

MENGIDENTIFIKASI POLA KECELAKAAN LALU LINTAS DENGAN K-MEANS CLUSTERING

by Masnawati Fkti

Submission date: 20-Jun-2025 11:11AM (UTC+0700)

Submission ID: 2702683490

File name: Artikel_TiKomSiN.pdf (631.2K)

Word count: 3448

Character count: 21542

MENGIDENTIFIKASI POLA KECELAKAAN LALU LINTAS DENGAN K-MEANS CLUSTERING

Victor Bandhaso¹⁾; Masna Wati²⁾; Haviluddin³⁾

¹⁾²⁾³⁾Program Studi Informatika, Universitas Mulawarman

¹⁾bandhasovictor@gmail.com; ²⁾masnawati@fkti.unmul.ac.id; ³⁾haviluddin@unmul.ac.id

ABSTRACT

Traffic accidents are a complex problem that have broad impacts on the social and economic aspects of society. This study aims to identify temporal patterns of traffic accident occurrences based on temporal and demographic attributes using the K-Means Clustering algorithm on 9,659 accident data in Central Java Province in 2024. The time attribute was converted into decimal format, and the profession of the involved parties was transformed into numerical form through encoding. With the help of the Elbow method, four main clusters were identified, indicating that accidents most frequently occur in the morning and afternoon. Interestingly, the professions most often involved are military/police personnel and housewives, two groups with high mobility during peak hours. These findings suggest that the risk of accidents also threatens officers who are supposed to maintain traffic order. Therefore, a more comprehensive prevention approach is needed, such as driving safety training for officers, adjustment of operational schedules, and data-driven public education. This study demonstrates that K-Means Clustering is an effective method for uncovering hidden patterns in accident data in a simple yet meaningful way.

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. Data Kepolisian Republik Indonesia menunjukkan ribuan insiden terjadi setiap tahun, baik di wilayah perkotaan maupun pedesaan [3].

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 [4]. Akibatnya, risiko kecelakaan terus meningkat dan berpotensi mengganggu kesejahteraan masyarakat secara luas.

Penelitian sebelumnya umumnya bersifat deskriptif dan terbatas pada klasifikasi kecelakaan berdasarkan lokasi, jenis kendaraan, atau kondisi cuaca [5]. 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.

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.

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 mengenai penerapan algoritma K-Means Clustering dalam pengelompokan data kecelakaan lalu lintas telah banyak dilakukan dalam berbagai konteks dan pendekatan analisis [6]. Secara umum, algoritma ini 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 [7]. Dalam penelitian tersebut, data waktu kejadian dikonversi ke dalam format desimal guna mempermudah pemrosesan numerik. Hasilnya menunjukkan pembagian data ke dalam tiga kluster utama, yaitu pagi, sore, dan malam hari, masing-masing dengan tingkat intensitas kecelakaan yang berbeda. Informasi ini digunakan oleh otoritas setempat dalam menyusun regulasi jam operasional kendaraan berat, sehingga diharapkan dapat menurunkan angka kecelakaan pada waktu-waktu rawan.

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 [8]. Temuan tersebut dijadikan dasar dalam penjadwalan ulang waktu kerja petugas lalu lintas serta penempatan infrastruktur pengatur arus kendaraan pada jam-jam sibuk.

Sementara itu, penelitian lain mengembangkan pendekatan yang lebih kompleks dengan menggabungkan variabel waktu dan cuaca dalam model klusterisasi. Dengan menggunakan K-Means Clustering, peneliti menemukan bahwa kecelakaan cenderung meningkat pada saat hujan ringan di malam hari. Namun, representasi waktu dalam studi tersebut masih menggunakan kategori umum seperti pagi, siang, dan malam, yang kurang presisi dalam konteks numerik.

Dalam konteks pengelolaan jalur tol, studi meneliti kecelakaan di ruas tol Trans Jawa. Dengan mengonversi data waktu ke dalam format angka desimal, penelitian tersebut membentuk empat kluster utama, yang menunjukkan bahwa waktu puncak kecelakaan terjadi pada pukul 07.00–08.00 dan 17.00–18.00 [9]. Temuan ini memberikan

kontribusi terhadap manajemen lalu lintas kendaraan logistik di wilayah tersebut.

Penelitian lainnya membandingkan performa algoritma K-Means dengan DBSCAN dalam pengelompokan data berdimensi tunggal seperti waktu. Hasilnya menunjukkan bahwa K-Means lebih unggul dalam hal efisiensi komputasi dan interpretabilitas hasil klusterisasi. Hal ini memperkuat argumentasi bahwa K-Means sangat sesuai untuk digunakan dalam pengolahan data linier seperti waktu kejadian kecelakaan.

Kajian lain mengombinasikan atribut spasial dan temporal untuk mengidentifikasi pola kecelakaan. Dengan mengintegrasikan koordinat geografis dan waktu kejadian, penelitian ini berhasil memetakan titik rawan kecelakaan dalam konteks spasial-temporal. Meskipun demikian, aspek waktu dalam studi ini belum dianalisis secara numerik mendalam, sehingga akurasi dan presisi dalam representasi waktu kurang optimal.

Studi lainnya menerapkan K-Means dalam mengelompokkan data berdasarkan jenis kendaraan dan waktu kejadian. Penelitian ini menghasilkan tiga kluster utama yang mencerminkan pola distribusi kecelakaan menurut jenis kendaraan. Namun, fokus utamanya lebih kepada variabel jenis kendaraan daripada pendalaman pada aspek temporal [10].

Penelitian lain juga menyoroti distribusi kecelakaan berdasarkan hari dalam seminggu. Studi ini menunjukkan bahwa kecelakaan cenderung meningkat pada hari Senin dan Jumat. Meski demikian, analisisnya tidak mencakup detail distribusi waktu dalam satu hari secara granular, sehingga masih menyisakan ruang eksplorasi pada dimensi temporal harian [11].

Berdasarkan kajian pustaka tersebut, dapat disimpulkan bahwa meskipun algoritma K-Means Clustering telah banyak digunakan dalam menganalisis data kecelakaan lalu lintas, sebagian besar penelitian masih terbatas pada pengelompokan berbasis waktu dalam kategori umum dan belum memanfaatkan representasi numerik yang lebih presisi. Selain itu, belum terdapat studi yang secara khusus menelaah distribusi waktu kecelakaan lalu lintas di Provinsi Jawa Tengah menggunakan data tahun berjalan.

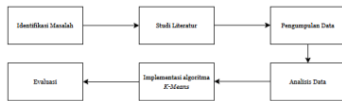
Penelitian ini hadir untuk mengisi kekosongan tersebut dengan memanfaatkan pendekatan K-Means Clustering terhadap

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.

Dengan pendekatan ini, penelitian bertujuan untuk menyajikan model analisis yang lebih presisi terhadap dimensi waktu sebagai variabel utama, yang hasilnya diharapkan dapat digunakan sebagai dasar dalam perumusan kebijakan berbasis bukti. Misalnya, hasil pengelompokan dapat digunakan untuk menentukan waktu optimal penempatan petugas lalu lintas, pemasangan rambu otomatis pada jam rawan, serta perencanaan sistem manajemen lalu lintas berbasis waktu. Keunggulan penelitian ini terletak pada penerapan metode numerik yang terstruktur serta penggunaan data aktual, yang belum banyak disentuh dalam studi-studi sebelumnya, terutama yang berfokus pada wilayah Provinsi Jawa Tengah.

III. METODE PENELITIAN

Tahapan penelitian yang sebagai berikut:



Gambar 1. Diagram Alir

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 Literatur

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 mengombinasikan analisis waktu dan profesi dalam proses klusterisasi.

3.3 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yg 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 Analisis Data

Analisis data dalam penelitian ini dilakukan untuk mengelompokkan kejadian kecelakaan lalu lintas berdasarkan waktu kejadian dan profesi menggunakan algoritma K-means Clustering. Proses ini bertujuan untuk mengidentifikasi pola waktu rawan kecelakaan serta profesi yang paling sering terlibat, untuk mendukung kebijakan keselamatan lalu lintas yang lebih efektif.

Data dianalisis 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, ditentukan menggunakan metode Elbow. Hasil kluster dianalisis secara deskriptif berdasarkan median jam kejadian dan rata-rata label profesi untuk melihat karakteristik masing-masing kelompok.

Dengan pendekatan ini, proses analisis data dilakukan secara sistematis, terukur, dan dapat memberikan gambaran konkret

mengenai jam-jam rawan kecelakaan serta kelompok yang rentan, sehingga dapat digunakan dalam perencanaan strategi keselamatan lalu lintas.

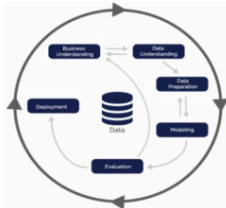
3.5 Implementasi Algoritma K-Means

Setelah 5 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 kalster (centroid). Hasil klusterisasi ditambahkan sebagai kluster laber kluster pada dataset.

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 klasternya) terhadap jumlah kluster (K) [12].

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



Gambar 2. Metode CRISP-DM

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

Tabel 1. Deskripsi Proses CRISP-DM

No	Fase	Deskripsi
1.	Business Understanding	Memahami tujuan dan kebutuhan analisis berdasarkan konteks permasalahan.
2.	Data understanding	Mengumpulkan dan mengenali struktur data untuk memahami informasi awal.

3.	Data Preparation	Menyiapkan data agar siap dianalisis, termasuk pembersihan dan tranformasi data
4.	Modelling	Membangun model analisis menggunakan algoritma yang sesuai
5.	Evaluation	Mengevaluasi hasil model untuk memastikan kesesuaian dengan tujuan awal.
6.	Deployment	Menerapkan hasil analisis ke dalam bentuk yang dapat digunakan untuk pengampilan keputusan.

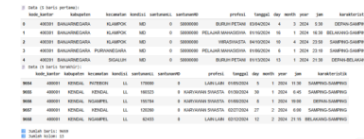
IV. HASIL DAN PEMBAHASAN

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 Gambar 3.



Gambar 3. Data Awal

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_numerik'] = 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, perubahan data tersebut dapat dilihat pada Gambar 4.

	jam	profesi	jam_numerik	profesi_encoded
0	10:00	1	10	1
1	11:00	2	11	2
2	12:00	3	12	3
3	13:00	4	13	4

Gambar 4. Perubahan Data

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 klaster untuk melihat distribusi data menggunakan kode berikut.

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

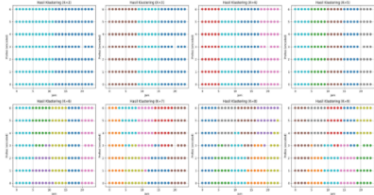
# Buat subplot 2 baris x 4 kolom
fig, axes = plt.subplots(2, 4, figsize=(20, 10))
axes = axes.flatten()

# 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()
```

Berdasarkan program tersebut nilai K dapat dilihat pada Gambar 5 berikut.



Gambar 5. Perubahan Data

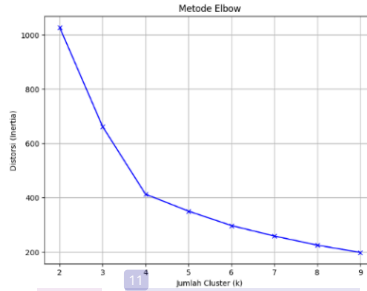
Hasil iterasi K disajikan dalam tabel dibawah ini dan distribusi anggota sebagai acuan pemilihan K yang paling sesuai.

Tabel 2. 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

Setelah proses klasterisasi, evaluasi dilakukan menggunakan metode Elbow untuk menentukan jumlah kluster yang optimal.



Gambar 5. Hasil Evaluasi menggunakan Metode Elbow

Hasil grafik Metode Elbow, untuk menentukan jumlah kluster 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 kluster optimal, karena setelah titik tersebut penambahan kluster 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  17.825974    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

Terdapat empat kluster yang terbentuk. Klaster 0 memiliki rata-rata waktu kejadian pada pukul 17.70 (sekitar pukul 17:42 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 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 06.79 (sekitar 06:47 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.

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. Informasi ini dapat dijadikan dasar untuk perencanaan dan implementasi kebijakan keselamatan lalu lintas yang lebih terarah dan efektif.

V. PENUTUP

Penelitian ini membuktikan bahwa algoritma K-Means Clustering mampu mengidentifikasi pola kecelakaan lalu lintas berdasarkan waktu kejadian dan profesi dengan efektif. Hasil analisis menunjukkan bahwa kecelakaan paling sering terjadi pada pagi dan sore hari, dengan profesi TNI/POLRI serta ibu rumah tangga sebagai pihak yang paling sering terlibat akibat tingginya mobilitas pada jam-jam sibuk tersebut. Hal ini dapat terjadi karena TNI/POLRI secara rutin melakukan patroli dan pengawasan lalu lintas terutama pada jam-jam sibuk guna menjaga ketertiban, sehingga risiko terlibat kecelakaan meningkat. Sementara itu, ibu rumah tangga cenderung melakukan aktivitas seperti mengantar anak ke sekolah, belanja, dan pekerjaan rumah tangga lainnya pada waktu-waktu tersebut, yang membuat mobilitas mereka juga tinggi dan rentan terhadap kecelakaan. Temuan ini dapat dimanfaatkan untuk menyusun kebijakan keselamatan lalu lintas yang lebih tepat sasaran. Disarankan agar hasil penelitian ini dijadikan acuan dalam

pengaturan operasional lalu lintas, pelatihan keselamatan berkendara bagi aparat, serta edukasi masyarakat yang lebih intensif. Penelitian lanjutan juga perlu dilakukan dengan menambahkan variabel lain seperti cuaca dan lokasi kejadian untuk mendapatkan hasil yang lebih komprehensif.

DAFTAR PUSTAKA

- [1] S. Gopalakrishnan, M. S. Kumar, dan R. S. R. Kumar, "Social and Economic Impact of Road Traffic Accidents on Patients: A Longitudinal Study at Tertiary Care Center," *International Journal of Community Medicine and Public Health*, vol. 7, no. 5, pp. 1858–1863, 2020. doi: <https://doi.org/10.18203/2394-6040.ijcmph20201981>.
- [2] J. Cardoso, E. Mota, L. Ferreira, dan P. Rios, "Productivity costs among people involved in traffic accidents," *Ciencia & Saude Coletiva*, vol. 25, no. 2, pp. 749–760, 2020. doi: <https://doi.org/10.1590/14138-232020252.15232018>.
- [3] A. Atalay, A. Ünal, dan M. Y. Çodur, "Transportation policies in increasing traffic safety," *International Advanced Researches and Engineering Journal*, vol. 2, no. 2, pp. 104–108, 2018.
- [4] A. Surya, D. Navianti, D. Mulyaningtyas, dan Y. Oktopianto, "Continuous Simulation on the Number of Traffic Accidents in Indonesia," *Jurnal Teknologi Transportasi dan Logistik*, vol. 5, no. 2, 2024. doi: <https://doi.org/10.52920/jttl.v5i2.94>.
- [5] R. Fitriani dan 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>.
- [6] A. Aljofey dan K. Al-Wagih, "Analysis of Accident Times for Highway Locations Using K-Means Clustering and Decision Rules Extracted from Decision Trees," *International Journal of Computer Applications Technology and Research*, vol. 7, no. 1, pp. 1–11, 2018. doi: <https://doi.org/10.7753/IJCATR0701.1001>.

- [7] W. Budiawan dan B. Purwanggono, "Clustering Analysis of Traffic Accident in Semarang City," *E3S Web of Conferences*, vol. 73, Art. no. 12001, 2018. doi: <https://doi.org/10.1051/E3SCONF/20187312001>.
- [8] A. Maulana, K. Ulfah, N. Muna, dan H. Asjtanto, "Pemetaan dan Analisis Tren Angka Kecelakaan di Kota Surabaya," *Sehat Rakyat: Jurnal Kesehatan Masyarakat*, vol. 2, no. 2, 2023. doi: <https://doi.org/10.54259/sehatrakyat.v2i2.1663>.
- [9] Y. Oktopianto dan S. Pangesty, "Analisis Daerah Lokasi Rawan Kecelakaan Jalan Tol Tangerang-Merak," *Jurnal Keselamatan Transportasi Jalan (Indonesian Journal of Road Safety)*, vol. 8, no. 1, 2021. doi: <https://doi.org/10.46447/kjtj.v8i1.301>.
- [10] I. F. Anshori dan 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, Feb. 2020. doi: <https://doi.org/10.51977/jti.v2i1.198>.
- [11] A. Zauardi dan 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, pp. 45–52, 2018. doi: <https://iptek.its.ac.id/index.php/jmaif/article/view/37670>.
- [12] 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 Journal on Wireless Communications and Networking*, vol. 2021, no. 1, pp. 1–10, 2021. doi: <https://doi.org/10.1186/s13638-021-01910-w>.

MENGIDENTIFIKASI POLA KECELAKAAN LALU LINTAS DENGAN K-MEANS CLUSTERING

ORIGINALITY REPORT

14%	11%	8%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	jurnal.umuslim.ac.id Internet Source	1%
2	ejournal.uksw.edu Internet Source	1%
3	Submitted to Sydney Polytechnic Institute Student Paper	1%
4	www.coursehero.com Internet Source	1%
5	Dwi Tatang Warianta, Paramesti Astagina, Richy Julianto, Florentina Yuni Arini. "Optimalisasi K-Means Menggunakan Algoritma Firefly Untuk Segmentasi Pelanggan pada E-commerce", JURNAL FASILKOM, 2025 Publication	1%
6	Submitted to University of the Pacific Student Paper	1%
7	digilib.unila.ac.id Internet Source	1%
8	p3m.sinus.ac.id Internet Source	1%
9	edoc.pub Internet Source	1%
10	ojs.trigunadharma.ac.id Internet Source	<1%

11	Eko Santoso, Agus Umar Hamdani. "SEGMENTASI PELANGGAN PADA PT. TAB MENGUNAKAN ALGORITMA K-MEANS CLUSTERING", JIKI (Jurnal Ilmu Komputer & Informatika), 2023 Publication	<1 %
12	ejournal.jak-stik.ac.id Internet Source	<1 %
13	Submitted to Athlone Institute of Technology Student Paper	<1 %
14	medium.com Internet Source	<1 %
15	Submitted to University of Wisconsin - Stout Student Paper	<1 %
16	docobook.com Internet Source	<1 %
17	fr.scribd.com Internet Source	<1 %
18	lingkungan.ft.unand.ac.id Internet Source	<1 %
19	repository.ub.ac.id Internet Source	<1 %
20	repository.utu.ac.id Internet Source	<1 %
21	Garnet Filemon Waluyono. "PERANCANGAN SISTEM UKURAN SEPATU UNTUK ANAK PEREMPUAN USIA 4-6 TAHUN BERDASARKAN DATA ANTROPOMETRI KAKI", Jurnal Aplikasi Ilmu Teknik Industri (JAPTI), 2020 Publication	<1 %
22	Iwan Irawan, Reza Rahman, Arief Wibowo. "PENGELOMPOKAN TRANSAKSI KARTU DEBIT	<1 %

PERBANKAN MENGGUNAKAN ALGORITMA K-MEANS", Jurnal Sistem Informasi dan Informatika (Simika), 2025

Publication

23 vdocuments.net <1 %
Internet Source

24 www.jurnal-umbuton.ac.id <1 %
Internet Source

25 www.scribd.com <1 %
Internet Source

26 123dok.com <1 %
Internet Source

27 jbc.bj.uj.edu.pl <1 %
Internet Source

28 www.journal.uad.ac.id <1 %
Internet Source

29 Agung Prayogo Bagustio, Ade Irma Purnamasari, Irfan Ali. "ANALISIS DATA PENJUALAN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING PADA TOKO KECANTIKAN PUTRI", PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer, 2024
Publication

30 Rosa Rosiana, Willy Prihartono, Fathurrohman Fathurrohman. "Implementation of K-Means Algorithm for Sub-district Grouping Based on Rice Plant Productivity in Cirebon Regency", Jurnal Informatika Terpadu, 2025
Publication

31 Asep Lukman Arip Hidayat, Helmi Zulqan, Gandung Triyono. "CLUSTERING OF INDONESIAN PROVINCES BASED ON LIFE

EXPECTANCY USING K-MEANS WITH ELBOW
METHOD EVALUATION", JSil (Jurnal Sistem
Informasi), 2025

Publication

32

zombiedoc.com

Internet Source

<1%

Exclude quotes Off

Exclude matches Off

Exclude bibliography On