

## Identifikasi Pola Kecelakaan Lalu Lintas Dengan K-Means Clustering

Victor Bandhaso<sup>1)</sup>; Masna Wati<sup>2)</sup>; Havaluddin<sup>3)</sup>

<sup>1) 2) 3)</sup>Program Studi Informatika, Universitas Mulawarman Samarinda

<sup>1)</sup>bandhasovictor@gmail.com; <sup>2)</sup>masnawati@fkti.unmul.ac.id; <sup>3)</sup>havaluddin@unmul.ac.id

### ABSTRACT

*Traffic accidents represent a complex issue with significant social and economic impacts. This study aims to identify temporal patterns of traffic accidents based on temporal and demographic attributes using the K-Means Clustering algorithm applied to 9,659 accident records in Central Java Province in 2024. Time attributes were converted to decimal format, while occupational data for the involved parties were transformed into numerical codes to enable clustering analysis. The K-Means Clustering algorithm was then employed to generate cluster models. Cluster 0 is characterized by an afternoon peak in incident time around 18.10, with the closest encoded occupational category corresponding to TNI–POLRI personnel. Cluster 1 consists of an average incident occurring at 06.26, predominantly involving homemakers. Cluster 2 is dominated by homemakers, with incidents generally occurring around 17.03. Cluster 3 shows the dominance of TNI–POLRI personnel, with incidents most frequently occurring at 07.19. These findings indicate that the most frequently involved occupational groups are military/police personnel and homemakers, both of which exhibit high mobility during peak hours and also threaten officers who are supposed to maintain traffic order.*

**Keywords:** Traffic accidents, K-Means, Temporal attributes, Profession, Central Java.

### I. PENDAHULUAN

Kecelakaan lalu lintas merupakan persoalan krusial dalam sistem transportasi yang berdampak signifikan terhadap aspek sosial dan ekonomi masyarakat [1].

Selain menimbulkan korban jiwa dan luka, kecelakaan juga menyebabkan kerugian materiil serta beban psikologis bagi korban dan keluarga [2]. Di Indonesia, angka kecelakaan masih menunjukkan tren yang memprihatinkan, dimana pada tahun 2024 terjadi peningkatan hampir 8 kali lipat dari tahun sebelumnya, mencapai 1.150.000 kejadian dengan 27.000 korban jiwa. Data Kepolisian Republik Indonesia menunjukkan sekitar 3-4 orang meninggal setiap jam akibat kecelakaan sebagai penyebab kematian tertinggi ketiga di Indonesia [3], [4], [5].

Kondisi tersebut menunjukkan bahwa permasalahan kecelakaan tidak hanya berkaitan dengan jumlah kejadian, tetapi juga dengan kompleksitas faktor penyebab yang saling berinteraksi [6], [7]. Salah satu aspek yang masih kurang mendapat perhatian adalah pola distribusi waktu kejadian kecelakaan secara rinci serta keterkaitannya dengan karakteristik sosial pengguna jalan. Padahal, pemahaman terhadap pola waktu yang spesifik dapat memberikan dasar yang lebih kuat dalam merancang strategi keselamatan yang bersifat preventif dan tepat sasaran [8], [9].

Salah satu daerah dengan tingkat kecelakaan lalu lintas yang cukup tinggi adalah Provinsi Jawa Tengah. Tingginya mobilitas penduduk dan peningkatan jumlah kendaraan belum sepenuhnya diimbangi dengan pengelolaan transportasi dan kesadaran berlalu lintas yang memadai [10]. Akibatnya, risiko kecelakaan terus meningkat dan berpotensi mengganggu kesejahteraan masyarakat secara luas.

Namun demikian, meskipun Jawa Tengah merupakan wilayah dengan tingkat risiko yang signifikan, kajian yang secara khusus menganalisis pola kecelakaan berbasis dimensi waktu yang presisi dan dikaitkan dengan karakteristik demografis masih terbatas. Sebagian besar analisis yang ada belum memanfaatkan data secara optimal untuk menggali pola tersembunyi yang dapat menjelaskan kapan dan pada kelompok sosial mana risiko kecelakaan paling tinggi terjadi.

Penelitian sebelumnya umumnya bersifat deskriptif dan terbatas pada klasifikasi kecelakaan berdasarkan lokasi, jenis kendaraan, atau kondisi cuaca [11]. Pendekatan tersebut belum sepenuhnya mampu mengungkap pola tersembunyi berdasarkan dimensi waktu dan karakteristik sosial seperti profesi, yang justru memiliki potensi besar dalam mengidentifikasi faktor pemicu kecelakaan secara lebih mendalam.

Selain itu, dalam banyak penelitian, diantaranya [12], [13], [14], variabel waktu masih direpresentasikan dalam bentuk kategori umum seperti pagi, siang, dan malam, sehingga mengurangi tingkat presisi analisis. Padahal, representasi waktu dalam bentuk numerik kontinu memungkinkan proses pengelompokan yang lebih akurat dalam mengidentifikasi pola distribusi kejadian kecelakaan. Keterbatasan ini menunjukkan adanya kesenjangan dalam pemanfaatan metode analisis data yang mampu mengolah dimensi temporal secara lebih rinci.

Penelitian ini bertujuan untuk mengidentifikasi pola distribusi kecelakaan lalu lintas di Provinsi Jawa Tengah berdasarkan dimensi temporal dan demografis menggunakan pendekatan K-Means Clustering. Variabel yang digunakan mencakup waktu kejadian yang dikonversi ke format numerik serta profesi pihak yang terlibat dalam kecelakaan. Melalui proses pengelompokan tersebut, diharapkan ditemukan kelompok waktu dan karakteristik sosial tertentu yang memiliki risiko kecelakaan lebih tinggi.

Dengan demikian, penelitian ini tidak hanya berfokus pada pengelompokan data, tetapi juga menempatkan dimensi waktu sebagai variabel utama yang dianalisis secara kuantitatif. Pendekatan ini diharapkan mampu mengisi kesenjangan penelitian sebelumnya dengan memberikan pemahaman yang lebih mendalam mengenai struktur pola kecelakaan berbasis waktu dan karakteristik sosial.

Hasil penelitian ini diharapkan dapat memberikan informasi yang relevan untuk mendukung perumusan kebijakan keselamatan lalu lintas berbasis data. Temuan tersebut dapat dimanfaatkan dalam penguatan pengawasan pada waktu rawan, pemasangan rambu peringatan, serta kampanye keselamatan yang menasar kelompok profesi berisiko tinggi.

## II. TINJAUAN PUSTAKA

Penelitian yang mengkaji pola pada data kecelakaan lalu lintas telah banyak dilakukan dalam berbagai konteks dan pendekatan analisis, termasuk *clustering* [6]. Secara umum, algoritma *clustering* digunakan untuk mengidentifikasi pola tersembunyi dalam data, khususnya dalam hal distribusi waktu kejadian kecelakaan, guna mendukung pengambilan kebijakan yang lebih tepat sasaran.

Salah satu studi yang relevan meneliti distribusi temporal kecelakaan di wilayah Kota

Semarang [15]. Penelitian tersebut menunjukkan bahwa konversi waktu ke format desimal mampu menghasilkan tiga kluster utama (pagi, sore, malam) yang berguna dalam penyusunan kebijakan operasional kendaraan berat. Meskipun demikian, penelitian tersebut berhenti pada level segmentasi umum, pendekatan numerik yang digunakan belum dimanfaatkan untuk menggali pola waktu yang lebih halus. Penelitian ini mengambil posisi untuk menganalisis distribusi waktu secara lebih detail berbasis centroid numerik.

Penelitian serupa menganalisis data kecelakaan di Kota Surabaya dengan menerapkan metode elbow dalam menentukan jumlah kluster optimal. Studi ini mengungkap bahwa insiden kecelakaan paling banyak terjadi pada rentang waktu pukul 06.00–08.00 dan 16.00–18.00 [16]. Temuan tersebut dijadikan dasar dalam penjadwalan ulang waktu kerja petugas lalu lintas serta penempatan infrastruktur pengatur arus kendaraan pada jam-jam sibuk. Namun, fokus utama studi tersebut terletak pada identifikasi *peak hours*, bukan pada struktur distribusi kluster itu sendiri. Dengan demikian, analisis yang dihasilkan cenderung deskriptif terhadap frekuensi, bukan eksploratif terhadap pola pembentukan kluster. Berbeda dengan hal tersebut, penelitian ini menitikberatkan pada interpretasi posisi centroid dan karakteristik masing-masing kluster sebagai representasi pola data.

Dalam konteks jalan tol Trans Jawa 16 penggunaan format waktu desimal menghasilkan empat kluster dengan identifikasi jam puncak kecelakaan. Walaupun sudah menggunakan representasi numerik, penelitian tersebut masih berorientasi pada identifikasi waktu puncak, bukan pada analisis struktur distribusi waktu secara menyeluruh. Penelitian ini memperluas pendekatan tersebut dengan tidak hanya mengidentifikasi puncak, tetapi juga menganalisis sebaran dan kedekatan antar kluster sebagai dasar interpretasi pola risiko.

Studi lainnya menerapkan *K-Means Clustering* dalam pengelompokan data berdasarkan jenis kendaraan dan waktu kejadian [7]. Hasil penelitian menunjukkan terbentuknya tiga kluster utama yang merepresentasikan pola kecelakaan menurut jenis kendaraan. Meskipun memasukkan variabel waktu, fokus analisis didominasi oleh karakteristik kendaraan, sehingga eksplorasi pola temporal belum menjadi perhatian utama.

Selanjutnya, penelitian 18 meninjau distribusi kecelakaan berdasarkan hari dalam seminggu dan menemukan peningkatan kejadian pada hari tertentu, seperti Senin dan Jumat. Namun, analisis tersebut tidak mengkaji distribusi waktu dalam skala harian secara rinci, sehingga belum menggambarkan pola temporal intrahari secara komprehensif.

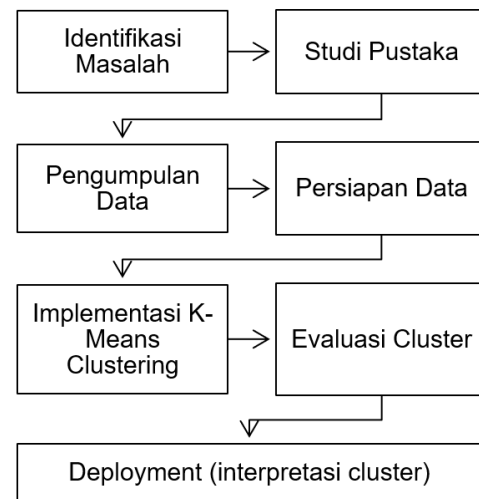
Berdasarkan kajian tersebut, dapat diidentifikasi bahwa meskipun teknik *Clustering* telah banyak digunakan, sebagian besar penelitian variabel waktu sebagai variabel sekunder. Akibatnya, pemanfaatan representasi numerik waktu untuk mengungkap pola distribusi yang lebih rinci masih perlu dilakukan. Selain itu, belum terdapat penelitian yang secara spesifik memfokuskan analisis pada struktur kluster temporal berbasis data terkini di wilayah Provinsi Jawa Tengah.

Penelitian ini hadir untuk mengisi kekosongan tersebut dengan memanfaatkan pendekatan K-Means Clustering untuk pengelompokan data waktu kejadian kecelakaan lalu lintas yang direpresentasikan secara numerik dalam format desimal. Dataset yang digunakan merupakan data kecelakaan di Provinsi Jawa Tengah sepanjang tahun 2024. Jumlah kluster ditentukan melalui metode elbow, dan analisis dilakukan terhadap nilai centroid serta distribusi data antar kluster.

Melalui pendekatan ini, penelitian bertujuan menghasilkan model analisis yang lebih presisi dengan menempatkan waktu sebagai variabel utama. Hasil yang diperoleh diharapkan dapat mendukung perumusan kebijakan berbasis bukti, seperti optimasi penempatan petugas, pengaturan lalu lintas adaptif, serta penentuan waktu kritis berbasis pola kluster. Keunggulan penelitian ini terletak pada penggunaan representasi numerik waktu secara konsisten dan analisis struktur kluster yang lebih mendalam, khususnya dalam konteks data aktual di Provinsi Jawa Tengah.

### III. METODE PENELITIAN

Penelitian ini terdiri dari beberapa tahapan yang disajikan pada Gambar 1.



Gambar 1. Tahapan Pengelompokan Pola Kecelakaan Lalu Lintas

#### 3.1 Identifikasi Masalah

Tahapan awal dalam penelitian ini dilakukan dengan mengidentifikasi permasalahan utama, yaitu tingginya jumlah kecelakaan lalu lintas serta belum diterapkannya pengelompokan waktu kejadian secara sistematis. Selain itu, data demografis seperti profesi pihak yang terlibat dalam kecelakaan juga belum dimanfaatkan dalam analisis untuk mengungkap pola kejadian.

#### 3.2 Studi Pustaka

Studi literatur dilakukan untuk meninjau berbagai penelitian sebelumnya yang menerapkan algoritma K-Means Clustering dalam analisis data kecelakaan lalu lintas. Beberapa penelitian menunjukkan bahwa metode ini efektif untuk mengelompokkan waktu kejadian kecelakaan menjadi beberapa kluster berdasarkan tingkat intensitas kejadian. Namun, sebagian besar studi hanya fokus pada waktu kejadian tanpa mempertimbangkan atribut sosial seperti profesi. Oleh karena itu, penelitian ini bertujuan menggabungkan analisis waktu dan profesi dalam proses klusterisasi.

#### 3.3 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari dataset kecelakaan lalu lintas di Provinsi Jawa Tengah tahun 2024, yang tersedia di platform Kaggle. Dataset ini terdiri dari 9.659 baris data.

Atribut yang terdapat dalam dataset ini antara lain: "kode\_kantor", "kabupaten", "kecamatan", "kondisi", "santunanLL", "santunanMD", "profesi", "tanggal", "day", "month", "year", "jam", dan "karakteristik". Penelitian ini tidak menggunakan data primer karena fokus utama diarahkan pada analisis data historis yang telah tersedia secara publik,

khususnya yang berkaitan dengan atribut waktu dan profesi yang terlibat dalam kecelakaan.

### 3.4 Persiapan Data

Persiapan data bertujuan untuk membuat dataset berdasarkan hasil pengumpulan data dimana dataset ini siap dimodelkan menggunakan algoritma K-means Clustering. Data diolah menggunakan Google Colab dengan bahasa pemrograman Python. Tahapan dimulai dengan seleksi data, kemudian pembersihan data, lalu konversi waktu ke format desimal, dan encoding data profesi ke bentuk numerik menggunakan Label Encoding. Setelah data siap, algoritma K-means diterapkan dengan jumlah kluster yang bervariasi.

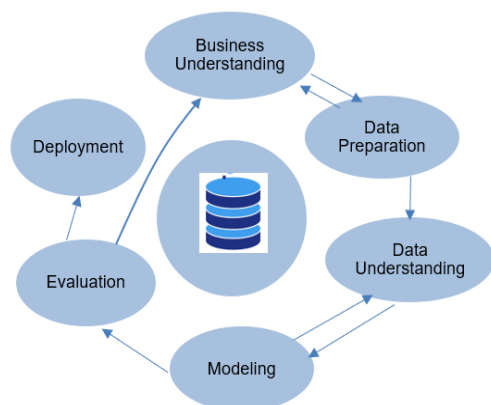
### 3.5 Implementasi K-Means Clustering

Setelah jumlah kluster optimal ditentukan, algoritma K-Means diterapkan untuk mengelompokkan data berdasarkan kemiripan waktu kejadian dan profesi. Setiap data akan dikelompokkan ke dalam kluster tertentu berdasarkan jarak terdekat ke pusat kluster (centroid).

### 3.6 Evaluasi

Sebagai langkah terakhir, dilakukan evaluasi kualitas kluster menggunakan metode Elbow. Metode ini bertujuan untuk menentukan jumlah kluster optimal dengan cara memvisualisasikan nilai inerti (total jarak kuadrat antara data dan centroid klusternya) terhadap jumlah kluster (K).

Tahapan penelitian ini menggunakan pendekatan data mining dengan mengacu pada proses standar CRISP-DM yang disajikan pada Gambar 2.



Gambar 2. Penerapan CRISP-DM

Adapun deskripsi pada fase CRISP-DM, dijelaskan pada Tabel 1.

Tabel 1. Deskripsi Proses CRISP-DM

No	Fase	Deskripsi
1.	<i>Business Understanding</i>	Memahami tujuan dan kebutuhan analisis berdasarkan konteks permasalahan.
2.	<i>Data understanding</i>	Mengumpulkan dan mengenali struktur data untuk memahami informasi awal.
3.	<i>Data Preparation</i>	Menyiapkan data agar siap dianalisis, termasuk pembersihan dan tranformasi data
4.	<i>Modelling</i>	Membangun model analisis menggunakan algoritma yang sesuai
5.	<i>Evaluation</i>	Mengevaluasi hasil model untuk memastikan kesesuaian dengan tujuan awal.
6.	<i>Deployment</i>	Menerapkan hasil analisis ke dalam bentuk yang dapat digunakan untuk pengampilan keputusan.

## IV. HASIL DAN PEMBAHASAN

### 4.1 Business Understanding

Tahapan ini bertujuan untuk memahami masalah utama, yaitu tingginya angka kecelakaan lalu lintas dan belum adanya pemetaan waktu kejadian berdasarkan profesi, sehingga diperlukan analisis untuk mendukung kebijakan keselamatan yang lebih tepat sasaran.

### 4.2 Data Understanding

Pada tahap ini, data awal kecelakaan lalu lintas Jawa Tengah tahun 2024 dikumpulkan yg berisi 9.659 baris data dan 13 atribut. Visualisasi data awal disajikan pada Tabel 2.

Tabel 2. Data Kecelakaan Lalu Lintas Jawa Tengah Tahun 2024

Kode	Kabupaten	...	Kondisi	Tanggal	Jam
400301	Banjarnegara	...	MD	03/04/2024	5:30
400301	Banjarnegara	...	MD	01/16/2024	16:30
400301	Banjarnegara	...	MD	04/10/2024	23:50

Kode	Kabupaten	...	Kondisi	Tanggal	Jam
400301	Banjarnegara	...	MD	01/06/2024	23:10
400201	Banjarnegara	...	MD	01/13/2024	21:30
...	...	...	...	...	...
400001	Kendal	...	LL	01/05/2024	11:30
400001	Kendal	...	LL	01/30/2024	6:45
400001	Kendal	...	LL	01/08/2024	19:00
400001	Kendal	...	LL	02/27/2024	6:00
400001	Kendal	...	LL	02/12/2024	21:15

### 4.3 Data Preparation

Selanjutnya data awal dilakukan seleksi dan diambil 2 atribut yaitu "jam" dan "profesi" menggunakan kode berikut.

```
df_selected = df[['profesi', 'jam']].copy()
# Tampilkan 5 baris pertama dari DataFrame yang dipilih
print("Data yang dipilih (profesi, jam):")
print(df_selected.head())
print("Data yang dipilih (profesi, jam):")
print(df_selected.tail())
```

Atribut "jam" dikonversi ke format numerik, sedangkan "profesi" diencoding ke dalam bentuk angka untuk keperluan analisis menggunakan kode berikut.

```
df_selected['jam_datetime'] = pd.to_datetime(df_selected['jam'], format='%H:%M', errors='coerce').dt.time
df_selected['jam_numerik'] = df_selected['jam_datetime'].apply(lambda x: x.hour if pd.notna(x) else np.nan)
le = LabelEncoder()
df_selected['profesi_encoded'] = le.fit_transform(df_selected['profesi']).astype(str).fillna('')
```

Data dibersihkan dengan menghapus duplikat dan mengisi nilai NaN pada kolom jam menggunakan median serta pada kolom profesi menggunakan rata-rata secara sistematis menggunakan kode berikut.

```
# Hapus duplikat
df_selected.drop_duplicates(inplace=True)
# Cek nilai kosong
```

```
print(df_selected[['jam_numerik', 'profesi_encoded']].isnull().sum())

# Imputasi NaN
median_jam = df_selected['jam_numerik'].median()
mode_profesi = df_selected['profesi_encoded'].mode()[0] if not df_selected['profesi_encoded'].mode().empty else 0
df_selected[['jam_numerik', 'profesi_encoded']] = df_selected[['jam_numerik', 'profesi_encoded']].fillna({'jam_numerik': median_jam, 'profesi_encoded': mode_profesi})
```

Setelah sebanyak tiga kali dilakukan preprocessing data dimana perubahan data awal dapat dilihat pada Gambar 3.

TABEL RINGKASAN PERUBAHAN DATA SELAMA PROSES CLEANING

Tahapan Cleaning Data	Jumlah Baris	NaN di kolom "jam" (es1)	NaN di kolom "profesi" (es1)	NaN di kolom "jam_numerik" (es2)	NaN di kolom "profesi_encoded" (es2)
1. DATA AWAL	969	0	0	0	0
2. SETELAH SELEKSI KOLOM & TRANSFORMASI	836	0	0	0	0
3. SETELAH PENGHAPUSAN DUPLIKAT	836	0	0	0	0
4. SETELAH TRANSFORMASI & DUPLIKAT DIHAPUS	836	0	0	0	0
5. SETELAH IMPUTASI	836	0	0	0	0

Gambar 3. Perubahan Dataset

### 4.4 Modelling

Pada tahap permodelan, algoritma K-Means Clustering digunakan untuk mengelompokkan data yang telah diproses berdasarkan kemiripan fitur. Permodelan dilakukan dengan mencoba nilai K = 2 hingga K = 10, untuk setiap K, dicatat jumlah anggota pada masing-masing kluster untuk melihat distribusi data menggunakan kode berikut.

```
# Ambil fitur
features_for_clustering = ['jam_numerik', 'profesi_encoded']
X = df_selected[features_for_clustering].dropna()

# Buat subplot 2 baris x 4 kolom
fig, axes = plt.subplots(2, 4, figsize=(20, 10))
axes = axes.flatten()

# Loop untuk K dari 2 sampai 9
for idx, k in enumerate(range(2, 10)):
    kmeans = KMeans(n_clusters=k, random_state=42, n_init=10)
    labels = kmeans.fit_predict(X)
```

```
# Plot jam vs profesi, warna
berdasarkan cluster
axes[idx].scatter(X['jam_numerik
'], X['profesi_encoded'], c=labels,
cmap='tab10', s=50)
axes[idx].set_title(f'Hasil
Klastering (K={k})')
axes[idx].set_xlabel('Jam')
axes[idx].set_ylabel('Profesi
(encoded)')
axes[idx].grid(True)
# Rapikan layout
plt.tight_layout()
plt.show()
```

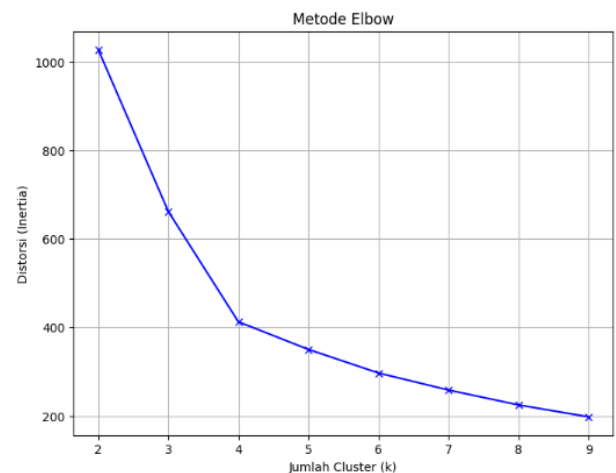
Hasil iterasi setiap nilai K yang digunakan dalam algoritma K-Means Clustering dalam mengelompokkan data disajikan dalam Tabel 3 dan distribusi anggota sebagai acuan pemilihan K yang paling sesuai.

**Tabel 3. Iterasi Nilai K**

K	Distribusi anggota	K	Distribusi anggota
2	Cluster 0 = 429 Cluster 1 = 407 Total = 836	3	Cluster 0 = 256 Cluster 1 = 277 Cluster 2 = 303 Total = 836
4	Cluster 0 = 187 Cluster 1 = 214 Cluster 2 = 231 Cluster 3 = 204 Total = 836	5	Cluster 0 = 213 Cluster 1 = 148 Cluster 2 = 120 Cluster 3 = 196 Cluster 4 = 159 Total = 836
6	Cluster 0 = 150 Cluster 1 = 146 Cluster 2 = 131 Cluster 3 = 147 Cluster 4 = 134 Cluster 5 = 128 Total = 836	7	Cluster 0 = 132 Cluster 1 = 140 Cluster 2 = 85 Cluster 3 = 127 Cluster 4 = 137 Cluster 5 = 90 Cluster 6 = 125 Total = 836
8	Cluster 0 = 101 Cluster 1 = 89 Cluster 2 = 129 Cluster 3 = 142 Cluster 4 = 94 Cluster 5 = 111 Cluster 6 = 90 Cluster 7 = 80 Total = 836	9	Cluster 0 = 106 Cluster 1 = 80 Cluster 2 = 81 Cluster 3 = 137 Cluster 4 = 84 Cluster 5 = 109 Cluster 6 = 84 Cluster 7 = 80 Cluster 8 = 75 Total = 836

#### 4.5 Evaluation

Tahap ini merupakan tahap pengujian hasil cluster. Proses klasterisasi dataset menggunakan algoritma K-means Clustering dimana dibentuk empat model pengelompokan data dengan variasi nilai k yang dapat dilihat pada Tabel 3. Selanjutnya, hasil cluster tersebut dievaluasi menggunakan metode Elbow untuk menentukan jumlah klaster yang optimal.



**Gambar 4. Hasil Evaluasi menggunakan Metode Elbow**

Grafik Metode Elbow pada Gambar 4 untuk menentukan jumlah klaster optimal pada algoritma K-Means. Grafik ini memperlihatkan penurunan nilai inertia yang tajam hingga k=4, kemudian melandai. Titik belok pada k=4 menunjukkan jumlah klaster optimal, karena setelah titik tersebut penambahan klaster tidak memberikan perbaikan signifikan. Oleh karena itu, k=4 dipilih karena dianggap paling seimbang antara kompleksitas dan hasil pengelompokan. Hasil nilai centroid dari k=4 ditunjukkan pada Gambar 5.

```
DataFrame Centroid dalam Skala Asli:
jam_numerik  profesi_encoded
0    17.695187    4.893048
1     6.261682    1.373832
2    17.025974    1.406926
3     6.789216    4.892157

Interpretasi Centroid dalam Skala Asli:
Klaster 0:
Rata-rata Jam: 17.70
Rata-rata Profesi (Encoded): 4.89
Profesi Terdekat (Interpretasi): TNI POLRI
Klaster 1:
Rata-rata Jam: 6.26
Rata-rata Profesi (Encoded): 1.37
Profesi Terdekat (Interpretasi): IBU RUMAH TANGGA
Klaster 2:
Rata-rata Jam: 17.03
Rata-rata Profesi (Encoded): 1.41
Profesi Terdekat (Interpretasi): IBU RUMAH TANGGA
Klaster 3:
Rata-rata Jam: 6.79
Rata-rata Profesi (Encoded): 4.89
Profesi Terdekat (Interpretasi): TNI POLRI
```

**Gambar 5. Hasil Nilai Centroid**

### 3.6 Deployment

Tahap deployment bertujuan untuk menerapkan hasil analisis klasterisasi menjadi informasi yang berguna dalam mendukung pengambilan keputusan. Hasil clustering menggunakan algoritma K-Means mengidentifikasi empat klaster waktu kejadian kecelakaan, yang masing-masing berkaitan dengan kategori profesi tertentu.

Klaster 0 memiliki rata-rata waktu kejadian pada pukul 18.10 (sekitar pukul 17:42-18.20 atau sore hari) dan profesi yang paling dekat dengan nilai encoded-nya adalah TNI POLRI, yang menunjukkan bahwa kelompok ini banyak mengalami kecelakaan pada waktu pulang kerja. Klaster 1 berisi kejadian yang rata-rata terjadi pada pukul 06.26 (sekitar 06:15-06.35 pagi), dan profesi dominannya adalah ibu rumah tangga, mengindikasikan kecelakaan saat aktivitas pagi hari seperti mengantar anak atau belanja. Klaster 2 juga didominasi oleh ibu rumah tangga, dengan rata-rata waktu kejadian di pukul 17.03, menandakan waktu sibuk di sore hari. Sementara itu, Klaster 3 kembali menunjukkan dominasi profesi TNI POLRI dengan waktu rata-rata kejadian pukul 07.19 (sekitar 06:47-07.30 pagi). Secara keseluruhan, hasil ini menunjukkan bahwa waktu-waktu paling rawan kecelakaan adalah pagi dan sore hari, dan profesi yang paling sering terlibat adalah TNI POLRI serta ibu rumah tangga, karena pada jam-jam tersebut mereka cenderung melakukan aktivitas mobilitas tinggi. Informasi ini dapat dijadikan dasar untuk perencanaan dan implementasi kebijakan keselamatan lalu lintas yang lebih terarah dan efektif.

## V. PENUTUP

Penelitian ini menunjukkan bahwa algoritma K-Means Clustering mampu mengidentifikasi pola kecelakaan lalu lintas berdasarkan dimensi waktu dan profesi. Berdasarkan metode *Elbow*, jumlah klaster optimal diperoleh pada  $k=4$  yang bermakna data paling sesuai dikelompokkan menjadi empat klaster. Hasil analisis centroid menunjukkan bahwa Klaster 0 memiliki rata-rata waktu kejadian pada pukul 18.10 dengan dominasi profesi TNI/POLRI. Klaster 1 terjadi pada pukul 06.26 dan didominasi oleh ibu rumah tangga. Klaster 2 didominasi oleh ibu rumah tangga dengan rata-rata waktu kejadian pukul 17.03. Sementara itu, Klaster 3 didominasi oleh TNI/POLRI dengan rata-rata waktu kejadian pukul 07.19.

Secara keseluruhan, pola ini menunjukkan bahwa kecelakaan cenderung terjadi pada jam sibuk pagi dan sore hari, dengan kelompok profesi TNI/POLRI serta ibu rumah tangga sebagai pihak yang paling sering terlibat. Temuan ini mengindikasikan bahwa tingginya mobilitas pada jam-jam sibuk berkontribusi terhadap peningkatan risiko kecelakaan, termasuk pada aparat yang berperan dalam pengawasan lalu lintas. Oleh karena itu, hasil penelitian ini dapat dimanfaatkan sebagai dasar dalam perumusan kebijakan keselamatan lalu lintas yang lebih terarah, seperti penguatan pengawasan pada waktu rawan, pelatihan keselamatan berkendara bagi aparat, serta edukasi masyarakat berbasis kelompok berisiko.

Untuk pengembangan selanjutnya, disarankan agar penelitian memasukkan variabel tambahan seperti kondisi cuaca dan lokasi kejadian guna menghasilkan analisis yang lebih komprehensif dan mendalam.

## DAFTAR PUSTAKA

- [1] S. P. Rathod and R. S. Hadaye, "Social and economic impact of road traffic accidents on patients: A longitudinal study at tertiary care center," *Int. J. Community Med. Public Health*, vol. 14, no. 2, pp. 17–21, 2020, doi: <https://doi.org/10.18203/2394-6040.ijcmph20201981>.
- [2] J. P. Cardoso, E. L. A. Mota, L. N. Ferreira, and P. A. A. Rios, "Productivity Costs Among People Involved in Traffic Accidents," *Cien. Saude Colet.*, vol. 25, pp. 749–760, 2020, doi: <https://doi.org/10.1590/14138-232020252.15232018>.
- [3] B. Hermanto, L. S. Putranto, and D. M. Ma'soem, "Peranan Pengemudi Dalam Kecelakaan Lalu Lintas Jalan: Literature Review," *JMTS: Jurnal Mitra Teknik Sipil*, pp. 597–606, 2022, doi: <https://doi.org/10.24912/jmts.v5i3>.
- [4] E. Marwita and M. D. R. Putri, "Analisis Karakteristik Kecelakaan Lalu Lintas Dengan Menggunakan Metode AEK Dan BKA Pada Ruas Jalan Raya Bogor," in *Prosiding Seminar Nasional Teknik Sipil*, 2025, pp. 543–553. doi: <https://doi.org/10.57203/jriteks.v4i1.2025.43-48>.
- [5] A. A. B. O. K. Surya, D. R. Navianti, D. O. Mulyaningtyas, and Y. Oktopianto, "Continuous Simulation on the Number of Traffic Accidents In Indonesia," *Jurnal*

- Teknologi Transportasi dan Logistik*, vol. 5, no. 2, pp. 235–242, 2024, doi: <https://doi.org/10.52920/jttl.v5i2.94>.
- [6] R. N. Fitriani and N. Khasanah, “Analysis of Traffic Accident Correspondence in Central Java Province,” *Jurnal Fourier*, vol. 11, no. 2, pp. 78–87, 2022, doi: <https://doi.org/10.14421/fourier.2022.112.78-87>.
- [7] M. N. I. Amanah, R. Syarifuddin, and H. Hakim, “Analisis Situational Awareness Pada Pengemudi Ojek Online Di Kota Makassar Dengan Metode Situational Awareness Rating Technique (SART),” *Journal Industrial Engineering and Management (JUST-ME)*, vol. 6, no. 01, pp. 137–150, 2025, doi: <https://doi.org/10.47398/just-me.v6i01.141>.
- [8] T. K. Titus and M. Jajuli, “Clustering Data Kecelakaan Lalu Lintas di Kecamatan Cileungsi Menggunakan Metode K-Means,” *Generation Journal*, vol. 6, no. 1, pp. 1–12, 2022, doi: <https://doi.org/10.29407/gj.v6i1.16103>.
- [9] W. Budiawan and B. Purwanggono, “Clustering analysis of traffic accident in Semarang City,” in *E3S Web of Conferences*, EDP Sciences, 2018, p. 12001. doi: <https://doi.org/10.1051/E3SCONF/20187312001>.
- [10] A. R. Maulana, K. U. N. El Muna, and H. Asjtanto, “Pemetaan dan analisis tren angka kecelakaan di Kota Surabaya,” *Sehat Rakyat: Jurnal Kesehatan Masyarakat*, vol. 2, no. 2, pp. 250–257, 2023, doi: <https://doi.org/10.54259/sehatrakyat.v2i2.1663>.
- [11] Y. Oktopianto and S. Pangesty, “Analisis Daerah Lokasi Rawan Kecelakaan Jalan Tol Tangerang-Merak,” *Jurnal Keselamatan Transportasi Jalan (Indonesian Journal of Road Safety)*, vol. 8, no. 1, pp. 26–37, 2021, doi: <https://doi.org/10.46447/ktj.v8i1.301>.
- [12] I. F. Anshori and Y. Nuraini, “Pengelompokan Data Kecelakaan Lalu Lintas di Kota Tasikmalaya Menggunakan Algoritma K-Means,” *Jurnal Responsif: Riset Sains dan Informatika*, vol. 2, no. 1, pp. 118–127, 2020, doi: <https://doi.org/10.51977/jti.v2i1.198>.
- [13] A. Zauardi and H. Suprayitno, “Analisa Karakteristik Kecelakaan Lalu Lintas di Jalan Ahmad Yani Surabaya Melalui Pendekatan Knowledge Discovery in Database,” *Jurnal Manajemen Aset Infrastruktur & Fasilitas*, vol. 2, no. 1, 2018, doi: <https://doi.org/10.46447/ktj.v8i1.301>.
- [14] T. A. Permana, O. S. Bachri, and R. M. H. Bhakti, “Pemetaan Wilayah Rawan Kecelakaan Lalu Lintas di Kabupaten Brebes Menggunakan Algoritma K-Means,” *Elkom: Jurnal Elektronika dan Komputer*, vol. 18, no. 1, pp. 230–241, 2025, doi: <https://doi.org/10.51903/elkom.v18i1.2929>.
- [15] M. A. Aldi and Z. Fatah, “Implementasi K-means Clustering Dalam Pengelompokan Data Kunjungan Wisatawan Asing di Indonesia,” *Jurnal Ilmiah Multidisiplin Ilmu*, vol. 2, no. 1, pp. 13–19, 2025, doi: <https://doi.org/10.69714/3hhfj353>.
- [16] F. Fitriah, A. Eviyanti, and H. Hindarto, “Pengelompokan Pelanggaran Lalu Lintas Menggunakan Algoritma K-Means pada Data CCTV,” *SMATIKA JURNAL*, vol. 15, no. 02, pp. 442–453, 2025, doi: <https://doi.org/10.32664/smatika.v15i02.1739>.
- [17] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, and J. Liu, “A Quantitative Discriminant Method of Elbow Point for The Optimal Number of Clusters in Clustering Algorithm,” *EURASIP J. Wirel. Commun. Netw.*, vol. 2021, no. 1, p. 31, 2021, doi: <https://doi.org/10.1186/s13638-021-01910-w>.
- [18] R. R. Aria, “Implementasi Algoritma K-Means untuk Pengelompokan Data Imunisasi Balita dengan Metode CRISP-DM,” *REMIK: Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 9, no. 1, pp. 189–197, 2025, doi: [10.33395/remik.v9i1.14391](https://doi.org/10.33395/remik.v9i1.14391).