i> Denda	60
Tabel 4.38	Hasil <i>Clustering</i> Form	60

DAFTAR GAMBAR

Gambar 2.1	KDD (<i>Knowledge Discovery in Databases</i>).....	9
Gambar 2.2	Contoh Grafik Metode <i>Elbow</i>	18
Gambar 3.1	Teknik Analisis Data	28
Gambar 4.1	Pengecekan <i>Missing value</i>	35
Gambar 4.2	Data Siap Diolah.....	35
Gambar 4.3	<i>Centroid</i> Untuk $K = 1$	46
Gambar 4.4	<i>Centroid</i> Untuk $K = 3$	47
Gambar 4.5	<i>Centroid</i> Untuk $K = 4$	48
Gambar 4.6	<i>Centroid</i> Untuk $K = 5$	49
Gambar 4.7	<i>Centroid</i> Untuk $K = 5$	51
Gambar 4.8	<i>Centroid</i> Untuk $K = 7$	52
Gambar 4.9	<i>Elbow Method</i>	54
Gambar 4.10	Proses <i>K-Prototype</i>	56
Gambar 4.11	<i>Centroid</i>	56
Gambar 4.12	Hasil <i>Clustering</i>	57
Gambar 4.13	<i>Elbow Method</i> Pada Data Asli.....	58
Gambar 4.14	Grafik Hasil <i>Cluster 0</i> Pada Atribut (Form, Pasal, Jenis Kendaraan, Barang Bukti, Dan Denda).....	63
Gambar 4.15	Grafik Hasil <i>Cluster 1</i> Pada Atribut (Form, Pasal, Kendaraan, Barang Bukti, Dan Denda)	65
Gambar 4.16	Grafik Hasil <i>Cluster 2</i> Pada Atribut (Form, Pasal, Jenis Kendaraan, Barang Bukti, Dan Denda).....	67
Gambar 4.17	Grafik Hasil <i>Cluster 3</i> Pada Atribut (Form, Pasal, Jenis Kendaraan, Barang Bukti, Dan Denda).....	69

DAFTAR SIMBOL

$d(X_j, Z_i)$	: Jarak euclidean antara data X_j dan Z_i
m_N	: Jumlah atribut numerik
x_{jl}^N	: Nilai data ke j pada atribut numerik ke l
z_{il}^N	: Nilai <i>Centroid</i> ke i pada atribut numerik ke l
m_c	: Jumlah atribut kategorikal
x_{jl}^C	: Nilai data ke j pada atribut kategorik ke l
z_{il}^C	: Nilai <i>Centroid</i> ke i pada atribut kategorik ke l
γ	: Parameter penyeimbang
x'	: Data yang akan dinormalisasi
x	: Nilai yang dipilih dari data
x_{min}	: Nilai Minimum dari data
x_{max}	: Nilai Maksimum dari data
v_i	: <i>Centroid</i> klaster ke- i
N_i	: Total anggota dalam klaster ke- i
x_{ji}	: Nilai data ke j dalam klaster ke- i

DAFTAR LAMPIRAN

Lampiran 1	Data Pelanggaran Lalu Lintas Di kota Malang Tahun 2021	79
Lampiran 2	Data Setelah Dilakukan Transformasi Data dan Preprocessig Data.....	79
Lampiran 3	Hasil <i>Clustering K-prototype</i>	79
Lampiran 4	Source Code Algoritma <i>K-Prototype</i>	79

ABSTRAK

Sabrani, Ivan Bintang. 2024. **Implementasi Algoritma K-Prototype pada Pengelompokan Data Pelanggaran Lalu Lintas di Kota Malang**. Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Hisyam Fahmi, M.Kom. (II) Evawati Alisah, M.Pd.

Kata Kunci: *K-Prototype, Clustering, Pelanggaran Lalu lintas, Kota Malang, Data mining.*

Penelitian ini bertujuan mengelompokkan data pelanggaran lalu lintas di Kota Malang tahun 2021. *Data mining* menjadi salah satu teknik penting dalam analisis data skala besar untuk mengidentifikasi pola-pola berharga yang dapat mendukung pengambilan keputusan. Salah satu metode dalam *data mining* adalah *clustering*, yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik. Algoritma K-Prototype diterapkan untuk melakukan *clustering* pada data pelanggaran lalu lintas, yang melibatkan atribut numerik dan kategorikal. Data yang digunakan meliputi 7.328 kasus dengan lima atribut utama: pasal yang dilanggar, barang bukti, jenis kendaraan, formulir tilang, dan jumlah denda. Proses penelitian meliputi transformasi data, normalisasi atribut numerik, dan implementasi algoritma K-Prototype. Evaluasi menggunakan *Within-Cluster Sum of Squares* (WCSS) untuk memastikan kualitas *clustering*. Dihasilkan klaster terbaik sebanyak empat *euclidian*. Hasil pengelompokan ini menghasilkan klaster-klaster yang menggambarkan pola pelanggaran spesifik. Klaster 0 (4117 kasus) dan klaster 1 (1230 kasus) didominasi pelanggaran sepeda motor, dengan pelanggaran terbanyak pada Pasal 287(1) di klaster 0 dan Pasal 285(1) di klaster 1. klaster 2 (1337 kasus) dan klaster 3 (510 kasus) didominasi kendaraan roda empat, dengan pelanggaran terbanyak pada Pasal 287(1) di klaster 2 dan Pasal 286 di klaster 3. Hasilnya diharapkan menjadi referensi untuk upaya peningkatan kesadaran berlalu lintas serta perencanaan kebijakan yang lebih efektif dalam mengurangi angka pelanggaran dan kecelakaan di masa depan.

ABSTRACT

Sabrani, Ivan Bintang. 2024. **Implementation of the K-Prototype Algorithm in Clustering Traffic Violation Data in Malang City**. Thesis. Mathematics Study Program, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University of Malang. Advisors: (I) Hisyam Fahmi, M.Kom. (II) Evawati Alisah, M.Pd.

Keywords: K-Prototype, Clustering, Traffic Violations, Malang City, Data mining.

This research aims to group traffic violation data in Malang City in 2021. Data mining is one of the important techniques in large-scale data analysis to identify valuable patterns that can support decision making. One method in data mining is clustering, which is used to group data based on similar characteristics. The K-Prototype algorithm is applied to cluster traffic violation data, which involves numeric and categorical attributes. The data used includes 7.328 cases with five main attributes: the violated article, evidence, type of vehicle, ticket form, and amount of fine. The research process includes data transformation, normalization of numeric attributes, and implementation of the K-Prototype algorithm. Evaluation using Within-Cluster Sum of Squares (WCSS) to ensure clustering quality. The best clusters produced were four clusters. The results of this clustering produce clusters that describe specific violation patterns. Cluster 0 (4117 cases) and cluster 1 (1230 cases) are dominated by motorcycle violations, with the most violations being Article 287(1) in cluster 0 and Article 285(1) in Cluster 1. Cluster 2 (1337 cases) and Cluster 3 (510 cases) are dominated by four-wheeled vehicles, with the most violations being Article 287(1) in Cluster 2 and Article 286 in Cluster 3. The results are expected to be a reference for efforts to increase traffic awareness and more effective policy planning in reducing the number of violations and accidents in the future.

مستخلص البحث

صبرني ، إيفان بينتانه. ٢٠٢٤. تنفيذ خوارزمية *K-Prototype* على تجميع بيانات المخالفات المرورية في مدينة مالانغ. البحث الجامعي. قسم الرياضيات ، كلية العلوم والتكنولوجيا ، جامعة مولانا مالك إبراهيم الإسلامية الحكومية بمالانغ. المشرف: (١) حسيام فهمي، الماجستير (٢) إيفواتي أليسة، الماجستير.

الكلمات المفتاحية: النماذج الأولية K ، استخراج البيانات ، المخالفات المرورية ، مالانغ ، نعين البيانات.

يهدف هذا البحث إلى تجميع بيانات المخالفات المرورية في مدينة مالانغ في عام ٢٠٢١. يعد التنقيب عن البيانات أحد التقنيات المهمة في تحليل البيانات على نطاق واسع لتحديد الأنماط القيمة التي يمكن أن تدعم عملية صنع القرار. إحدى طرق التنقيب عن البيانات هي التجميع ، والذي يستخدم لتجميع البيانات بناء على خصائص التشابه. يتم تطبيق خوارزمية *K-Prototype* على بيانات انتهاك حركة مرور نظام المجموعة، والتي تتضمن سمات رقمية وفئوية. تضمنت البيانات المستخدمة ٧٣٢٨ حالة بخمس سمات رئيسية: المواد المنتهكة ، والأدلة ، ونوع السيارة ، واستمارة التذكرة ، ومقدار الغرامات. تتضمن عملية البحث تحويل البيانات ، وتطبيع السمات العددية ، وتنفيذ خوارزميات *K-Prototype* يستخدم التقييم مجموع مربعات داخل نظام المجموعة (*WCSS*) لضمان جودة التجميع. تم إنتاج أفضل المجموعات ما يصل إلى أربعة مجموعات. تنتج نتائج هذا التجميع مجموعات تصف أنماط انتهاك محددة. وتضمن مخالفات الدراجات النارية على المجموعة ٠ (٤١١٧ حالة) والمجموعة ١ (١٢٣٠ حالة)، حيث ترتكب معظم الانتهاكات في المادة ٢٨٧ (١) في المجموعة ٠ والمادة ٢٨٥ (١) في المجموعة ١. وتضمن المركبات ذات الأربع عجلات على المجموعة ٢ (١٣٣٧ حالة) والمجموعة ٣ (٥١٠ حالات)، ومعظم الانتهاكات هي المادة ٢٨٧ (١) في المجموعة ٢ والمادة ٢٨٦ في المجموعة ٣. ومن المتوقع أن تكون النتائج مرجعا للجهود المبذولة لزيادة الوعي المروري بالإضافة إلى تخطيط سياسات أكثر فعالية في الحد من عدد المخالفات والحوادث في المستقبل.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data mining adalah salah satu metode yang banyak digunakan dalam mengolah dan menganalisis data yang sangat besar untuk mendapatkan informasi berharga dan digunakan dalam pengambilan keputusan. Sebagaimana yang dinyatakan oleh Susanto & Sudiyatno (2014), *data mining* sangat berguna untuk kasus *database* berukuran besar dan membantu dalam menyusun informasi yang relevan untuk pengambilan keputusan. Seiring perkembangan teknologi, jumlah data yang dihasilkan dan disimpan semakin meningkat secara signifikan. Oleh karena itu, diperlukan teknik dan metode khusus untuk menggali informasi dari data tersebut.

Dalam *data mining*, salah satu pendekatan yang populer adalah *clustering*, yaitu teknik untuk mengelompokkan data berdasarkan pola atau karakteristik tertentu. *Clustering* memungkinkan data yang beragam menjadi lebih terorganisir dan mudah dianalisis (Irawan dkk., 2020). Salah satu algoritma yang terdapat pada *clustering* adalah *K-Prototype*. Algoritma ini beroperasi dengan mengelompokkan data ke dalam sejumlah klaster dengan meminimalkan jarak antara setiap data dan *centroid* dari masing-masing klaster. Dengan menggunakan pendekatan matematis, algoritma ini menentukan data ke dalam klaster yang paling tepat. Salah satu keunggulan *K-Prototype* adalah kemampuannya untuk menangani *dataset* berukuran sangat besar dan dapat mengolah data yang bersifat campuran (numerik atau kategorik).

Merujuk pada penelitian sebelumnya, seperti yang dilakukan Ferbyansyah (2023) tentang data pelanggaran lalu lintas di Kotabumi Kabupaten Lampung Utara. Yang menunjukkan bahwa, analisis dan keefektifan pengelompokan data menggunakan algoritma *clustering K-Prototype* di Kotabumi terbukti efektif. Hal ini dibuktikan dengan dapat mengelompokkan pola karakteristik pelanggaran lalu lintas yang terjadi, dengan dibaginya menjadi tiga klaster (Ferbyansyah, 2023).

Penelitian tersebut menunjukkan bahwa algoritma *K-Prototype* cukup efektif dalam memahami pola pelanggaran lalu lintas. Hal ini menjadi relevan bagi Kota Malang, sebagai salah satu kota besar di Indonesia yang mengalami perkembangan pesat dalam berbagai aspek, termasuk jumlah kendaraan dan lalu lintas yang padat. Menurut Data Badan Pusat Statistika atau BPS mencatat bahwa tahun 2021 ada kendaraan sepeda motor sebanyak 283.581. Tidak hanya sepeda motor, data pada BPS pada tahun 2021 mencatat bahwa kendaraan lainnya juga mengalami peningkatan diantaranya ada mobil penumpang sebanyak 78.127, bis sebanyak 793, dan truk sebanyak 4.777. Kutzbach (2010) menyatakan bahwa peningkatan kendaraan ini akan memberikan dampak signifikan pada sektor transportasi, termasuk peningkatan polusi, kemacetan, serta kecelakaan lalu lintas yang dapat menimbulkan korban jiwa (Ferbyansyah, 2023).

Kesadaran masyarakat akan pentingnya menaati peraturan lalu lintas menjadi aspek penting dalam menjaga ketertiban, keamanan, dan keselamatan di jalan raya. Hal ini tidak hanya mencerminkan keselamatan, tetapi juga tanggung jawab moral, dan etika sebagai warga negara. Aturan yang dibuat bukan semata-mata untuk membatasi, melainkan untuk menjaga ketertiban, keamanan, dan keselamatan bersama. Menghormati hukum lalu lintas adalah bagian dari kewajiban

sosial yang mendukung terciptanya ketertiban di masyarakat. Dalam setiap sistem hukum, penting untuk menghormati otoritas yang ditetapkan. Hal ini sejalan dengan ajaran agama yang menekankan taat kepada Allah dan Rasul-Nya, serta pemegang urusan yang bertanggung jawab atas penerapan hukum. Di jalan raya, pengendara diwajibkan untuk taat pada aturan yang berlaku, sebagaimana ditegaskan dalam Qs. An-Nisa ayat 59 yang berbunyi (Kemenag, 2024):

يَا أَيُّهَا الَّذِينَ ءَامَنُوا أَطِيعُوا اللَّهَ وَأَطِيعُوا الرَّسُولَ وَأُولَى الْأَمْرِ مِنْكُمْ فَإِن تَنَزَعْتُمْ فِي شَيْءٍ فَرُدُّوهُ إِلَى اللَّهِ وَالرَّسُولِ إِن كُنتُمْ تُؤْمِنُونَ بِاللَّهِ وَالْيَوْمِ ءَآخِرِ ؕ ذَٰلِكَ خَيْرٌ وَأَحْسَنُ تَأْوِيلًا

“Hai orang-orang yang beriman, taatilah Allah dan taatilah Rasul (Nya), dan ulil amri di antara kamu. Kemudian jika kamu berlainan pendapat tentang sesuatu, maka kembalikanlah ia kepada Allah (Al Quran) dan Rasul (sunnahnya), jika kamu benar-benar beriman kepada Allah dan hari kemudian. Yang demikian itu lebih utama (bagimu) dan lebih baik akibatnya.”

Menurut tafsir Wajiz dapat disimpulkan bahwa ayat tersebut menekankan pentingnya taat kepada otoritas yang sah, dalam hal ini Allah dan Rasul-Nya, serta pemegang urusan di antara umat (Nugroho, 2019). Dalam konteks lalu lintas, otoritas diwakili oleh hukum dan peraturan lalu lintas yang ditetapkan oleh pemerintah. Sama seperti pentingnya taat kepada Allah dan Rasul-Nya, penting juga untuk patuh terhadap aturan lalu lintas. Sejalan dengan itu, diperlukan metode yang tepat untuk menangani pelanggaran lalu lintas dengan efektif agar membuat pelanggar tersebut tidak mengulangnya lagi.

Sebagaimana yang telah diuraikan sebelumnya, dengan begitu penulis menggunakan algoritma *clustering K-Prototype* untuk pelanggaran lalu lintas yang sering terjadi di Kota Malang, sehingga data dapat dikelompokkan menjadi berbagai klaster berdasarkan kemiripan karakteristiknya. Penelitian ini bertujuan

untuk mengelompokkan dan menganalisis data pelanggaran lalu lintas di Kota Malang dengan menggunakan algoritma *K-Prototype*. Pengelompokan ini diharapkan dapat membantu pihak instansi untuk memahami pola dan karakteristik pelanggaran lalu lintas yang terjadi di Kota Malang.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, rumusan masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana hasil pengelompokan algoritma *K-Prototype* untuk mengelompokkan data pelanggaran lalu lintas di Kota Malang?
2. Bagaimana pola dan karakteristik pelanggaran lalu lintas yang sering terjadi di Kota Malang?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah di uraikan, maka tujuan penelitian ini dapat ditulis yaitu sebagai berikut:

1. Mengetahui hasil algoritma *K-Prototype* untuk mengelompokkan data pelanggaran lalu lintas di Kota Malang.
2. Mengetahui pola dan karakteristik pelanggaran lalu lintas yang sering terjadi di Kota Malang.

1.4 Manfaat Penelitian

Berdasarkan tujuan penelitian yang telah dijelaskan di atas, beberapa manfaat yang dapat diperoleh dari penelitian ini, yaitu sebagai berikut:

1. Manfaat Teoritis

Secara teoritis, penelitian ini diharapkan dapat memberikan ilmu dalam *data mining*, khususnya pada penerapan algoritma *K-Prototype* untuk mengelompokkan data yang bersifat campuran (numerik dan kategori). Penelitian ini juga memperkaya literatur tentang metode *clustering* dengan studi kasus spesifik pada pelanggaran lalu lintas, sehingga dapat membantu memahami pola-pola pelanggaran secara lebih terstruktur dan sistematis.

2. Manfaat praktis

- a. Untuk peneliti, penelitian ini diharapkan dapat menambah wawasan serta pengetahuan mengenai pola dan karakteristik penerapan algoritma *K-Prototype* pada pengelompokan data pelanggaran lalu lintas di Kota Malang, beserta tingkat keefektifannya.
- b. Untuk pembaca, penelitian ini diharapkan dapat memberikan wawasan terkait karakteristik implementasi algoritma *K-Prototype* pada pengelompokan data pelanggaran lalu lintas di Kota Malang.
- c. Untuk akademi, penelitian ini diharapkan dapat menjadi sumber referensi yang berguna sekaligus dapat menginspirasi akademi untuk melakukan penelitian lebih lanjut.
- d. Untuk pihak instansi, penelitian ini diharapkan dapat dipergunakan dalam merancang strategi sosialisasi kepada masyarakat untuk menanggulangi tingkat pelanggaran lalu lintas yang sering dilanggar oleh masyarakat.

1.5 Batasan Masalah

Terdapat beberapa batasan masalah pada yang perlu diperhatikan dalam penelitian ini, yaitu sebagai berikut:

1. Data yang digunakan adalah data pelanggaran lalu lintas hanya data yang tercatat oleh pihak Kejaksaan Negeri Kota Malang selama satu tahun yaitu pada tahun 2021.
2. Data yang akan diolah ialah data dengan pelanggar yang terkena pasal tunggal.

1.6 Definisi Istilah

Terdapat beberapa istilah yang digunakan pada penelitian ini sebagai berikut:

1. Algoritma *K-Prototype* : Algoritma *clustering* yang mengelompokkan data campuran (numerik dan kategorik) dengan menggabungkan jarak *euclidean* untuk fitur numerik dan *dissimilarity* untuk fitur kategori.
2. *Data mining* : Proses mengeksplorasi dan menemukan pola atau informasi berharga dari sejumlah besar data.
3. *Clustering* : Adalah teknik dalam analisis data yang digunakan untuk mengelompokkan data ke dalam beberapa grup atau *klaster* berdasarkan kesamaan karakteristik atau fitur tertentu.

4. Klaster : Kelompok hasil pengelompokan data yang memiliki kesamaan karakteristik dalam analisis data *clustering*.
5. *Centroid* : Titik tengah atau pusat dari suatu klaster yang digunakan dalam algoritma *K-Prototype* untuk menentukan kelompok data.
6. WCSS (*Within-cluster Sum of Squares*) : Jumlah kuadrat dari perbedaan antara titik-titik data dan *centroid* mereka. Ini adalah ukuran untuk menilai kualitas klaster.
7. Pasal : Setiap pasal mengandung aturan, norma, atau ketentuan yang harus dipatuhi oleh masyarakat.
9. *Iterasi* : Proses perulangan dalam algoritma.

BAB II

KAJIAN TEORI

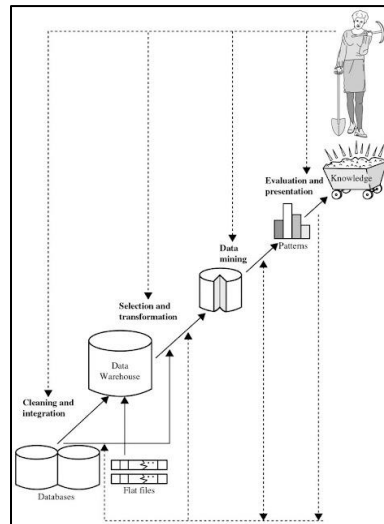
2.1 Teori Pendukung

2.1.1 *Data mining*

Perkembangan zaman makin hari semakin canggih begitu pun dengan data. Data berkembang dengan cepat dan jumlahnya yang sangat besar, sehingga hal ini disebut sebagai era data. Pertumbuhan ini memunculkan kebutuhan akan alat dan metode yang dapat secara otomatis mengungkap atau mengekstraksi informasi dari sejumlah besar data, serta mengubahnya menjadi pengetahuan yang bermanfaat. Hal inilah yang mendorong munculnya *data mining*.

Data mining adalah proses ekstraksi pengetahuan yang tersembunyi, pola-pola menarik, atau informasi yang bermanfaat dari sekumpulan data yang besar dan kompleks. Menurut Pramudiono, *data mining* merupakan proses analisis otomatis pada kumpulan data yang besar atau kompleks untuk mengidentifikasi pola atau kecenderungan penting yang mungkin sebelumnya tidak tampak (Ria, 2018).

Data mining sering kali dianggap dengan istilah KDD (*Knowledge Discovery in Database*) padahal *data mining* merupakan salah satu bagian dari KDD. KDD merupakan proses yang lebih luas dan mencakup beberapa tahap penting untuk mengekstrak pengetahuan dari data. Menurut Jiawei, KDD dibagi menjadi beberapa tahapan dan digambarkan pada Gambar 2.1.



Gambar 2.1 KDD (Knowledge Discovery in Databases)

1. *Data selection*

Data selection/Pemilihan Data melibatkan pemilihan data yang relevan dari sumber data yang lebih besar. Data yang dipilih harus mendukung tujuan analisis yang telah ditetapkan.

2. *Pre-processing Data*

Pemrosesan data, Pada tahap ini data dibersihkan dan dipersiapkan untuk analisis lebih lanjut. Ini termasuk penanganan data yang hilang (*missing value*), menghilangkan *noise*, dan koreksi kesalahan dalam data.

3. *Data Transformation*

Transformasi Data yaitu data yang telah diproses kemudian diubah menjadi format yang lebih sesuai untuk *data mining*. Transformasi ini dapat mencakup normalisasi, penggabungan, atau reduksi dimensi data.

4. *Data mining*

Tahap inti dari KDD, di mana teknik statistik, pembelajaran mesin, dan algoritma lainnya digunakan untuk mengekstrak pola, tren, atau model dari data yang telah diproses. *Data mining* bertujuan untuk mengidentifikasi

hubungan yang tidak terduga dan menyajikan data dalam bentuk ringkasan yang baru.

5. *Pattern Evaluation*

Evaluasi Pola atau model yang dihasilkan dari *data mining* dievaluasi untuk memastikan validitas dan kegunaannya. Evaluasi ini melibatkan penilaian apakah pola tersebut benar-benar signifikan dan relevan untuk kebutuhan pengguna.

6. *Knowledge Presentation*

Presentasi Pengetahuan biasa disajikan dalam format yang dapat dipahami oleh pengguna akhir, seperti laporan, visualisasi, atau model interpretatif. Tahap ini memastikan bahwa hasil KDD dapat diimplementasikan dan dimanfaatkan dengan baik oleh pemilik data.

2.1.2 *Clustering*

Clustering adalah salah satu teknik yang ada dalam *data mining* dan digunakan untuk mengelompokkan sekumpulan objek atau menjadi beberapa kelompok atau klaster yang memiliki kesamaan tertentu. Sedangkan Menurut Tan (2006), *Clustering* adalah langkah untuk mengelompokkan data ke dalam kelompok atau klaster yang berbeda-beda, sehingga data dalam satu klaster memiliki kesamaan karakteristik. Tujuan *clustering* adalah untuk menemukan pola atau struktur alami yang terdapat dalam data, tanpa adanya label atau informasi sebelumnya mengenai klasifikasi objek-objek tersebut (Wahyuni & Arianto, 2019).

Clustering memiliki banyak manfaat dalam memprediksi dan menganalisis masalah bisnis tertentu. Contohnya, dalam segmentasi pasar, pemasaran, pemetaan wilayah, manajemen *marketing*. Dalam hal ini *clustering* dapat digunakan secara efektif. Selain itu, dalam bidang lain seperti *computer vision* dan pemrosesan citra, *clustering* juga berguna untuk mengidentifikasi objek.

Hasil *clustering* yang bagus pasti akan menghasilkan kesamaan internal yang tinggi antara data di dalamnya, sementara data yang diluar klaster itu akan memiliki perbedaan yang signifikan. Namun, kualitas hasil *clustering* sangat dipengaruhi oleh metode yang digunakan. Ada banyak metode yang dapat digunakan untuk mengukur jarak, salah satunya adalah *euclidean distance* (Pribadi dkk., 2022).

Euclidean distance merupakan metrik jarak yang digunakan dalam ruang *euclidean* untuk mengukur jarak antara dua titik dalam dimensi yang lebih dari satu ruang *euclidean* dapat meliputi berbagai dimensi, termasuk bidang *euclidean* dua atau lebih. Berikut adalah rumus *euclidian distance*:

$$d(X_j, Z_i) = \sqrt{\sum_{l=1}^{m_N} (x_{jl}^N - z_{il}^N)^2} \quad (2.1)$$

Keterangan:

$d(X_j, Z_i)$ = Jarak *euclidean* antara data X_j dan *centroid* Z_i

m_N = Jumlah atribut numerik

x_{jl}^N = Nilai data ke j pada atribut numerik ke l

z_{il}^N = Nilai *Centroid* ke i pada atribut numerik ke l

Metode *clustering* atau klasterisasi dibagi menjadi dua yaitu *hierarchical clustering* dan *partitional clustering* (Tan, 2011). Namun ada beberapa metode tambahan diantaranya *Grid-based clustering*, *model-based clustering*, *Density-based clustering*, dan *counstranint-based clustering*.

Metode *partitional clustering* adalah mengelompokkan data ke dalam beberapa klaster untuk membentuk struktur hierarki tertentu dan setiap klaster memiliki titik pusat (*centroid*). Metode ini mencakup teknik-teknik seperti *K-Means*, *K-Medoids*, *K-Prototype*, *Fuzzy*, dan *mixturing modelling*.

Hierarchical clustering adalah metode pengelompokan data dengan membentuk struktur hierarki, di mana data dikelompokkan melalui diagram atau bagan yang menggambarkan urutan tingkatan dari klaster, mulai dari klaster individu hingga klaster yang lebih besar, di mana prosesnya melibatkan penggabungan atau pemisahan kelompok data dalam setiap iterasinya. Metode ini menghasilkan struktur hierarki dari klaster yang dapat divisualisasikan dalam bentuk dendrogram, yang menunjukkan bagaimana klaster dibentuk secara bertahap, terdapat 4 contoh metode *hierarchy clustering*, yaitu: *Single Linkage*, *Complete Linkage*, *Average Group Linkage*, *Average Linkage* (Annisa Ekasetya & Jananto, 2020).

2.1.3 Algoritma *K- Prototype*

Salah satu metode klasterisasi yang menggabungkan kemampuan algoritma *K-Means* dan *K-Modes* ialah *K-Prototype*. Algoritma ini cocok digunakan pada data yang bersifat campuran, yaitu data yang memiliki fitur numerik dan kategorikal (Kim, 2017). *K-Prototype* bekerja dengan

memperhitungkan jarak *euclidean* untuk data numerik dan frekuensi kesamaan untuk data kategorikal, sehingga algoritma ini dapat mengelompokkan data berdasarkan karakteristik yang bervariasi.

Algoritma ini relatif efisien dalam hal komputasi karena tidak menggunakan perhitungan yang kompleks, memungkinkan untuk diterapkan pada dataset yang berukuran besar. Algoritma *K-Prototype* menggunakan proses iteratif untuk meminimalkan jarak antara anggota kluster dan *centroid* yang dipilih. Algoritma *K-Prototype* berfungsi secara iteratif dan berbasis pada perhitungan numerik dan kategorikal. Algoritma ini tergolong dalam kategori pembelajaran tanpa pengawasan (*unsupervised learning*), yang berarti bahwa ia tidak memerlukan label yang sudah ditentukan sebelumnya untuk mengelompokkan data. Dengan pendekatan ini, algoritma *K-Prototype* mampu menemukan pola dan struktur dalam data secara mandiri.

Menurut Nooraeni (2019) untuk menggabungkan ukuran ketidaksamaan yang digunakan dalam Algoritma *K-Prototype* diperoleh dengan menggabungkan konsep jarak *euclidean* dengan ukuran *dissimilarity* untuk menangani data campuran. Pada data numerik, ukuran ketidaksamaan dihitung menggunakan *euclidean distance*, yang mengukur jarak antar titik dalam ruang numerik dan memungkinkan perhitungan jarak kuadrat antar nilai atribut sebagaimana pada persamaan (2.1). Sementara itu, pada data kategorikal, *dissimilarity* dihitung berdasarkan perbedaan antar kategori. Dalam hal ini, ketidaksamaan antar data kategorikal diberikan nilai 1 jika atributnya berbeda dan 0 jika sama, yang dikenal sebagai ukuran jarak *hamming* untuk data kategorikal (Nooraeni & Supriadi,

2019). Sebagaimana ukuran kesamaan data kategorikal dinyatakan sebagai berikut:

$$d(X_j, Z_i) = \gamma \sum_{l=m_N+1}^{m_C} \delta(x_{jl}^C, z_{il}^C) \quad (2.2)$$

$$\text{dissimilarity } \delta(x_{jl}^C, z_{il}^C) = \begin{cases} 0 & x_{jl}^C = z_{il}^C \\ 1 & x_{jl}^C \neq z_{il}^C \end{cases} \quad (2.3)$$

Keterangan:

- $d(X_j, Z_i)$ = Jarak antara data ke X_j dan *centroid* Z_i
- γ = Parameter penyeimbang
- $\sum_{l=m_N+1}^{m_C} \delta(x_{jl}^C, z_{il}^C)$ = Penjumlahan dari $(l = m_N + 1)$ hingga m_C
- m_C = Jumlah atribut kategorik
- x_{jl}^C = Nilai data ke j pada atribut kategorik ke l
- z_{il}^C = Nilai *Centroid* ke i pada atribut kategorik ke l
- $\delta(x_{jl}^C, z_{il}^C)$ = Fungsi *dissimilarity* antara komponen kategorikal
 x_{jl}^C dan z_{il}^C akan bernilai 1 jika atribut tidak sama
 dan akan bernilai 0 jika atribut sama.

Menurut Pradhana (2022) γ yang akan digunakan menggunakan *default* yang sudah ditetapkan dari *library* yaitu menggunakan persamaan berikut (Pradhana, 2022).

$$\gamma = 0,5 \times \text{std}(X_{\text{Numerikal}}) \quad (2.4)$$

Keterangan:

$\text{std}(X_{\text{Numerikal}})$ = Standar deviasi dari atribut numerik dalam data.

Kemudian akan digabungkan antara persamaan 2.1 dan persamaan 2.2 sehingga akan membentuk kesamaan untuk data yang beratribut numerik dan beratribut kategorik. Akan ditunjukkan sebagai berikut:

$$d(X_j, Z_i) = \sqrt{\sum_{l=1}^{m_N} (x_{jl}^N - z_{il}^N)^2 + \gamma \sum_{l=m_N+1}^{m_C} \delta(x_{jl}^C, z_{il}^C)} \quad (2.5)$$

Keterangan:

$$\sum_{l=1}^{m_N} (x_{jl}^N - z_{il}^N)^2 = \text{Data yang beratribut numerik}$$

$$\sum_{l=m_N+1}^{m_C} \delta(x_{jl}^C, z_{il}^C) = \text{Data yang beratribut kategorikal}$$

Adapun lima tahapan utama pada algoritma *K-Prototype* antara lain:

1. Memilih k buah objek secara acak dari sebuah *dataset* sebagai *centroid* dalam data dengan minimal $K = 2$ dan maksimal $\sqrt{n}$ atau $\frac{n}{2}$, dimana n adalah banyak data.
2. Menghitung jarak antara setiap objek dan titik tengah dihitung menggunakan pada persamaan 2.5
3. Mengelompokkan setiap data ke dalam kluster berdasarkan kedekatan data dengan titik tengahnya.
4. Menentukan titik tengah baru dengan melakukan perhitungan modus untuk data kategorikal dan untuk data numerik menggunakan rata-rata dari data pada setiap kluster dengan menggunakan persamaan sebagai berikut:

$$v_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ji} \quad (2.6)$$

Keterangan:

v_i = Centroid klaster ke- i

N_i = Total anggota dalam klaster ke- i

x_{ji} = Nilai data ke j dalam klaster ke- i

5. Menghitung kembali nilai titik pusat dan penempatan data dalam klaster dihitung lagi sampai nilai titik pusat dari semua klaster yang dibentuk tidak berubah. Jika tidak ada perubahan maka proses *clustering* akan dianggap selesai.

2.1.4 Normalisasi Data

Proses pengolahan data yang bertujuan untuk mengubah skala data ke dalam rentang tertentu tanpa mengubah pola distribusi data aslinya disebut dengan normalisasi data. Proses ini penting dalam analisis data, terutama ketika menggunakan algoritma *machine learning* dan *data mining*, karena data yang memiliki rentang atau skala yang berbeda dapat memengaruhi hasil analisis dan performa model (Han dkk., 2011).

Algoritma *K-Prototype* bekerja dengan menghitung jarak antara data dan titik pusat menggunakan metrik. Jika data memiliki skala yang berbeda, maka fitur dengan rentang nilai yang lebih besar akan lebih mendominasi perhitungan jarak, sehingga menyebabkan bias dalam pembentukan klaster (Han dkk., 2011). Oleh karena itu, normalisasi data diperlukan untuk memastikan bahwa semua fitur memiliki pengaruh yang seimbang dalam analisis *clustering*. Dengan menormalisasi yang berarti mengubah data ke rentang $[0, 1]$ dan akan dirumuskan sebagai berikut:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2.7)$$

Keterangan :

x' = Nilai yang akan dinormalisasi

x = Nilai yang dipilih dari data

x_{min} = Nilai Minimum dari data

x_{max} = Nilai Maksimum dari data

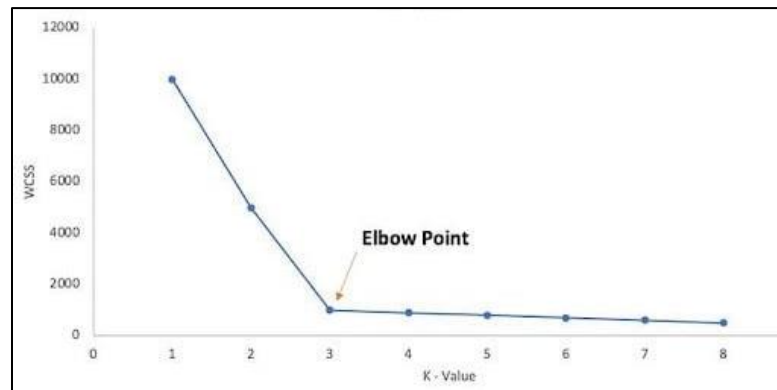
2.1.5 Metode *Elbow*

Metode ini merupakan salah satu teknik yang digunakan dalam analisis kluster khususnya dalam menentukan jumlah kluster (K) terbaik biasa digunakan dalam sebuah algoritma *clustering*. Teknik ini bertujuan untuk mengidentifikasi kluster yang tepat yang dapat memberikan interpretasi yang paling baik terhadap data dengan membandingkan total kluster yang membentuk siku (Ayu dkk., 2019)

Pada metode ini, dilakukan perhitungan terhadap nilai *Within-Cluster Sum of Squares* (WCSS), untuk menentukan jumlah kluster optimal. WCSS adalah jumlah kuadrat jarak antara setiap titik data dengan titik tengah klasternya. Semakin besar total atau jumlah kluster K , maka akan semakin kecil nilai dari WCSS yang dihasilkan. Metode *elbow* mencari titik di mana penurunan WCSS yang signifikan mulai melambat dan membentuk sudut siku, yang menandakan jumlah kluster yang optimal. Adapun perhitungan WCSS sebagai berikut:

$$d(X_j, Z_i) = \sum_{l=1}^{m_N} (x_{jl}^N - z_{il}^N)^2 + \gamma \sum_{l=m_N+1}^{m_C} \delta(x_{jl}^C, z_{il}^C) \quad (2.8)$$

Setelah dilakukan perhitungan nilai WCSS, jumlah kluster (K) dapat mengalami perubahan yaitu penurunan. Dimana nilai K akan menurun secara bertahap hingga mencapai titik di mana nilai tersebut menjadi stabil. Berikut adalah contoh grafik yang dihasilkan dari metode ini:



Gambar 2.2 Contoh Grafik Metode *Elbow*

Grafik pada Gambar 2.2 menggambarkan jumlah kluster (K) pada sumbu x dan *Within-Cluster Sum of Squares* (WCSS) sebagai metrik kesalahan pada sumbu y . Terlihat bahwa pada Gambar 2.2, terjadi penurunan yang signifikan dari nilai WCSS ketika nilai K berubah dari dua ke tiga, namun setelah itu, penurunan WCSS menjadi lebih lambat setelah mencapai nilai $K = 3$, sehingga grafik membentuk sudut yang hampir tajam atau siku-siku. Oleh karena itu, nilai optimal dari K adalah tiga.

2.1.6 Pelanggaran Lalu Lintas

Kejadian ketika orang tidak mengikuti aturan yang telah ditetapkan pada aturan lalu lintas disebut pelanggar lalu lintas. Hal Ini melibatkan pengemudi dan pengguna jalan yang dapat menyebabkan dampak buruk, seperti kecelakaan dan kerugian materil atau bahkan korban jiwa. Pelanggaran ini dapat mengancam

keselamatan pengguna jalan lainnya dan membentuk budaya melanggar aturan membuat masyarakat kurang memahami pentingnya ketertiban di jalan raya (Dewi dkk., 2019). Keberadaan pelanggaran ini mengganggu kenyamanan dan keamanan dalam berkendara. Di kota-kota besar termasuk Kota Malang, masalah kedisiplinan dalam berlalu lintas masih sering terjadi berdasarkan catatan kepolisian Kota Malang. Adapun jenis pelanggaran yang sering terjadi dan tertulis berdasarkan Undang-undang lalu lintas dan angkutan jalan tahun 2009 yang intinya sebagai berikut:

1. Pasal 278: Setiap pengendara roda empat atau lebih yang tidak memiliki perlengkapan seperti ban, segitiga pengaman, dongkrak, pembuka roda, dan peralatan pertolongan pertama untuk kecelakaan.
2. Pasal 280: Setiap Kendaraan bermotor yang tidak dilengkapi Tanda Nomor Kendaraan Bermotor (TNKB).
3. Pasal 281: Setiap pengendara yang tidak memiliki SIM.
4. Pasal 282: Setiap pengguna jalan tidak mengikuti perintah yang diberikan petugas Polri.
5. Pasal 285(1): Pengendara sepeda motor tidak memenuhi persyaratan teknis dan kondisi kendaraan yang baik di jalan.
6. Pasal 285(2): Persyaratan teknis kendaraan yang tidak memenuhi ketentuan, termasuk: kaca spion, klakson, lampu utama, lampu mundur, lampu batas dimensi kendaraan, lampu gandengan, lampu rem, lampu penunjuk arah, alat pemantul cahaya, alat pengukur kecepatan, kedalaman alur ban, kaca depan, sepatbor, bumper, penggandengan, penempelan, atau penghapus kaca.

7. Pasal 286: Kekuatan mesin kendaraan harus sesuai dengan berat kendaraan.
8. Pasal 287(1): Pengemudi yang melanggar rambu-rambu lalu lintas
9. Pasal 287(2): Melanggar perintah atau larangan yang disampaikan melalui Alat Pemberi Isyarat Lalu Lintas.
10. Pasal 288(1): Pengemudi kendaraan bermotor yang tidak memiliki Surat Tanda Nomor Kendaraan atau biasa disebut STNK.
11. Pasal 288(2): Pengendara bermotor yang mempunyai SIM tetapi tidak dapat menunjukkan saat diminta.
12. Pasal 288(3): Kendaraan yang tidak dilengkapi dengan surat keterangan uji berkala dan tanda lulus uji berkala.
13. Pasal 289: Penumpang yang berada di samping pengemudi yang tidak memakai sabuk pengaman.
14. Pasal 291(1): Tidak menggunakan helm standar nasional.
15. Pasal 291(2): Pengemudi yang membiarkan penumpangnya tidak mengenakan helm.
16. Pasal 293(1): Tidak menghidupkan lampu utama pada malam hari.
17. Pasal 293(2): Tidak menghidupkan lampu utama pada siang hari
18. Pasal 297: Terlibat dalam balapan liar di jalan umum.
19. Pasal 300: Tidak menggunakan jalur yang telah ditentukan.
20. Pasal 303: Mengangkut penumpang dengan kendaraan yang seharusnya hanya digunakan untuk barang.
21. Pasal 307: Tidak mematuhi aturan tentang tata cara pemuatan dan daya angkut kendaraan.

Adapun jenis form bagi pelanggar yang merupakan bentuk surat tilang yang mengindikasikan pelanggaran tertentu, diantaranya :

1. Form merah : untuk pelanggaran yang tidak bisa diselesaikan di tempat.
2. Form biru : untuk pelanggaran yang bisa langsung diselesaikan dengan membayar denda.

Terdapat juga tipe kendaraan yang digunakan oleh pelanggar saat terjadi pelanggaran lalu lintas. Identifikasi jenis kendaraan penting dalam pencatatan dan analisis pelanggaran, karena masing-masing tipe kendaraan memiliki aturan dan ketentuan yang berbeda. Adapun jenis kendaraannya sebagai berikut:

1. Sepeda Motor : Kendaraan roda dua, biasanya mencakup berbagai jenis atau merek sepeda motor.
2. Mobil Pick up: Kendaraan roda empat atau lebih yang dirancang untuk mengangkut barang, biasanya terbuka di bagian belakang.
3. Mobil Penumpang Pribadi: Kendaraan roda empat yang digunakan untuk transportasi pribadi,
4. Mobil Penumpang Umum: Kendaraan yang digunakan untuk angkutan umum, seperti angkot atau taksi.
5. Bus: Kendaraan besar yang berfungsi untuk mengangkut penumpang dalam jumlah banyak, biasanya untuk transportasi umum atau antar kota.
6. Truk: Kendaraan besar yang digunakan untuk mengangkut barang atau kargo.

Adapun dokumen atau barang yang disita oleh pihak berwenang sebagai bukti pelanggaran. Barang bukti ini berfungsi sebagai pendukung dalam proses

hukum dan membantu memastikan pelanggar mempertanggungjawabkan tindakannya. Diantaranya sebagai berikut:

1. Buku KIR: Buku yang membuktikan kendaraan telah lulus uji kelayakan (KIR) dan memenuhi standar keselamatan.
2. Kendaraan: Kendaraan yang dikendarai oleh pelanggar dapat disita sebagai barang bukti.
3. SIM A: Surat Izin Mengemudi (SIM) untuk pengendara mobil penumpang pribadi.
4. SIM BI: SIM untuk pengemudi kendaraan dengan total berat maksimal 3.500 kg berupa mobil bus dan barang perseorangan.
5. SIM BI Umum: SIM untuk pengemudi kendaraan dengan total berat maksimal 3.500 kg berupa mobil bus dan barang umum.
6. SIM BII Umum: SIM untuk pengemudi kendaraan berupa alat berat, kendaraan penarik, dan kendaraan dengan menarik kereta tempelan atau gandengan umum dengan berat maksimal 1.000 kg.
7. SIM C: SIM untuk pengendara sepeda motor.
8. STNK: Surat Tanda Nomor Kendaraan, dokumen resmi yang menunjukkan kepemilikan dan legalitas kendaraan.

Banyak kasus pelanggaran yang terjadi untuk itu perlu dilakukan upaya pencegahan atau pengurangan pelanggaran lalu lintas merupakan hal yang penting bagi berkehidupan bangsa Indonesia agar menciptakan masyarakat yang patuh terhadap hukum. Beberapa upaya yang dapat dilakukan antara lain: Edukasi dan sosialisasi, penegakan hukum yang tegas, peningkatan infrastruktur lalu lintas. penerapan teknologi (seperti sistem tilang elektronik), dan program pelatihan

berkendara. Dengan menerapkan upaya-upaya tersebut, diharapkan tingkat pelanggaran lalu lintas bisa berkurang secara signifikan, sehingga tercipta lingkungan yang lebih aman, nyaman, dan tertib bagi seluruh pengguna jalan di Indonesia.

2.2 Kajian Integrasi Topik dalam Al-Qur'an/Hadits

Islam merupakan agama yang sempurna dalam mengatur berbagai aspek kehidupan manusia. Dalam hal ini mencakup berbagai dimensi, termasuk aspek sosial, ekonomi, dan hukum yang berlaku dalam masyarakat. Dengan begitu, ajaran Islam memberikan pedoman yang jelas dan terstruktur bagi umatnya untuk menjalani kehidupan yang harmonis dan adil di tengah masyarakat.

Dengan demikian, ketaatan kepada hukum dan peraturan merupakan salah satu prinsip yang telah ditekankan dalam ajaran Islam. Al-Qur'an dan Hadist menekankan pentingnya ketaatan kepada Allah SWT, Rasul-Nya, dan otoritas yang telah ditetapkan. Sebagaimana disebutkan dalam Qs. An-Nisa ayat 59 (Kemenag, 2024):

يَا أَيُّهَا الَّذِينَ ءَامَنُوا أَطِيعُوا اللَّهَ وَأَطِيعُوا الرَّسُولَ وَأُولِيَ الْأَمْرِ مِنْكُمْ

“Hai orang-orang yang beriman, taatilah Allah dan taatilah Rasul (Nya), dan ulil amri di antara kamu”

Ayat ini memerintahkan seseorang untuk wajib taat kepada ulil amri selama, ketaatan tersebut tidak bertentangan dengan perintah Allah dan Rasul-Nya. Menurut penjelasan Ibnu Katsir, istilah *ulil amri* bersifat luas dan mencakup baik itu para pemimpin pemerintahan maupun ulama. Namun dalam konteks lalu lintas,

Pemerintahan inilah yang merupakan *ulil amri* itu sendiri dan yang menetapkan peraturan lalu lintas demi menjaga keselamatan dan keteraturan di jalan raya. Pelanggaran terhadap aturan lalu lintas dapat dilihat sebagai tindakan yang tidak sejalan dengan prinsip ketaatan yang diajarkan oleh Islam.

Selain itu, pentingnya mematuhi hukum juga tercermin dalam berbagai hadis Rasulullah SAW yang mengajarkan tentang perlindungan terhadap diri sendiri dan orang lain. Sebagaimana yang dinyatakan dalam sebuah hadist yang diriwayatkan oleh Imam Ahmad, Rasulullah SAW bersabda:

“Barang siapa yang mencelakakan seorang muslim, maka Allah akan mencelakakannya, dan barang siapa yang menyulitkan seorang muslim, maka Allah akan menyulitkannya.” (Hadist Riwayat Tirmidzi No. 1940)

Hadist ini mengajarkan bahwa setiap tindakan yang dapat menyakiti atau membahayakan orang lain sangat dilarang dalam islam, termasuk seorang pelanggar lalu lintas yang dapat menyebabkan kecelakaan atau kerugian. Ketika seseorang melanggar aturan lalu lintas, mereka tidak hanya membahayakan diri sendiri, tetapi juga orang lain yang berada di sekitarnya.

Selain itu, prinsip Islam mengenai pentingnya keselamatan dan perlindungan terhadap sesama manusia juga bisa dilihat dalam hadist lainnya, di mana Rasulullah SAW bersabda:

“Tidak boleh membahayakan diri sendiri dan tidak boleh membahayakan orang lain.” (Hadist Riwayat Ibnu Majah dan Ad-Daruqutni No. 3079)

Hadist ini menegaskan bahwa setiap muslim harus bertindak dengan cara yang tidak merugikan diri sendiri maupun orang lain. Dengan demikian, ketika

seorang pengendara melanggar aturan lalu lintas, seperti melaju dengan kecepatan tinggi, menerobos lampu merah, atau tidak menggunakan helm, maka tindakan tersebut bertentangan dengan ajaran Islam karena dapat menimbulkan resiko bagi diri sendiri maupun pengguna jalan lainnya.

Kajian integrasi pelanggaran lalu lintas dengan ajaran agama islam menunjukkan bahwa setiap pelanggaran merupakan bentuk ketidak patuhan terhadap prinsip ketaatan, keselamatan, dan tanggung jawab. Dengan demikian, penegakan hukum lalu lintas bukan hanya masalah sosial, tetapi juga memiliki dimensi religius. Oleh karena itu, metode seperti *K- Prototype* dapat digunakan sebagai pendekatan analisis untuk mengidentifikasi pola pelanggaran lalu lintas yang dapat memberikan solusi efektif dalam menegakkan aturan sesuai dengan ajaran Islam.

Integrasi ajaran agama dalam analisis pelanggaran lalu lintas ini diharapkan dapat meningkatkan kesadaran masyarakat tentang pentingnya mematuhi aturan, bukan hanya demi ketertiban dan, tetapi juga sebagai bentuk keselamatan diri dan implementasi nilai-nilai religius dalam kehidupan sehari-hari.

2.3 Kajian Topik Dengan Teori Pendukung

Pada penelitian ini terdapat beberapa tahapan yang akan dilakukan untuk mengelompokkan data pelanggaran lalu lintas di Kota Malang dengan menggunakan algoritma *K-Prototype*. Tahap pertama adalah pengumpulan data pelanggaran lalu lintas dan seleksi atribut data mana saja yang akan dipilih.

Tahap berikutnya adalah transformasi data, yaitu proses inisialisasi pada data yang beratribut kategorik guna mempermudah pemrograman/perhitungan dan

akan dilakukan normalisasi data yang beratribut numerik ke dalam interval $[0,1]$. Normalisasi ini penting untuk menghindari bias akibat perbedaan skala, oleh karena itu akan dilakukan perhitungan seperti pada persamaan (2.7), sehingga algoritma *K-Prototype* dapat bekerja dengan lebih akurat. Setelah data ditransformasikan, tahap selanjutnya adalah *preprocessing* data, dimana data akan dicek apakah ada data terdapat *missing value* atau tidak. Jika ada maka data akan diolah seperti diabaikan atau menggantinya dengan rata-rata parameter.

Tahap selanjutnya adalah mengimplementasikan *K-Prototype* sebagaimana pada persamaan (2.5). Dimana data akan bersifat campuran yaitu data kategorik dan data numerik. Setelah itu data akan dimasukkan ke dalam klaster yang jaraknya paling dekat. Setelah semua data dimasukkan ke dalam klaster terdekat, pusat klaster dihitung ulang berdasarkan anggota klaster saat ini sebagaimana pada persamaan (2.6) jika klaster tidak ada perubahan maka proses iterasi akan berakhir.

Tahap terakhir adalah evaluasi jumlah klaster (K) yang optimal dengan menggunakan WCSS dan perhitungan yang akan dilakukan menggunakan persamaan (2.8). Dengan metode ini, titik optimal pada grafik akan digunakan untuk menentukan total klaster terbaik untuk mengelompokkan data pelanggaran.

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Pada Penelitian ini menggunakan metode kuantitatif, di mana data disajikan dalam bentuk numerik dan kategorik untuk dianalisis menggunakan metode algoritma *K-Prototype*. Penelitian ini dilakukan untuk memahami pola dan karakteristik pelanggaran lalu lintas yang terjadi di Kota Malang.

3.2 Data dan Sumber Data

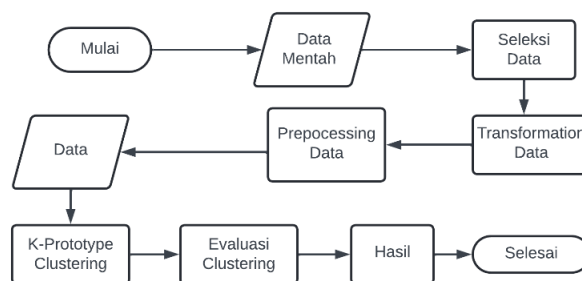
Pada Penelitian ini menggunakan data pelanggaran lalu lintas di Kota Malang sepanjang tahun 2021 yang berasal dari Kejaksaan Negeri Kota Malang. Data ini mencatat sebanyak 7.328 kasus pelanggaran dengan tujuh atribut, yaitu nama pelanggar, plat nomor kendaraan pelanggar, form, pasal, barang bukti, jenis kendaraan, dan denda, sebagaimana pada lampiran 1. Dikarenakan nama dan plat nomor kendaraan pelanggar merupakan identitas pribadi yang bersifat unik, maka penelitian ini hanya berfokus pada lima variabel, yaitu:

1. Form (X_1) – Digunakan untuk mencatat pelanggaran yang dilakukan.
2. Pasal (X_2) – Menggambarkan pasal yang dilanggar oleh pelanggar.
3. Barang Bukti (X_3) – Merupakan barang yang disita oleh petugas saat pelanggaran terjadi.
4. Jenis Kendaraan (X_4) – Menunjukkan jenis kendaraan yang terlibat saat pelanggaran terjadi.

5. Denda (X_5) – Menunjukkan besaran denda yang harus dibayarkan oleh pelanggar atas pelanggaran yang telah dilakukan pelanggar.

3.3 Teknik Analisis Data

Dalam penelitian ini, teknik analisis data yang digunakan melibatkan beberapa langkah utama untuk memastikan bahwa hasil yang diperoleh dari pengelompokan data pelanggaran lalu lintas di Kota Malang akurat dan informatif. Teknik-teknik tersebut bisa dilihat pada Gambar 3.1.



Gambar 3.1 Teknik Analisis Data

1. Pengumpulan data mentah terkait pelanggaran lalu lintas yang diperoleh dari Kejaksaan Negeri Kota Malang.
2. Seleksi Data: setelah melakukan pengumpulan data, lalu akan dipilih atribut yang digunakan dalam penelitian dan menghapus yang tidak digunakan dalam penelitian.
3. Transformasi Data : Proses ini meliputi perubahan format atau tipe data untuk data kategorikal menggunakan inisialisasi guna mempermudah perhitungan dan menormalisasikan data pada atribut numerik menggunakan *Min-Max Normalization*: Mengubah nilai ke dalam rentang [0, 1] menggunakan persamaan 2.7 dikarenakan rentang data yang terlalu besar

agar tidak terdapat bias dalam data, kemudian data siap untuk diolah lebih lanjut.

4. *Preprocessing data* : pada tahap ini, data akan dilakukan pengecekan terhadap *missing value*, kemudian akan dilakukan penghapusan *missing value* jika ada.
5. Implementasi algoritma *K-Prototype* untuk melakukan pengelompokan data pelanggaran lalu lintas berdasarkan kesamaan antar karakteristik data. Proses ini melibatkan:
 - a. Memilih k buah objek secara acak dari sebuah dataset sebagai *centroid* dalam data.
 - b. Menghitung jarak antara setiap objek dan titik tengah dihitung menggunakan pada persamaan 2.5
 - c. Mengelompokkan setiap data ke dalam klaster berdasarkan kedekatan data dengan titik tengahnya
 - d. Menentukan titik tengah baru dengan melakukan perhitungan modus untuk data kategorik dan untuk data numerik dilakukan perhitungan rata-rata dari data pada setiap klaster dengan menggunakan persamaan 2.6.
 - e. Langkah-langkah di atas akan terus diulang sampai konvergensi tercapai. Yaitu ketika tidak ada perubahan signifikan dalam keanggotaan klaster atau ketika perbaikan jarak sudah stabil. Konvergensi ini menandakan bahwa klaster sudah optimal.
6. Evaluasi Klaster: Evaluasi hasil klasterisasi dilakukan untuk menilai kualitas pengelompokan. Akan digunakan evaluasi menggunakan WCSS digunakan untuk menilai K optimal dengan menggunakan persamaan 2.8.

7. Penyajian hasil: Hasil pengelompokan akan dianalisis untuk mengidentifikasi karakteristik utama dari setiap klaster. Hasil ini kemudian disajikan dalam bentuk tabel dan diagram batang yang mendetail.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Seleksi Data (*Data selection*)

Tahap ini merupakan tahapan awal untuk menyeleksi data yang akan digunakan dalam penelitian. Proses seleksi data dilakukan dengan tujuan untuk memastikan bahwa hanya data yang relevan dan mendukung analisis yang dipilih. Dari keseluruhan data yang tersedia, hanya lima atribut yang dipilih untuk memenuhi kebutuhan penelitian ini. Pemilihan lima atribut ini didasarkan pada pertimbangan relevansi terhadap tujuan penelitian dan efektivitas dalam penerapan algoritma *K-Prototype*. Atribut-atribut yang dipilih meliputi Pasal, Barang Bukti, Jenis Kendaraan, Denda, dan Form. Dimana atribut-atribut ini dianggap dapat memberikan informasi penting dalam analisis klusterisasi data pelanggaran lalu lintas. Sebagaimana bisa dilihat pada Tabel 4.1 dan terdapat 7.328 data yang akan dilakukan *clustering*.

Tabel 4.1 Seleksi Data

NO	FORM	PASAL	BARANG BUKTI	JENIS KENDARAAN	DENDA
1	Merah	285(1)	STNK	Sepeda Motor	74000
2	Merah	285(1)	STNK	Sepeda Motor	74000
3	Merah	285(1)	STNK	Sepeda Motor	74000
4	Merah	285(1)	STNK	Sepeda Motor	74000
⋮	⋮	⋮	⋮	⋮	⋮
7328	Merah	287(1)	SIM A	Mobil Penumpang Pribadi	99000

4.2 Transformasi Data

Tahap ini merupakan proses penting dalam persiapan data sebelum dilakukan analisis klusterisasi. Pada tahap ini, data yang telah diseleksi kemudian diubah ke

dalam format atau bentuk yang sesuai agar dapat digunakan oleh algoritma *K-Prototype*. Proses transformasi bertujuan untuk menyelaraskan tipe data, mengatasi perbedaan skala, dan memastikan konsistensi dalam representasi data, terutama karena *K-Prototype* bekerja dengan atribut numerik dan kategorikal.

Pada atribut kategorikal akan dilakukan inisialisasi data sebelum diterapkan pada algoritma *K-Prototype*. Pada tahap ini, atribut-atribut kategorikal yang akan dilakukan inisialisasi yaitu pada form, pasal, dan jenis kendaraan. Hal ini dilakukan agar data menjadi dalam bentuk yang lebih sederhana dan konsisten. Adapun tahap inisialisasi sebagai berikut:

Tabel 4.2 Inisialisasi Atribut Pasal

Pasal	Keterangan	Inisialisasi
278	Setiap pengendara mobil yang tidak dilengkapi dengan perlengkapan berupa ban cadangan, segitiga pengaman, dongkrak, pembuka roda, dan peralatan pertolongan pertama pada kecelakaan	A
280	kendaraan bermotor yang tidak dipasang Tanda Nomor Kendaraan Bermotor (TNKB)	B
281	Tidak memiliki SIM	C
282	Setiap Pengguna Jalan Tidak mematuhi perintah yang diberikan petugas Polri	D
285(1)	Pengemudi Sepeda Motor tidak memenuhi persyaratan teknis dan baik di jalan	E
285(2)	Persyaratan Teknis Ranmor tidak memenuhi persyaratan teknis meliputi : Kaca Spion, Klakson, Lampu utama, Lampu mundur, lampu batas tanda batas Dimensi badan kendaraan, lampu gandengan, lampu Rem, lampu penunjuk arah, alat pemantul cahaya, alat pengukur kecepatan, kedalaman alur ban, kaca depan, spakbor, bumper, penggandengan, penempelan, atau penghapus kaca.	F
286	Kesesuaian daya mesin penggerak terhadap berat kendaraan	G
287(1)	Melanggar rambu lalu lintas	H
287 (2)	Melanggar aturan perintah atau melanggar yang dinyatakan dengan Alat Pemberi Isyarat Lalu Lintas	I
288 (1)	Tidak Memiliki STNK	J

288 (2)	Pengendara Sepeda Motor yang memiliki SIM namun tidak dapat menunjukkannya	K
288 (3)	Kendaraan tidak dilengkapi dengan surat keterangan uji berkala dan tanda lulus uji berkala	L
289	Setiap pengemudi atau penumpang yang duduk di samping pengemudi mobil tidak menggunakan sabuk keselamatan	M
291 (1)	Tidak mengenakan helm standar nasional	N
291 (2)	pengemudi yang membiarkan penumpangnya tidak mengenakan helm	O
293 (1)	Tanpa menyalakan Lampu utama pada malam hari	P
293 (2)	Tanpa menyalakan Lampu utama pada siang hari	Q
297	Balapan liar di Jalanan	R
300	Penggunaan jalur atau lajur. Tidak menggunakan lanjut yang telah ditentukan	S
303	Pengemudi mobil barang untuk mengangkut orang	T
307	Tidak mematuhi ketentuan mengenai tata cara pemuatan, daya angkut	U

Tabel 4.3 Inisialisasi Atribut Formulir

Formulir	Inisialisasi
Merah	Me
Biru	Bi

Tabel 4.4 Inisialisasi Atribut Jenis Kendaraan

Jenis Kendaraan	Inisialisasi
Sepeda Motor	SM
Mobil pick up	MP
Mobil penumpang pribadi	MPP
Mobil penumpang umum	MPU
Bus	BUS
Truck	TRUK

Kemudian untuk atribut numerik, yaitu pada atribut denda, akan dilakukan normalisasi agar nilai-nilainya berada dalam rentang yang sama. Normalisasi ini bertujuan untuk mencegah dominasi atribut dengan skala besar terhadap atribut lainnya saat penghitungan jarak. Metode normalisasi yang digunakan adalah *min-max normalization*, yang mengubah nilai ke dalam rentang [0, 1] menggunakan persamaan 2.7, sebagaimana akan ditunjukkan pada Tabel 4.5.

Tabel 4.5 Normalisasi Atribut Denda

Denda	Normalisasi
47000	0
49000	0,001377
74000	0,018595
99000	0,035813
101000	0,037190
119000	0,049587
149000	0,070248
199000	0,104683
249000	0,139118
299000	0,173554
990000	0,649449
999000	0,655647
1499000	1

Di bawah ini merupakan proses perhitungan dari normalisasi data berdasarkan persamaan 2.7.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Akan dilakukan perhitungan normalisasi secara manual untuk salah satu nilai, misalnya untuk $x = 199000$. Maka

$$x_{max} = 1499000$$

$$x_{min} = 47000$$

$$x = 199000$$

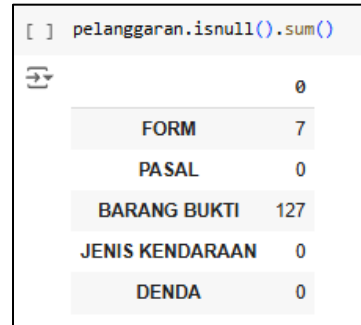
Maka:

$$x' = \frac{199000 - 47000}{1499000 - 47000} = \frac{152000}{1452000} = 0,104683$$

4.3 Preprocessing Data

Tahap *preprocessing* data bertujuan untuk mempersiapkan data agar berkualitas dan siap digunakan dalam analisis klusterisasi. Langkah yang dilakukan

adalah pengecekan *missing value* atau data yang tidak terisi pada data dan ditemukan sebanyak 134 data yang kosong yang terdiri dari tujuh pada atribut form dan 127 pada atribut barang bukti, sebagaimana bisa dilihat pada Gambar 4.1.



```
[ ] pelanggaran.isnull().sum()
```

	0
FORM	7
PASAL	0
BARANG BUKTI	127
JENIS KENDARAAN	0
DENDA	0

Gambar 4.1 Pengecekan *Missing value*

Lalu 134 data kosong ini akan dihapus dikarenakan untuk menghindari potensi bias atau kesalahan dalam analisis. Mengisi data kosong dengan nilai estimasi (seperti rata-rata atau modus) dianggap kurang tepat, karena dapat mengubah distribusi data dan mengaburkan pola yang sebenarnya. Dengan *preprocessing* ini, data menjadi lebih bersih, konsisten, dan siap untuk diimplementasikan menggunakan algoritma *K-Prototype*, sehingga diharapkan hasil klusterisasi menjadi lebih optimal. Sebagaimana ditunjukkan pada Gambar 4.2, data telah siap diolah menggunakan Algoritma *K-Prototype* dengan jumlah data sebanyak 7.194.

	FORM	PASAL	BARANG BUKTI	JENIS KENDARAAN	DENDA
0	Me	E	STNK	SM	0.018595
1	Me	E	STNK	SM	0.018595
2	Me	E	STNK	SM	0.018595
3	Me	E	STNK	SM	0.018595
4	Me	H	STNK	SM	0.018595
...	...	...	...	...	...
7323	Bi	P	STNK	SM	0.035813
7324	Bi	P	STNK	SM	0.035813
7325	Bi	P	STNK	SM	0.018595
7326	Bi	P	STNK	SM	0.018595
7327	Me	Q	SIM C	SM	0.001377

7194 rows x 5 columns

Gambar 4.2 Data Siap Diolah

4.4 Implementasi Algoritma *K-Prototype*

Tahap ini adalah bagian utama dari penelitian ini, dimana dilakukan klasterisasi data pelanggaran lalu lintas menggunakan algoritma *K-Prototype*. Algoritma ini dipilih karena kemampuannya mengelompokkan data campuran yang mencakup atribut numerik dan kategorikal, seperti yang terdapat pada Gambar 4.2. *K-Prototype* memiliki kelebihan dalam kesederhanaan proses algoritmanya, membuatnya lebih efisien untuk menangani *dataset* yang berukuran besar.

Adapun beberapa tahap yang akan dilakukan pada proses *clustering* menggunakan algoritma *K-Prototype* dimulai dengan pemilihan k buah objek secara acak dari *dataset* sebagai *centroid*. Setelah *centroid* ditentukan, langkah selanjutnya adalah menghitung jarak antara setiap objek dalam *dataset* dengan *centroid* yang telah ditetapkan, menggunakan persamaan 2.5. Setiap data kemudian dikelompokkan berdasarkan jarak terdekatnya dengan *centroid*, sehingga terbentuk klaster-klaster awal. Selanjutnya, titik tengah baru ditentukan dengan menghitung rata-rata untuk atribut numerik dan modus untuk atribut kategorik yang terdapat dalam setiap klaster. Proses ini diulang dengan menghitung kembali nilai *centroid* dan menempatkan data ke dalam klaster, hingga tidak ada perubahan pada nilai *centroid* (anggota setiap klaster tidak berubah). ketika tidak terdapat perubahan lebih lanjut, proses *clustering* dianggap selesai.

4.4.1 Implementasi Data Sampel Perhitungan Manual Algoritma *K-Prototype*

Data yang akan digunakan untuk perhitungan manual algoritma *K-Prototype* merupakan *dataset* yang terdiri dari berbagai atribut, baik numerik

maupun kategorikal, yang merepresentasikan karakteristik dari objek yang akan dikelompokkan. Adapun sampel yang diambil dari beberapa pelanggaran lalu lintas yang terjadi sebanyak 10 data, sebagaimana bisa dilihat pada Tabel 4.6.

Tabel 4.6 Data Sampel Pelanggaran Lalu Lintas

NO	FORM	PASAL	BARANG BUKTI	JENIS KENDARAAN	DENDA
1	Me	H	SIM A	MPP	0,035813
2	Me	H	STNK	Truk	0,104683
3	Me	E	STNK	SM	0,018595
4	Me	C	STNK	MP	0,070248
5	Me	C	STNK	MP	0,070248
6	Me	H	SIM A	MPP	0,035813
7	Me	H	STNK	MP	0,035813
8	Me	I	STNK	SM	0,018595
9	Me	I	SIM A	MPP	0,001377
10	Me	H	SIM A	MPP	0,035813

1. Menentukan jumlah K kluster

Dalam penelitian ini, jumlah kluster yang ditetapkan adalah sebanyak 2.

Penentuan jumlah kluster dilakukan dengan mempertimbangkan kebutuhan analisis yang relevan dalam penelitian ini.

2. Menentukan *Centroid* awal

Pada tahap ini, akan dilakukan pemilihan *centroid* awal untuk setiap kluster.

Dan *centroid* ini dipilih secara acak dari data sampel. Penentuan *centroid* ini diambil dari banyak K kluster sebagaimana pada Tabel 4.7.

Tabel 4.7 *Centroid* Sampel Data

NO	FORM	PASAL	BARANG BUKTI	JENIS KENDARAAN	DENDA
3	Me	E	STNK	SM	0,018595
4	Me	C	STNK	MP	0,070248

3. Proses perhitungan jarak antar data

Selanjutnya akan dilakukan perhitungan dengan menghitung jarak seluruh data ke titik pusat kluster dengan menggunakan rumus pada persamaan 2.5.

Kemudian akan dikelompokkan berdasarkan klaster terpendek pada setiap klaster. Untuk gamma akan dihitung menggunakan *default library* yaitu menggunakan persamaan 2.4. Sehingga dituliskan sebagai berikut:

$$\begin{aligned}
 \text{Mean} &= \frac{\text{Jumlah data Numerik}}{\text{Banyak data Numerik}} \\
 &= \frac{(0,035813+0,104683+0,018595+\dots+0,035813)}{10} \\
 &= 0,0426998
 \end{aligned}$$

Setelah ditemukan mean, kemudian akan dihitung Nilai variansi sebagai berikut:

$$\text{Variansi} = \frac{\sum(x_i - \text{Mean})^2}{\text{Banyak data}}$$

Untuk setiap data x_i akan dihitung $(x_i - \text{Mean})^2$

$$\begin{aligned}
 x_1 &= (0,035813 - 0,0426998)^2 \\
 &= 0,00004743
 \end{aligned}$$

Perhitungan selanjutnya dapat dilihat pada Tabel 4.8.

Tabel 4.8 Data Denda dan Hasil Kuadrat Selisih terhadap Mean

No	DENDA	$(x_i - \text{Mean})^2$
1	0,035813	0,00004743
2	0,104683	0,00384192
3	0,018595	0,00058104
4	0,070248	0,00075890
5	0,070248	0,00075890
6	0,035813	0,00004743
7	0,035813	0,00004743
8	0,018595	0,00058104
9	0,001377	0,00170757
10	0,035813	0,00004743
Total		0,00841909

Kemudian akan dihitung standar deviasi namun akan dihitung variansinya terlebih dahulu.

$$\begin{aligned}
 \text{Variansi} &= \frac{\sum(x_i - \text{Mean})^2}{\text{Banyak data}} \\
 &= \frac{0,00841909}{10} \\
 &= 0,000841909 \\
 \text{Std} &= \sqrt{\text{Var}} \\
 &= \sqrt{0,000841909} \\
 &= 0,029015672
 \end{aligned}$$

Maka nilai dari γ adalah

$$\begin{aligned}
 \gamma &= 0,5 \times \text{std}(X_{\text{Numerikal}}) \\
 &= 0,5 \times 0,029015672 \\
 &= 0,0145078
 \end{aligned}$$

Kemudian akan dilakukan perhitungan pada iterasi pertama. Pada tahap ini, dilakukan penghitungan jarak antara setiap data dalam *dataset* dengan *centroid* pertama (C0) dan *centroid* kedua (C1). C0 menunjukkan *centroid* pertama, sedangkan C1 menunjukkan *centroid* kedua. Perhitungan ini bertujuan untuk mengidentifikasi kedekatan data dengan masing-masing *centroid*, sehingga mempermudah proses penentuan keanggotaan kluster. Rumus $d(X_j, Z_i)$, yang menunjukkan jarak antara data X_j dan *centroid* Z_i , dapat disederhanakan menjadi $d_{j,i}$. Misalnya, $d_{1,1}$ berarti jarak antara data ke-1 dengan *centroid* ke-1. Penulisan ini lebih ringkas dan mempermudah identifikasi jarak antara data dan *centroid*.

Iterasi pertama pada C0:

$$\begin{aligned}
 d_{1,1} &= \sqrt{(0,035813 - 0,018595)^2 + \gamma(d(Me, Me) + d(H, E) + d(SIM A, STNK) + d(MPP, SM))} \\
 &= \sqrt{0,00029646 + 0,0145078 \times (0 + 1 + 1 + 1)} \\
 &= 0,209332195 \\
 d_{2,1} &= \sqrt{(0,104683 - 0,018595)^2 + \gamma(d(Me, Me) + d(H, E) + d(STNK, STNK) + d(Truk, SM))} \\
 &= \sqrt{0,00741114 + 0,0145078 \times (0 + 1 + 0 + 1)} \\
 &= 0,190858104 \\
 &\vdots \\
 d_{10,1} &= \sqrt{(0,035813 - 0,018595)^2 + \gamma(d(Me, Me) + d(H, E) + d(SIM A, STNK) + d(MPP, SM))} \\
 &= \sqrt{0,00029646 + 0,0145078 \times (0 + 1 + 1 + 1)} \\
 &= 0,209332195
 \end{aligned}$$

Iterasi pertama pada C1:

$$\begin{aligned}
 d_{1,2} &= \sqrt{(0,035813 - 0,070248)^2 + \gamma(d(Me, Me) + d(H, C) + d(SIM A, STNK) + d(MPP, MP))} \\
 &= \sqrt{0,00118577 + 0,0145078 \times (0 + 1 + 1 + 1)} \\
 &= 0,211445685 \\
 d_{2,2} &= \sqrt{(0,104683 - 0,070248)^2 + \gamma(d(Me, Me) + d(H, C) + d(STNK, STNK) + d(Truk, MP))} \\
 &= \sqrt{0,00118577 + 0,0145078 \times (0 + 1 + 0 + 1)} \\
 &= 0,173785619 \\
 &\vdots \\
 d_{10,2} &= \sqrt{(0,035813 - 0,070248)^2 + \gamma(d(Me, Me) + d(H, C) + d(SIM A, STNK) + d(MPP, MP))} \\
 &= \sqrt{0,00118577 + 0,0145078 \times (0 + 1 + 1 + 1)} \\
 &= 0,211445685
 \end{aligned}$$

Hasil perhitungan jarak akan dilakukan ke semua data. Hasil perhitungan jarak antara *centroid* dengan data akan ditunjukkan pada Tabel 4.9.

Tabel 4.9 Hasil Klaster Iterasi 1

No	Jarak C0	Jarak C1	Klaster
1	0,209332195	0,211445685	0
2	0,190858104	0,173785619	1
3	0	0,17799917	0
4	0,17799917	0	1
5	0,17799917	0	1
6	0,209332195	0,211445685	0
7	0,171207861	0,125274121	1
8	0,120448479	0,17799917	0
9	0,209332195	0,219696889	0
10	0,209332195	0,211445685	0

Setelah iterasi pertama selesai, langkah berikutnya adalah menghitung ulang posisi *centroid* baru untuk setiap klaster berdasarkan data yang telah dikelompokkan. Caranya, hitung nilai rata-rata untuk atribut numerik dan tentukan nilai yang paling sering muncul untuk atribut kategorikal. Sebagaimana pada Tabel 4.10 dan Tabel 4.11.

Tabel 4.10 Pengelompokan Untuk Klaster 0 Iterasi 1

NO	FORM	PASAL	BARANG BUKTI	JENIS KENDARAAN	DENDA
1	Me	H	SIM A	MPP	0,03581300
3	Me	E	STNK	SM	0,01859500
6	Me	H	SIM A	MPP	0,03581300
8	Me	I	STNK	SM	0,01859500
9	Me	I	SIM A	MPP	0,00137700
10	Me	H	SIM A	MPP	0,03581300
Centroid baru	Me	H	SIM A	MPP	0,02433433

Tabel 4.11 Pengelompokan Untuk Klaster 1 Iterasi 1

NO	FORM	PASAL	BARANG BUKTI	JENIS KENDARAAN	DENDA
2	Me	H	STNK	Truk	0,104683
4	Me	C	STNK	MP	0,070248
5	Me	C	STNK	MP	0,070248
7	Me	H	STNK	MP	0,035813
Centroid baru	Me	H	STNK	MP	0,070248

Tabel 4.12 *Centroid* Baru dari Iterasi 1

FORM	PASAL	BARANG BUKTI	JENIS KENDARAAN	DENDA
Me	H	SIM A	MPP	0,02433433
Me	H	STNK	MP	0,07024800

Setelah titik *centroid* yang baru didapat, langkah selanjutnya yaitu melakukan perhitungan untuk iterasi kedua dengan menghitung jarak setiap data ke titik pusat *centroid* baru pada Tabel 4.12. Setiap data akan diperiksa kembali jaraknya untuk menentukan apakah perlu berpindah ke kluster yang berbeda berdasarkan titik *centroid* terbaru. Setelah selesai, *centroid* akan dihitung ulang berdasarkan komposisi data di setiap kluster yang baru jika kluster tidak mengalami perubahan maka proses *clustering* akan dianggap selesai. Untuk perhitungan bisa dilihat pada Tabel 4.13.

Tabel 4.13 Hasil Kluster Iterasi ke-2

No	Jarak C0	Jarak C1	Kluster
1	0,011479	0,173786	0
2	0,188339	0,125274	1
3	0,208702	0,177999	1
4	0,213615	0,120448	1
5	0,213615	0,120448	1
6	0,011479	0,173786	0
7	0,170726	0,034435	1
8	0,208702	0,177999	1
9	0,122617	0,219697	0
10	0,011479	0,173786	0

Bisa dilihat hasil kluster pada iterasi pertama berbeda dengan hasil pada iterasi kedua, artinya pembagian kluster masih akan berlanjut sampai mencapai konvergensi atau tidak ada perubahan pada pusat kluster atau anggota kluster tidak ada perubahan. Kemudian akan dicari *centroid* baru dari hasil iterasi kedua.

Tabel 4.14 Pengelompokan Untuk Klaster 0 Iterasi ke-2

NO	FORM	PASAL	BARANG BUKTI	JENIS KENDARAAN	DENDA
1	Me	H	SIM A	MPP	0,03581300
6	Me	H	SIM A	MPP	0,03581300
9	Me	I	SIM A	MPP	0,00137700
10	Me	H	SIM A	MPP	0,03581300
<i>Centroid baru</i>	Me	H	SIM A	MPP	0,027204

Tabel 4.15 Pengelompokan Untuk Klaster 1 Iterasi ke-2

NO	FORM	PASAL	BARANG BUKTI	JENIS KENDARAAN	DENDA
2	Me	H	STNK	Truk	0,10468300
3	Me	E	STNK	SM	0,01859500
4	Me	C	STNK	MP	0,07024800
5	Me	C	STNK	MP	0,07024800
7	Me	H	STNK	MP	0,03581300
8	Me	I	STNK	SM	0,01859500
<i>Centroid baru</i>	Me	H	STNK	MP	0,053030333

Tabel 4.16 *Centroid* Baru Dari Iterasi ke-2

FORM	PASAL	BARANG BUKTI	JENIS KENDARAAN	DENDA
Me	H	SIM A	MPP	0,02720400
Me	H	STNK	MP	0,05303033

Setelah titik *centroid* yang baru didapat pada Tabel 4.16, kemudian melakukan perhitungan untuk iterasi ketiga dengan menghitung jarak setiap data ke titik pusat *centroid* baru, sebagaimana pada Tabel 4.17.

Tabel 4.17 Hasil Klaster Iterasi ke-3

No	Jarak C0	Jarak C1	Klaster
1	0,008609	0,209332	0
2	0,187133	0,177999	1
3	0,208800	0,173786	1
4	0,213017	0,017218	1
5	0,213017	0,017218	1
6	0,008609	0,209332	0
7	0,170557	0,121673	1
8	0,208800	0,173786	1
9	0,123186	0,214922	0
10	0,008609	0,209332	0

Bisa dilihat hasil kluster pada iterasi ketiga (Tabel 4.17) memiliki hasil yang sama pada iterasi kedua (Tabel 4.13), oleh karena itu proses *clustering* selesai pada iterasi ke-3 dengan nilai *centroid* pada Tabel 4.18.

Tabel 4.18 *Centroid* Baru Dari Iterasi ke-3

FORM	PASAL	BARANG BUKTI	JENIS KENDARAAN	DENDA
Me	H	SIM A	MPP	0,02720400
Me	H	STNK	MP	0,05303033

Pada Tabel 4.18 hasil *centroid* sama pada Tabel 4.16 hal ini menandakan bahwa hasil *centroid* sudah optimal dan kluster telah mencapai konvergensi oleh karena itu proses iterasi berhenti sampai pada iterasi ke-3.

Setelah proses iterasi selesai, langkah selanjutnya adalah menghitung nilai WCSS untuk mengevaluasi performa *clustering*. WCSS merupakan metrik yang menunjukkan total jarak antara setiap data dengan pusat klasternya, yang digunakan untuk menilai kualitas kluster yang terbentuk. Dengan menggunakan persamaan 2.8 maka akan didapatkan WCSS untuk $K = 2$ didapatkan sebagai berikut:

$$\begin{aligned}
 d_{1,1} &= (0,035813 - 0,027204)^2 + \gamma(d(Me, Me) + d(H, H) + d(SIM A, SIM A) + d(MPP, MPP)) \\
 &= 0,0000741149 + 0,0145078 \times (0 + 0 + 0 + 0) \\
 &= 0,0000741149 \\
 d_{2,2} &= (0,104683 - 0,05303033)^2 + \gamma(d(Me, Me) + d(H, C) + d(STNK, STNK) + d(Truk, MP)) \\
 &= 0,002668 + 0,0145078 \times (0 + 1 + 0 + 1) \\
 &= 0,03168367
 \end{aligned}$$

Perhitungan selebihnya dapat dilihat pada Tabel 4.19.

Tabel 4.19 WCSS Untuk $K = 2$

No	WCSS
1	0,000074
2	0,031684
3	0,030201
4	0,000296
5	0,000296
6	0,000074
7	0,014804
8	0,030201
9	0,015175
10	0,000074
Total	0,122881

Setelah menghitung dan memperoleh hasil *clustering* untuk $K = 2$ secara manual, langkah selanjutnya adalah memperluas analisis dengan menggunakan kode python untuk menghitung dan mengevaluasi kluster hingga K stabil. Proses ini akan memungkinkan pengujian berbagai jumlah kluster secara otomatis, cepat dan efisien, sehingga dapat diperoleh pemahaman yang lebih lengkap tentang bagaimana data terbagi ke dalam kluster-kluster yang terbaik. Tujuannya adalah untuk menemukan jumlah kluster yang paling sesuai, di mana data dalam setiap kluster memiliki kesamaan yang lebih tinggi dan lebih terpisah dengan jelas dari kluster lainnya.

Selanjutnya untuk $K = 1$, dilakukan dengan cara meminimalkan nilai *Within-Cluster Sum of Squares* (WCSS). Metode ini memastikan bahwa pembagian data ke dalam kluster menghasilkan jarak intra-kluster yang terkecil, sehingga setiap data dalam kluster memiliki kesamaan yang tinggi terhadap pusat *centroid* yang ditentukan. Hasilnya, pusat *centroid* terbaik yang diperoleh pada Gambar 4.3.

```
print(kproto.cluster_centroids_)
[['0.0426998' 'Me' 'H' 'STNK' 'MPP']]
```

Gambar 4.3 Centroid Untuk $K = 1$

Selanjutnya akan dihitung jarak semua data terhadap *centroid* sebagaimana dapat dilihat pada Gambar 4.3, sehingga didapatkan hasil pada Tabel 4.20.

Tabel 4.20 Hasil Jarak Untuk $K = 1$

No	Jarak C0	Klaster
1	0,120645	0
2	0,135461	0
3	0,172037	0
4	0,172553	0
5	0,172553	0
6	0,120645	0
7	0,120645	0
8	0,172037	0
9	0,175280	0
10	0,120645	0

Kemudian akan dihitung nilai WCSS untuk $K = 1$, akan menghasilkan jarak intra-klaster yang terkecil sebagaimana bisa dilihat pada Tabel 4.21.

Tabel 4.21 WCSS Untuk $K = 1$

No	WCSS
1	0,014555
2	0,018350
3	0,029597
4	0,029775
5	0,029775
6	0,014555
7	0,014555
8	0,029597
9	0,030723
10	0,014555
Total	0,226037

Selanjutnya, untuk $K = 3$, pusat *centroid* terbaik telah ditentukan oleh sistem. Pemilihan ini dilakukan dengan meminimalkan nilai WCSS guna

memastikan bahwa pembagian data ke dalam kluster menghasilkan jarak intra-kluster yang terkecil. Hasil akhir proses penentuan *centroid* ini ditunjukkan pada Gambar 4.4, yang menampilkan koordinat *centroid* untuk masing-masing kluster.

```
print(kproto.cluster_centroids_)

[['0.027204' 'Me' 'H' 'SIM A' 'MPP']
 ['0.018595' 'Me' 'E' 'STNK' 'SM']
 ['0.070248' 'Me' 'C' 'STNK' 'MP']]
```

Gambar 4.4 Centroid Untuk $K = 3$

Selanjutnya akan dihitung jarak semua data terhadap *centroid* sebagaimana dapat dilihat pada Gambar 4.4, sehingga didapatkan hasil pada Tabel 4.22.

Tabel 4.22 Hasil Jarak Untuk $K = 3$

No	Jarak C0	Jarak C1	Jarak C2	Kluster
1	0,008609	0,209331936	0,211445428	0
2	0,187132561	0,190857915	0,173785411	2
3	0,208800179	0	0,177998967	1
4	0,213016868	0,177998967	0	2
5	0,213016868	0,177998967	0	2
6	0,008609	0,209331936	0,211445428	0
7	0,170557072	0,17120765	0,125273977	3
8	0,208800179	0,120448329	0,177998967	1
9	0,123186176	0,209331936	0,219696642	0
10	0,008609	0,209331936	0,211445428	0

Kemudian akan dihitung nilai WCSS untuk $K = 3$ akan menghasilkan jarak intra-kluster yang terkecil sebagaimana bisa dilihat pada Tabel 4.23.

Tabel 4.23 WCSS Untuk $K = 3$

No	WCSS
1	0,0000741149
2	0,0302013690
3	0
4	0
5	0
6	0,0000741149
7	0,0156935690
8	0,0145078000
9	0,0151748340
10	0,0000741149
Total	0,0757999170

Selanjutnya, untuk $K = 4$, pusat *centroid* terbaik telah ditentukan oleh Sistem. Pemilihan ini dilakukan dengan meminimalkan nilai WCSS guna memastikan bahwa pembagian data ke dalam kluster menghasilkan jarak intra-kluster yang terkecil. Hasil akhir proses penentuan *centroid* ini ditunjukkan pada Gambar 4.5, yang menampilkan koordinat *centroid* untuk masing-masing kluster.

```
print(kproto.cluster_centroids_)

[['0.027204' 'Me' 'H' 'SIM A' 'MPP']
 ['0.05876966666666665' 'Me' 'C' 'STNK' 'MP']
 ['0.104683' 'Me' 'H' 'STNK' 'Truk']
 ['0.018595' 'Me' 'E' 'STNK' 'SM']]
```

Gambar 4.5 Centroid Untuk $K = 4$

Selanjutnya akan dihitung jarak ke semua data sehingga didapatkan hasil pada Tabel 4.24.

Tabel 4.24 Hasil Jarak Untuk $K = 4$

No.	Jarak C0	Jarak C1	Jarak C2	Jarak C3	Kluster
1	0,008609	0,209881892	0,183735345	0,20933193	0
2	0,187132561	0,176418917	0	0,19085791	2
3	0,208800179	0,175013153	0,190857915	0	3
4	0,213016868	0,011478333	0,173785411	0,17799896	1
5	0,213016868	0,011478333	0,173785411	0,17799896	1
6	0,008609	0,209881892	0,183735345	0,20933196	0
7	0,170557072	0,12261651	0,138747529	0,1712076	1
8	0,208800179	0,175013153	0,190857915	0,12044832	3
9	0,123186176	0,216373099	0,232799333	0,20933193	0
10	0,008609	0,209881892	0,183735345	0,20933193	0

Kemudian akan dihitung nilai WCSS untuk $K = 4$ akan menghasilkan jarak intra-kluster yang terkecil sebagaimana bisa dilihat pada Tabel 4.25.

Tabel 4.25 WCSS Untuk $K = 4$

No	WCSS
1	0,0000741149
2	0
3	0
4	0,0001320000
5	0,0001320000
6	0,0000741149
7	0,0150350000
8	0,0145080000
9	0,0151750000
10	0,0000741149
Total	0,0452040000

Selanjutnya, untuk $K = 5$, pusat *centroid* terbaik telah ditentukan oleh Sistem. Pemilihan ini dilakukan dengan meminimalkan nilai WCSS guna memastikan bahwa pembagian data ke dalam kluster menghasilkan jarak intra-kluster yang terkecil. Hasil akhir proses penentuan *centroid* ini ditunjukkan pada Gambar 4.6, yang menampilkan koordinat *centroid* untuk masing-masing kluster.

```
print(kproto.cluster_centroids_)
[['0.035813' 'Me' 'H' 'SIM A' 'MPP']
 ['0.05876966666666665' 'Me' 'C' 'STNK' 'MP']
 ['0.018595' 'Me' 'E' 'STNK' 'SM']
 ['0.001377' 'Me' 'I' 'SIM A' 'MPP']
 ['0.104683' 'Me' 'H' 'STNK' 'Truk']]
```

Gambar 4.6 Centroid Untuk $K = 5$

Selanjutnya akan dihitung jarak semua data terhadap *centroid* sehingga didapatkan hasil pada Tabel 4.26.

Tabel 4.26 Hasil Jarak Untuk $K = 5$

No	Jarak C0	Jarak C1	Jarak C2	Jarak C3	Jarak C4	Kluster
1	0	0,2099	0,209	0,1253	0,1837	0
2	0,1837	0,1764	0,191	0,2328	0	4
3	0,2093	0,175	5E-06	0,2093	0,1909	2
4	0,2114	0,0115	0,178	0,2197	0,1738	1
5	0,2114	0,0115	0,178	0,2197	0,1738	1
6	0	0,2099	0,209	0,1253	0,1837	0
7	0,1703	0,1226	0,171	0,2114	0,1387	1
8	0,2093	0,175	0,12	0,1712	0,1909	2
9	0,1253	0,2164	0,209	0	0,2328	3
10	0	0,2099	0,209	0,1253	0,1837	1

Kemudian akan dihitung nilai WCSS untuk $K = 5$ akan menghasilkan jarak intra-kluster yang terkecil sebagaimana bisa dilihat pada Tabel 4.27.

Tabel 4.27 WCSS Untuk $K = 5$

No	WCSS
1	0
2	0
3	0,0000000000025
4	0,000132000000
5	0,000132000000
6	0
7	0,015035000000
8	0,014508000000
9	0
10	0
Total	0,029806000000

Selanjutnya, untuk $K = 6$, pusat *centroid* terbaik telah ditentukan oleh Sistem. Pemilihan ini dilakukan dengan meminimalkan nilai WCSS guna memastikan bahwa pembagian data ke dalam kluster menghasilkan jarak intra-kluster yang terkecil. Hasil akhir proses penentuan *centroid* ini ditunjukkan pada Gambar 4.7, yang menampilkan koordinat *centroid* untuk masing-masing kluster.

```
print(kproto.cluster_centroids_)

[['0.104683' 'Me' 'H' 'STNK' 'Truk']
 ['0.035813' 'Me' 'H' 'SIM A' 'MPP']
 ['0.001377' 'Me' 'I' 'SIM A' 'MPP']
 ['0.018595' 'Me' 'E' 'STNK' 'SM']
 ['0.035813' 'Me' 'H' 'STNK' 'MP']
 ['0.070248' 'Me' 'C' 'STNK' 'MP']]
```

Gambar 4.7 Centroid Untuk $K = 6$

Selanjutnya akan dihitung jarak semua data terhadap *centroid* sehingga didapatkan hasil pada Tabel 4.28.

Tabel 4.28 Hasil Jarak Untuk $K = 6$

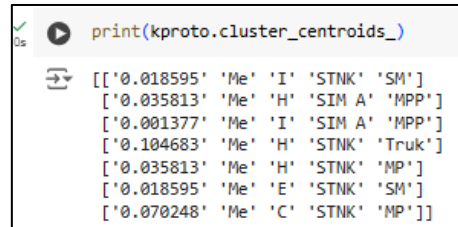
No	Jarak C0	Jarak C1	Jarak C2	Jarak C3	Jarak C4	Jarak C5	Klaster
1	0,1837	0	0,12527	0,2093	0,17034	0,2114	1
2	0	0,1837	0,2328	0,1909	0,13875	0,1738	0
3	0,1909	0,2093	0,20933	0	0,17121	0,178	3
4	0,1738	0,2114	0,2197	0,178	0,12527	0	5
5	0,1738	0,2114	0,2197	0,178	0,12527	0	5
6	0,1837	0	0,12527	0,2093	0,17034	0,2114	1
7	0,1387	0,1703	0,21145	0,1712	0	0,1253	4
8	0,1909	0,2093	0,17121	0,1204	0,17121	0,178	3
9	0,2328	0,1253	0	0,2093	0,21145	0,2197	2
10	0,1837	0	0,12527	0,2093	0,17034	0,2114	1

Kemudian akan dihitung nilai WCSS untuk $K = 6$ akan menghasilkan jarak intra-klaster yang terkecil sebagaimana bisa dilihat pada Tabel 4.29.

Tabel 4.29 WCSS Untuk $K = 6$

No	WCSS
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0,014508
9	0
10	0
Total	0,014508

Selanjutnya, untuk $K = 7$, pusat *centroid* terbaik telah ditentukan oleh Sistem. Sama seperti langkah sebelumnya. Pusat *centroid* ditunjukkan pada Gambar 4.8.



```
print(kproto.cluster_centroids_)

[[0.018595, 'Me', 'I', 'STNK', 'SM'],
 [0.035813, 'Me', 'H', 'SIM A', 'MPP'],
 [0.001377, 'Me', 'I', 'SIM A', 'MPP'],
 [0.104683, 'Me', 'H', 'STNK', 'Truk'],
 [0.035813, 'Me', 'H', 'STNK', 'MP'],
 [0.018595, 'Me', 'E', 'STNK', 'SM'],
 [0.070248, 'Me', 'C', 'STNK', 'MP']]
```

Gambar 4.8 Centroid Untuk $K = 7$

Selanjutnya akan dihitung jarak semua data sehingga didapatkan hasil pada Tabel 4.30.

Tabel 4.30 Hasil Jarak Untuk $K = 7$

No	Jarak C0	Jarak C1	Jarak C2	Jarak C3	Jarak C4	Jarak C5	Jarak C6	Klaster
1	0,2093	0	0,1253	0,1837	0,1703	0,2093	0,211	1
2	0,1909	0,1837	0,2328	0	0,1387	0,1909	0,174	3
3	0,1204	0,2093	0,2093	0,1909	0,1712	0	0,178	5
4	0,178	0,2114	0,2197	0,1738	0,1253	0,178	0	6
5	0,178	0,2114	0,2197	0,1738	0,1253	0,178	0	6
6	0,2093	0	0,1253	0,1837	0,1703	0,2093	0,211	1
7	0,1712	0,1703	0,2114	0,1387	0	0,1712	0,125	4
8	0	0,2093	0,1712	0,1909	0,1712	0,1204	0,178	0
9	0,1712	0,1253	0	0,2328	0,2114	0,2093	0,22	2
10	0,2093	0	0,1253	0,1837	0,1703	0,2093	0,211	1

Kemudian akan dihitung nilai WCSS untuk $K = 7$ akan menghasilkan jarak intra-klaster yang terkecil, sebagaimana bisa dilihat pada Tabel 4.31.

Tabel 4.31 WCSS Untuk $K = 7$

No	WCSS
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
Total	0

Dikarenakan nilai WCSS untuk $K = 7$ sudah 0, maka dapat dipastikan bahwa nilai WCSS untuk K seterusnya juga akan bernilai 0. Hal ini menunjukkan bahwa proses *clustering* sudah selesai sampai dengan $K = 7$. Dengan demikian, hasil analisis nilai WCSS dari $K = 1$ hingga $K = 7$ sudah dapat diketahui dan dapat digunakan sebagai dasar untuk evaluasi jumlah klaster yang optimal.

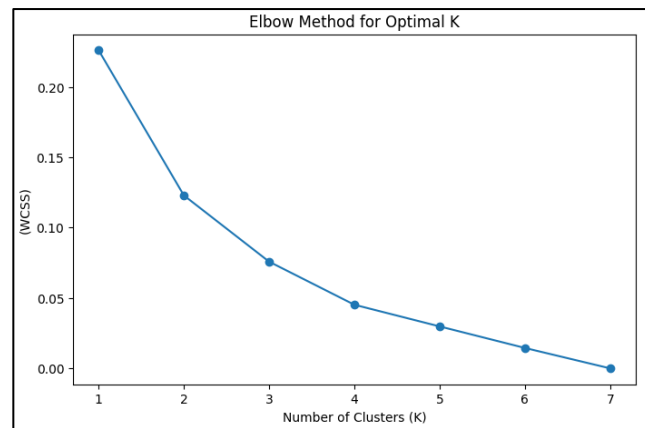
Berdasarkan penelitian yang dilakukan oleh Maori (2023), penentuan jumlah klaster terbaik menggunakan metode *elbow*, di mana nilai dari *elbow* dilihat dari nilai WCSS dengan selisih terbesar. Namun, nilai WCSS dengan selisih terbesar tidak selalu menunjukkan hasil *clustering* yang terbaik. Penurunan yang signifikan pada titik siku grafik lebih cocok digunakan untuk menentukan jumlah klaster yang optimal dan nilai klaster akan stabil setelah titik siku tersebut (Maori, 2023). Dalam penelitian ini, jumlah klaster optimal terlihat dari penurunan WCSS yang signifikan, seperti ditunjukkan pada Tabel 4.32.

Tabel 4.32 WCSS Setiap Klaster Pada Data Sampel

Nilai	WCSS	Selisih
$K = 1$	0,233625667	-
$K = 2$	0,122880983	0,110744684
$K = 3$	0,075799917	0,047081066
$K = 4$	0,045204000	0,030595917
$K = 5$	0,029806000	0,015398000
$K = 6$	0,014508000	0,015298000
$K = 7$	0	0,014508000

Setelah nilai WCSS (*Within-Cluster Sum of Squares*) untuk setiap jumlah klaster (K) diketahui, langkah selanjutnya adalah menggambarkan nilai-nilai tersebut dalam bentuk grafik. Grafik ini akan menampilkan hubungan antara jumlah klaster (K) dan nilai WCSS. Pola penurunan nilai

WCSS dapat digunakan untuk menentukan jumlah kluster yang optimal, dengan mengamati titik siku pada grafik, yang menandai perubahan signifikan dalam laju penurunan nilai WCSS.



Gambar 4.9 *Elbow Method*

Dari Gambar 4.9, grafik menunjukkan bahwa penurunan yang paling signifikan terjadi pada saat nilai WCSS $K = 2$ dengan selisih sebesar 0,11074468. Namun, merujuk pada penelitian yang dilakukan oleh Mayangsari (2022), selisih tertinggi antar kluster belum tentu menunjukkan hasil yang optimal. Penurunan tersebut tidak dianggap optimal karena setelah $K = 2$, berdasarkan Tabel 4.32, nilai WCSS mulai menunjukkan kestabilan setelah $K = 4$, yang ditandai dengan perbedaan selisih yang hampir sama, yaitu sekitar 0,015. Oleh karena itu, nilai K optimal adalah 4, dengan WCSS sebesar 0,045204 dan selisih 0,030596, yang membentuk sudut siku seperti terlihat pada Gambar 4.9. Hal ini juga didukung oleh grafik WCSS, di mana penurunan mulai menjadi lebih stabil setelah $K = 4$. Dengan demikian, kluster optimal dalam analisis ini ditetapkan pada $K = 4$ (Mayangsari, 2022).

4.4.2 Implementasi Algoritma *K-Prototype* Menggunakan Python

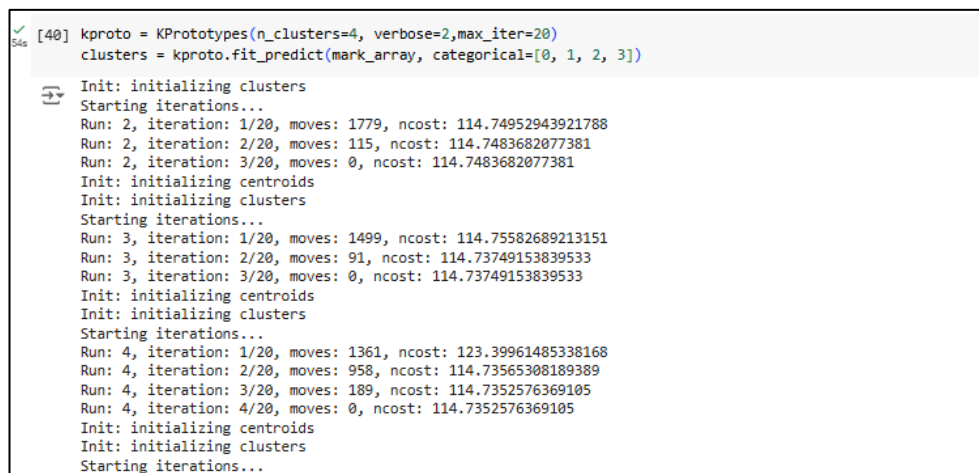
Pada uji sebelumnya (Sub bab 4.4.1), metode yang digunakan untuk *clustering* dilakukan secara manual pada data sampel. Namun, pada bagian ini, implementasi algoritma *K-Prototype* dilakukan secara penuh menggunakan Python dengan data sebanyak 7.194 sebagaimana pada Gambar 4.1 yang telah melewati proses *Pre-processing*. Dengan menggunakan pustaka *K-Modes*, yang menyediakan fungsi untuk algoritma *K-Prototype*, data yang mengandung atribut numerik dan kategorikal dapat diproses secara otomatis dan efisien. Penerapan algoritma ini memungkinkan pengolahan data dalam skala yang lebih besar dan dengan hasil yang lebih akurat oleh karena itu akan dihitung nilai WCSS untuk setiap kluster menggunakan sistem atau *library* yang ada. Hasil WCSS bisa dilihat pada Tabel 4.33.

Tabel 4.33 WCSS Setiap Kluster Pada Data Asli

Nilai	WCSS	Selisih
$K = 1$	176,5725	0
$K = 2$	152,4449	24,127600
$K = 3$	134,0986	18,346280
$K = 4$	114,7353	19,363320
$K = 5$	107,9009	6,834403
$K = 6$	98,47599	9,424865
$K = 7$	89,90412	8,571874
$K = 8$	82,07422	7,829900
$K = 9$	77,42863	4,645583

Seperti yang dapat dilihat pada Tabel 4.33, penurunan WCSS yang signifikan terjadi dari $K = 1$ hingga $K = 4$. Namun, setelah $K = 4$, nilai WCSS tidak mengalami perubahan yang terlalu besar. Hal ini dibuktikan dengan selisih yang relatif kecil setelah $K = 4$. Oleh karena itu, jumlah kluster optimal ditentukan pada $K = 4$. Hal ini didukung oleh data, di mana setelah $K = 4$,

penurunan WCSS mulai melambat secara signifikan. Pada $K = 4$, nilai WCSS tercatat sebesar 114,7352576 dengan selisih sebesar 19,36332. Berdasarkan analisis ini, $K = 4$ merupakan jumlah kluster yang optimal. Dengan demikian, dapat disimpulkan bahwa jumlah kluster optimal untuk data ini adalah $K = 4$. Kemudian, untuk proses *K-Prototype*, algoritma ini dapat diimplementasikan menggunakan *library*, seperti yang ditunjukkan pada Gambar 4.10.

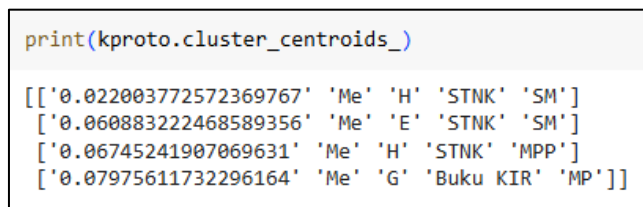


```
[40]: kproto = KPrototypes(n_clusters=4, verbose=2, max_iter=20)
clusters = kproto.fit_predict(mark_array, categorical=[0, 1, 2, 3])

Init: initializing clusters
Starting iterations...
Run: 2, iteration: 1/20, moves: 1779, ncost: 114.74952943921788
Run: 2, iteration: 2/20, moves: 115, ncost: 114.7483682077381
Run: 2, iteration: 3/20, moves: 0, ncost: 114.7483682077381
Init: initializing centroids
Init: initializing clusters
Starting iterations...
Run: 3, iteration: 1/20, moves: 1499, ncost: 114.75582689213151
Run: 3, iteration: 2/20, moves: 91, ncost: 114.73749153839533
Run: 3, iteration: 3/20, moves: 0, ncost: 114.73749153839533
Init: initializing centroids
Init: initializing clusters
Starting iterations...
Run: 4, iteration: 1/20, moves: 1361, ncost: 123.39961485338168
Run: 4, iteration: 2/20, moves: 958, ncost: 114.73565308189389
Run: 4, iteration: 3/20, moves: 189, ncost: 114.7352576369105
Run: 4, iteration: 4/20, moves: 0, ncost: 114.7352576369105
Init: initializing centroids
Init: initializing clusters
Starting iterations...
```

Gambar 4.10 Proses *K-Prototype*

Setelah proses klusterisasi menggunakan algoritma *K-Prototype*, nilai *centroid* terbaik berhasil ditentukan. *Centroid* ini merepresentasikan pusat dari masing-masing kluster, baik dari segi atribut numerik maupun kategori. Dalam hasil yang ditampilkan pada Gambar 4.11.



```
print(kproto.cluster_centroids_)

[['0.022003772572369767' 'Me' 'H' 'STNK' 'SM']
 ['0.060883222468589356' 'Me' 'E' 'STNK' 'SM']
 ['0.06745241907069631' 'Me' 'H' 'STNK' 'MPP']
 ['0.07975611732296164' 'Me' 'G' 'Buku KIR' 'MP']]
```

Gambar 4.11 *Centroid*

Setelah *Centroid* nya berhasil dibentuk maka untuk hasil *clustering* bisa ditampilkan sebagaimana pada Gambar 4.12.

pelanggaran							
	FORM	PASAL	BARANG BUKTI	JENIS KENDARAAN	DENDA	cluster	
0	Me	E	STNK	SM	0.018595	1	
1	Me	E	STNK	SM	0.018595	1	
2	Me	E	STNK	SM	0.018595	1	
3	Me	E	STNK	SM	0.018595	1	
4	Me	H	STNK	SM	0.018595	0	
...	...	...	...	...	...	...	
7323	Bi	P	STNK	SM	0.035813	0	
7324	Bi	P	STNK	SM	0.035813	0	
7325	Bi	P	STNK	SM	0.018595	0	
7326	Bi	P	STNK	SM	0.018595	0	
7327	Me	Q	SIM C	SM	0.001377	0	

7194 rows x 6 columns

Gambar 4.12 Hasil *Clustering*

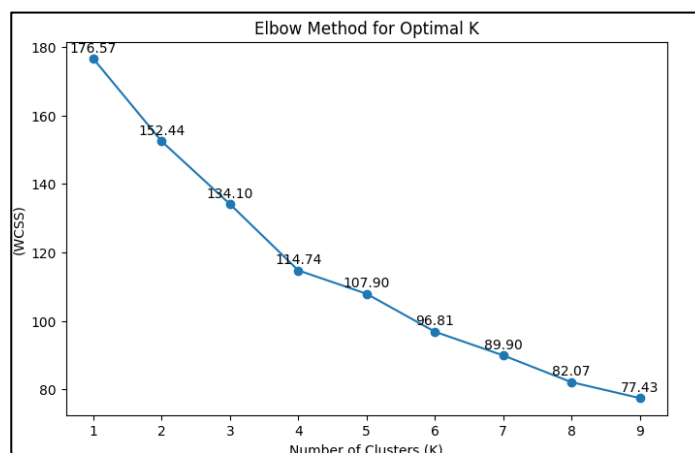
4.5 Evaluasi *Clustering*

Evaluasi *clustering* merupakan langkah penting untuk menilai kualitas hasil klusterisasi yang dilakukan oleh algoritma *K-Prototype*. Dalam konteks ini, evaluasi dilakukan dengan menggunakan *Within-Cluster Sum of Squares* (WCSS), yang digunakan untuk mengukur sejauh mana data dalam setiap kluster terdistribusi dengan baik.

Untuk menentukan jumlah kluster optimal, dilakukan analisis terhadap nilai WCSS untuk berbagai jumlah kluster (K). Biasanya, nilai WCSS akan menurun seiring dengan bertambahnya jumlah kluster. Namun, setelah mencapai titik tertentu, penurunan WCSS akan semakin kecil dan mendatar. Titik ini dikenal sebagai titik *elbow*, yang menandakan jumlah kluster yang optimal. Nilai WCSS bisa dilihat pada Tabel 4.33.

Setelah nilai WCSS untuk setiap jumlah kluster (K) diketahui, langkah selanjutnya adalah menggambarkan nilai-nilai tersebut dalam bentuk grafik. Grafik ini akan menampilkan hubungan antara jumlah kluster (K) dan nilai WCSS. Pola penurunan nilai WCSS dapat digunakan untuk menentukan jumlah kluster yang

optimal, dengan mengamati titik siku pada grafik, yang menandai perubahan signifikan dalam laju penurunan nilai WCSS.



Gambar 4.13 *Elbow Method* Pada Data Asli

Pada penelitian ini, analisis WCSS dilakukan untuk berbagai nilai K , dan hasilnya akan digambarkan menggunakan *elbow method*, sebagaimana pada Gambar 4.13 yang menunjukkan bahwa jumlah kluster optimal ditemukan pada $K = 4$ dimana terjadi patahan, selain itu penurunan nilai WCSS mulai melambat setelah $K = 4$ sebagaimana terdapat pada Tabel 4.33. Hal ini menunjukkan bahwa pembagian data menjadi empat kluster menghasilkan K yang paling optimal.

4.6 Hasil Clustering *K-Prototype*

Pada penelitian ini, data pelanggaran lalu lintas diolah menggunakan algoritma *K-Prototype*, yang merupakan metode *clustering* untuk data campuran (numerik dan kategorikal). Algoritma ini memungkinkan pengelompokan data dengan mempertimbangkan atribut kategorikal (jenis kendaraan, pasal, jenis kendaraan, dan form) serta atribut numerik (denda). Proses *clustering* menghasilkan empat kluster (0, 1, 2, dan 3) yang memiliki karakteristik unik berdasarkan jenis kendaraan yang terlibat, pasal, barang bukti, denda, dan form. Sebagaimana yang ditunjukkan sebagai berikut:

Tabel 4.34 Hasil *Clustering* Jenis Kendaraan

Kendaraan	Klaster 0	Klaster 1	Klaster 2	Klaster 3
BUS	1	0	4	16
Mobil Pick Up	124	0	141	350
Mobil Penumpang Pribadi	0	0	939	10
Mobil Penumpang Umum	1	0	1	8
Sepeda Motor	3945	1230	0	1
Truk	46	0	252	125
Total	4117	1230	1337	510

Tabel 4.35 Hasil *Clustering* Pasal

Pasal	Klaster 0	Klaster 1	Klaster 2	Klaster 3
278	1	0	0	12
280	118	2	10	3
281	350	21	79	3
282	203	0	7	3
285(1)	0	1147	0	0
285(2)	6	8	69	14
286	5	6	77	329
287(1)	1760	0	848	0
287(2)	107	0	77	7
288(1)	210	32	47	40
288(2)	32	2	0	0
288(3)	11	0	45	33
289	4	0	10	0
291(1)	444	8	0	0
291(2)	144	1	0	0
293(1)	388	0	0	0
293(2)	320	2	0	1
297	1	0	1	0
300	1	0	10	1
303	0	1	5	4
307	12	0	52	60
Total	4117	1230	1337	510

Tabel 4.36 Hasil *Clustering* Barang Bukti

Barang Bukti	Klaster 0	Klaster 1	Klaster 2	Klaster 3
Buku KIR	0	0	0	405
KENDARAAN	3	38	0	1
SIM A	39	4	348	45
SIM BI	3	0	22	9
SIM BI UMUM	6	0	30	14
SIM BII UMUM	4	0	17	5
SIM C	584	208	8	2
STNK	3478	980	912	29
Total	4117	1230	1337	510

Tabel 4.37 Hasil *Clustering* Denda

Denda	Klaster 0	Klaster 1	Klaster 2	Klaster 3
47000	0	0	1	0
49000	1459	57	160	7
74000	448	85	1	0
99000	2194	203	378	12
101000	1	0	0	0
119000	0	5	0	0
149000	6	819	431	389
199000	1	10	155	77
249000	8	50	209	23
299000	0	1	1	0
990000	0	0	0	1
999000	0	0	1	0
1499000	0	0	0	1
Total	4117	1230	1337	510

Tabel 4.38 Hasil *Clustering* Form

Form	Klaster 0	Klaster 1	Klaster 2	Klaster 3
Bi	393	139	66	2
Me	3724	1091	1271	508
Total	4117	1230	1337	510

Berdasarkan hasil analisis data pada Tabel 4.34 sampai pada Tabel 4.38, setiap klaster memiliki pola pelanggaran tertentu yang unik. Dalam analisis ini, keempat klaster (0, 1, 2, dan 3) akan dibandingkan untuk memahami karakteristik data secara menyeluruh. Setiap klaster mewakili kelompok pelanggaran dengan ciri-ciri spesifik, baik berdasarkan jenis kendaraan, pasal yang dilanggar, barang bukti yang disita, maupun jumlah denda yang dikenakan.

Untuk memberikan gambaran yang lebih jelas, visualisasi berikut menampilkan distribusi data pada masing-masing klaster (0, 1, 2, dan 3) berdasarkan atribut-atribut penting tersebut. Visualisasi ini bertujuan untuk mengidentifikasi pola dominan pada setiap klaster. Berikut adalah hasil visualisasi analisis keempat klaster.

1. Klaster 0

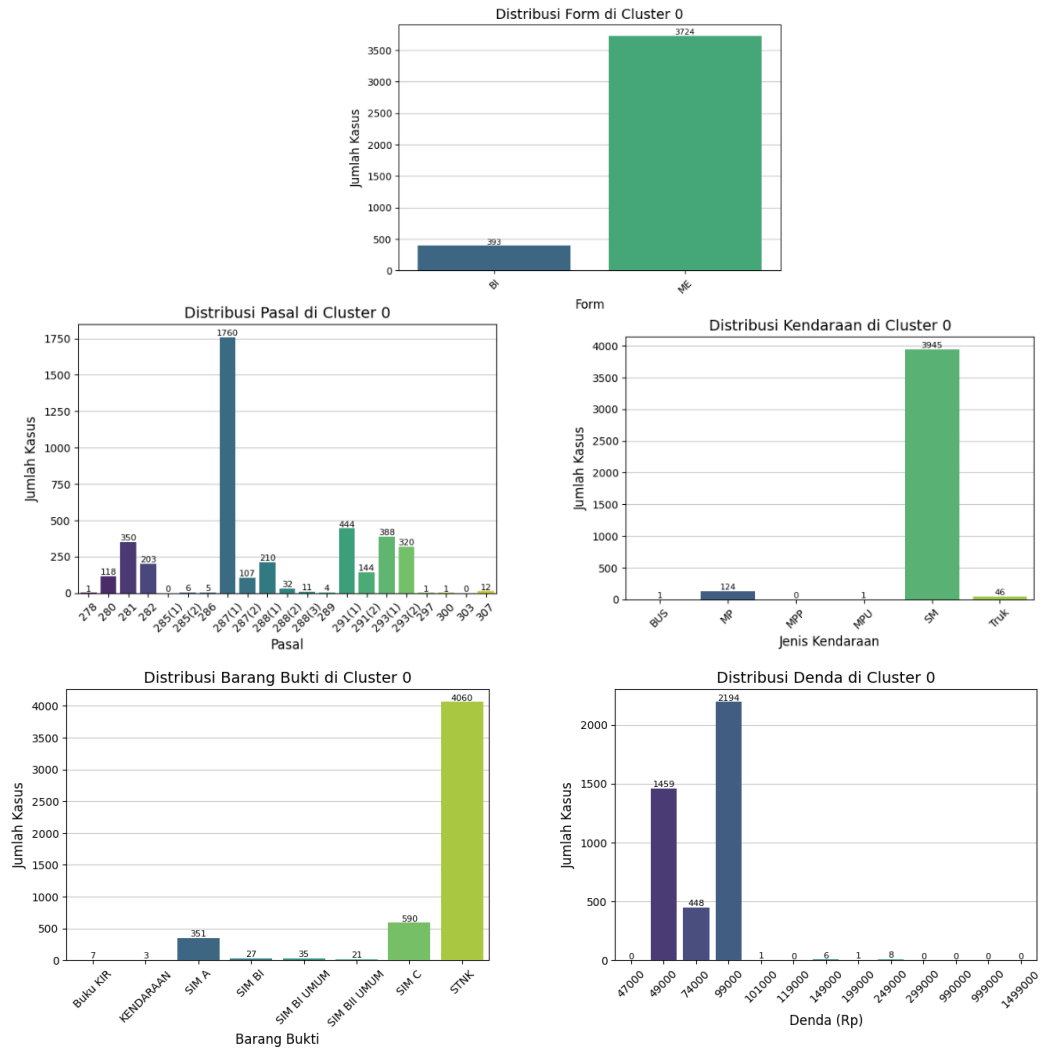
Pada klaster 0 kelompok dengan jumlah pelanggaran terbesar di antara seluruh klaster yaitu dengan 4117 kasus pelanggaran. Berdasarkan analisis data pasal yang dilanggar, Pasal 287(1) menjadi pelanggaran yang paling dominan dengan 1.760 kasus. Selain itu, Pasal 291(1), tercatat sebagai pasal kedua terbanyak dilanggar dengan 444 kasus. Pasal 293(1) sebanyak 388 kasus. Diikuti oleh Pasal 281 tercatat 350 kasus. Pasal 293(2) tercatat dengan 320 kasus. Pasal 288(1) mencatatkan 210 kasus, Pasal 282 tercatat sebanyak 203 kasus. Pasal 287(2) memiliki 107 kasus, dan Pasal 291(2) tercatat dengan 144 kasus. Pasal 280 tercatat dengan 118 kasus, sementara Pasal 288(2) tercatat sebanyak 32 kasus. Pasal 288(3) mencatatkan 11 kasus, diikuti oleh Pasal 307 dengan 12 kasus. Pasal 285(2) memiliki 6 kasus, sementara Pasal 286 mencatat 5 kasus. Pasal 289 tercatat dengan 4 kasus, dan pelanggaran pada Pasal 278, Pasal 297, serta Pasal 300 masing-masing hanya memiliki 1 kasus.

Berdasarkan jenis kendaraan, pelanggaran pada Klaster 0 hampir seluruhnya dilakukan oleh Sepeda Motor (SM) yang mencapai 3.945 kasus, di ikuti oleh Mobil Pick up (MP) dengan jumlah 124 kasus, Truk berjumlah 26 kasus, untuk Mobil Penumpang Umum (MPU) dan bus menunjukkan 1 kasus. Kemudian dengan barang bukti dalam pelanggaran di klaster 0 adalah STNK dengan jumlah 3.478 kasus, diikuti oleh SIM C dengan 584 kasus. Jenis barang bukti lainnya memiliki frekuensi yang jauh lebih rendah, SIM A dengan jumlah 39 kasus, SIM BI Umum sebanyak 6 kasus, SIM BII

Umum sebanyak 4 kasus, sedangkan kendaraan dan SIM BI sebanyak 3 kasus.

Dari sisi denda, sebagian besar pelanggar di Klaster 0 dikenakan denda sebesar 99.000, dengan 2.194 kasus, diikuti oleh denda sebesar 49.000 sebanyak 1.459 kasus. Nominal denda lainnya, seperti 74.000, memiliki frekuensi yang jauh lebih kecil sebanyak 448 kasus, denda 249.000 sebanyak 8 kasus, denda 149.000 sebanyak 6 kasus, sedangkan denda 19.000 dan 101.000 sebanyak 1 kasus. Adapun form merah (Me) yang dikenakan kepada si pelanggar sebanyak 3724 Kasus, sedangkan untuk form Biru (Bi) sebanyak 393 kasus.

Klaster ini menunjukkan bahwa pola pelanggaran di klaster 0 mayoritas melibatkan kendaraan roda dua dengan jenis pelanggaran yang berkaitan dengan pasal 287(1) tentang melanggar rambu lalu lintas dengan barang buktinya adalah STNK dan SIM C, serta didominasi oleh denda paling sering dikenakan kisaran nominal 49.000 sampai dengan 99.000.



Gambar 4.14 Grafik Hasil *Cluster 0* Pada Atribut (Form, Pasal, Jenis Kendaraan, Barang Bukti, Dan Denda)

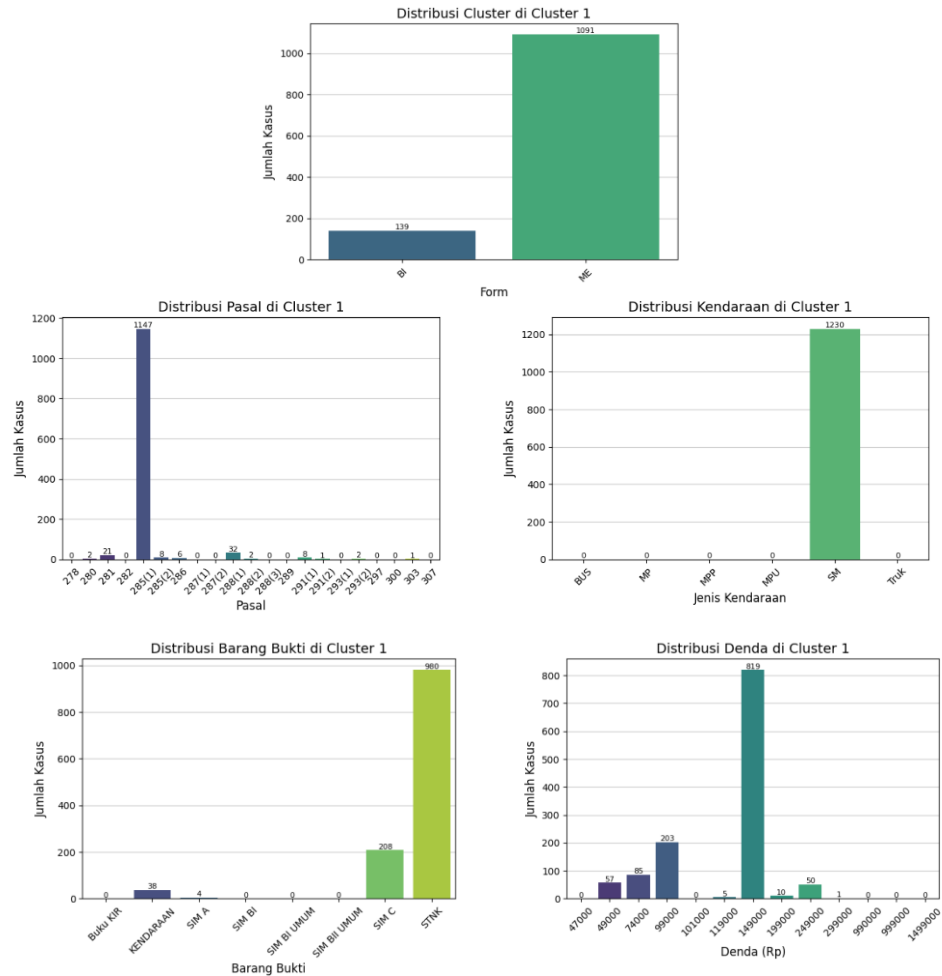
2. Klaster 1

Pada klaster 1 terdapat 1230 kasus pelanggaran. Berdasarkan analisis data pada pasal; Pasal 285(1) menjadi pelanggaran yang paling dominan dengan 1.147 kasus, kemudian diikuti oleh Pasal 281 dengan 21 kasus dan Pasal 288(1) sebanyak 32 kasus. Selanjutnya, Pasal 285(2) dan Pasal 291(1) masing-masing mencatatkan 8 kasus. Pasal 288(2), Pasal 293(2), dan Pasal 280 masing-masing memiliki 2 kasus, sedangkan Pasal 303 dan Pasal 291(2) mencatatkan 1 kasus. Adapun pasal-pasal lainnya, seperti Pasal 278, Pasal

282, Pasal 287(1), Pasal 287(2), Pasal 288(3), Pasal 289, Pasal 293(1), Pasal 297, Pasal 300, dan Pasal 307, tidak memiliki kasus pada klaster ini. Terkait dengan jenis kendaraan, mayoritas pelanggaran terjadi pada Sepeda Motor (SM) dengan jumlah pelanggaran mencapai 1.230 kasus. Sementara itu, kendaraan lainnya dengan roda empat atau lebih tidak memiliki kasus.

Barang bukti STNK tercatat sebagai barang bukti yang paling sering ditemukan, dengan total 980 kasus, diikuti oleh SIM C sebanyak 208 kasus dan kendaraan sebanyak 38 kasus, SIM A sebanyak 4 kasus. Dalam hal denda, sebagian besar pelanggaran dikenakan denda sebesar 149.000 dengan 819 kasus, disusul dengan denda 99.000 dengan 203 kasus dan 74.000 dengan 85 kasus. Sedangkan denda 49.000 sebanyak 57 kasus diikuti oleh denda 249.000 sebanyak 50 kasus. Adapun denda dengan jumlah kasus lebih kecil diantaranya denda 119.000 sebanyak 5 kasus, denda 199.000 sebanyak 10 kasus, dan denda 249.000 sebanyak 1 kasus. Adapun form yang sering dikenakan kepada pelanggar yaitu form Merah (Me) sebanyak 1091 kasus dan form Biru (Bi) sebanyak 139 kasus.

Klaster ini menunjukkan bahwa pengendara sepeda motor merupakan kelompok yang paling sering melanggar peraturan lalu lintas. Sebagian besar pelanggaran terlibat dalam pasal 285(1) tentang pengemudi sepeda motor tidak memenuhi persyaratan yang teknis dan baik di jalan. Dengan barang bukti yang sering ditemui adalah STNK dan SIM C, dengan denda paling sering ialah 149.000.



Gambar 4.15 Grafik Hasil *Cluster 1* Pada Atribut (Form, Pasal, Jenis Kendaraan, Barang Bukti, Dan Denda)

3. Klaster 2

Pada klaster 2 terdapat 1337 kasus pelanggaran. Berdasarkan analisis data pada pasal; Pasal 287(1) menjadi pelanggaran yang paling dominan dengan 848 kasus, kemudian diikuti oleh Pasal 281 dengan 79 kasus, serta Pasal 286 dan Pasal 287(2) yang masing-masing mencatatkan 77 kasus. Pasal 285(2) menyusul dengan 69 kasus, diikuti oleh Pasal 307 sebanyak 52 kasus dan Pasal 288(1) sebanyak 47 kasus. Pasal 288(3) tercatat dengan 45 kasus, sementara Pasal 280, Pasal 289, dan Pasal 300 masing-masing memiliki 10 kasus. Selanjutnya, Pasal 282 mencatat 7 kasus, diikuti oleh

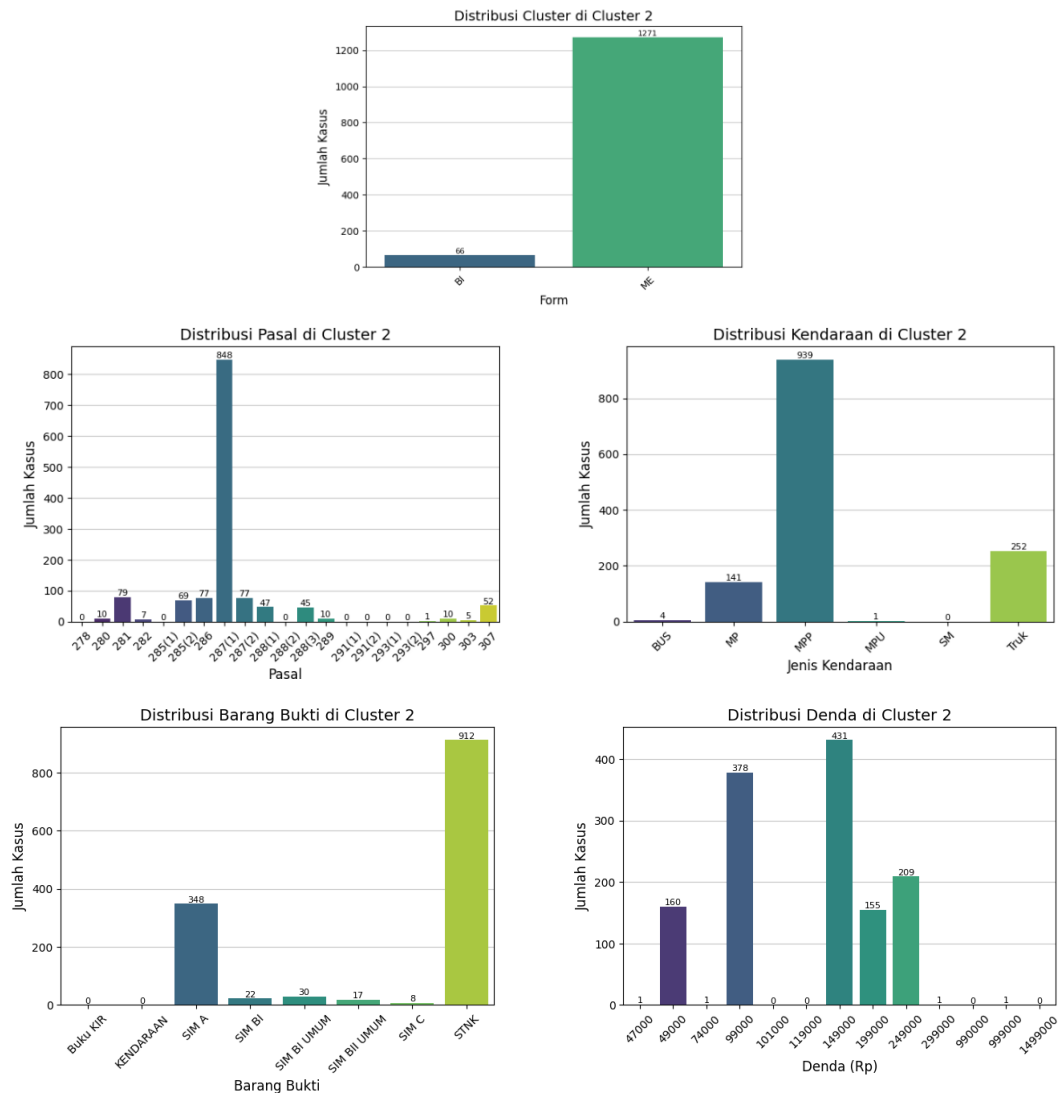
Pasal 303 dengan 5 kasus, dan Pasal 297 yang memiliki 1 kasus. Adapun pasal-pasal lainnya, seperti Pasal 278, Pasal 285(1), Pasal 288(2), Pasal 291(1), Pasal 291(2), Pasal 293(1), dan Pasal 293(2), tidak ditemukan pelanggaran pada klaster ini.

pelanggaran paling banyak terjadi pada kendaraan jenis kendaraan roda empat atau lebih seperti pada Mobil Penumpang Pribadi (MPP) sebanyak 939 kasus, diikuti oleh Truk 252 kasus, Mobil Pick up (MP) sebanyak 141 kasus, Bus 4 kasus dan Mobil Penumpang Umum (MPU) 1 kasus. Dalam hal barang bukti, STNK menjadi barang bukti yang paling sering dijumpai dengan 912 kasus, diikuti oleh SIM A dengan 348 kasus, sedangkan SIM BI sebanyak 22 kasus, SIM BI Umum sebanyak 30 kasus, SIM BII Umum sebanyak 17 kasus, dan SIM C sebanyak 8 kasus.

Terkait denda, nominal sebesar 149.000 diterapkan pada 431 kasus, menjadikannya denda yang paling sering dikenakan, diikuti oleh denda sebesar 99.000 dengan 378 kasus. Kemudian denda dengan nominal 249.000 sebanyak 209 kasus, denda 49.000 sebanyak 160 kasus, diikuti denda 199.000 sebanyak 155 kasus. Adapun denda lainnya, seperti 47.000, 74.000, 299.000 dan 990.000 yang masing-masing memiliki 1 kasus. Dan untuk form Merah (Me) sebanyak 1271 kasus dan untuk form Biru (Bi) sebanyak 66 kasus.

Dari klaster ini, terlihat bahwa pasal 287(1) sering dilanggar oleh pelanggar yaitu dengan melanggar rambu lalu lintas. Adapun kendaraan yang sering dijumpai ialah kendaraan dengan roda empat atau lebih, dan

barang bukti utama berupa STNK dan SIM A dengan denda aling sering dikenakan berada di 149.000 dengan form merah.



Gambar 4.16 Grafik Hasil *Cluster 2* Pada Atribut (Form, Pasal, Jenis Kendaraan, Barang Bukti, Dan Denda)

4. Klaster 3

Pada klaster 3 kelompok dengan jumlah pelanggaran terkecil di antara seluruh klaster yaitu dengan 510 kasus pelanggaran. Berdasarkan analisis data pasal; Pasal 286 menjadi pelanggaran yang paling dominan dengan 329 kasus, diikuti oleh Pasal 307 sebanyak 60 kasus, serta Pasal 288(1) dan Pasal 288(3) yang masing-masing mencatatkan 40 dan 33 kasus. Pasal 285(2)

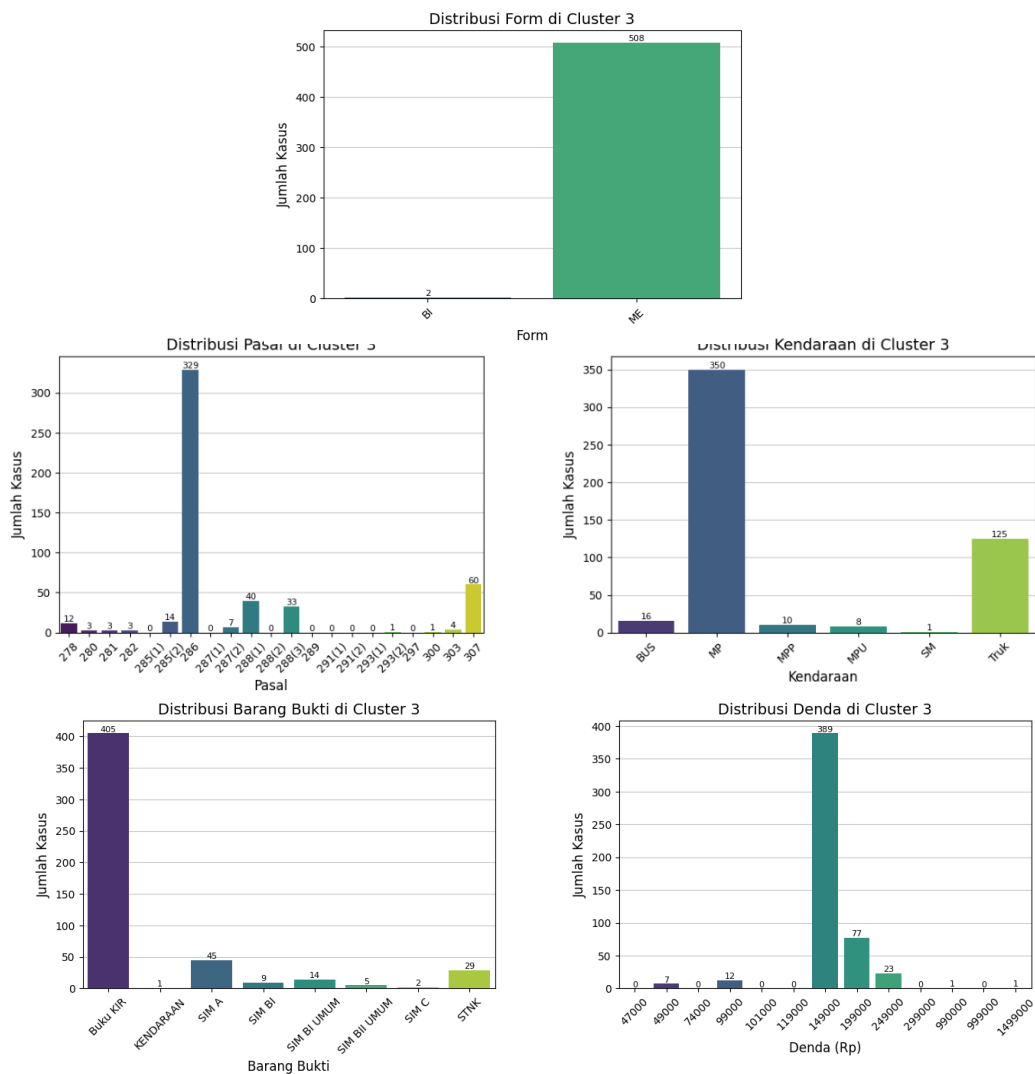
dengan 14 kasus, diikuti oleh Pasal 278 dengan 12 kasus dan Pasal 287(2) sebanyak 7 kasus. Pasal 303 mencatatkan 4 kasus, sedangkan Pasal 280, Pasal 281, dan Pasal 282 masing-masing memiliki 3 kasus. Selanjutnya, Pasal 293(2) dan Pasal 300 masing-masing mencatatkan 1 kasus. Adapun pasal-pasal lainnya, seperti Pasal 285(1), Pasal 287(1), Pasal 288(2), Pasal 289, Pasal 291(1), Pasal 291(2), Pasal 293(1), dan Pasal 297, tidak memiliki pelanggaran pada klaster ini.

Dari sisi jenis kendaraan, klaster ini didominasi oleh kendaraan roda empat atau lebih, dengan kendaraan Mobil Pick up (MP) mencatat 350 pelanggaran, diikuti oleh Truk sebanyak 125 kasus dan Bus sebanyak 16 kasus. Mobil Penumpang Pribadi (MPP) sebanyak 10 kasus dan Mobil Penumpang Umum (MPU) sebanyak 8 kasus. Hanya ada 1 pelanggaran yang melibatkan Sepeda Motor (SM).

Barang bukti yang paling banyak ditemukan adalah Buku KIR, dengan 405 kasus. Selain itu, barang bukti lain seperti SIM A dengan 45 kasus, SIM BI UMUM sebanyak 14 kasus, dan STNK 29 kasus. Adapun barang bukti, SIM BI terdapat 9 kasus, SIM BII UMUM sebanyak 5 kasus, SIM C sebanyak 2 kasus dan Kendaraan sebanyak 1 kasus. Dari sisi denda, sebagian besar pelanggaran dikenakan denda sebesar Rp 149.000, dengan total 389 kasus, diikuti oleh denda Rp 199.000 sebanyak 77 kasus, serta 249.000 sebanyak 23 kasus. Kemudian denda 99.000 sebanyak 12 kasus, denda 49.000 sebanyak 7 kasus. Ada pula beberapa pelanggaran dengan denda mencapai Rp 990.000 dan Rp 1.499.000, masing-masing tercatat sebanyak satu kasus. Adapun form yang diberikan kepada pelanggar seperti

pada form Merah (Me) sebanyak 508 kasus sedangkan form Biru (Bi) sebanyak 2 kasus.

Dari klaster ini, terlihat bahwa pelanggaran pada klaster 3 lebih sering terjadi pada kendaraan roda empat atau lebih, terutama yang terkait dengan pelanggaran pasal 286 yaitu berkaitan dengan kesesuaian daya mesin terhadap berat kendaraan. Dengan barang bukti paling banyak terdapat pada Buku KIR dan denda sering dikenakan ialah 149.000 dengan denda tertinggi 1.499.000 dan form kebanyakan pada form merah.



Gambar 4.17 Grafik Hasil *Cluster 3* Pada Atribut (Form, Pasal, Jenis Kendaraan, Barang Bukti, Dan Denda)

4.7 Kajian Keislaman Dengan Hasil Penelitian

Dalam Islam, ketaatan terhadap aturan dan hukum yang berlaku, termasuk hukum lalu lintas, merupakan bagian dari perwujudan *mashlahah* (kepentingan) dan *taat terhadap ulil amri* (pemimpin yang sah). Integrasi keislaman dalam analisis klaster pelanggaran lalu lintas dapat dilakukan dengan mengaitkan prinsip-prinsip Islam dalam kehidupan bermasyarakat dan pentingnya menaati aturan sebagai bagian dari ibadah. Islam mendorong umatnya untuk menjalankan kehidupan yang tertib, termasuk dalam hal berlalu lintas, sebagai wujud penghormatan terhadap hak orang lain dan tanggung jawab sosial. Hasil analisis klaster berhasil mengungkap pola pelanggaran yang sering terjadi di setiap klasternya. Setiap klaster memberikan gambaran tentang jenis pelanggaran yang paling sering dilakukan oleh masyarakat Kota Malang. Berikut adalah pola pelanggaran yang ditemukan pada masing-masing klaster:

1. Pelanggaran Rambu Lalu Lintas (Pasal 287(1))

Pelanggaran terhadap rambu lalu lintas, yang menjadi pelanggaran dominan dilanggar baik oleh roda dua maupun roda empat/lebih yang terdapat pada klaster 0 dan klaster 2, hal ini mencerminkan kurangnya kesadaran akan pentingnya menaati aturan yang telah ditetapkan untuk keteraturan dan keselamatan bersama. Dalam Islam, menaati aturan yang ditetapkan oleh pemerintah adalah bagian dari prinsip taat kepada *ulil amri* (pemimpin), sebagaimana disebutkan dalam Al-Qur'an (Kemenag, 2024):

"Hai orang-orang yang beriman, taatilah Allah dan taatilah Rasul (Nya), dan ulil amri di antara kamu..." (QS. An-Nisa: 59).

Rambu lalu lintas adalah bagian dari aturan yang diciptakan oleh pemerintah untuk mengatur kehidupan bermasyarakat, menjaga keselamatan pengguna jalan, dan mencegah terjadinya kecelakaan yang dapat merugikan banyak pihak. Melanggar rambu lalu lintas bukan hanya menunjukkan ketidakpatuhan terhadap hukum negara, tetapi juga merupakan bentuk pelanggaran terhadap amanah yang diberikan oleh pemimpin untuk menjaga keteraturan. Dalam Islam, ketertiban dalam berlalu lintas adalah bagian dari tanggung jawab sebagai seorang Muslim untuk menjaga keselamatan diri dan orang lain, yang merupakan salah satu bentuk ibadah. Rasulullah SAW bersabda:

"Setiap Muslim terhadap Muslim lainnya haram darahnya, hartanya, dan kehormatannya." (Hadist Riwayat Muslim No. 2564).

Pada Hadist tersebut haram darahnya berarti tidak boleh dibunuh atau ditumpahkan darahnya. Haram hartanya maksudnya adalah harta seorang muslim tidak boleh diambil. Sedangkan maksud dari haram kehormatannya adalah seorang muslim tidak boleh dituduh dengan tuduhan yang bisa merusak kehormatannya. Pelanggaran rambu lalu lintas yang mengakibatkan kecelakaan dapat berdampak pada kerugian fisik, materi, hingga jiwa, yang bertentangan dengan prinsip Islam untuk menjaga nyawa dan harta benda. Lebih dari itu, tidak mematuhi rambu lalu lintas dapat mencerminkan sikap abai terhadap hak-hak orang lain yang juga menggunakan jalan raya. Maka, menaati aturan lalu lintas, termasuk rambu-rambu, bukan hanya kewajiban sosial, tetapi juga bagian dari implementasi

nilai-nilai Islam dalam kehidupan sehari-hari, sebagai wujud ketaatan kepada Allah, Rasul-Nya, dan pemimpin yang sah.

2. Pengemudi sepeda motor tidak memenuhi persyaratan teknis dan baik di jalan (Pasal 285(1))

Pelanggaran ini sering terjadi pada klaster 1 dengan jenis kendaraan sepeda motor. Pelanggaran ini terjadi ketika pengemudi sepeda motor tidak memastikan kendaraannya memenuhi persyaratan teknis dan layak jalan, seperti kelengkapan lampu, klakson, sistem pengereman, dan kondisi ban. Mengabaikan aspek-aspek tersebut tidak hanya membahayakan pengemudi itu sendiri, tetapi juga dapat membahayakan pengguna jalan lainnya. Kendaraan yang tidak layak jalan meningkatkan risiko kecelakaan dan berpotensi merugikan orang lain. Dalam perspektif Islam, menjaga keselamatan diri sendiri dan orang lain merupakan kewajiban yang tidak boleh diabaikan. Rasulullah SAW bersabda:

"Tidak boleh membahayakan diri sendiri maupun orang lain"
(Hadist Riwayat Ibnu Majah No. 2340 dan 2341).

Hadist ini mengajarkan untuk menjaga diri sendiri maupun orang lain oleh karena itu untuk memastikan kendaraan dalam kondisi baik adalah bentuk tanggung jawab seorang Muslim terhadap amanah yang diberikan. Kendaraan yang digunakan merupakan fasilitas yang harus dirawat agar tidak menimbulkan mudarat (merugikan) bagi diri sendiri maupun masyarakat. Islam juga mengajarkan pentingnya menghormati hak orang

lain, termasuk dalam penggunaan jalan raya, sehingga menjaga keselamatan bersama menjadi bagian dari amal saleh yang bernilai ibadah.

3. Tidak kesesuaian daya mesin penggerak terhadap berat kendaraan (Pasal 286)

Pelanggaran ini sering terjadi pada klaster 3, terjadi terutama pada kendaraan roda empat atau lebih. Pelanggaran ini terjadi ketika kendaraan tidak memiliki daya mesin yang sesuai dengan beban yang dibawa, yang dapat mengurangi performa kendaraan dan berisiko menyebabkan kecelakaan. Dalam banyak kasus, kendaraan yang tidak sesuai spesifikasi daya mesin dengan berat beban cenderung mengalami penurunan efisiensi, masalah mekanis, dan bahkan kegagalan teknis yang dapat berujung pada kecelakaan. Hal ini tentunya membahayakan pengemudi, penumpang, dan pengguna jalan lainnya.

Dalam pandangan Islam, setiap tindakan yang dilakukan harus memenuhi prinsip kehati-hatian dan tanggung jawab, termasuk dalam hal berkendara. Islam mengajarkan pentingnya menjaga keselamatan diri dan orang lain, sebagaimana dalam Hadist Nabi Muhammad SAW yang mengatakan.

“Tidaklah seseorang itu dianggap beriman hingga ia mencintai untuk saudaranya apa yang ia cintai untuk dirinya sendiri.” (Hadist Riwayat Bukhari No.13).

Hadits ini mengajarkan untuk peduli terhadap keselamatan orang lain sebagaimana seseorang peduli terhadap keselamatan dirinya sendiri. Dalam berkendara, ini berarti pengemudi harus memastikan kendaraannya dalam kondisi yang layak dan aman, tidak hanya demi keselamatan pribadi, tetapi

juga demi keselamatan orang lain di jalan. Jika daya mesin kendaraan tidak sesuai dengan berat beban, hal itu dapat meningkatkan risiko kecelakaan yang dapat merugikan orang lain, yang tentu saja bertentangan dengan ajaran Islam untuk tidak membahayakan sesama.

Dengan demikian, analisis klaster ini dapat menjadi bahan refleksi bahwa kepatuhan terhadap aturan lalu lintas tidak hanya berdampak pada keselamatan fisik tetapi juga merupakan implementasi dari ajaran Islam untuk hidup bertanggung jawab, menghormati sesama, dan menjaga keselamatan sebagai bentuk ibadah. Hal ini juga menjadi pengingat pentingnya dakwah tentang kesadaran berlalu lintas yang berlandaskan nilai-nilai agama.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan Hasil dari penelitian data pelanggaran lalu lintas di Kejaksaan Negeri Kota Malang telah dilakukan beberapa kesimpulan sebagai berikut:

1. Dari data yang telah diproses menggunakan algoritma *K-Prototype* dengan menggunakan *library* Python berdasarkan form, jenis pelanggaran, jenis kendaraan, denda, dan barang bukti didapatkan hasil *clustering* sebanyak empat klaster. Hasil *clustering* ini menunjukkan kemampuan algoritma *K-Prototype* dalam mengelompokkan data pelanggaran lalu lintas di Kota Malang.
2. Berdasarkan data pelanggaran lalu lintas di Kota Malang telah diterapkannya dengan Algoritma *K-Prototype*. Dimana dapat diketahui pola dan karakteristik pelanggaran sebagai berikut:
 - a. Klaster 0 (4.117 kasus pelanggaran): Mayoritas pelanggaran oleh Sepeda motor, terkait pasal 287(1) (melanggar rambu lalu lintas). Barang bukti: STNK, SIM C. Denda dominan: Rp49.000–Rp99.000.
 - b. Klaster 1 (1.230 kasus pelanggaran): Pelanggaran dominan oleh sepeda motor, pasal 285(1) (tidak memenuhi persyaratan teknis). Barang bukti: STNK, SIM C. Denda sering: Rp149.000.
 - c. Klaster 2 (1.337 kasus pelanggaran): Kendaraan roda empat atau lebih, melanggar pasal 287(1) (melanggar rambu lalu lintas). Barang bukti: STNK, SIM A. Denda dominan: Rp149.000.

- d. Klaster 3 (510 kasus pelanggaran): Kendaraan roda empat atau lebih, pelanggaran pasal 286 (daya mesin tidak sesuai). Barang bukti: Buku KIR. Denda dominan: Rp149.000.

5.2 Saran

Saran yang dapat diberikan oleh penulis untuk penelitian selanjutnya sebagai berikut:

1. Pada penelitian ini penulis menggunakan lima atribut yaitu empat atribut kategorik dan satu atribut numerik yang diharapkan pada penelitian selanjutnya atribut yang digunakan lebih banyak dibandingkan dengan atribut yang penulis gunakan pada penelitian ini.
2. Penelitian ini hanya menggunakan data pelanggaran lalu lintas di Kota Malang dalam periode tertentu. Pada penelitian selanjutnya, disarankan untuk memperluas cakupan data, dari segi lokasi (misalnya mencakup seluruh wilayah Jawa Timur).

DAFTAR PUSTAKA

- 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.
- Ayu, D., Dewi, I. C., & Pramita, K. (2019). Analisis Perbandingan Metode Elbow dan Sillhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *JURNAL MATRIX*, 9(3), 102.
- 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. *Jurnal Sistem Informasi Dan Teknologi Informasi*, 8(1).
- Ferbyansyah, R. (2023). *Analisis Klasterisasi Pelanggaran Lalu Lintas Menggunakan Algoritma K-Prototype Traffic Violation Clustering Analysis of using Algorithm K-Prototype*.
- Riska, F. (2023). *Hadist Riwayat Tirmidzi No. 1940*.
- Laila. (2021). *Hadits Riwayat Bukhari No.13*.
- Ramala, H. (2023). *Hadits Riwayat Ibnu Majah dan Ad-Daruqutni No. 3079*.
- Ramala, H. (2023). *Hadits Riwayat Ibnu Majah No. 2340 dan 2341*.
- Hairul, H. (2020). *Hadits Riwayat Muslim No. 2564*.
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.
- Irawan, V., Rizal, A., & Purnamasari, I. (2020). Penerapan Algoritma K-Mean Clustering Pada Kecelakaan Lalu Lintas Berdasarkan Kabupaten/Kota Di Provinsi Jawa Tengah Tahun 2020. *Jurnal Ilmiah Wahana Pendidikan*, 8(10), 293–300. <https://doi.org/10.5281/zenodo.6820090>
- Kemenag. (2024). *Al-qur'an*. <https://quran.nu.or.id/an-nisa%27/59>.
- Kim, B. (2017). A fast K-prototypes algorithm using partial distance computation. *Symmetry*, 9(4). <https://doi.org/10.3390/sym9040058>
- Maori, N. A. (2023). Metode Elbow Dalam Optimasi Jumlah Cluster Pada K-Means Clustering. *Jurnal SIMETRIS*, 14.
- Mayangsari, Y. (2022). *Analisis K-Means pada Pengelompokan Kabupaten-Kota Provinsi Jawa Timur Berdasarkan Kasus Kesembuhan dan Kasus Kematian COVID-19*.

- Nooraeni, R., & Supriadi, J. (2019). K-Prototype untuk Pengelompokan Data Campuran. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 13, 9–16.
- Nugroho, B. (2019). Ketaatan Berlalu Lintas Sebagai Penerapan Nilai-Nilai Islam. *Jurnal Tarjih dan Pengembangan Pemikiran Islam*, 16(1). <https://jurnal.tarjih.or.id/index.php/tarjih/article/view/16.107>
- Pradhana, O. J. (2022). *Penerapan Clustering Pada Tipe Data Campuran Menggunakan K-Prototype Pada Perusahaan Multifinance*. 9(Juni).
- Pribadi, W. W., Yunus, A., & Sartika Wiguna, A. (2022). Perbandingan Metode K-Means Euclidean Distance Dan Manhattan Distance Pada Penentuan Zonasi COVID-19 Di Kabupaten Malang. *Jurnal Mahasiswa Teknik Informatika*, 6(2).
- Ria, D. (2018). Analisa Penerapan *Data mining* Pada Penerimaan Mahasiswa Politeknik Negeri Lhokseumawe Menggunakan Algoritma K-Means. *Jaringan Cerdas Multimedia*, 208–209.
- Wahyuni, D., & Arianto, A. (2019). Analisa Clustering Pada Data Pelanggaran Lalu Lintas Di Pengadilan Negeri Dumai Dengan Menggunakan Metode K-Means. *JISKa*, 4(2), 76–87.

LAMPIRAN

Lampiran 1 Data Pelanggaran Lalu Lintas Di kota Malang Tahun 2021

https://drive.google.com/file/d/1AuAMbwCVysYWA1vOm_zcO05QuS8xIWJf/view?usp=sharing



Lampiran 2 Data Setelah Dilakukan Transformasi Data dan Preprocessig Data

<https://docs.google.com/spreadsheets/d/1lTUh5YMDrVwUD9W7JL7dpDSVGYDiuRax/edit?usp=sharing&ouid=104738692458319655037&rtpof=true&sd=true>



Lampiran 3 Hasil *Clustering K-prototype*

<https://docs.google.com/spreadsheets/d/1XXyv1WB6HKP6d8-bjNbbQ4ytXDhaA3J6/edit?usp=sharing&ouid=104738692458319655037&rtpof=true&sd=true>



Lampiran 4 *Source Code* Algoritma *K-Prototype*

```
!pip install kmodes
import numpy as np
import pandas as pd
from kmodes.kprototypes import KPrototypes
from sklearn.preprocessing import MinMaxScaler
import matplotlib.pyplot as plt

from google.colab import drive
drive.mount('/content/drive')
pelanggaran = pd.read_csv('/content/drive/MyDrive/DATA PELANGGARAN/Lampiran/data-mentah.csv')

pelanggaran = pelanggaran.drop(columns=['NAMA', 'NOMOR POLISI'])

jenis_kendaraan_mapping = {
    'Sepeda Motor': 'SM',
    'Mobil Pick Up': 'MP',
    'Mobil Penumpang Pribadi': 'MPP',
    'Mobil Penumpang Umum': 'MPU',
    'Bus': 'BUS',
    'Truck': 'TRUK'
}
```

```

FORM_mapping = {
    'Merah': 'Me',
    'BIRU': 'Bi'
}

# Ubah kolom 'PASAL' menjadi string dan hilangkan spasi tambahan
pelanggaran['PASAL'] = pelanggaran['PASAL'].astype(str).str.strip()

# Buat mapping nilai pasal ke nomor urut
pasal_mapping = {
    '278': 'A', # Changed to string literals
    '280': 'B', # Changed to string literals
    '281(1)': 'C', # Changed to string literals
    '282': 'D', # Changed to string literals
    '285(1)': 'E', # Changed to string literals
    '285(2)': 'F', # Changed to string literals
    '286': 'G', # Changed to string literals
    '287(1)': 'H', # Changed to string literals
    '287(2)': 'I', # Changed to string literals
    '288(1)': 'J', # Changed to string literals
    '288(2)': 'K', # Changed to string literals
    '288(3)': 'L', # Changed to string literals
    '289': 'M', # Changed to string literals
    '291(1)': 'N', # Changed to string literals
    '291(2)': 'O', # Changed to string literals
    '293(1)': 'P', # Changed to string literals
    '293(2)': 'Q', # Changed to string literals
    '297': 'R', # Changed to string literals
    '300': 'S', # Changed to string literals
    '303': 'T', # Changed to string literals
    '307': 'U' # Changed to string literals
}

pelanggaran['PASAL'] = pelanggaran['PASAL'].map(pasal_mapping)
pelanggaran['FORM'] = pelanggaran['FORM'].map(FORM_mapping)
pelanggaran['JENIS KENDARAAN'] = pelanggaran['JENIS KENDARAAN'].map(jenis_kendaraan_mapping)
# Menampilkan 5 baris pertama
print(pelanggaran.head())
# Ambil kolom DENDA untuk normalisasi
denda_column = pelanggaran[['DENDA']]

# Ambil kolom DENDA untuk normalisasi
denda_column = pelanggaran[['DENDA']]
# Inisialisasi MinMaxScaler dan lakukan normalisasi
scaler = MinMaxScaler()
normalized_denda = scaler.fit_transform(denda_column)
# Simpan hasil normalisasi kembali ke kolom DENDA di dataframe
pelanggaran['DENDA'] = normalized_denda
print(pelanggaran)

pelanggaran.isnull().sum()
pelanggaran = pelanggaran.dropna()
pelanggaran.isnull().sum()

mark_array=pelanggaran.values

```

```

mark_array[:, 4] = mark_array[:, 4].astype(float)
mark_array

cost = []
K_range = range(2, 10)
for K in K_range:
    kproto = KPrototypes(n_clusters=K,)
    clusters = kproto.fit_predict(mark_array, categorical=[0, 1,
2, 3])
    cost.append(kproto.cost_)
    print(f'K = {K}, Cost = {kproto.cost_}')

kproto = KPrototypes(n_clusters=4, verbose=2,max_iter=20)
clusters = kproto.fit_predict(mark_array, categorical=[0, 1, 2, 3])

cluster_dict=[]
for c in clusters:
    cluster_dict.append(c)
cluster_dict

pelanggaran['cluster']=cluster_dict
pelanggaran

cost = []
K_range = range(2, 10) # Coba beberapa nilai K
for K in K_range:
    kproto = KPrototypes(n_clusters=K)
    kproto.fit_predict(mark_array, categorical=[0, 1, 2, 3])
    cost.append(kproto.cost_)
# Plot Elbow Curve
plt.figure(figsize=(8, 5))
plt.plot(K_range, cost, marker='o')
plt.title('Elbow Method for Optimal K')
plt.xlabel('Number of Clusters (K)')
plt.ylabel('Cost (WCSS)')

for i, txt in enumerate(cost):
    plt.annotate(f'{txt:.2f}', (K_range[i], cost[i]),
textcoords="offset points", xytext=(0,5), ha='center')

plt.show()

```

RIWAYAT HIDUP



Ivan Bintang Sabrani, lahir di Malang pada 1 Januari 2002. Penulis merupakan anak kedua dari tiga bersaudara, dengan seorang kakak laki-laki dan seorang adik laki-laki, dari pasangan Alm. Bambang Budi Wahono dan Ibu Widiyawati. Masa kecil penulis dihabiskan di lingkungan yang hangat bersama keluarga yang penuh kasih sayang, meskipun harus menghadapi kehilangan sang ayah di usia muda, yang menjadi salah satu pengalaman penting dalam membentuk kepribadian dan semangat hidup penulis. Pendidikan formal penulis dimulai di TK Kartika IV-1. Penulis kemudian melanjutkan pendidikan dasar di SDN Kesatrian 2 dan berhasil menyelesaikan pendidikan tersebut pada tahun 2014. Kemudian penulis melanjutkan pendidikan menengah pertama di SMPN 10 Malang, di mana penulis berhasil lulus pada tahun 2017 dengan prestasi yang membanggakan. Pada jenjang menengah atas, penulis melanjutkan pendidikan di SMAN 6 Malang, dan menyelesaikan pendidikan menengah atas pada tahun 2020, penulis memutuskan untuk melanjutkan pendidikan tinggi. Pada tahun 2021, penulis diterima di Universitas Islam Negeri Maulana Malik Ibrahim Malang, memilih Program Studi Matematika di Fakultas Sains dan Teknologi. Selama masa perkuliahan, penulis menunjukkan semangat yang besar dalam mengeksplorasi dunia akademik dan non-akademik. Penulis aktif mengikuti berbagai seminar, pelatihan, dan kegiatan organisasi yang mendukung pengembangan wawasan dan keterampilan. Selain itu, penulis juga pernah berkontribusi dalam kepanitiaan acara-acara kampus, yang menjadi pengalaman berharga dalam mengelola kegiatan dan bekerja sama dalam tim.



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

BUKTI KONSULTASI SKRIPSI

Nama : Ivan Bintang Sabrani
NIM : 210601110076
Fakultas / Jurusan : Sains dan Teknologi / Matematika
Judul Skripsi : Implementasi Algoritma *K-Prototype* Pada Pengelompokan Data Pelanggaran Lalu Lintas Di Kota Malang
Pembimbing I : Hisyam Fahmi, M.Kom
Pembimbing II : Evawati Alisah, M.Pd.

No	Tanggal	Hal	Tanda Tangan
1.	20 Juni 2024	Konsultasi Topik dan Data	1.
2.	22 Juli 2024	Konsultasi Bab I, II, dan III	2.
3.	30 September 2024	Konsultasi Bab I, II, dan III	3.
4.	04 Oktober 2024	ACC Bab I, II, dan III	4.
5.	04 Oktober 2024	Konsultasi Kajian Agama Bab I dan II	5.
6.	07 Oktober 2024	ACC Kajian Agama Bab I dan II	6.
7.	08 Oktober 2024	ACC Seminar Proposal	7.
8.	28 Oktober 2024	Konsultasi Revisi Seminar Proposal	8.
9.	01 November 2024	Konsultasi Revisi Seminar Proposal	9.
10.	04 November 2024	Konsultasi Bab IV	10.
11.	05 November 2024	Konsultasi Bab IV	11.
12.	06 November 2024	Konsultasi Kajian Agama Bab IV	12.
13.	07 November 2024	Konsultasi Bab V	13.
14.	07 November 2024	ACC Kajian Agama Bab IV	14.
15.	08 November 2024	Konsultasi Bab IV dan V	15.



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

16.	08 November 2024	ACC Bab IV dan V	16.
17.	11 November 2024	ACC Seminar Hasil	17.
18.	28 November 2024	Konsultasi Revisi Seminar Hasil	18.
19.	05 Desember 2024	ACC Sidang Skripsi	19.
20.	20 Desember 2024	ACC Keseluruhan	20.

Malang, 20 Desember 2024

Mengetahui,

Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc.

NIP. 19741129 200012 2 005