# PENERAPAN ALGORITMA K-MEANS UNTUK PENGELOMPOKAN TINGKAT KEMACETAN LALU LINTAS

Fadia Aulia Azizah[1*], Dwi Hartanti[2], Pramono[3]

[1]Teknik Informatika
Universitas Duta Bangsa
[1*]220103013@mhs.udb.ac.id

[2]Teknik Informatika
Universitas Duta Bangsa
[2]dwihartanti@udb.ac.id

[3]Teknik Informatika
Universitas Duta Bangsa
[3]pramono@udb.ac.id

*Abstrak*— **Peningkatan jumlah kendaraan yang tidak diimbangi dengan perkembangan infrastruktur jalan memicu lonjakan kemacetan lalu lintas. Untuk mengatasi masalah ini, diperlukan analisis data yang terstruktur guna memetakan pola kepadatan kendaraan. Penelitian ini menerapkan teknik *data mining* dengan pendekatan *clustering* melalui algoritma K-Means untuk mengelompokkan tingkat kemacetan berdasarkan data transportasi. Metodologi penelitian mengikuti tahapan CRISP-DM, yang meliputi *business understanding, data understanding, data preparation, modeling, evaluation,* hingga *deployment*. Data yang digunakan bersumber dari Kaggle dengan variabel *start point, end point, travel time, distance,* dan waktu perjalanan. Validasi dan evaluasi kualitas klaster diuji menggunakan *Silhouette Score* dan *Davies-Bouldin Index* (DBI). Hasil akhir dari penelitian ini diharapkan mampu membentuk klaster tingkat kemacetan yang representatif dengan nilai pengujian validitas klaster yang optimal. Melalui pendekatan kuantitatif ini, penelitian dapat memberikan gambaran pola kepadatan yang valid serta menjadi bahan rujukan strategis dalam pengelolaan lalu lintas yang lebih efektif.**
*Kata kunci*— Kemacetan lalu lintas, data mining, clustering, algoritma K-Means

*Abstract*— **The continuous increase in the number of vehicles, unmatched by infrastructure development, has triggered a surge in traffic congestion. To address this issue, a structured data analysis is required to map vehicle density patterns. This study applies data mining techniques using a clustering approach via the K-Means algorithm to group congestion levels based on transportation data. The research methodology follows the CRISP-DM framework, which encompasses business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The dataset utilized is sourced from Kaggle, featuring variables such as start point, end point, travel time, distance, and travel duration. The validation and quality evaluation of the clusters are tested using the Silhouette Score and Davies-Bouldin Index (DBI). Ultimately, this research is expected to generate representative congestion level clusters with optimal cluster validity scores. Through this quantitative approach, the study provides a valid overview of density patterns and serves as a strategic reference for more effective traffic management.**
***Keywords***— Traffic Congestion, Data Mining, Clustering, K-Means Algorithm

## I. PENDAHULUAN

Kemacetan lalu lintas telah menjadi isu transportasi krusial di berbagai wilayah perkotaan dengan tren yang terus melonjak tiap tahunnya. Fenomena ini dipicu oleh laju pertumbuhan jumlah kendaraan yang tidak selaras dengan perluasan kapasitas infrastruktur jalan. Akibatnya, kepadatan yang ditimbulkan memicu berbagai dampak negatif, seperti keterlambatan waktu tempuh, pemborosan konsumsi bahan bakar, hingga degradasi kualitas lingkungan akibat polusi. Situasi ini menghadirkan tantangan besar bagi pemerintah dan otoritas terkait dalam merumuskan manajemen lalu lintas yang efisien [1]. Lebih jauh lagi, tingginya intensitas kemacetan ini turut mengoreksi tingkat produktivitas masyarakat serta memperburuk kualitas mobilitas di area perkotaan [3].

Pendekatan *data mining* kini banyak diadopsi dalam analisis sistem transportasi berkat kemajuan teknologi informasi. Sebagai metode pengolahan data, *data mining* mampu mengungkap pola maupun insight tersembunyi dari data bervolume besar. Dalam konteks transportasi, teknik ini digunakan untuk mengevaluasi situasi jalan raya, yang hasilnya dapat menjadi panduan bagi otoritas terkait dalam menetapkan kebijakan [8]. Dari berbagai

teknik yang tersedia, *clustering* atau pengelompokan data berdasarkan kemiripan atribut menjadi salah satu metode yang paling sering diterapkan [5].

K-Means menjadi algoritma *clustering* yang sangat populer lantaran mekanismenya yang sederhana, mudah diterapkan, dan efisien untuk pengolahan data bervolume besar [2]. Metode ini mengelompokkan data berdasarkan kemiripan jarak terdekat guna mengungkap pola yang tersembunyi [9]. Dalam kasus analisis lalu lintas di kota Surabaya, metode ini sukses mengklasifikasikan kondisi jalan menjadi kategori sepi, lancar, padat, dan macet total menggunakan variabel derajat kejenuhan serta volume kendaraan [1]. Keandalan algoritma K-Means ini diperkuat oleh sejumlah penelitian lain yang menggunakannya untuk membedah pola kemacetan berbasis data GPS dan deret waktu (*time series*) [6], [11].

Penggunaan algoritma K-Means dalam bidang transportasi telah banyak diterapkan karena mampu membantu proses identifikasi pola kepadatan lalu lintas secara lebih terstruktur. Kumar dan Singh menjelaskan bahwa penerapan K-Means dalam analisis lalu lintas cerdas dapat membantu proses pengelompokan kondisi jalan berdasarkan tingkat kemacetan sehingga mendukung pengelolaan transportasi yang lebih efektif [7]. Selain itu, penelitian Garcia dan Lopez menunjukkan bahwa metode K-Means memiliki performa yang cukup baik dalam pengelompokan data lalu lintas dibandingkan beberapa metode clustering lainnya [10].

Meskipun demikian, pengelolaan data transportasi pada beberapa instansi masih dilakukan secara konvensional dan belum memanfaatkan proses analisis clustering secara optimal. Data lalu lintas umumnya hanya digunakan sebagai laporan statistik tanpa dilakukan pengelompokan untuk menemukan pola kemacetan tertentu [4]. Akibatnya, proses identifikasi wilayah atau waktu dengan tingkat kemacetan serupa menjadi lebih sulit dilakukan sehingga pengambilan keputusan terkait pengaturan lalu lintas kurang efektif.

Berpijak pada tantangan tersebut, penelitian ini menerapkan metode *clustering* K-Means untuk mengelompokkan tingkat kemacetan berdasarkan atribut data transportasi. Dataset yang digunakan mengintegrasikan parameter seperti lokasi asal, lokasi tujuan, waktu tempuh, jarak, dan waktu perjalanan. Proses pengelompokan ini ditargetkan dapat mengungkap karakteristik dan pola kepadatan lalu lintas secara terstruktur. Hasil akhir dari analisis ini diharapkan mampu memberikan pemahaman yang lebih baik mengenai dinamika jalan raya serta menjadi rujukan strategis bagi otoritas terkait dalam mengelola sistem transportasi.

## II. METODOLOGI PENELITIAN

### 2.1 Jenis dan Sumber Data

Penelitian ini menggunakan jenis kuantitatif berupa data transportasi yang berkaitan dengan tingkat kemacetan lalu lintas. Data yang digunakan meliputi beberapa variabel, yaitu start point, end point, travel time, distance, dan waktu perjalanan. Data tersebut digunakan untuk menganalisis pola tingkat kepadatan lalu lintas berdasarkan karakteristik perjalanan kendaraan.

Penelitian ini memanfaatkan dataset transportasi dari platform Kaggle sebagai sumber data utama untuk mengelompokkan tingkat kepadatan lalu lintas melalui algoritma K-Means. Di samping itu, landasan teori dalam penelitian ini diperkuat oleh penggunaan data sekunder. Informasi sekunder tersebut diperoleh melalui penelusuran pustaka terhadap jurnal ilmiah, buku, dan laporan penelitian sebelumnya yang relevan dengan bidang *data mining*, metode *clustering*, serta optimalisasi algoritma K-Means.

Adapun teknik pengumpulan data dalam penelitian ini meliputi:

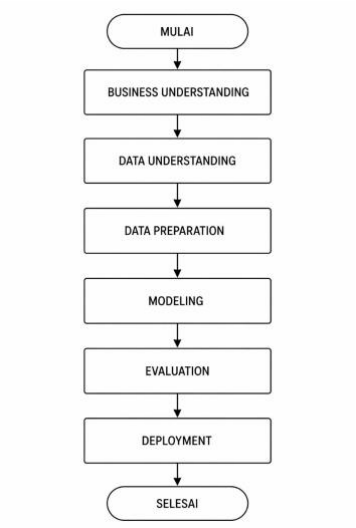
1. Studi Literatur, Tahapan ini dilaksanakan dengan menelaah berbagai referensi ilmiah, termasuk jurnal penelitian, buku teks, artikel

akademis, dan literatur relevan lainnya yang mengkaji bidang data mining, metode clustering, algoritma K-Means, serta fenomena kemacetan lalu lintas. Kegiatan studi literatur ini diproyeksikan untuk membangun fondasi teoretis yang kuat guna mendukung seluruh rangkaian penelitian sekaligus menjadi pijakan utama dalam melakukan analisis data.

2. Dokumentasi Dataset, Metode ini dilakukan dengan mengunduh dan menyusun dataset transportasi dari Kaggle. Di dalam dataset tersebut, terdapat informasi mendetail mengenai aktivitas perjalanan kendaraan yang meliputi titik asal, titik tujuan, durasi komputasi waktu, jarak geografis, serta waktu tempuh kendaraan. Data-data tersebut dikondisikan sebagai korpus data utama yang akan diolah untuk memetakan dan mengelompokkan pola kepadatan lalu lintas.

## 2.2 Tahapan Penelitian

Penelitian ini mengadopsi metode *data mining* berbasis klasterisasi menggunakan algoritma K-Means. Untuk memperoleh hasil pemetaan tingkat kepadatan lalu lintas yang optimal, metodologi penelitian dijalankan melalui prosedur yang terstruktur dan sistematis. Rincian mengenai tahapan-tahapan penelitian tersebut adalah sebagai berikut:




Gambar 1. Alur Tahapan Penelitian

1. *Business Understanding*
Tahap ini berfokus pada identifikasi problem kemacetan lalu lintas serta penetapan sasaran yang ingin dicapai dalam penelitian. Isu sentral yang melatarbelakangi studi ini adalah tingginya arus kepadatan lalu lintas yang menghambat visualisasi pola kemacetan secara terstruktur. Sebagai solusinya, pendekatan *clustering* dengan algoritma K-Means diterapkan untuk mengelompokkan strata kemacetan dengan memanfaatkan basis data transportasi.

2. *Data Understanding*
Pada tahap ini, dataset transportasi yang menjadi objek penelitian dikumpulkan untuk kemudian dipelajari karakteristiknya. Proses penelaahan ini meliputi pemetaan variabel, penentuan tipe data, serta analisis kondisi awal data secara menyeluruh. Atribut data yang digunakan dalam model ini bersandar pada beberapa parameter kunci, yakni *start point, end point, travel time, distance,* serta waktu perjalanan.

3. *Data Preparation*
Pada fase ini, dataset ditransformasikan terlebih dahulu agar siap dimasukkan ke dalam model klasterisasi. Aktivitas pra-pemrosesan ini mencakup penyaringan data dari kotoran (*cleaning*), penghapusan baris data yang berulang, penanganan nilai yang kosong (*missing value*), serta penyesuaian format struktur data. Hal ini dilakukan demi menjamin kualitas dan validitas data saat diproses menggunakan algoritma K-Means.

4. *Modeling*
Tahap pemodelan ini merealisasikan proses klasterisasi dengan mengimplementasikan algoritma K-Means melalui bantuan bahasa pemrograman Python. Komputasi K-Means diterapkan untuk menyaring dan mengelompokkan data transportasi ke dalam beberapa klaster spesifik. Penentuan kelompok tersebut didasarkan pada tingkat kemiripan karakteristik data yang diukur

menggunakan formula perhitungan jarak Euclidean (*Euclidean distance*).

5. *Evaluation*
   Tahap ini analisis dilakukan untuk menilai keandalan dan kualitas struktur klaster hasil pemodelan. Proses validasi tersebut bersandar pada dua parameter uji, yaitu *Silhouette Score* dan *Davies-Bouldin Index* (DBI). Keluaran dari evaluasi ini tidak hanya merepresentasikan performa dari algoritma *clustering*, melainkan juga mempermudah penentuan jumlah kelompok terbaik dalam pemetaan data.

6. *Deployment*
   Tahap akhir adalah penyajian hasil analisis clustering dalam bentuk visualisasi dan interpretasi data. Hasil pengelompokan digunakan untuk memberikan informasi mengenai pola tingkat kemacetan lalu lintas sehingga dapat membantu dalam memahami kondisi kepadatan lalu lintas secara lebih terstruktur.

## III. HASIL DAN PEMBAHASAN

### 3.1 Hasil Pengolahan Data

Penelitian ini memanfaatkan dataset transportasi yang bersumber dari platform Kaggle, dengan proses pengolahan data yang dieksekusi menggunakan bahasa pemrograman Python. Sebelum memasuki tahap klasterisasi, dataset terlebih dahulu diwajibkan melalui fase pra-pemrosesan (*preprocessing*). Langkah ini mengintegrasikan serangkaian prosedur seperti inspeksi nilai yang hilang (*missing value*), eliminasi data ganda (*duplikat*), serta transformasi atribut data agar selaras dengan kebutuhan komputasi algoritma K-Means. Berdasarkan hasil akhir dari tahapan *preprocessing* tersebut, dataset dikonfirmasikan telah memenuhi syarat dan berada dalam kondisi siap pakai untuk memetakan strata kemacetan lalu lintas.

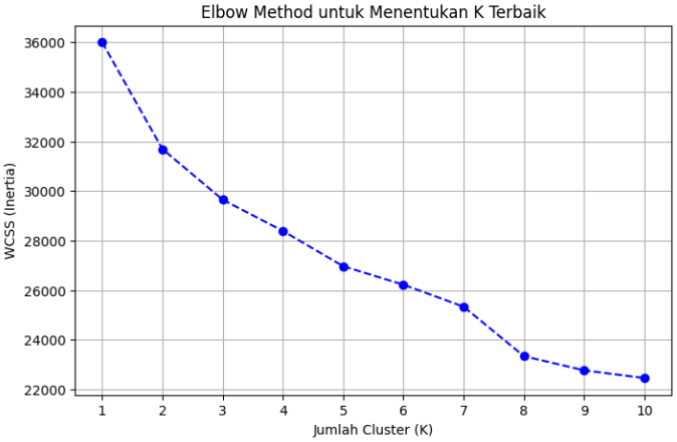
Pratinjau 5 Data Teratas Setelah Preprocessing:

| | start_point | end_point | time_of_day | day_of_week | traffic_condition | event_count | is_holiday | vehicle_density | population_density | weather | public_transport_availability | historical_delay_factor |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1.671691 | -1.087233 | 0.424095 | -0.485781 | -0.802685 | -0.333515 | 1.008032 | 0.745849 | -0.780496 | 0.172957 | 0.749969 | 1.232879 |
| 1 | 0.790462 | -1.087233 | -0.475682 | -0.485781 | -0.802685 | -0.250455 | 1.008032 | -0.586025 | 1.535516 | -0.881660 | 0.749969 | 1.222961 |
| 2 | -0.971995 | 0.595356 | 0.424095 | -1.492930 | 1.042922 | -0.333515 | -0.992032 | -1.917899 | -0.780496 | 0.172957 | -0.761049 | 1.202120 |
| 3 | 1.671691 | 0.595356 | 1.323873 | 1.528516 | -0.802685 | -0.333515 | 1.008032 | -0.586025 | -0.780496 | 1.227575 | -0.761049 | -0.830503 |
| 4 | -0.971995 | -0.245939 | 0.424095 | 1.024942 | -0.802685 | -0.250455 | 1.008032 | -1.917899 | -0.780496 | 0.172957 | 0.749969 | -0.830503 |

Gambar 2. Hasil Preprocessing Data

### 3.2 Penentuan Jumlah Cluster

Untuk mengidentifikasi jumlah klaster terbaik, penelitian ini mengandalkan *Elbow Method*. Teknik ini bekerja dengan menetapkan nilai $K$ paling efisien berdasarkan parameter *Within Cluster Sum of Squares* (WCSS). Selanjutnya, tren penurunan nilai WCSS tersebut dipetakan ke dalam bentuk kurva atau grafik *elbow* guna menentukan titik patahan (*elbow point*) yang tepat.




Gambar 3. Grafik Elbow Method

Merujuk pada visualisasi *Elbow Method*, titik patahan (*elbow point*) yang krusial mulai terbentuk secara konsisten pada nilai $K = 3$. Berpijak pada karakteristik kurva tersebut, penelitian ini menetapkan penggunaan tiga klaster dalam mengeksekusi proses pengelompokan tingkat kepadatan lalu lintas.

### 3.3 Hasil Clustering Menggunakan Algoritma K-Means

Setelah jumlah kelompok ditetapkan, algoritma K-Means dijalankan dengan menetapkan parameter $K$ sebanyak tiga klaster. Output dari proses pengelompokan ini menghasilkan tiga partisi data terpisah, yang masing-masing memiliki atribut unik dalam menggambarkan tingkat kepadatan lalu lintas.

Tabel 1. Jumlah Data Pada Setiap Cluster

| Cluster | Jumlah Data |
|---|---|
| 0 | 1239 |
| 1 | 1082 |
| 2 | 679 |

Implementasi algoritma K-Means berhasil membagi dataset menjadi tiga klaster utama, yang secara berurutan diberi label klaster 0, klaster 1, dan klaster 2. Pengelompokan ini didasarkan pada tingkat kesamaan atribut lalu lintas yang melekat pada setiap entitas data perjalanan, di mana data dengan karakteristik serupa disatukan ke dalam kelompok yang sama.

```
--- 5 Data Teratas Hasil Pengelompokan Kemacetan ---
                start_point                  end_point  \
0    West Jakarta (Jakarta Barat)     East Jakarta (Jakarta Timur)
1  South Jakarta (Jakarta Selatan)     East Jakarta (Jakarta Timur)
2  Central Jakarta (Jakarta Pusat)  South Jakarta (Jakarta Selatan)
3    West Jakarta (Jakarta Barat)  South Jakarta (Jakarta Selatan)
4  Central Jakarta (Jakarta Pusat)    North Jakarta (Jakarta Utara)

   traffic_condition  Cluster_Kemacetan
0                5.0                  0
1                5.0                  0
2                9.0                  2
3                5.0                  1
4                5.0                  2
```

Gambar 4. Hasil Pengelompokan Data Kemacetan

Data pada Gambar 4 menunjukkan bahwa seluruh sampel perjalanan telah mendapatkan label kelompok yang mengidentifikasi tingkat kemacetan spesifik. Contohnya, pergerakan rute dari Jakarta Barat ke Jakarta Timur dengan parameter `traffic_condition` senilai 5 masuk ke dalam kategori Klaster 0. Di sisi lain, variasi rute lainnya tersebar ke Klaster 1 dan Klaster 2 sesuai dengan kedekatan atribut yang dimiliki. Hasil pengelompokan ini menegaskan keandalan metode K-Means dalam memetakan dataset transportasi berdasarkan kemiripan pola situasi jalan raya.

Berdasarkan hasil clustering, setiap cluster menunjukkan karakteristik kondisi lalu lintas yang berbeda. Cluster dengan nilai traffic condition dan faktor keterlambatan yang rendah dapat diinterpretasikan sebagai kondisi lalu lintas lancar. Cluster dengan nilai sedang menunjukkan kondisi lalu lintas padat, sedangkan cluster dengan nilai tertinggi menunjukkan kondisi lalu lintas macet.


Gambar 5. Visualisasi Hasil Clustering

Hasil pemetaan *scatter plot* membuktikan bahwa metode K-Means berhasil memisahkan data perjalanan ke dalam tiga kelompok berdasarkan kemiripan pola atributnya. Perbedaan distribusi data pada setiap klaster tersebut memberikan kontribusi penting dalam mengidentifikasi sekaligus memetakan tipologi kepadatan lalu lintas.

## 3.4 Evaluasi Hasil Clustering

Pengujian performa model diimplementasikan dengan memanfaatkan metode *Silhouette Score* serta *Davies-Bouldin Index* (DBI). Kedua metrik validasi tersebut diaplikasikan sebagai instrumen ukur untuk menilai tingkat keandalan, kerapatan, dan kualitas dari struktur klaster yang telah dihasilkan oleh algoritma.

Tabel 2. Hasil Evaluasi Clustering

| Metode Evaluasi | Nilai |
|---|---|
| Silhouette Scorel | 0.1181 |
| Davies-Bouldin Index | 2.5221 |

Tingkat kedekatan objek dalam suatu kelompok dan efektivitas pemisahannya dengan kelompok lain ditandai oleh nilai *Silhouette Score* yang mendekati angka 1. Di sisi lain, keandalan model berdasarkan *Davies-Bouldin*

*Index* diukur dari semakin rendahnya perolehan nilai indeks, yang mengindikasikan bahwa setiap klaster memiliki karakteristik yang unik dan minim tumpang tindih.

3.5 Pembahasan

Hasil analisis menunjukkan bahwa metode K-Means berhasil membagi data transportasi menjadi tiga kelompok berdasarkan karakteristik kemacetan jalan raya. Melalui pemanfaatan variabel-variabel operasional perjalanan, algoritma ini mampu menciptakan partisi data dengan tingkat kepadatan yang kontras satu sama lain. Ditinjau dari fase visualisasi dan evaluasi metrik, K-Means menunjukkan performa yang cukup baik dalam memetakan pola situasi lalu lintas secara akurat berdasarkan atribut proksimitas datanya.

Struktur klaster yang dihasilkan selanjutnya dapat diaplikasikan untuk menyajikan informasi mengenai kondisi riil di berbagai rute perjalanan guna mendukung sistem manajemen transportasi cerdas. Melalui interpretasi karakteristik tiap kelompok, pola penumpukan kendaraan pada jam-jam tertentu dapat diidentifikasi dengan lebih mudah. Hasil ini juga selaras dengan rujukan ilmiah dari Annisa dan Putra, yang menyatakan bahwa pendekatan K-Means sangat efektif untuk mengelompokkan kondisi lalu lintas berdasarkan kemiripan atribut datanya.

## IV. KESIMPULAN

Rangkaian uji coba dalam penelitian ini membuktikan keberhasilan penerapan metode K-Means *clustering* untuk mengelompokkan tingkat kemacetan lalu lintas ke dalam tiga tingkatan situasi, yakni Lancar, Padat, dan Macet Total. Melalui metrik validasi eksternal, performa model ini menunjukkan perolehan indeks *Silhouette Score* di angka 0,1181 serta skor *Davies-Bouldin Index* (DBI) mencapai 2,5221.

Output dari evaluasi numerik ini merepresentasikan bahwa dataset transportasi urban memiliki pola penyebaran yang sangat variatif dengan tingkat interpenetrasi (*overlapping*) antar-kelas yang masif. Kondisi tersebut memicu pembentukan struktur klaster yang bersifat cair dan kurang rigid, terutama saat dihadapkan pada parameter spasio-temporal mentah berupa metrik jarak (*distance*) serta waktu tempuh (*travel_time*).

Secara aplikatif, hasil pengelompokan ini tetap memberikan nilai guna sebagai referensi awal bagi otoritas terkait dalam mengidentifikasi pola kepadatan arus lalu lintas. visualisasi kelompok yang dihasilkan dapat mempermudah pemetaan titik-titik rawan kemacetan secara spasial. Hal ini penting untuk mendukung pengambilan keputusan yang cepat terkait manajemen rekayasa lalu lintas di lapangan.

Sementara itu, untuk mengatasi kelemahan model saat ini, penelitian berikutnya disarankan untuk memprioritaskan optimasi data sebelum memasuki algoritma klasterisasi. Beberapa langkah taktis yang dapat diambil meliputi eliminasi nilai *outliers* secara ketat serta transformasi skala fitur memanfaatkan *MinMax Scaler* atau *Robust Scaler*. Pendekatan ini diharapkan mampu mereduksi efek *noise* pada parameter spasio-temporal sehingga kualitas pemisahan antar-klaster dapat ditingkatkan secara optimal.

## REFERENSI

[1] R. Annisa and D. Putra, “Penerapan K-Means clustering untuk analisis kondisi lalu lintas di Surabaya,” Jurnal Transportasi Indonesia, vol. 7, no. 2, pp. 120–130, 2025.

[2] M. Siregar and A. Wijaya, “Perbandingan algoritma K-Means dan K-Medoids pada data pelanggaran lalu lintas,” Jurnal Sains dan Teknologi Informasi, vol. 9, no. 1, pp. 45–53, 2024.

[3] Y. Pratama and B. Nugroho, “Analisis kemacetan lalu lintas pada ruas Jalan Daan Mogot Jakarta,” Jurnal Teknik Transportasi Terapan, vol. 6, no. 1, pp. 55–63, 2025.

[4] H. Saputra and F. Kurniawan, “Analisis kinerja lalu lintas pada simpang bersinyal berdasarkan PKJI 2023,” Jurnal Transportasi, vol. 25, no. 1, pp. 1–10, 2025.

[5] A. Rahman and T. Hidayat, “Klasterisasi wilayah rawan kriminalitas menggunakan algoritma K-Means,” Jurnal INSTEK, vol. 10, no. 1, pp. 88–96, 2025.

[6] L. Zhang and X. Chen, “Traffic congestion clustering using K-Means based on GPS data,” IEEE Access, vol. 11, pp. 12345–12355, 2023.

[7] R. Kumar and P. Singh, “Smart traffic analysis using K-Means clustering algorithm,” Procedia Computer Science, vol. 230, pp. 321–330, 2024.

[8] S. Hidayati and R. Prakoso, “Penerapan data mining untuk analisis kemacetan di DKI Jakarta,” Jurnal Teknik Komputer, vol. 8, no. 2, pp. 100–108, 2023.

[9] D. Utami and E. Saputri, “Analisis kepadatan lalu lintas menggunakan algoritma K-Means,” Jurnal Teknologi dan Manajemen, vol. 5, no. 2, pp. 77–85, 2023.

[10] M. Garcia and J. Lopez, “Comparison of DBSCAN and K-Means in traffic data clustering,” Soft Computing, vol. 28, pp. 5678–5689, 2024.

[11] Y. Wang and H. Li, “Traffic pattern clustering based on time series using K-Means,” Sustainability, vol. 17, no. 3, p. 1234, 2025.

[12] A. R. Lestari and M. I. S. Sembiring, “Implementasi algoritma K-Means untuk pemetaan wilayah tingkat kemacetan lalu lintas pada kota besar,” Jurnal Komputer dan Teknologi Informasi, vol. 11, no. 2, pp. 142–150, 2024.

[13] K. Adi and S. Wahyuni, “Analisis klasterisasi arus lalu lintas menggunakan metode K-Means berdasarkan data sensor kendaraan,” Jurnal Informatika dan Sistem Pengambilan Keputusan, vol. 6, no. 1, pp. 89–98, 2025.

[14] T. Nguyen and V. Pham, “Urban traffic congestion classification using an optimized K-Means clustering approach,” International Journal of Intelligent Transportation Systems Research, vol. 22, no. 2, pp. 210–221, 2024.

[15] F. Fahrezi and R. Wardoyo, “Pemanfaatan data floating car untuk pengelompokan pola kemacetan jalan arteri menggunakan K-Means,” Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 11, no. 3, pp. 455–464, 2024.

[16] M. Rossi and G. Bianchi, “Evaluating road traffic density and congestion levels through K-Means data mining techniques,” Transportation Research Procedia, vol. 75, pp. 812–821, 2025.