INSTITUT TEKNOLOGI-PLN

SKRIPSI

ANALISIS SEGMENTASI RUAS JALAN BERDASARKAN KARAKTERISTIK KECEPATAN LALU LINTAS DI DKI JAKARTA MENGGUNAKAN K-MEANS CLUSTERING

Disusun Oleh:

Dea Enjelika

Nim : 202231048

PROGRAM STUDI S1 TEKNIK INFORMATIKA

FAKULTAS TELEMATIKA ENERGI

INSTITUT TEKNOLOGI PLN

JAKARTA 2026

LEMBAR PERSETUJUAN PROPOSAL SKRIPSI

Judul Proposal Skripsi	: Analisis Segmentasi Ruas Jalan Berdasarkan Karakter Kecepatan Lalu Lintas Di DKI Jakarta Menggunakan K-Means Clustering

Identitas Penelitian	
a. Nama Mahasiswa	: Dea Enjelika
b. NIM	202231048
c. Program studi	: S1 Teknik Informatika
d. Fakultas/Sekolah	: Fakultas Telematika Dan Energi
e. No.HP	081368193107
f. Email	: Dea2231048@itpln.ac.id
Jangka Waktu Penelitian	
a. Waktu Mulai	: 01 April 2026
b. Waktu Selesai	: 10 Juli 2026
Lokasi Penelitian	
a. Nama Instansi	: Institut Teknologi Perusahaan Listrik Negara
b. Alamat Instansi	: Menara PLN, Jl. Lingkar Luar Barat, Duri Kosambi, Cengkareng, Jakarta Barat, DKI Jakarta, 11750
Nama dosen pembimbing	: Pritasari Palupiningsih S.Kom., M.Kom

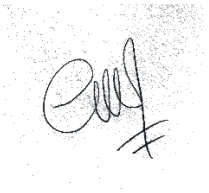
Jakarta, 12 April 2026

Disetujui, Nama Pembimbing



(Pritasari Palupiningsih S.Kom., M.Kom)

Nama Mahasiswa



(Dea Enjelika)

Mengetahui
Kepala Program Studi



(Sely Karmila, S.Kom, M.Si)

PERNYATAAN KEASLIAN SKRIPSI

Nama Mahasiswa	: Dea Enjelika
NIM	: 202231048
Program Studi	: S1 Teknik Informatika
Fakultas/Sekolah	: Telematika Energi
Judul Tugas Akhir	: Analisis segmentasi ruas jalan berdasarkan karakterristik kecepatan lalu lintas di DKI Jakarta menggunakan K-Means Clustering

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar Sarjana baik di lingkungan Institut Teknologi PLN maupun di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka. Pernyataan ini dibuat dengan penuh kesadaran dan rasa tanggung jawab serta bersedia bertanggung jawab segala resiko jika ternyata pernyataan ini tidak benar.

Jakarta, 23 Juli 2026



Dea Enjelika

LEMBAR PENGESAHAN SKRIPSI

Judul tugas akhir	: Analisis Segmentasi Ruas Jalan Berdasarkan Karakteristik Kecepatan lalu lintas di DKI Jakarta Menggunakan K-Means Clustering
Nama mahasiswa	: Dea Enjelika
NIM	202231048
Program studi	: Teknik Informatika
Fakultas	: Telematika Energi

Disetujui oleh

Pembimbing 1

Pritasari Palupiningsih, S.Kom., M.Kom

NIDN : 0301027703

Diketahui Oleh

Kepala Program Studi

Sely Karmila, S.Kom., M.Si

NIDN : 0320047601

Dekan

Dr. Meilia Nur Susanti, S.T., M.Kom

NIDN : 0319057601

LEMBAR PENGESAHAN DOSEN PENGUJI

Judul tugas akhir : Analisis Segmentasi Ruas Jalan Berdasarkan Karakteristik Kecepatan lalu lintas di DKI Jakarta Menggunakan K-Means Clustering

Nama mahasiswa : Dea Enjelika

NIM 202231048

Program studi : Teknik Informatika

Fakultas : Telematika Energi

Disetujui oleh

Ketua Penguji

Dr. Puji Catur Siswipraptini, S.T., MTI

NIDN : 0301027703

Anggota Penguji I

Karina Djunaidi, S.T., MTI

NIDN : 0313058902

Anggota Penguji II/Pembimbing

Pritasari Palupiningsih, S.Kom., M.Kom

NIDN :0328018803

Diketahui oleh

Kepala Program Studi

Sely Karmila, S.Kom., M.Si

NIDN : 0320047601

KATA PENGANTAR

Alhamdulillah, dengan segala puji dan syukur saya panjatkan kepada Allah SWT yang telah melimpahkan segala rahmat dan karunia-Nya, sehingga skripsi yang berjudul

"**Analisis Segmentasi Ruas Jalan Berdasarkan Karakteristik Kecepatan Lalu lintas Di DKI Jakarta Menggunakan K-Means Clustering**" dapat terselesaikan, Shalawat dan salam senantiasa tercurah kepada Nabi Muhammad SAW, yang telah membawa umat manusia menuju jalan ilmu pengetahuan dan kebaikan.

Adapun tujuan dari Skripsi ini adalah untuk memenuhi salah satu persyaratan

akademis guna menyelesaikan Program Studi Strata Satu (S1) Teknik Informatika Fakultas Telematika Energi Institut Teknologi PLN Jakarta.

Skripsi ini tidak akan terselesaikan tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, penulis menyampaikan rasa terima kasih kepada berbagai pihak atas arahan, bimbingan, serta dukungan selama penyusunan Skripsi ini. Pada kesempatan ini, penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. Iwa Garniwa Mulyana K., MT selaku Rektor di Institut Teknologi PLN Jakarta.
2. Ibu Meilia Nur Indah S., ST., M.Kom selaku Dekan Fakultas Telematika Energi

   Intitut Teknologi PLN Jakarta dan Ibu Sely Karmila, S.Kom, M.Si selaku Kepala Program Studi S1 Teknik Informatika Institut Teknologi PLN Jakarta.
3. Ibu Pritasari Palupiningsih S.Kom., M.kom. Selaku Dosen Pembimbing
4. Seluruh dosen Fakultas Telematika Energi Institut Teknologi PLN Jakarta yang

   telah memberikan ilmu yang bermanfaat selama penulis menimba ilmu di bangku perkuliahan.

5. Allah SWT yang selalu memberi penulis hikmat, kelancaran, kesehatan, kekuatan, dan kebahagiaan untuk menyelesaikan penelitian ini.

6. Kepada Keluarga Mama, Papa, dan Keponakan yang telah memberikan dukungan perhatian, Kebutuhan, kondisi, selama Penulis melakukan proses mengerjakan Skripsi.
7. Sahabat dari The Mungiels (Mica, Luthfi, Adel, Meta, Thanap), yang mana telah memberikan Penulis semangat dan termotivasi.
8. Sahabat dari *Noxy* (Kak Septi, Vala, Anin, najel, Yoci) yang telah memberikan penulis Kebahagian, meluangkan waktu dan bermain Game, serta mendengarkan penulis menghadapi rasa lelah saat menyelesaikan penelitian ini.
9. Terakhir kepada Sahabat *Racing* (Renata, Widy, Ajeng, Rasti, Faza), yang telah memberikan semangat, motivasi, saran dan kebersamaan selama penulis menempuh pendidikan. Terima Kasih atas setiap perjalanan yang telah dilalui bersama.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan belum sepenuhnya sempurna. Oleh karena itu, penulis terbuka terhadap kritik dan saran yang membangun sebagai bahan perbaikan di masa mendatang. Penulis berharap skripsi ini dapat memberikan manfaat bagi pembaca serta menjadi tambahan referensi dalam pengembangan penelitian.

Akhir kata, semoga seluruh proses penyusunan skripsi ini menjadi pengalaman yang bermakna dan dapat menjadi bekal bagi penulis untuk terus belajar, berkembang, serta memberikan kontribusi yang bermanfaat pada masa yang akan datang.

ANALISIS SEGMENTASI RUAS JALAN BERDASARKAN KARAKTERISTIK KECEPATAN LALU LINTAS DI DKI JAKARTA MENGGUNAKAN K-MEANS CLUSTERING

Dea Enjelika, 202231048

Di bawah bimbingan: Pritasari Palupiningsih., S.Kom., M.Kom

ABSTRAK

Kemacetan lalu lintas merupakan masalah utama di kawasan perkotaan, terutama di Jakarta, yang dipicu oleh tingkat mobilitas yang tinggi serta ketidakseimbangan antara jumlah kendaraan dan kapasitas jalan. Salah satu indikator utama untuk mengidentifikasi tingkat kemacetan adalah kecepatan lalu lintas, di mana penurunan kecepatan mencerminkan peningkatan kepadatan lalu lintas. Berdasarkan tinjauan pustaka, berbagai penelitian menunjukkan bahwa metode pengelompokan, khususnya K-Means, efektif dalam mengelompokkan kondisi lalu lintas dan mengungkap pola kemacetan yang tersembunyi. Selain itu, penggunaan metode evaluasi seperti Elbow Method dan Silhouette Score telah terbukti mampu menentukan jumlah kluster yang optimal dan mengukur kualitas hasil pengelompokan.Penelitian ini bertujuan untuk menganalisis segmentasi ruas jalan di DKI Jakarta berdasarkan karakteristik kecepatan lalu lintas menggunakan metode K-Means Clustering. Data kecepatan lalu lintas dikumpulkan, diproses, dan diubah menjadi fitur analitis seperti kecepatan rata-rata, minimum, maksimum, dan variasinya. Selanjutnya, proses pengelompokan dilakukan untuk mengelompokkan ruas jalan berdasarkan kesamaan karakteristik kecepatan. Jumlah kluster optimal ditentukan menggunakan metode Elbow, sedangkan kualitas kluster dievaluasi menggunakan Silhouette Score.Hasil penelitian ini diharapkan dapat menghasilkan klasifikasi ruas jalan yang mencerminkan tingkat kemacetan lalu lintas, seperti kondisi lalu lintas lancar, padat, dan macet total. Dengan demikian, penelitian ini dapat berkontribusi pada pemahaman yang lebih terstruktur mengenai pola lalu lintas serta menjadi landasan bagi pengambilan keputusan yang bertujuan untuk meningkatkan efisiensi mobilitas dan sistem transportasi di Jakarta.

Kata Kunci : Kemacetan Lalu Lintas, Kecepatan Lalu Lintas, *K-Means Clustering,* Segmentasi Ruas Jalan, *Elbow Method, Silhouette Score*

ANALYSIS OF ROAD SEGMENTATION BASED ON TRAFFIC SPEED CHARACTERISTICS IN DKI JAKARTA USING K-MEANS CLUSTERING

Dea Enjelika, 202231048

Under guidance: Pritasari Palupiningsih., S.Kom., M.Kom

ABSTRACT

Traffic congestion is a major problem in urban areas, particularly in Jakarta, caused by high levels of mobility and an imbalance between the number of vehicles and road capacity. One of the key indicators for identifying the level of congestion is traffic speed, where a decrease in speed reflects an increase in traffic density. Based on a literature review, various studies show that clustering methods, particularly K-Means, are effective in grouping traffic conditions and uncovering hidden congestion patterns. Additionally, the use of evaluation methods such as the Elbow Method and Silhouette Score has proven effective in determining the optimal number of clusters and measuring the quality of clustering results. nhnThis study aims to analyze the segmentation of road segments in DKI Jakarta based on traffic speed characteristics using the K-Means Clustering method. Traffic speed data was collected, processed, and converted into analytical features such as average, minimum, and maximum speeds, as well as their variation. Next, a clustering process was performed to group road segments based on similarities in speed characteristics. The optimal number of clusters was determined using the Elbow method, while cluster quality was evaluated using the Silhouette Score. The results of this study are expected to produce a classification of road segments that reflects traffic congestion levels, such as smooth, heavy, and total gridlock conditions. Thus, this study can contribute to a more structured understanding of traffic patterns and serve as a foundation for decision-making aimed at improving mobility efficiency and the transportation system in Jakarta.

Keywords: Traffic Congestion, Traffic Speed, K-Means Clustering, Road Segment Segmentation, Elbow Method, Silhouette Score

DAFTAR ISI

LEMBAR PERSETUJUAN PROPOSAL SKRIPSI i
KATA PENGANTAR ii
ABSTRAK iv
ABSTRACT v
DAFTAR ISI vii
DAFTAR TABEL viii
DAFTAR GAMBAR ix
DAFTAR LAMPIRAN x
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 2
1.3 Tujuan Penelitian 2
1.4 Manfaat Penelitan 2
1.5 Ruang Lingkup 3
1.6 Kebaharuan 3
BAB II TINJAUAN PUSTAKA 4
2.1 Studi Literatur 4
2.2 Landasan Teori 7
2.2.1 Kemacetan Lalu Lintas ***8***
2.2.2 Clustering ***10***
2.2.3 Algoritma K-Means ***11***
2.2.4 Penentuan Jumlah Cluster ***12***
2.2.5 Metod Evaluasi Cluster ***12***
2.2.6 Crisp DM ***13***
2.2.7 Unifiead Modeling Language ***14***
2.3 Kerangka Pemikiran 14
BAB III METODE PENELITIAN 17
3.1 Lokasi dan Waktu Penelitian 17
3.1.1 Lokasi Penelitian ***17***
3.1.2 Waktu Penelitian ***17***
3.2 Desain Penelitian 17

3.2.1 Business Understanding *18*
3.1.2 Data Understanding *19*
3.1.3 Data Preparation *20*
3.1.4 Modelling *24*
3.1.5 Evaluation Model *30*
BAB IV 32
4.1 Hasil 32
4.1.1 Deskripsi Dataset *32*
4.1.2 Hasil Data Preparation *33*
4.1.3 Hasil Elbow Method *39*
4.1.4 Hasil Silhouette Score *42*
4.1.6 Visualisasi *47*
4.2 Pembahasan 51
4.2.1 Interpretasi Cluster *51*
4.2.2 Analisis Pola *51*
4.2.3 Alasan Pemilihan Nilai K *53*
BAB V PENUTUP 54
5.1 Kesimpulan 54
5.2 Saran 55
DAFTAR PUSTAKA 57
DAFTAR RIWAYAT HIDUP 60
LAMPIRAN 61

DAFTAR TABEL

Tabel 2. 1 Matriks Penelitain 4
Tabel 3. 1 Jumlah Observasi Berdasarkan Wilayah 20
Tabel 3. 2 Data Cleaning 21
Tabel 3. 3 Normalization scaler 24
Tabel 3. 4 Inisialisasi Data dan Centroid 26
Tabel 3. 5 Table Simulation SSE 27
Tabel 3. 6 Result Silhoutte score 30
Tabel 4. 1 Deskripsi Dataset 32
Tabel 4. 2 Grafik Elbow 39
Tabel 4. 3 silhouette Score 42
Tabel 4. 4 Ringkasan Hasil Clustering 45
Tabel 4. 5 Pola Temporal variabelitas kecepatan 52

DAFTAR GAMBAR

Gambar 2. 1 CrispDM 13
Gambar 2. 3 Kerangka Pemikiran 15
Gambar 3. 1 Desain Research 18
Gambar 3. 2 Data Segmen Ruas Jalan II 22
Gambar 3. 3 Data Segemen Ruas Jalan II 22
Gambar 3. 4 Elbow Method 25
Gambar 3. 5 Shiloutte Score Grafik 31
Gambar 4. 1 Missing value Dataset 34
Gambar 4. 2 Anomali Value 34
Gambar 4. 3 Parsing Periode 35
Gambar 4. 4 EDA Dataset 36
Gambar 4. 5 Grafik Elbow Method 41
Gambar 4. 6 Grafik Silhouette Score untuk Berbagai Nilai k 44
Gambar 4. 7 PCA visualization 48
Gambar 4. 8 Visualization Categori Clsuter 49
Gambar 4. 9 Visulization Scater 50

DAFTAR LAMPIRAN

Lampiran 1 Lembar Bimbingan Skripsi 61
Lampiran 2 Pemohonan Mengikuti Sidang Skripsi 62
Lampiran 3 Surat Keterangan Persyaratan Akademik dan Keuangan 63
Lampiran 4 Surat Keterangan Bebas Administrasi 64
Lampiran 5 Surat Keterangan Bebas Pinjam Alat Laboratorium 65

SKRIPSI

Analisis Segmentasi Ruas Jalan Berdasarkan Karakteristik Kecepatan lalu lintas di DKI Jakarta menggunakan Kinerja K-Means Clustering

Disusun Oleh :

Dea Enjelika

NIM : 202231048

Diajukan untuk memenuhi persyaratan

PROGRAM STUDI S1 TEKNIK

INFORMATIKA FAKULTAS TELEMATIKA

ENERGI INSTITUT TEKNOLOGI PLN

Jakarta, 23 Juli 2026

Mengetahui	**Disetujui**
Kepala Program Studi S1 Teknik Informatika	**Dosen Pembimbing Utama**
	sidang 24/7/26 24/7
Sely Karmila, S.Kom., M.Si	**Pritasari Palupiningsih S.Kom., M.Kom**
NIDN : 0320047601	**NIDN : 0328018803**

BAB I
PENDAHULUAN

1.1 Latar Belakang

Kemacetan lalu lintas merupakan salah satu masalah paling umum di kota-kota besar, termasuk Jakarta. Seiring dengan pertumbuhan populasi dan aktivitas perkotaan, permintaan akan transportasi pun meningkat. Namun, pertumbuhan ini tidak selalu diimbangi dengan kapasitas jalan yang memadai, sehingga sering kali menyebabkan kemacetan, terutama pada jam-jam sibuk . Kemacetan lalu lintas tidak hanya memengaruhi waktu tempuh, tetapi juga berdampak pada berbagai aspek kehidupan masyarakat. Aktivitas sehari-hari menjadi kurang efisien akibat terganggunya mobilitas, baik untuk bekerja, bersekolah, maupun kegiatan lainnya. Selain itu, kemacetan berdampak pada sektor logistik, di mana distribusi barang menjadi lebih lambat dan kurang efektif. Hal ini tak terhindarkan meningkatkan biaya operasional dan berdampak pada perekonomian secara keseluruhan (Oce Ovalina Yumame et al., 2025).

Di sisi lain, tingginya penggunaan kendaraan pribadi juga merupakan salah satu penyebab utama kemacetan lalu lintas. Jumlah kendaraan yang besar, yang tidak sebanding dengan kapasitas jalan, membuat kondisi lalu lintas semakin padat. Situasi ini semakin rumit akibat tingginya tingkat aktivitas masyarakat yang terjadi hampir sepanjang hari (Kilang Permatasari & Kristiyanto, 2025). Dalam menganalisis kemacetan lalu lintas, salah satu indikator yang paling sering digunakan adalah kecepatan kendaraan. Ketika kecepatan kendaraan menurun, hal itu biasanya menandakan adanya kemacetan atau hambatan di ruas jalan tertentu. Oleh karena itu, data kecepatan lalu lintas dapat digunakan untuk mengidentifikasi pola kemacetan dengan lebih jelas (Li, 2024a).

Kemajuan teknologi kini memungkinkan analisis data lalu lintas dilakukan dengan lebih mudah dan akurat. Salah satu metode yang dapat digunakan adalah K-Means Clustering, yang mengelompokkan data berdasarkan karakteristik yang serupa. Dalam konteks penelitian ini, metode ini digunakan untuk mengelompokkan ruas jalan berdasarkan pola kecepatan kendaraan, sehingga dapat menentukan tingkat kemacetan pada setiap ruas jalan (Etemad et al., 2023). Penelitian sebelumnya telah menunjukkan bahwa metode pengelompokan dapat secara efektif membantu dalam memahami pola lalu lintas dan mengidentifikasi tingkat kemacetan. Namun, masih terdapat keterbatasan dalam studi-studi tersebut, terutama terkait belum dipertimbangkannya karakteristik fisik jalan (seperti jenis jalan arteri, kolektor, atau lokal) serta waktu pengambilan data (misalnya pada jam sibuk atau di luar jam sibuk) yang sangat memengaruhi dinamika kecepatan lalu lintas. Penelitian yang secara khusus mengelompokkan ruas jalan di Wilayah Metropolitan Jakarta dengan mempertimbangkan kedua faktor tersebut masih sangat terbatas.

Berdasarkan hal tersebut, penelitian ini akan mengisi celah tersebut dengan mengalihkan fokus analisis dari pendekatan konvensional berbasis wilayah administratif ke pendekatan berbasis ruas jalan secara lebih spesifik. Dengan memanfaatkan data kecepatan lalu lintas sebagai indikator utama, penelitian ini menerapkan metode K-Means Clustering untuk mengelompokkan ruas jalan berdasarkan kemiripan pola kecepatan yang teramati. Pendekatan ini memungkinkan terbentuknya segmentasi ruas jalan yang lebih terstruktur, di mana setiap kelompok cluster yang dihasilkan merepresentasikan tingkat kemacetan yang berbeda mulai dari kondisi lancar hingga padat. Hasil dari pengelompokan ini diharapkan tidak hanya memberikan gambaran spasial yang lebih rinci mengenai dinamika lalu lintas di DKI Jakarta, kepentingan dalam merumuskan kebijakan transportasi yang lebih tepat sasaran, guna meningkatkan mobilitas masyarakat dan karakteristik jalan umum di DKI Jakarta.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah.

1. Bagaimana karakteristik kecepatan lalu lintas pada ruas jalan di DKI Jakarta?
2. Bagaimana hasil pengelompokan (clustering) ruas jalan menggunakan metode K-Means berdasarkan data kecepatan lalu lintas?
3. Bagaimana interpretasi hasil clustering tersebut dalam menggambarkan tingkat kemacetan pada masing-masing ruas jalan?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah.

1. Menganalisis karakteristik kecepatan lalu lintas pada ruas jalan di DKI Jakarta.
2. Mengelompokkan ruas jalan berdasarkan kemiripan pola kecepatan menggunakan metode K-Means Clustering.
3. Menginterpretasikan hasil pengelompokan tersebut sebagai representasi tingkat kemacetan pada setiap ruas jalan.

1.4 Manfaat Penelitan

Manfaat penelitian ini Meliputi.

1. Penelitian ini diharapkan dapat menambah referensi dalam bidang data mining, khususnya penerapan metode K-Means Clustering pada analisis transportasi dan kemacetan lalu lintas.
2. Memberikan gambaran mengenai tingkat kemacetan pada ruas jalan di

DKI Jakarta berdasarkan data kecepatan.
3. Membantu pihak terkait (pemerintah DKI Jakarta) dalam memahami pola kemacetan secara lebih terstruktur.

1.5 Ruang Lingkup

Ruang Lingkup dari penelitian ini adalah sebagai berikut.

1. Penelitian difokuskan pada wilayah DKI Jakarta.
2. Data yang digunakan berupa data kecepatan lalu lintas pada ruas jalan.
3. Objek penelitian adalah ruas jalan, bukan wilayah administratif seperti kecamatan atau kota.
4. Metode yang digunakan adalah K-Means Clustering untuk melakukan pengelompokan data.
5. Analisis dilakukan untuk menggambarkan tingkat kemacetan berdasarkan karakteristik kecepatan.
6. Penelitian ini tidak mempertimbangkan faktor lain seperti volume kendaraan, kondisi cuaca, maupun kejadian insidental (kecelakaan, dll).

1.6 Kebaharuan

Kebaruan dalam penelitian ini terletak pada pendekatan yang digunakan dalam menganalisis kemacetan lalu lintas, yaitu dengan.

1. Menggunakan data kecepatan lalu lintas sebagai indikator utama untuk merepresentasikan tingkat kemacetan.
2. Menerapkan metode K-Means Clustering untuk mengelompokkan ruas jalan berdasarkan pola kecepatan, sehingga menghasilkan segmentasi yang lebih terstruktur.
3. Mengalihkan fokus analisis dari pendekatan konvensional berbasis wilayah menjadi berbasis ruas jalan, sehingga hasil analisis lebih spesifik dan relevan terhadap kondisi nyata di lapangan.
4. Menghasilkan klasifikasi ruas jalan berdasarkan tingkat kemacetan yang dapat digunakan sebagai dasar untuk memahami pola mobilitas dan potensi permasalahan transportasi di DKI Jakarta.

BAB II
TINJAUAN PUSTAKA

2.1 Studi Literatur

Pada bagian *studi literatur review* dipaparkan beberapa penelitian terdahulu yang memiliki keterkaitan dengan topik penelitian. Penelitian-penelitian tersebut digunakan sebagai bahan kajian untuk memahami perkembangan topik, pendekatan yang telah digunakan, serta menemukan celah penelitian yang masih dapat dikembangkan. Berikut detail pemabahasan *studi literatur* yang saya gunakan, pada tabel 1.1 matriks penelitian berikut.

Tabel 2. 1 Matriks Penelitain

No	Tahun	Judul	Metodologi	Hasil Penelitian	Relevansi	Rumus	Kesimpulan
1	(Gupta et al., 2025)	Analysis of Traffic Flow Congestion by Integrating Supervised ML with K-Means Clustering	Hybrid K-Means, LR, DT, GB, RF, SVR	Integrasi K-Means meningkatkan performa signifikan - RMSE GB turun dari 2180.85 620.84 - RF 2186.32 632.07 - DT 2188.05 640.80 - SVR 1531.82 795.05 - R^2 meningkat hingga 0.92 - MAE terbaik SVR 301.46 - LR performa buruk karena tidak mampu menangani non-linear data	Sangat relevan karena membuktikan bahwa K-Means mampu menangkap pola kemacetan tersembunyi sebelum analisis lanjutan	K-Means (minimasi jarak) $\sum(xi - \mu k)^2$	K-Means efektif sebagai tahap awal segmentasi trafik, terutama jika dikombinasikan dengan model prediksi
2	(Istiqomah et al., 2025)	Penerapan K-Means Clustering untuk Analisis Kondisi Lalu Lintas di	K-Means, Silhouette Coefficient	- Menghasilkan 4 cluster kondisi lalu lintas - Volume motor >4700 dan DS >0.45 = kondisi padat	Sangat relevan karena sama-sama mengelompokkan kondisi kemacetan berdasarkan indikator lalu lintas	Euclidean Distance $d(x, y) = \sqrt{\sum(xi - yi)^2}$ Silhouette s=(b-a)/max(a, b)	K-Means efektif untuk mengelompokkan kondisi lalu lintas, namun kualitas cluster tergantung fitur

		Surabaya		- Nilai Silhouette = 0.63 (struktur sedang) - Data awal 503 → 364 setelah preprocessing			
3	(Kai Jiang et al., 2025)	Time Period Analysis of Traffic Flow using K-Mean	K-Means + Elbow Method + SSE	- Optimal cluster = 4 waktu - Periode 1 traffic rendah (malam) - Periode 2 traffic tinggi (peak hingga ~3000 kendaraan) - SSE digunakan untuk menentukan K	Relevan karena menunjukkan bahwa pola kemacetan juga bisa dilihat dari dimensi waktu	Volume $Q = N/\Delta t$	K-Means berhasil mengelompokkan pola lalu lintas berdasarkan waktu menjadi 4 cluster yang mencerminkan kondisi off-peak hingga peak, sehingga efektif untuk analisis kemacetan berbasis waktu
4	(Etemad et al., 2023)	Clustering of Urban Patterns by K-Means & DTW	K-Means, Dynamic Time Warping	- Dataset: 2.4 juta data time series - Mengelompokkan jalan menjadi 3 tipe: 1. Jalan kecil 2. Jalan utama 3. Jalan sedang - Digunakan untuk imputasi data kecepatan & visualisasi traffic	Sangat relevan karena menggunakan data kecepatan ruas jalan	DTW distance + K-Means Kompleksitas: O(NTK)	K-Means + DTW efektif untuk data time series seperti kecepatan lalu lintas
5	(Li, 2024b)	Smart City Traffic Data Analysis using Weighted K-Means	Weighted K-Means, Holt Exponential Smoothing	- Analisis 100.000 data → 98.548 valid - Mengidentifikasi pola kemacetan di 5 wilayah - Prediksi traffic akurat (contoh 9.81	Relevan karena menunjukkan pengembangan K-Means untuk data kompleks dan dinamis	Objective function $\sum\sum(CVm(U))^2$ Holt: xt+h = lt + trend	Weighted K-Means meningkatkan akurasi dalam data lalu lintas kompleks

				pcu/h mendekati real) - Mengatasi kelemahan K-Means dengan pembobotan			
6	(filki. widyatami, 2025)	Analisis Kemacetan Lalu Lintas di Daan Mogot	PKJI 2023 (analisis kapasitas & DS)	- Volume max 3084 smp/jam - DS 1.46 – 1.70 (kategori F) - Kecepatan arus bebas 57 km/jam - Menunjukkan kondisi macet parah	Relevan sebagai baseline analisis kemacetan konvensional (tanpa clustering)	Derajat kejenuhan DS = Q/C	Metode konvensional mampu mengukur kemacetan, namun tidak mampu melakukan segmentasi seperti K-Means
7	(Sohail et al., 2024)	Data-driven insights: Unravelling traffic dynamics with k-means clustering and vehicle type differentiation	K-Means Clustering, Multiple Linear Regression	mengidentifikasi 2 cluster optimal dengan nilai *silhouette coefficient* 0.41–0.44. Cluster 0 menunjukkan kondisi pra-macet/macet, sedangkan Cluster 1 adalah arus bebas (*free flow*).	menggunakan K-Means untuk membedakan fase lalu lintas (macet vs lancar) berdasarkan kecepatan dan volume kendaraan.	∑∑(	K-Means efektif mengklasifikasi kondisi lalu lintas perkotaan secara dinamis dan mengungkap pola tersembunyi melampaui statistik rata-rata manual.
8	(Satria Ardi Perdana et al., 2022)	Analisis Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Aplikasi Alfagift	CRISP-DM, K-Means, Elbow Method	Menghasilkan 3 cluster optimal berdasarkan nilai SSE sebesar 1119.28. Segmentasi didasarkan pada umur, jenis kelamin, frekuensi, tipe pembayaran, dan kota.	Relevan sebagai referensi optimasi jumlah cluster (K) menggunakan metode Elbow untuk mendapatkan hasil pengelompokan yang paling stabil.	SSE = $\sum_{(d)}$ 2	Metode *Elbow* sangat membantu menentukan jumlah K terbaik agar strategi yang diambil (pemasaran/kebijakan) menjadi lebih tepat sasaran.
9	(Hafidz Ardan	Segmentasi Pelanggan	Metode Deskriptif, K-Means, RFM Analysis,	Menghasilkan 2 cluster optimal dengan skor silhouette tertinggi	Relevan untuk teknik validasi kualitas	Silhouette Score & RFM (Recenc	Kombinasi analisis RFM dan K-Means memberikan

	a et al., 2024)	Penjualan Online Menggunakan Metode K-means Clustering	Silhouette Analysis	0.54. Cluster 0 memiliki total transaksi tinggi, sementara Cluster 1 memiliki frekuensi transaksi tinggi.	cluster menggunakan Silhouette Analysis guna memastikan segmentasi yang terbentuk bersifat informatif.	y, Frequency, Monetary)	wawasan mendalam mengenai perilaku subjek yang diteliti untuk personalisasi strategi.
10	(Akbar et al., 2025)	Segmentasi Tunggakan Pelanggan Menggunakan Algoritma K-Means Cluster pada Perusahaan Air Minum Daerah	K-Means Clustering, Slovin Formula	Mengelompokkan data ke dalam 3 cluster: C0 (Rendah - 327 data), C1 (Tinggi - 6 data), dan C2 (Sedang - 59 data). Proses konvergen pada iterasi ke-4.	Relevan dalam hal manajemen data besar dan penggunaan parameter numerik (jumlah bulan & total tunggakan) untuk menentukan tindakan tegas.	Relevan dalam hal manajemen data besar dan penggunaan parameter numerik (jumlah bulan & total tunggakan) untuk menentukan tindakan tegas.	Algoritma K-Means terbukti mampu memetakan pola perilaku (pembayaran/ ketaatan) untuk mendukung pengambilan keputusan yang bersifat administratif.

2.2 Landasan Teori

Landasan teori digunakan sebagai dasar konseptual dalam menjelaskan variabel, istilah, dan pendekatan yang digunakan dalam penelitian. Teori-teori yang dibahas pada penelitian ini.

2.2.1 Segmentasi

Segmentasi merupakan suatu pendekatan yang digunakan untuk membagi sekumpulan data menjadi kelompok-kelompok yang berbeda secara lebih spesifik berdasarkan karakteristik dan perilaku yang serupa di dalamnya. Tujuan utama dari penerapan segmentasi ini adalah untuk mempermudah analisis suatu pola, memahami kecenderungan setiap entitas, serta mengidentifikasi kelompok-kelompok yang memiliki atribut atau kebutuhan yang sejenis. Identifikasi kelompok yang terpisah ini pada akhirnya merupakan langkah awal yang akan sangat membantu dalam mengembangkan rencana aksi maupun melakukan analisis lanjutan yang lebih relevan dan terarah sesuai dengan karakteristik masing-masing

kelompok. Dalam penerapannya, salah satu metode yang sangat efektif untuk memfasilitasi segmentasi ini adalah algoritma *K-Means clustering*, yakni sebuah metode pengelompokan berjenis non-hierarkis yang memiliki keunggulan pada waktu komputasi yang relatif cepat. Algoritma ini beroperasi secara sistematis dengan cara membagi data menjadi beberapa cluster yang berbeda murni berdasarkan tingkat kemiripan atribut yang melekat pada data tersebut. Proses pembentukannya dilakukan dengan mengelompokkan satuan data berdasarkan jarak terdekat dengan nilai *centroid* titik pusat awal yang pada mulanya dipilih secara acak. Setiap data tersebut kemudian dihitung jarak terdekatnya umumnya menggunakan rumus *Euclidean Distance* di mana data yang berdekatan dengan centroid tertentu akan bergabung membentuk satu *cluster*, dan proses iterasi ini akan terus berjalan berulang hingga tidak ada lagi perubahan anggota pada setiap kelompoknya(Hafidz Ardana et al., 2024).

2.2.1 Kemacetan Lalu Lintas

Kemacetan lalu lintas adalah kondisi di mana arus kendaraan pada suatu ruas jalan tidak lagi bergerak secara optimal karena volume kendaraan mendekati atau bahkan melebihi kapasitas jalan. Dalam situasi seperti itu, kecepatan kendaraan cenderung menurun, waktu tempuh menjadi lebih lama, dan arus lalu lintas menjadi tidak stabil. Dalam studi lalu lintas di Surabaya, kondisi kemacetan ditandai oleh volume kendaraan yang tinggi disertai dengan tingkat saturasi yang meningkat, yang menunjukkan bahwa kemacetan bukan hanya soal jumlah kendaraan, tetapi juga kapasitas jalan untuk menampung arus tersebut (Istiqomah et al., 2025). Dalam konteks transportasi perkotaan, kemacetan bukan hanya gangguan perjalanan, tetapi juga berdampak pada efisiensi mobilitas, biaya operasional transportasi, dan kualitas lingkungan perkotaan. Studi arus lalu lintas berbasis data menunjukkan bahwa seiring meningkatnya kepadatan, pola lalu lintas bergeser dari keadaan stabil menuju keadaan puncak, dan berpotensi memasuki fase kemacetan yang lebih parah. Dalam penelitian dinamika lalu lintas, pengelompokan berdasarkan jumlah kendaraan, kecepatan, dan jenis kendaraan mengungkapkan bahwa perbedaan pola lalu lintas dapat diidentifikasi melalui kombinasi variabel-variabel tersebut, termasuk penurunan kecepatan sebagai indikator utama perubahan kondisi arus lalu lintas (Sohail et al., 2024).

Kecepatan kendaraan merupakan salah satu indikator paling langsung untuk membaca tingkat kemacetan. Semakin tinggi kecepatan rata-rata kendaraan pada suatu ruas jalan, biasanya semakin lancar kondisi lalu lintasnya. Sebaliknya, ketika kecepatan menurun secara konsisten, hal tersebut menunjukkan adanya hambatan, kepadatan, atau beban lalu lintas yang meningkat. Dalam studi urban *traffic pattern* berbasis data kecepatan, pola kecepatan ruas jalan digunakan sebagai fitur utama untuk membedakan karakteristik jalan dan kondisi lalu lintasnya. Ini menunjukkan bahwa kecepatan tidak hanya berguna sebagai ukuran teknis, tetapi juga dapat menjadi dasar segmentasi kondisi lalu lintas (Etemad et al., 2023). Pada penelitian analisis periode waktu lalu lintas, periode dengan arus rendah seperti malam hingga dini hari ditandai oleh lalu lintas yang relatif stabil, sedangkan periode siang hingga jam sibuk menunjukkan lonjakan volume hingga mendekati 3000 kendaraan pada arah tertentu. Temuan ini mendukung bahwa saat arus meningkat, kecepatan cenderung tertekan dan kondisi jalan bergerak menuju kemacetan (Kai Jiang et al., 2025).

Secara umum, hubungan antara kecepatan dan kondisi lalu lintas dapat dijelaskan sebagai berikut.

1. Kecepatan tinggi lalu lintas lancar
   Ruas jalan masih mampu melayani arus kendaraan dengan baik. Pergerakan kendaraan relatif stabil, hambatan kecil, dan waktu tempuh lebih efisien.
2. Kecepatan sedang lalu lintas padat
   Arus kendaraan mulai meningkat dan ruang gerak kendaraan mulai terbatas. Pada tahap ini lalu lintas belum sepenuhnya macet, tetapi sudah menunjukkan gejala kepadatan.
3. Kecepatan rendah lalu lintas macet
   Kendaraan bergerak lambat, sering berhenti atau mengantre, dan efisiensi perjalanan menurun. Dalam praktiknya, kondisi ini sering berasosiasi dengan volume tinggi, derajat kejenuhan besar, atau bottleneck pada ruas jalan (Istiqomah et al., 2025).

Berdasarkan uraian tersebut, kecepatan dapat digunakan sebagai indikator utama dalam penelitian ini untuk merepresentasikan tingkat kemacetan pada ruas jalan.

2.2.2 Clustering

Pengelompokan (*clustering*) adalah proses mengelompokkan data ke dalam beberapa kluster berdasarkan tingkat kesamaan karakteristiknya. Objek-objek dalam satu kluster memiliki kesamaan yang lebih besar satu sama lain daripada dengan objek-objek di kluster lain 11. Dalam penelitian mengenai pengelompokan pola lalu lintas perkotaan, pengelompokan dijelaskan sebagai proses membagi data menjadi beberapa kelompok sehingga data dalam kelompok yang sama lebih mirip satu sama lain dan lebih berbeda dari kelompok lain. Secara umum, *clustering* adalah bentuk pembelajaran tanpa pengawasan sebuah teknik yang beroperasi tanpa label kelas yang telah ditentukan sebelumnya. Ini berarti bahwa data tidak ditugaskan ke kategori tertentu sejak awal, sebaliknya, pola cluster diidentifikasi langsung dari struktur data yang melekat (Septia Dini Arifta et al., n.d.).

Secara umum, pengelompokan (*clustering*) dibagi menjadi dua pendekatan utama.

1. *Partioning Clustering*
   Pendekatan ini membagi data secara langsung ke dalam sejumlah kluster tertentu yang telah ditentukan sebelumnya. Salah satu metode paling populer dalam kategori ini adalah *K-Means*. Dalam metode ini, setiap titik data ditempatkan pada kluster dengan pusat kluster (centroid) terdekat. Pendekatan ini cocok digunakan ketika peneliti sudah memiliki jumlah kluster tertentu dalam pikiran dan ingin memperoleh hasil segmentasi yang jelas (Ikotun et al., 2023).
2. *Hierarchical Clustering*
   Pendekatan ini membagi data secara langsung ke dalam sejumlah kluster tertentu yang telah ditentukan sebelumnya. Salah satu metode paling populer dalam kategori ini adalah *K-Means*(Safitri et al., 2024). Dalam metode ini, setiap titik data ditempatkan pada kluster dengan pusat kluster (*centroid*) terdekat. Pendekatan ini cocok digunakan ketika peneliti sudah memiliki jumlah kluster tertentu dalam pikiran dan ingin memperoleh hasil segmentasi yang jelas (M Khatami & HB Agtriadi, n.d.).

Tujuan utama pengelompokan adalah untuk mengidentifikasi pola, struktur, atau segmen tersembunyi dalam data sehingga data yang kompleks menjadi lebih mudah dipahami. Dalam konteks lalu lintas, pengelompokan dapat membantu K3.

a) Mengidentifikasi perbedaan tingkat kepadatan lalu lintas antar segmen jalan,
b) Mengidentifikasi pola lalu lintas berdasarkan waktu,
c) Membedakan karakteristik segmen jalan berdasarkan kecepatan atau volume,
d) Serta mendukung pengambilan keputusan dalam manajemen transportasi.

Dalam studi ini, pengelompokan digunakan untuk mengelompokkan segmen jalan di Jakarta berdasarkan kesamaan karakteristik kecepatannya, sehingga menghasilkan segmentasi tingkat kemacetan yang lebih terstruktur.

2.2.3 Algoritma K-Means

K-Means adalah salah satu algoritma pengelompokan yang paling umum digunakan karena sederhana, mudah diterapkan, dan cukup efisien untuk kumpulan data yang besar. Dalam berbagai penelitian yang Anda kutip, *K-Means* telah digunakan untuk menganalisis pola lalu lintas, mengelompokkan periode lalu lintas, mengidentifikasi kondisi lalu lintas, dan bahkan mengelompokkan perilaku pengguna transportasi 1. Konsep dasar *K-Means* adalah membagi data menjadi K kluster berdasarkan kedekatan data dengan titik pusat kluster, yang dikenal sebagai centroid. Setiap titik data ditugaskan ke centroid terdekat, dan centroids kemudian diperbarui berdasarkan rata-rata anggota kluster. Proses ini diulang hingga pengelompokan menjadi stabil atau tidak lagi berubah secara signifikan (Li, 2024).

Secara umum, langkah kerja algoritma K-Means adalah sebagai berikut.

1. Menentukan jumlah cluster (K)
   Peneliti menentukan lebih dulu berapa banyak kelompok yang ingin dibentuk. Nilai K biasanya dicoba beberapa alternatif, lalu dipilih yang paling sesuai berdasarkan evaluasi cluster.
2. Inisialisasi centroid
   Titik pusat cluster awal dipilih, biasanya secara acak dari data yang tersedia. Nilai centroid awal ini akan menjadi acuan pembagian data pada iterasi pertama
3. Menghitung jarak setiap data ke centroid
   Setiap data dihitung jaraknya ke seluruh centroid. Pada banyak studi, termasuk studi lalu lintas Surabaya, jarak yang digunakan adalah Euclidean Distance. Data kemudian ditempatkan pada cluster dengan jarak minimum.
4. Update centroid
   Setelah semua data masuk ke cluster tertentu, centroid baru dihitung sebagai rata-rata dari anggota cluster tersebut. Proses ini membuat pusat cluster semakin merepresentasikan data anggotanya.
5. Iterasi hingga konvergen

Langkah perhitungan jarak dan pembaruan centroid diulang sampai hasil cluster stabil, artinya perpindahan anggota cluster sudah sangat kecil atau tidak berubah lagi.

2.2.4 Penentuan Jumlah Cluster

Evaluasi kluster diperlukan untuk menilai apakah jumlah kluster yang dipilih sudah tepat dan apakah hasil pengelompokan yang dihasilkan memuaskan. Dua metode evaluasi yang paling relevan dengan penelitian ini adalah Metode Elbow dan Skor Silhouette(Amal et al., n.d.).

1. Metode Elbow
   Metode *Elbow* digunakan untuk membantu menentukan jumlah kluster yang optimal. Prinsipnya melibatkan perbandingan nilai Sum of Squared Errors (SSE) pada berbagai jumlah kluster. Seiring dengan meningkatnya nilai K, SSE biasanya menurun karena data dibagi menjadi subkelompok yang lebih kecil. Namun, pada titik tertentu, penurunan SSE mulai melambat. Titik belok ini disebut "*elbow*" dan dianggap sebagai jumlah kluster yang paling wajar (Hendrastuty, 2024).
   Rumus SSE secara umum dituliskan sebagai

$$SSE = \sum_{i=1}^{k} \sum_{p \in C_i} | p - m_i | 2$$

Pada rumus di atas C_i adalah cluster ke-i, p adalah titik data dalam cluster tersebut, dan $m_i$adalah centroid cluster. Rumus ini digunakan untuk melihat seberapa rapat anggota cluster terhadap pusatnya.

2.2.5 Metode Evaluasi Cluster

1. Silhouette

Nilai *Silhouette* digunakan untuk mengukur kualitas hasil pengelompokan. Nilainya berkisar antara -1 hingga 1. Semakin mendekati 1, semakin baik, karena hal ini menunjukkan bahwa titik-titik data lebih dekat dengan kelompoknya sendiri daripada dengan kelompok lain. Nilai yang mendekati 0 menunjukkan bahwa batas antar kelompok tidak jelas, sedangkan nilai negatif mengindikasikan kemungkinan terjadinya kesalahan klasifikasi (Abdurrasyid et al., 2025).

Secara matematis, silhouette untuk satu data ditulis.

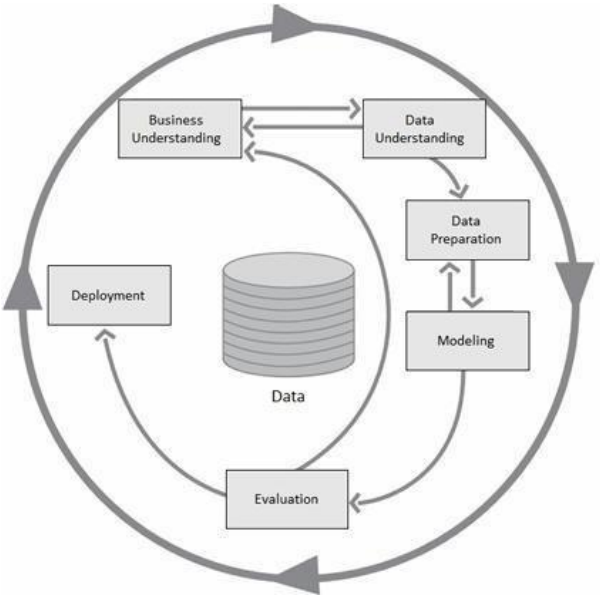
$$s(i) = max(a(i), b(i))b(i) - a(i)$$

dengan $a(i)$sebagai rata-rata jarak data ke anggota cluster yang sama, dan

$b(i)$sebagai rata-rata jarak data ke cluster terdekat lainnya. Dalam penelitian ini, *Elbow* Metode akan membantu menentukan jumlah cluster terbaik, sedangkan *Silhouette* Score dapat digunakan untuk mengecek seberapa baik kualitas segmentasi ruas jalan yang dihasilkan(Nugraha et al., 2024).

2.2.6 Crisp DM

Metodologi Cross-Industry Standard Process for Data Mining (CRISP-DM) adalah kerangka kerja sistematis yang digunakan dalam penelitian berbasis penambangan data untuk mengelola proses analisis data secara terstruktur (Shimaoka et al., 2024). Metodologi ini terdiri dari enam tahapan utama Pemahaman Bisnis, Pemahaman Data, Persiapan Data, Pemodelan, Evaluasi, dan Implementasi, yang saling terkait dan berulang. Pendekatan ini memungkinkan penyesuaian antar tahapan jika terjadi perubahan data atau kebutuhan analisis, sehingga meningkatkan fleksibilitas dan keandalan proses. Dalam studi ini, *CRISP-DM* digunakan untuk menganalisis segmentasi jalan berdasarkan karakteristik kecepatan lalu lintas, mulai dari mengidentifikasi masalah kemacetan, pengumpulan dan eksplorasi data, hingga proses pembersihan data dan pembentukan fitur seperti rata-rata, deviasi standar, dan variasi kecepatan. Selanjutnya, tahap pemodelan dilakukan menggunakan algoritma K-Means Clustering untuk mengelompokkan segmen jalan berdasarkan karakteristik yang serupa.



Gambar 2. 1 CrispDM

source: (Cui, 2019)

Gambar 2.1 menampilkan alur *CRISP-DM* dalam bentuk siklus, menekankan bahwa proses penambangan data bersifat iteratif dan saling terkait antar tahapan.

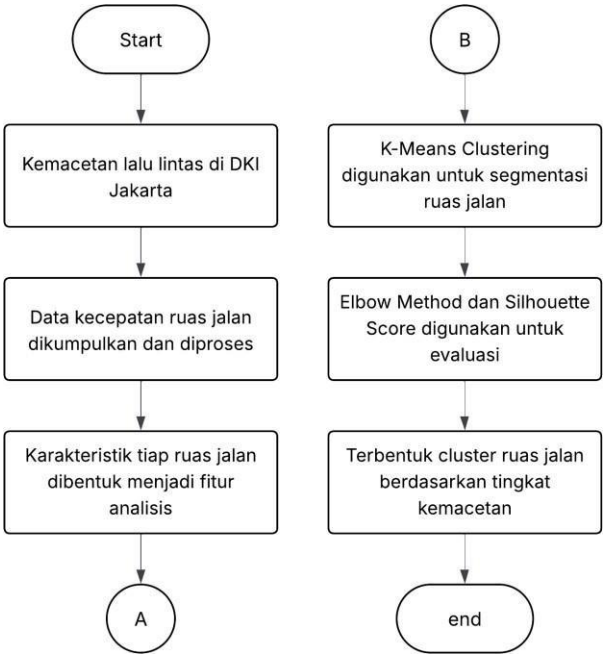
Data ditempatkan di pusat seluruh proses, menunjukkan bahwa setiap tahapan berfokus pada pengolahan dan pemanfaatan data. Hubungan antara tahapan Persiapan Data dan Pemodelan dijelaskan oleh hubungan dua arah, menunjukkan bahwa proses pemodelan seringkali memerlukan penyesuaian ulang data. Tahapan Evaluasi berperan dalam menilai kualitas hasil pengelompokan yang telah dilakukan dan berfungsi sebagai dasar untuk melakukan perbaikan jika hasil yang diperoleh tidak optimal. Selanjutnya, tahapan Penerapan adalah tahapan implementasi hasil analisis yang dapat digunakan sebagai dasar pengambilan keputusan, seperti mengkategorikan kondisi lalu lintas ke dalam kategori lancar, sedang, dan padat. Dengan demikian, *CRISP-DM* menyediakan pendekatan yang fleksibel dan adaptif dalam menghasilkan analisis yang tidak hanya akurat tetapi juga relevan dengan kebutuhan penelitian.

2.3 Kerangka Pemikiran

Kemacetan lalu lintas di Jakarta merupakan masalah yang erat kaitannya dengan tingginya tingkat mobilitas masyarakat dan aktivitas perkotaan yang padat. Salah satu indikator yang dapat digunakan untuk menilai kondisi kemacetan adalah kecepatan lalu lintas. Ketika kecepatan di suatu ruas jalan tinggi, arus lalu lintas cenderung lancar. Sebaliknya, ketika kecepatan rendah, ruas jalan tersebut cenderung mengalami kemacetan. Hubungan ini didukung oleh berbagai studi lalu lintas yang menggunakan kecepatan, volume, dan pola waktu sebagai variabel utama dalam menilai kondisi jalan. Berdasarkan pemahaman ini, penelitian ini memanfaatkan data kecepatan lalu lintas sebagai dasar untuk melakukan segmentasi ruas jalan. Namun, data mentah tidak digunakan secara langsung. Data harus melalui tahap prapemrosesan untuk memastikan data lebih bersih dan konsisten, diikuti dengan rekayasa fitur untuk menangkap karakteristik setiap ruas jalan. Selanjutnya, metode *K-Means Clustering* digunakan untuk mengelompokkan ruas jalan berdasarkan kesamaan karakteristik kecepatannya.

Jumlah klaster yang optimal ditentukan dengan menggunakan Metode *Elbow*, dan kualitas hasil pengelompokan dapat dievaluasi menggunakan Skor *Silhouette*. Proses ini menghasilkan beberapa kelompok ruas jalan dengan pola

kecepatan yang berbeda-beda. Kelompok-kelompok ini kemudian diinterpretasikan sebagai indikator kondisi lalu lintas, seperti ruas jalan yang cenderung lancar, padat, atau macet total (Faktor et al., 2022).



Gambar 2. 2 Kerangka Pemikiran

Deskripsi kerangka pemikiran.

1. Kemacetan Lalu Lintas di Jakarta

Bagian ini menjelaskan latar belakang utama penelitian ini: kemacetan lalu lintas yang sering terjadi di Jakarta akibat tingginya tingkat aktivitas dan banyaknya jumlah kendaraan. Masalah ini menjadi dasar untuk analisis lebih lanjut

2. Pengumpulan dan Pengolahan Data Kecepatan Segmen Jalan

Pada tahap ini, data kecepatan lalu lintas dari berbagai segmen jalan dikumpulkan. Data tersebut kemudian diproses melalui langkah-langkah prapemrosesan seperti pembersihan data, normalisasi, dan penyesuaian format untuk mempersiapkannya bagi analisis.

3. Karakteristik segmen jalan diubah menjadi fitur analisis

Data yang telah diproses kemudian diubah menjadi fitur yang mewakili karakteristik masing-masing segmen jalan, seperti kecepatan rata-rata, kecepatan minimum, kecepatan maksimum, dan variasi kecepatan. Fitur-fitur ini akan menjadi dasar bagi proses pengelompokan.

4. Pengelompokan K-Means digunakan untuk segmentasi ruas jalan

Pada tahap ini, metode K-Means digunakan untuk mengelompokkan ruas jalan berdasarkan kesamaan karakteristik kecepatannya. Tujuannya adalah untuk membentuk kelompok ruas jalan dengan kondisi lalu lintas yang serupa.

5. Metode Elbow dan Skor Silhouette digunakan untuk evaluasi

Setelah proses pengelompokan dilakukan, hasilnya dievaluasi menggunakan

a) Metode Elbow untuk menentukan jumlah kelompok yang optimal

b) Skor Silhouette untuk mengukur kualitas hasil pengelompokan

Langkah ini sangat penting untuk memastikan bahwa hasil pengelompokan valid dan optimal.

6. Segmen jalan dikelompokkan berdasarkan tingkat kemacetan lalu lintas Tahap akhir merupakan hasil dari seluruh proses: pembentukan beberapa kelompok segmen jalan yang mewakili tingkat kemacetan lalu lintas yang berbeda, seperti lancar, padat, dan macet total. Hasil ini dapat digunakan untuk memahami pola lalu lintas dengan lebih baik.

BAB III

METODE PENELITIAN

3.1 Lokasi dan Waktu Penelitian

Institut Teknologi Perusahan Listrik Negara (ITPLN) dan *Hybrid*

3.1.1 Lokasi Penelitian

Penelitian ini dilakukan di kampus Institut Teknologi PLN Jakarta, yang berlokasi di Menara PLN, Jl. Lkr. Luar Barat, Lantai 2, RT.1/RW.1, Duri Kosambi, Kecamatan Cengkareng, Kota Jakarta Barat, Daerah Khusus Ibukota Jakarta 11750.

Selain itu, subjek penelitian yang dianalisis dalam studi ini adalah ruas jalan di wilayah DKI Jakarta, dengan menggunakan data kecepatan lalu lintas sebagai dasar analisis segmentasi.

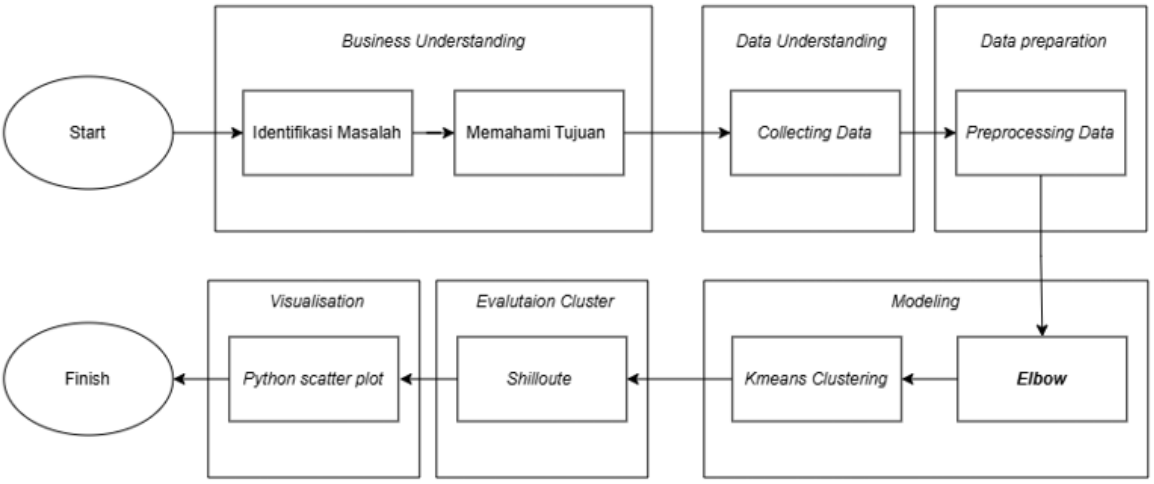
3.1.2 Waktu Penelitian

Pengumpulan data untuk penelitian ini dilakukan dengan menggunakan data sekunder berupa data kecepatan lalu lintas pada ruas jalan di DKI Jakarta, yang diperoleh melalui portal resmi pemerintah, *https://data.go.id/*. Data yang digunakan mencakup periode 2023–2025 dan kemudian diolah serta dianalisis dalam penelitian ini.

Periode penelitian berlangsung dari 31 Maret 2026 hingga 1 Juli 2026, mencakup tahap pengumpulan data, pembersihan data, pengolahan data, analisis menggunakan metode K-Means Clustering, dan penyusunan laporan penelitian.

3.2 Desain Penelitian

Pada bagian ini menjelaskan secara rinci tahapan-tahapan dalam desain penelitian dengan menggunakan metodelogi *CRISP-DM (Cross Industry Standard for Data mining),* mulai dari kerangka penelitian hingga metode apa yang akan di gunakan dalam penelitian ini untuk mendapatkan pengelompokan yang optimal, *Desain* Penelitian ini dapat dilihat pada *flowchart* dibawah ini.



Gambar 3. 1 Desain Research

Gambar 3.1 di atas menggambarkan tahapan dari awal hingga akhir. Dimulai dari *Business Understanding* (identifikasi masalah dan penetapan tujuan), dilanjutkan ke Data Understanding (pengumpulan dan *preprocessing* data). Selanjutnya masuk ke tahap Modelling menggunakan *K-Means Clustering* dengan metode *Elbow* untuk menentukan jumlah *cluster* optimal, lalu dievaluasi menggunakan *Silhouette Score*. Hasil pemodelan akhir divisualisasikan dengan *scatter* plot *Python* sebagai output akhir.

3.2.1 Business Understanding

Pada bagian *Business Understanding* merupakan fase awal dalam metodologi *CRISP-DM*, yang bertujuan untuk memahami konteks masalah, menganalisis persyaratan, dan mengidentifikasi tujuan yang ingin dicapai melalui studi ini. Dalam studi ini, fokus utama masalah terletak pada kondisi lalu lintas di wilayah DKI Jakarta, yang sering mengalami kemacetan dan antrean kendaraan di berbagai ruas jalan. Kemacetan ini berdampak pada berkurangnya mobilitas masyarakat, meningkatnya waktu tempuh, serta menurunnya efisiensi dalam kegiatan ekonomi dan distribusi logistik di wilayah perkotaan. Dalam konteks ini, salah satu indikator yang dapat digunakan untuk menggambarkan kondisi lalu lintas adalah kecepatan kendaraan, karena nilai kecepatan secara langsung mencerminkan tingkat arus lalu lintas atau kemacetan pada ruas jalan tertentu. Semakin rendah kecepatan rata-rata pada ruas jalan tertentu, semakin kuat indikasi bahwa ruas tersebut mengalami kemacetan.

kondisi tersebut, ialah masalah utama yang diidentifikasi dalam penelitian ini adalah kurangnya klasifikasi terstruktur terhadap ruas jalan di Jakarta berdasarkan karakteristik kecepatan lalu lintasnya. Oleh karena itu, diperlukan pendekatan analitis yang dapat mengolah data kecepatan menjadi informasi yang lebih terstruktur, sehingga pola kemacetan pada ruas jalan dapat diidentifikasi dengan lebih jelas. Dalam penelitian ini, pendekatan yang digunakan adalah

pengelompokan dengan algoritma *K-Means*, karena metode ini dapat mengelompokkan objek berdasarkan kesamaan karakteristik numeriknya. Data kecepatan lalu lintas yang diperoleh dari *portal resmi pemerintah* untuk dilakukkan analisis karakteristik kecepatan lalu lintas pada segmen jalan di Jakarta, kemudian mengelompokkan segmen-segmen tersebut ke dalam beberapa kluster berdasarkan kesamaan pola kecepatannya. Dengan demikian, hasil akhir yang diharapkan bukan sekadar sekumpulan kluster, tetapi juga interpretasi yang bermakna mengenai kondisi lalu lintas seperti kelompok segmen jalan yang cenderung lancar, sedang, atau macet.

Oleh karena itu, pada tahap Business Understanding ditetapkan bahwa penelitian ini berfokus pada segmentasi ruas jalan, bukan wilayah administratif, dan variabel utama yang digunakan adalah kecepatan lalu lintas. Penetapan fokus ini penting agar seluruh proses analisis pada tahap berikutnya, mulai dari pemahaman data, persiapan data, hingga pemodelan, tetap berada dalam arah yang sesuai dengan tujuan penelitian. Dengan demikian, tahap Business Understanding menjadi dasar konseptual yang menentukan keberhasilan keseluruhan proses data mining dalam penelitian ini.

3.1.2 Data Understanding

Pada bagian Data Understanding bertujuan untuk memahami karakteristik data yang digunakan dalam penelitian, baik dari sisi sumber data maupun struktur dan pola yang terkandung di dalamnya. Pada tahap ini dilakukan proses pengumpulan data serta eksplorasi awal terhadap data untuk memastikan kualitas dan kesiapan data sebelum masuk ke tahap pengolahan lebih lanjut.

3.1.2.1 Data collecting

Pengumpulan data dalam penelitian ini dilakukan dengan memanfaatkan data sekunder yang bersumber dari portal resmi Pemerintah Indonesia, yaitu Portal Satu Data Indonesia (data.go.id). Portal ini merupakan layanan data terbuka (open data) yang dikelola oleh Kementerian Perencanaan Pembangunan Nasional/BAPPENAS, sehingga menjamin aksesibilitas, validitas, dan kredibilitas data yang digunakan.

Data yang diambil berasal dari dataset "Kecepatan Lalu Lintas pada Ruas Jalan di DKI Jakarta" yang mencakup periode pengamatan dari tahun 2023 hingga 2025 (kurun waktu 3 tahun). Penggunaan data sekunder dari sumber terbuka ini bertujuan untuk memastikan *reproducibility* (keterulangan) penelitian, sekaligus memberikan gambaran kondisi lalu lintas yang aktual dan objektif berdasarkan rekaman resmi.

Data penelitian diperoleh dengan mengunduh dokumen dari website resmi.

Adapun struktur data awal yang diperoleh terdiri dari empat variabel utama, yaitu

1) periode_data : menunjukkan waktu pengumpulan data dalam format tahun dan bulan (YYYYMM), yang mencerminkan variasi temporal kondisi lalu lintas,
2) wilayah : menunjukkan lokasi administratif tingkat kota di DKI Jakarta,
3) ruas_jalan : menunjukkan nama spesifik segmen jalan yang menjadi objek pengamatan,
4) kecepatan : menunjukkan nilai kecepatan rata-rata kendaraan (dalam Km/jam) pada ruas dan periode terkait.

Pada tahap awal, berhasil dikumpulkan 1.201 baris data mentah dari portal tersebut. Namun, untuk menjamin kualitas analisis, data tersebut melalui serangkaian proses pembersihan *data cleaning* yang meliputi penanganan nilai hilang *missing values*, penghapusan data duplikat, penyaringan nilai kecepatan yang tidak logis (kecepatan ≤ 0 Km/jam atau > 120 Km/jam).

Setelah melalui tahap pra pemrosesan tersebut, diperoleh 1.177 baris data valid yang siap untuk dianalisis. Data valid ini mencakup 50 segmen jalan unik yang tersebar di berbagai wilayah administratif DKI Jakarta, serta mencakup 25 periode waktu yang berbeda, sehingga kaya akan variasi spasial dan temporal untuk mendukung analisis segmentasi dan pemodelan clustering. Kumpulan data inilah yang kemudian menjadi fondasi utama dalam tahap *Exploratory Data Analysis (EDA*) dan pemodelan *K-Means Clustering* untuk mengidentifikasi pola kemacetan berdasarkan karakteristik kecepatan di setiap ruas jalan.

3.1.2.2 Exploratory Data Analysis (EDA)

Setelah proses pengumpulan data dilakukan, tahap selanjutnya adalah eksplorasi data atau *Exploratory Data Analysis (EDA)* yang bertujuan untuk memahami karakteristik, distribusi, serta kualitas data yang digunakan dalam penelitian. Pada tahap ini dilakukan analisis terhadap jumlah observasi pada setiap wilayah, jumlah ruas jalan yang diamati, serta jumlah periode data yang tersedia.

Tabel 3. 1 Jumlah Observasi Berdasarkan Wilayah

No	Wilayah	Jumlah Observasi
1	Jakarta Pusat	347
2	Jakarta Selatan	360
3	Jakarta Barat	211
4	Jakarta Utara	24
5	Jakarta Timur	235

Berdasarkan hasil analisis EDA, distribusi jumlah pengamatan di setiap wilayah administratif disajikan pada Tabel 3.1. Secara keseluruhan, terdapat

1.177 titik data yang tersebar di lima wilayah administratif DKI Jakarta.

Tabel tersebut menunjukkan bahwa sebaran observasi tidak merata antar wilayah. Wilayah dengan jumlah pengamatan tertinggi adalah Jakarta Selatan dengan 360 titik data, diikuti oleh Jakarta Pusat dengan 347 titik data, Jakarta Timur dengan 235 titik data, dan Jakarta Barat dengan 211 titik data. Sementara itu, Jakarta Utara mencatatkan jumlah pengamatan paling rendah, yaitu hanya 24 titik data. Selain itu, pada tahap eksplorasi data awal ditemukan ketidakkonsistenan dalam penulisan nama wilayah, seperti variasi penggunaan huruf kapital serta adanya format "Kota Adm." yang tidak seragam, sehingga diperlukan proses standardisasi untuk memastikan keakuratan analisis.

3.1.3 Data Preparation

Pada bagian ini adalah prosedur pemrosesan data yang bertujuan untuk menghasilkan kumpulan data yang bersih, terstruktur, dan siap digunakan untuk proses pengelompokan. Tahap ini mencakup beberapa langkah utama, antara lain pembersihan data, transformasi data, rekayasa fitur, dan normalisasi data. Proses ini sangat penting karena kualitas data sangat memengaruhi hasil model.

1. *Cleaning Data*

   Cleaning data dilakukan untuk memastikan bahwa data yang digunakan bebas dari kesalahan, ketidakkonsistenan, dan nilai yang tidak valid. Berdasarkan hasil eksplorasi data pada tahap sebelumnya, ditemukan perbedaan dalam penulisan atribut wilayah, seperti penggunaan huruf kapital dan variasi penamaan, misalnya "***Kota Adm.*" *dan* "*Jakarta.*"** Hal ini menandakan adanya ketidakkonsistenan data yang berpotensi memengaruhi proses pengelompokan.

Selain itu, nilai kecepatan yang tidak valid juga dibersihkan, seperti nilai yang hilang atau nilai yang tidak masuk akal. Data yang tidak memenuhi kriteria ini dihapus atau disesuaikan agar tidak mengganggu proses analisis. Pada tahap ini, pemeriksaan juga dilakukan untuk mendeteksi duplikasi data dan konsistensi format di setiap atribut. Dengan demikian, data yang digunakan dalam penelitian ini telah melalui proses pembersihan, sehingga menjadi lebih akurat dan representatif.

Tabel 3. 2 Data Cleaning

No	Periode data	Wilayah	Ruas Jalan	Kecepatan
1	2024-01	Jakarta Barat	Jalan Jenderal S. Parman	19.62 km/jam
2	2024-01	Jakarta Barat	Jalan Jenderal S. Parman	22.45 km/jam
3	2024-02	Jakarta Barat	Jalan Jenderal S. Parman	16.80 km/jam
4	2024-01	Jakarta Barat	Jalan Pintu Besar Selatan	13.27 km/jam
5	2024-02	Jakarta Barat	Jalan Pintu Besar Selatan	13.28 km/jam
Dst…				

Pada tabel 3.2 menunjukkan data hasil proses data cleaning yang telah siap digunakan untuk tahap analisis selanjutnya. Data tersebut terdiri dari atribut periode_data, wilayah, ruas_jalan, dan kecepatan, di mana nilai kecepatan telah dikonversi ke dalam bentuk numerik serta dibersihkan dari nilai yang tidak valid.

2. *Tranform Data*

   Setelah data dibersihkan, langkah selanjutnya adalah melakukan transformasi data agar dapat diproses oleh algoritma K-Means. Transformasi dilakukan dengan mengubah format data ke dalam bentuk numerik, khususnya pada atribut kecepatan yang sebelumnya memiliki format tertentu. Selain itu, data yang awalnya berbentuk data mentah per periode diubah menjadi bentuk agregasi berdasarkan wilayah dan ruas jalan. Proses ini dilakukan dengan mengelompokkan data menggunakan fungsi ***groupby***, sehingga setiap ruas jalan memiliki representasi data yang lebih ringkas dalam bentuk nilai statistik. Hasil transformasi ini menghasilkan

dataset baru dengan ukuran ***75 baris dan 14 kolom (75,14),*** yang menunjukkan jumlah segmen ruas jalan yang telah diproses dan siap digunakan dalam tahap berikutnya.

3. *Feature Engineering*

   Proses feature engineering dilakukan untuk menciptakan variabel baru yang mampu menggambarkan karakteristik kecepatan lalu lintas secara lebih komprehensif untuk setiap segmen jalan. Berdasarkan data yang telah dibersihkan, proses feature engineering ini melibatkan penggabungan data menggunakan fungsi *groupby* berdasarkan atribut *region* dan *road_segment,* sehingga setiap segmen jalan direpresentasikan dalam satu baris data yang berisi ringkasan statistik dari semua pengamatan kecepatan yang tersedia. Hasil dari proses ini adalah kumpulan data baru yang disebut data segmen per ruas jalan, dengan total 75 baris dan 14 kolom, di mana setiap baris mewakili satu ruas jalan. Fitur yang dihasilkan terdiri dari beberapa ukuran statistik yang menggambarkan pola kecepatan lalu lintas, yaitu:

Tabel 3.1 Data Segmen Ruas Jalan II

Data Segment per Ruas Jalan

wilayah	ruas_jalan	rata	median	min	max	std
JAKARTA BARAT	Jl Jendral S.Parman	19.62	22.45	13.83	22.58	5.01471
JAKARTA BARAT	Jl Pintu Besar Selatan	13.2767	13.28	13.05	13.5	0.225019
JAKARTA BARAT	Jl Tomang Raya	26.3067	27.69	20.96	30.27	4.80669
JAKARTA PUSAT	Jl Balikpapan	24	24	24	24	0
JAKARTA PUSAT	Jl Gajah Mada	18.3933	18.73	13.87	22.58	4.36475

Berdasarkan data segmen ruas jalan diatas berikut detial penjealsanya :

a) rata_kecepatan (mean)
   Menunjukkan nilai rata-rata kecepatan kendaraan pada suatu ruas jalan. Fitur ini digunakan untuk merepresentasikan kondisi umum lalu lintas, apakah cenderung cepat atau lambat.
b) median_kecepatan
   Menunjukkan nilai tengah dari distribusi kecepatan, yang lebih tahan terhadap pengaruh nilai ekstrem (outlier) dibandingkan rata-rata.
c) min_kecepatan dan max_kecepatan
   Menggambarkan batas minimum dan maksimum kecepatan yang terjadi pada suatu ruas jalan. Kedua fitur ini menunjukkan kondisi terburuk dan terbaik dari lalu lintas.

d) std_kecepatan (standar deviasi)
   Menunjukkan tingkat variasi kecepatan. Nilai yang tinggi menunjukkan bahwa kecepatan sangat fluktuatif, sedangkan nilai rendah menunjukkan kondisi yang lebih stabil.
e) jumlah_observasi
   Menunjukkan jumlah data kecepatan yang tersedia untuk setiap ruas jalan, yang mencerminkan banyaknya pengamatan yang dilakukan.
f) jumlah_periode
   Menunjukkan jumlah periode waktu yang berbeda dalam pengambilan data, yang menggambarkan cakupan temporal dari data.
g) q1_kecepatan dan q3_kecepatan (kuartil)
   Digunakan untuk melihat distribusi data kecepatan. Q1 menunjukkan batas bawah distribusi, sedangkan Q3 menunjukkan batas atas distribusi.

Berikut fitur turunan yang memberikan informasi tambahan mengenai pola dataset

a) range_kecepatan
   Merupakan selisih antara kecepatan maksimum dan minimum. Fitur ini menunjukkan rentang perubahan kecepatan pada suatu ruas jalan.
b) iqr_kecepatan (Interquartile Range)
   Merupakan selisih antara Q3 dan Q1, yang menunjukkan sebaran data kecepatan tanpa dipengaruhi oleh outlier.
c) cv_kecepatan (Coefficient of Variation)
   Merupakan perbandingan antara standar deviasi dan rata-rata kecepatan. Fitur ini digunakan untuk mengukur tingkat variasi relatif terhadap nilai rata-rata, sehingga dapat membedakan ruas jalan yang stabil dan yang fluktuatif.

4. *Normalization Scaler*

   Langkah terakhir dalam persiapan data adalah normalisasi data menggunakan metode StandardScaler. Proses ini dilakukan untuk menyesuaikan skala setiap fitur agar memiliki distribusi yang seragam, sehingga tidak ada satu pun fitur yang mendominasi proses pengelompokan. Hasil normalization menunjukan mean mendakit 0 dan standar deviation 1.

Tabel 3. 3 Normalization scaler

Fitur	Mean (0)	Std Dev (1)
rata_kecepatan	2.22×10^{-16}	1.0067
median_kecepatan	-3.08×10^{-16}	1.0067
min_kecepatan	-1.44×10^{-16}	1.0067
max_kecepatan	-3.55×10^{-17}	1.0067
std_kecepatan	-1.47×10^{-16}	1.0067
range_kecepatan	1.35×10^{-16}	1.0067
iqr_kecepatan	1.66×10^{-16}	1.0067
cv_kecepatan	-2.72×10^{-16}	1.0067
Fitur	Mean (≈ 0)	Std Dev (≈ 1)
rata_kecepatan	2.22×10^{-16}	1.0067
median_kecepatan	-3.08×10^{-16}	1.0067

Tabel 3.3 di atas menampilkan hasil proses standardisasi (StandardScaler) yang diterapkan pada seluruh fitur numerik sebelum dilakukan pemodelan K-Means Clustering. Proses standardisasi ini bertujuan untuk menyeragamkan skala antar fitur, mengingat setiap fitur memiliki satuan dan rentang nilai yang berbeda—misalnya rata_kecepatan dalam Km/jam dan cv_kecepatan yang merupakan rasio tanpa satuan. Tanpa standardisasi, fitur dengan rentang nilai yang lebih besar akan mendominasi perhitungan jarak dalam algoritma K-Means, sehingga hasil pengelompokan menjadi bias.

Hasil pada tabel menunjukkan bahwa seluruh fitur telah berhasil dinormalisasi dengan baik, yang ditandai dengan nilai Mean yang mendekati 0 (dalam orde 10^{-16} hingga 10^{-17}) dan nilai Std Dev yang mendekati 1 (1.0067). Nilai mean yang mendekati nol mengindikasikan bahwa distribusi data telah dipusatkan di sekitar titik nol, sementara standar deviasi yang mendekati satu menunjukkan bahwa data telah memiliki varians yang seragam. Dengan demikian, seluruh fitur memiliki skala yang sebanding dan siap digunakan sebagai masukan dalam algoritma clustering. Selisih kecil dari nilai 0 dan 1 (misalnya 2.22×10^{-16} dan 1.0067) merupakan hal yang wajar akibat pembulatan presisi numerik dalam komputasi.

Tabel 3.2 Data sebelum & sesudah Feature engineering

Data Sebelum Feature Engineering (Data Per Periode per Ruas Jalan)

5 Data Asli :

periode_data	wilayah	ruas_jalan	kecepatan	tahun	bulan	tanggal_periode
202510	JAKARTA PUSAT	Jalan Jendral Sudirman	23.93	2025	10	2025-10-11 00:00:00
202407	JAKARTA SELATAN	Jalan Panglima Polim	20.85	2024	7	2024-07-01 00:00:00
202408	JAKARTA SELATAN	Jalan Panglima Polim	20.79	2025	8	2024-08-01 00:00:00
202501	JAKARTA SELATAN	Jalan Fatmawati	25.18	2025	1	2025-01-01 00:00:00
202511	JAKARTA PUSAT	Jalan Salemba Raya Sisi Barat dan Jalan Salemba Raya Sisi Timur	35.82	2025	11	2025-11-01 00:00:00

Data Sesudah Feature Engineering (Agregasi per Ruas Jalan)

5 Hasil Data Segmentasi :

wilayah	ruas_jalan	rata	median	min	max	std	obs	prd	q1	q3	range	iqr	cv
JAKARTA PUSAT	JL. M.H. THAMRIN	28.43	28.43	26.78	30.08	2.33345	2	2	27.605	29.255	3.3	1.65	0.0820771
JAKARTA SELATAN	JL SISIMANGARAJA	34.885	34.885	29.62	40.15	7.44583	2	2	32.2525	37.5175	10.53	5.265	0.213439
JAKARTA PUSAT	Jl Pramuka	26.7205	26.7205	20.68	32.76	4.54713	22	22	22.0875	31.2575	12.08	9.17	0.170174
JAKARTA SELATAN	Jl Sisimangaraja	21.7217	21.7217	10.34	33.67	5.48428	23	23	18.58	25.975	23.23	7.395	0.252479
JAKARTA PUSAT	JL SALEMBA RAYA SISI BARAT DAN JALAN SALEMBA RAYA SISI TIMUR	36.11	36.11	34.99	37.23	1.58392	2	2	35.55	36.67	2.24	1.12	0.0438637

Tabel 3.2 di atas menyajikan perbandingan antara struktur data sebelum dan sesudah dilakukan proses feature engineering pada penelitian ini. Terdapat dua bagian utama yang menggambarkan transformasi data dari bentuk mentah (raw data) menjadi bentuk agregasi yang siap untuk dianalisis lebih lanjut.

a). Data Sebelum *Feature Engineering*

Bagian atas gambar menampilkan 5 baris sampel data mentah sebelum dilakukan agregasi. Data ini merepresentasikan pengamatan kecepatan lalu lintas pada setiap periode waktu tertentu (bulanan) untuk setiap ruas jalan. Struktur data awal terdiri dari beberapa kolom utama, yaitu:

1) periode_data menunjukkan kode waktu pengambilan data dalam format tahun dan bulan (YYYYMM), misalnya 202510 yang berarti Oktober 2025.
2) wilayah menunjukkan lokasi administratif di DKI Jakarta, seperti JAKARTA PUSAT dan JAKARTA SELATAN.

3) ruas_jalan : nama spesifik segmen jalan yang menjadi objek pengamatan.
4) kecepatan nilai kecepatan rata-rata kendaraan dalam satuan Km/jam pada periode dan ruas jalan tertentu.
5) tahun dan bulan hasil ekstraksi dari kolom periode_data untuk memudahkan analisis temporal.
6) tanggal_periode representasi tanggal dalam format datetime yang menunjukkan awal periode pengamatan.

Dari sampel data yang ditampilkan, terlihat bahwa setiap baris mewakili satu pengamatan pada satu ruas jalan di satu periode waktu tertentu. Sebagai contoh, ruas jalan Jalan Jenderal Sudirman di wilayah Jakarta Pusat tercatat memiliki kecepatan 23,93 Km/jam pada periode Oktober 2025, sementara ruas Jalan Panglima Polim di Jakarta Selatan menunjukkan fluktuasi kecepatan dari 28,85 Km/jam (Juli 2024) menjadi 20,79 Km/jam (Agustus 2024). Data ini mencerminkan variasi temporal yang cukup signifikan, yang menunjukkan bahwa kondisi lalu lintas di setiap ruas jalan tidak statis dan dapat berubah dari waktu ke waktu.

b). Data Sesudah *Feature Engineering*

Bagian bawah gambar menampilkan 5 baris sampel data hasil agregasi, yang merupakan transformasi dari data mentah menjadi fitur-fitur statistik per ruas jalan. Proses ini dilakukan dengan mengelompokkan (group by) data berdasarkan kombinasi wilayah dan ruas jalan, kemudian menghitung berbagai metrik deskriptif untuk setiap segmen jalan. Kolom-kolom yang dihasilkan meliputi:

1) rata_kecepatan rata-rata kecepatan kendaraan pada suatu ruas jalan selama seluruh periode pengamatan.
2) median_kecepatan nilai tengah dari seluruh data kecepatan, yang lebih robust terhadap *outlier* dibandingkan rata-rata.
3) min_kecepatan dan max_kecepatan nilai kecepatan terendah dan tertinggi yang tercatat pada ruas jalan tersebut.
4) std_kecepatan : standar deviasi kecepatan, yang mengukur seberapa besar variasi atau fluktuasi kecepatan dari waktu ke waktu.
5) jumlah_observasi banyaknya pengamatan periode yang tersedia untuk ruas jalan tersebut.
6) jumlah_periode jumlah periode waktu unik yang tercakup dalam data
7) q1_kecepatan dan q3_kecepatan kuartil pertama dan ketiga, yang menggambarkan sebaran data kecepatan.
8) range_kecepatan selisih antara kecepatan maksimum dan minimum, yang menunjukkan rentang variasi kecepatan.

9) *iqr_kecepatan (interquartile range)* selisih antara kuartil ketiga dan pertama, yang mengukur sebaran data di bagian tengah.

10) *cv_kecepatan (coefficient of variation)* rasio antara standar deviasi dan rata-rata, yang menunjukkan tingkat variabilitas relatif kecepatan pada suatu ruas jalan.

Dari sampel data agregasi, dapat diamati karakteristik masing-masing ruas jalan secara lebih *komprehensif*. Sebagai contoh, ruas JALAN M.H. THAMRIN di Jakarta Pusat memiliki rata-rata kecepatan 28,43 Km/jam dengan standar deviasi yang relatif kecil (2,33), yang mengindikasikan kondisi lalu lintas yang cenderung stabil. Sementara itu, ruas JALAN SISIMANGARAJA di Jakarta Selatan menunjukkan rata-rata kecepatan 21,72 Km/jam dengan standar deviasi yang lebih tinggi (5,48) dan *coefficient of variation* (0,25), yang mengindikasikan fluktuasi kecepatan yang lebih besar kemungkinan mencerminkan kondisi lalu lintas yang lebih padat dan tidak menentu.

3.1.4 Modelling

Pada tahap modeling merupakan tahap inti dalam penelitian ini, di mana data yang telah melalui proses data preparation selanjutnya dianalisis menggunakan algoritma K-Means Clustering. Tujuan dari tahap ini adalah untuk mengelompokkan ruas jalan berdasarkan kemiripan karakteristik kecepatan lalu lintas. Proses clustering dilakukan dengan menentukan jumlah cluster optimal menggunakan metode Elbow, kemudian dilanjutkan dengan proses pembentukan cluster menggunakan algoritma K-Means. Dalam penelitian ini, jumlah cluster yang digunakan adalah K = 3, yang kemudian diinterpretasikan menjadi kategori kondisi lalu lintas yaitu macet, sedang, dan lancar.

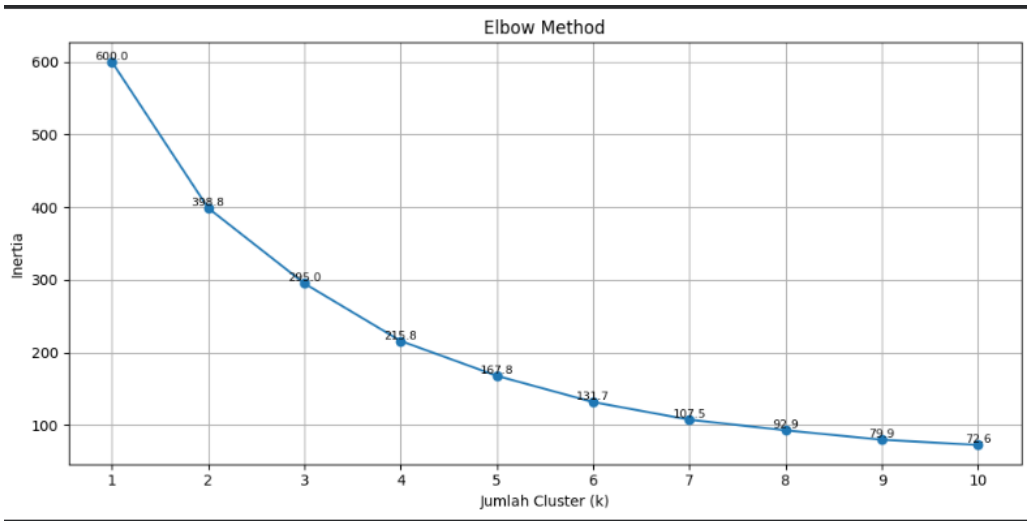
3.1.4.1 Elbow

Pada tahap ini, jumlah kluster yang optimal ditentukan sebelum proses pengelompokan K-Means dilakukan. Data yang digunakan terdiri dari data yang telah dinormalisasi (X_scaled) yang telah melalui proses persiapan data, sehingga memastikan bahwa semua fitur berada pada skala yang seragam dan siap digunakan dalam perhitungan pengelompokan.

Dalam penelitian ini, jumlah kluster ditentukan menggunakan metode Elbow dengan memanfaatkan nilai inersia yang dihasilkan oleh algoritma K-Means. Inertia adalah nilai yang menunjukkan jarak kuadrat total antara setiap titik data dan pusat kluster terdekat, yang secara konseptual setara dengan Sum of Squared Errors (SSE). Perhitungan nilai inertia dilakukan secara otomatis oleh sistem menggunakan pustaka scikit-learn, sehingga tidak diperlukan perhitungan manual untuk seluruh kumpulan data.

Proses perhitungan dilakukan dengan menguji berbagai jumlah kluster, khususnya dalam rentang:

$$K = 1 \text{ SAMPAI } 10$$



Gambar 3. 4 Elbow Method

Proses penentuan jumlah klaster diawali dengan mengevaluasi nilai inersia dari berbagai konfigurasi K melalui grafik Elbow. Melalui kurva tersebut, kita dapat melihat titik balik di mana penurunan inersia mulai melambat, yang mengindikasikan batas optimal penambahan jumlah klaster. Untuk menghindari subjektivitas dalam membaca grafik, rentang nilai K yang diperoleh kemudian diuji kembali menggunakan metrik Silhouette Score. Kombinasi kedua metode ini memastikan bahwa jumlah klaster yang dipilih memang mampu memisahkan data secara optimal.

3.1.4.2 K-means Clustering

Pada tahap ini dilakukan proses clustering menggunakan algoritma K-Means dengan jumlah cluster sebanyak 3. Proses dilakukan secara iteratif dengan menghitung jarak setiap data terhadap centroid dan mengelompokkan data berdasarkan jarak terdekat.

1) Inisialisasi Data dan Centroid

Tabel 3. 4 Inisialisasi Data dan Centroid

Data	rata_kecepatan
Data 1	19.62
Data 2	13.27

Centroid awaal
C0 = 20.00, C1 = 15.00, C2 = 30.00

a) Menghitung Jarak Menggunakan Euclidean Distance

$$D = \sqrt{(x - c)^2}$$

Ket :
x = Data
c = Centroid

1. Perhitungan Data ke-1
   Ke centroid C0:
   $\sqrt{(19.62 - 20\ )^2}$= 0.38
   Ke centroid C1:
   $\sqrt{(19.62 - 15\ )^2}$= 4.62
   Ke centroid C2:
   $\sqrt{(19.62 - 30\ )^2}$= 10.38

2. Perhitungan data ke-2
   Ke centroid C0:
   $\sqrt{(13.27 - 20\ )^2}$= 6.73
   Ke centroid C1:
   $\sqrt{(13.27 - 15\ )^2}$= 1.73
   Ke centroid C2:

$$\sqrt{(13.27 - 30)^2} = 16.73$$

a) Menentukan Cluster Terdekat
Data 1 → jarak terkecil = *0.38* → *masuk C0*
Data 2 → jarak terkecil = *1.73* → *masuk C1*

b) Menghitung Jarak Kuadrat (Sesuai SSE)
Setelah setiap data dikelompokkan ke dalam cluster terdekat berdasarkan jarak Euclidean, langkah selanjutnya adalah menghitung jarak kuadrat dari setiap data terhadap centroid cluster yang dipilih. Proses ini bertujuan untuk mengukur seberapa dekat setiap data dengan pusat clusternya.

Perhitungan jarak kuadrat ini merupakan bagian dari konsep Sum of Squared Errors (SSE), yang digunakan dalam algoritma K-Means untuk mengevaluasi kualitas clustering. Semakin kecil nilai SSE, maka semakin baik hasil pengelompokan, karena data berada lebih dekat dengan centroid cluster masing-masing.

Dalam implementasi menggunakan Python (scikit-learn), perhitungan SSE ini dilakukan secara otomatis melalui nilai inertia. Namun, untuk memberikan pemahaman yang lebih jelas, berikut ditampilkan contoh perhitungan manual menggunakan beberapa data.

Tabel 3. 5 Table Simulation SSE

Data	Nilai (rata_kecepatan)	Cluster	Jarak
Data 1	19.62	C0	0.38
Data 2	13.27	C1	1.73

Kuadrat Jarak
Data 1 = $(0.38)^2$ = 0.1444 => 0.14
Data 2 = $(1.73)^2$ = 2.9929 => 2.99
Interpretasi
Nilai *0.14* menunjukkan bahwa Data 1 sangat dekat dengan centroid C0.

Nilai *2.99* menunjukkan bahwa Data 2 masih cukup dekat dengan centroid C1, namun memiliki variasi yang lebih besar dibanding Data 1.

Akumulasi SSE

Jika seluruh 2 data simulasi tersebut di gabungkan:

$$SSE = \sum \quad (x_i - c)^2$$

$$\text{SSE} = 0.144 + 2.99 = 3.13$$

3. *Update Centroid*

Setelah seluruh data dikelompokkan ke dalam *cluster* berdasarkan jarak terdekat, langkah selanjutnya adalah melakukan pembaruan *centroid (update centroid).* Tujuan dari tahap ini adalah untuk mendapatkan posisi pusat *cluster* yang baru berdasarkan anggota cluster yang telah terbentuk pada iterasi sebelumnya.

Centroid baru dihitung dengan cara mencari rata-rata dari seluruh data dalam *cluster* tersebut.

$$c = \frac{\sum \quad x}{n}$$

Ket:

c = *centroid baru*

x = data dalam *cluster*

n =jumlah dalam *cluster*

cluster C0:

Data yang masuk: 19.62 dan 18.39

$$C0 = \frac{19.62 + 18.39}{2} = 19.005$$

$$C0 = \frac{38.01}{2} = 19.00$$

cluster C1:

$$C1 = \frac{13.27}{1} = 13.27$$

cluster C2:

$$C2 = \frac{26.30}{1} = 26.30$$

Interpretasi

Centroid C0 mengalami perubahan dari nilai awal menjadi *19.00* *Centroid* C1 dan C2 tidak berubah karena hanya memiliki satu data

a) Iterasi kedua

Setelah *centroid* diperbarui, langkah berikutnya adalah melakukan iterasi kedua, yaitu menghitung kembali jarak setiap data terhadap centroid terbaru. Tujuannya adalah untuk memastikan apakah terdapat perubahan keanggotaan *cluster*.

Data 1 (19.62)

Centorid C0:

$$\sqrt{(19.62 - 19.00)^2} = \sqrt{(0.62)^2} = 0.62$$

Centorid C1:

$$\sqrt{(19.62 - 13.27)^2} = \sqrt{(6.35)^2 = 6.35}$$

Centorid C2:

$$\sqrt{(19.62 - 26.30)^2} = \sqrt{(6.68)^2 = 6.68}$$

Jarak kecil = 0.62 (C0)

Data 2 (13.27

Centorid C0:

$$\sqrt{(13.27 - 19.00)^2} = \sqrt{(5.73)^2} = 5.73$$

Centorid C1:

$$\sqrt{(13.27 - 13.27)^2} = 0$$

Centorid C2:

$$\sqrt{(13.27 - 26.30)^2} = 13.04$$

Jarak kecil = 0 (C1)

Berdasarkan hasil perhitungan tidak terjadi perubahan posisi cluster pada Data 1 dan Data 2 Seluruh data tetap berada pada cluster sebelumnya *(konvergensi == berhenti)*

3.1.5 Evaluation Model

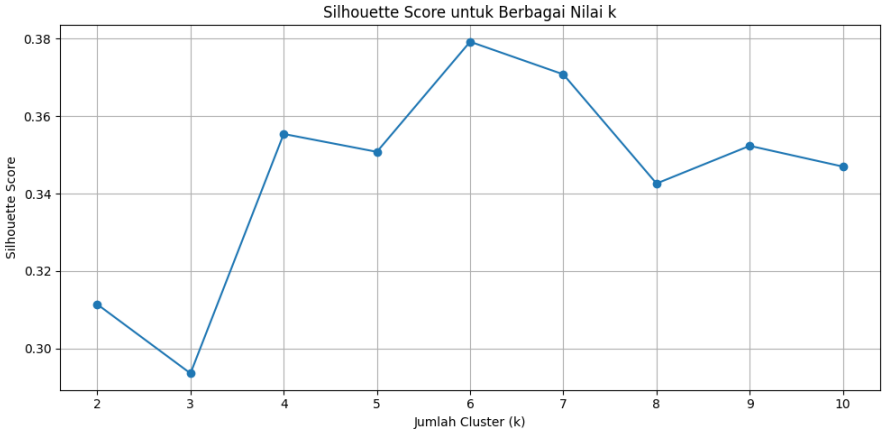
Selama tahap evaluasi model, kualitas hasil pengelompokan dinilai menggunakan metode Silhouette Score. Evaluasi ini bertujuan untuk mengukur seberapa baik setiap titik data cocok dengan klusternya dibandingkan dengan kluster lain. Perhitungan dilakukan dengan menguji berbagai jumlah kluster, mulai dari k = 2 hingga k = 10, menggunakan algoritma K-Means pada data yang telah dinormalisasi (X_scaled). Untuk setiap nilai k, model K-Means dijalankan untuk menghasilkan label kluster, kemudian Skor Silhouette dihitung menggunakan pustaka *scikit-learn.*

Tabel 3. 6 Result Silhoutte score

k	Inertia	Silhouette Score
2	398.768	0.311364
3	294.953	0.293514
4	215.848	0.355403
5	167.796	0.350783
6	131.696	0.379246
7	107.452	0.370816
8	92.850	0.342596
9	79.896	0.352337
10	72.644	0.346972
k	Inertia	Silhouette Score

Berdasarkan Tabel 3.6, terlihat bahwa nilai silhouette score mengalami peningkatan dari k = 3 ke k = 4, kemudian relatif stabil hingga mencapai nilai maksimum pada k = 6, yaitu sebesar 0.379246. Nilai ini menunjukkan bahwa pada k = 6, struktur cluster yang terbentuk memiliki tingkat pemisahan dan kepadatan yang paling baik dibandingkan jumlah cluster lainnya.

Selain itu, nilai inertia juga terus menurun seiring bertambahnya jumlah cluster, yang menunjukkan bahwa jarak antar data dengan centroid semakin kecil. Namun, penurunan inertia tidak selalu menunjukkan kualitas cluster yang lebih baik, sehingga diperlukan evaluasi tambahan menggunakan silhouette score.



Gambar 3. 5 Shiloutte Score Grafik

Berdasarkan grafik silhouette score:

1. Terjadi peningkatan kualitas cluster dari k = 3 ke k = 4
2. Nilai silhouette mencapai puncaknya pada k = 6
3. Setelah k = 6, nilai silhouette cenderung menurun
4. Hal ini menunjukkan bahwa penambahan cluster setelah k = 6 tidak meningkatkan kualitas cluster secara signifikan

Dengan demikian, secara evaluasi numerik, jumlah cluster terbaik adalah: K = 6

Berdasarkan hasil evaluasi menggunakan metode Silhouette Score, ditemukan bahwa jumlah kluster yang optimal secara numerik adalah k = 6 karena menghasilkan nilai Silhouette Score tertinggi. Namun, dalam penelitian ini, jumlah kluster yang digunakan adalah k = 3. Pilihan ini dibuat dengan mempertimbangkan interpretabilitas dan tujuan analisis, yaitu untuk mengelompokkan kondisi lalu lintas ke dalam kategori yang lebih mudah dipahami secara praktis, seperti lancar, sedang, dan macet. Dengan demikian, meskipun secara matematis k = 6 memberikan kualitas pemisahan kluster yang lebih baik, penggunaan k = 3 dianggap lebih relevan dalam mendukung interpretasi hasil dan pengambilan keputusan dalam konteks transportasi.

BAB IV
HASIL DAN PEMBAHASAAN

4.1 Hasil

Pada bagian sub bab ini, akan memaparkan hasil dari pengolahan data mulai dari tahap pengenalan data hinga hasil dari model clustering.

4.1.1 Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari data sekunder mengenai kecepatan lalu lintas pada ruas jalan di wilayah DKI Jakarta untuk periode 2023–2025. Data tersebut diperoleh dari portal data resmi pemerintah Indonesia dan digunakan sebagai dasar untuk segmentasi ruas jalan menggunakan metode *K-Means Clustering.*

Berdasarkan hasil eksplorasi dataset awal, ditemukan bahwa dataset tersebut terdiri dari 625 baris data dan 4 kolom atribut utama. Setiap baris mewakili data kecepatan lalu lintas pada ruas jalan tertentu selama periode tertentu. Struktur dataset yang digunakan dalam penelitian ini ditunjukkan pada tabel berikut.

Tabel 4. 1 Deskripsi Dataset

No	Nama Kolom	Tipe Data	Deskripsi
1	periode_data	int64	Menunjukkan periode waktu pengambilan data lalu lintas
2	wilayah	object	Menunjukkan wilayah administratif di DKI Jakarta
3	ruas_jalan	object	Menunjukkan nama ruas jalan yang diamati
4	kecepatan	object	Menunjukkan nilai kecepatan kendaraan pada ruas jalan

Berdasarkan hasil pemeriksaan kualitas data, seluruh atribut memiliki jumlah data valid sebanyak 625 data (non-null), sehingga tidak ditemukan nilai kosong (missing value) pada dataset awal. Selain itu, dataset terdiri dari atribut kategorikal dan numerik yang selanjutnya diproses pada tahap data preparation sebelum digunakan dalam proses clustering.

Contoh data awal yang digunakan dalam penelitian ditunjukkan pada beberapa ruas jalan di wilayah Jakarta Pusat, seperti Jalan Gajah Mada, Jalan Hayam Wuruk, Jalan Majapahit, Jalan Medan Merdeka Barat, dan Jalan M.H. Thamrin. Nilai kecepatan pada dataset menunjukkan variasi kondisi lalu lintas yang berbeda pada masing-masing ruas jalan.

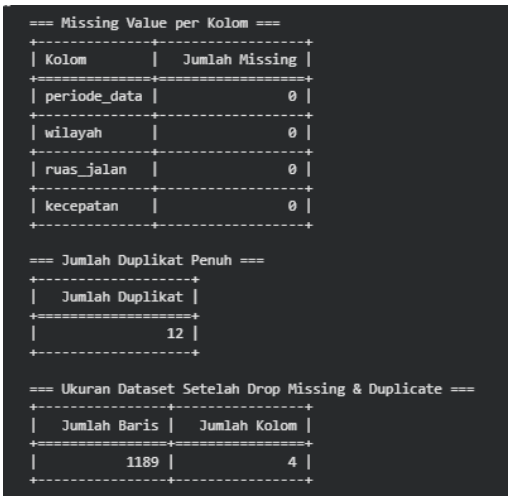
Hasil eksplorasi awal juga menunjukkan bahwa atribut kecepatan masih tersimpan dalam format object sehingga perlu dilakukan transformasi ke dalam bentuk numerik sebelum digunakan dalam proses analisis. Selain itu, data kemudian melalui tahap pembersihan dan normalisasi agar dapat digunakan secara optimal pada algoritma *K-Means Clustering.*

4.1.2 Hasil Data Preparation

Pada sub proses ini kita akan membahas 3 tahapan yaitu *cleaning data, feature engineering,* dan *normalize data.*

1. Pembersihan Data
   Pada tahap cleaning data dilakukan beberapa proses pembersihan dan validasi data untuk memastikan dataset berada dalam kondisi bersih, konsisten, dan siap digunakan dalam proses clustering menggunakan algoritma K-Means. Proses cleaning data pada penelitian ini dilakukan melalui beberapa tahapan sebagai berikut.
   a) Standarisasi Format dataset
      Tahap pertama dilakukan dengan membersihkan atribut bertipe string seperti wilayah dan ruas_jalan. Proses ini bertujuan untuk menghilangkan spasi kosong (whitespace) yang tidak diperlukan agar tidak terjadi perbedaan penulisan data yang sebenarnya sama. Selain itu, atribut kecepatan yang sebelumnya masih tersimpan dalam format object dikonversi ke dalam bentuk numerik menggunakan fungsi to_numeric(). Sebelum proses konversi dilakukan, tanda koma (,) pada nilai kecepatan terlebih dahulu diubah menjadi titik (.) agar sesuai dengan format numerik pada Python.

   b) Validasi Missing Value dan Data Duplikat
      Setelah proses standardisasi dilakukan, tahap berikutnya adalah pemeriksaan missing value dan data duplikat. Validasi ini bertujuan untuk memastikan tidak terdapat data kosong maupun data berulang yang dapat memengaruhi kualitas hasil clustering.
      1) atribut periode_data memiliki 0 missing value,
      2) atribut wilayah memiliki 0 missing value,
      3) atribut ruas_jalan memiliki 0 missing value,
      4) atribut kecepatan memiliki 0 missing value.



```
=== Missing Value per Kolom ===
+--------------+-----------------+
| Kolom        |   Jumlah Missing |
+==============+=================+
| periode_data |               0 |
+--------------+-----------------+
| wilayah      |               0 |
+--------------+-----------------+
| ruas_jalan   |               0 |
+--------------+-----------------+
| kecepatan    |               0 |
+--------------+-----------------+

=== Jumlah Duplikat Penuh ===
+-------------------+
|   Jumlah Duplikat |
+===================+
|                12 |
+-------------------+

=== Ukuran Dataset Setelah Drop Missing & Duplicate ===
+----------------+----------------+
|   Jumlah Baris |   Jumlah Kolom |
+================+================+
|           1189 |              4 |
+----------------+----------------+
```

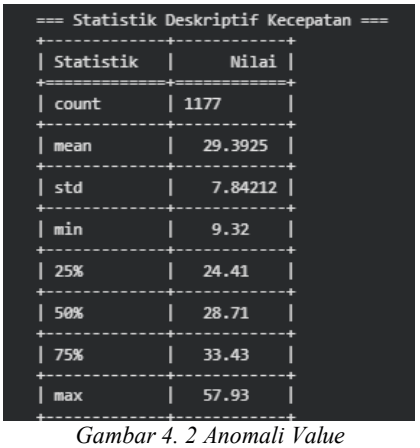
Gambar 4. 1 Missing value Dataset

Selain itu, hasil validasi juga menunjukkan bahwa dataset tidak memiliki data duplikat penuh. Oleh karena itu, jumlah dataset tetap sebanyak. ***1189 baris data dan 4 kolom atribut.***

Meskipun tidak ditemukan missing value namun duplikasi data ditemukan yaitu sebanyak 12 *row* data, proses pembersihan tetap dilakukan menggunakan metode *dropna*() dan *drop_duplicates()* untuk memastikan kualitas data tetap terjaga.

c) *Confersion* anomali Value

Tahap selanjutnya adalah proses filtering terhadap nilai kecepatan yang dianggap tidak logis atau tidak wajar dalam konteks lalu lintas perkotaan. Pada penelitian ini, nilai kecepatan kurang dari atau sama dengan nol (<= 0) dianggap tidak valid sehingga dihapus dari dataset.



```
=== Statistik Deskriptif Kecepatan ===
+-------------+------------+
| Statistik   |      Nilai |
+=============+============+
| count       | 1177       |
+-------------+------------+
| mean        |   29.3925  |
+-------------+------------+
| std         |    7.84212 |
+-------------+------------+
| min         |    9.32    |
+-------------+------------+
| 25%         |   24.41    |
+-------------+------------+
| 50%         |   28.71    |
+-------------+------------+
| 75%         |   33.43    |
+-------------+------------+
| max         |   57.93    |
+-------------+------------+
```

Gambar 4. 2 Anomali Value

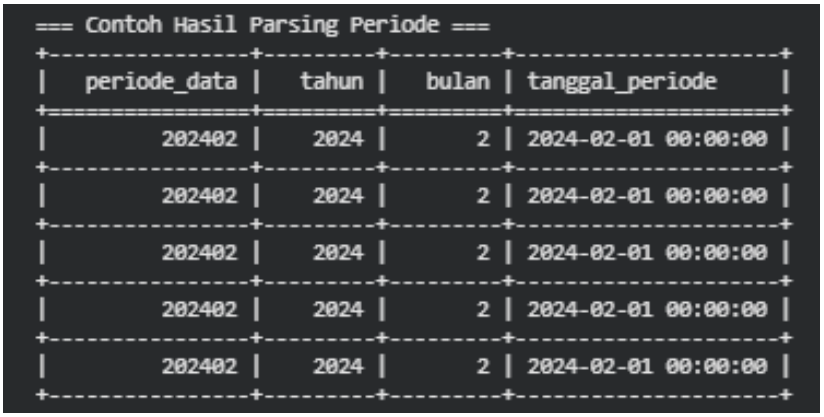
Pada gambar 4.2, dilakukan juga pembatasan terhadap nilai kecepatan ekstrem untuk mengurangi pengaruh outlier pada proses clustering. Nilai kecepatan maksimum dibatasi hingga 120 km/jam karena kecepatan di atas nilai tersebut dianggap kurang realistis untuk kondisi jalan perkotaan di DKI Jakarta.

1) jumlah data valid sebanyak 1189 data,
2) rata-rata kecepatan sebesar 29,39 km/jam,
3) standar deviasi sebesar 7,84 km/jam
4) kecepatan minimum sebesar 9,32 km/jam,
5) kecepatan maksimum sebesar 57,93 km/jam.

Hasil tersebut menunjukkan bahwa distribusi data kecepatan masih berada pada rentang yang wajar untuk kondisi lalu lintas perkotaan.

d) Standarisasi Format periode data

Tahap terakhir pada proses cleaning data adalah penyesuaian format atribut periode_data. Data periode yang semula berbentuk numerik dengan format YYYYMM dikonversi menjadi format string agar lebih mudah diproses.



```
=== Contoh Hasil Parsing Periode ===
+----------------+---------+---------+---------------------+
|   periode_data |   tahun |   bulan | tanggal_periode     |
+================+=========+=========+=====================+
|         202402 |    2024 |       2 | 2024-02-01 00:00:00 |
+----------------+---------+---------+---------------------+
|         202402 |    2024 |       2 | 2024-02-01 00:00:00 |
+----------------+---------+---------+---------------------+
|         202402 |    2024 |       2 | 2024-02-01 00:00:00 |
+----------------+---------+---------+---------------------+
|         202402 |    2024 |       2 | 2024-02-01 00:00:00 |
+----------------+---------+---------+---------------------+
|         202402 |    2024 |       2 | 2024-02-01 00:00:00 |
+----------------+---------+---------+---------------------+
```

Gambar 4. 3 Parsing Periode

Pada gambar 4.3 di atas dilakukan parsing data menggunakan fungsi to_datetime() untuk menghasilkan atribut, tahun, bulan, tanggal_periode. Penambahan atribut tersebut bertujuan untuk mempermudah proses analisis temporal terhadap pola lalu lintas pada periode tertentu.

2. Pembentukan Fitur

Pada tahap ini, fitur-fitur baru dibuat untuk menggambarkan karakteristik kecepatan lalu lintas pada setiap ruas jalan secara lebih komprehensif.

Proses ini dilakukan setelah tahap pembersihan data dan analisis data eksploratori (EDA), di mana hasil eksplorasi data digunakan sebagai dasar untuk menentukan fitur-fitur yang relevan bagi proses pengelompokan menggunakan algoritma *K-Means*. Berdasarkan hasil analisis data eksploratori (EDA), ditemukan bahwa dataset memiliki distribusi pengamatan yang berbeda di setiap kota administratif di Jakarta. Kota administratif dengan jumlah pengamatan terbanyak adalah Jakarta Pusat dengan 277 titik data, diikuti oleh Jakarta Selatan dengan 117 titik data, dan Jakarta Barat dengan 57 titik data. Selain itu, hasil eksplorasi juga menunjukan bahwa dataset terdiri dari 50 ruas jalan unik dan 25 peridode data unik.
Temuan ini menunjukkan bahwa data memiliki variasi temporal dan spasial yang cukup beragam sehingga diperlukan representasi fitur yang mampu menggambarkan pola kecepatan pada setiap ruas jalan secara lebih detail.

Selain itu, hasil statistik deskriptif kecepatan per wilayah menunjukkan bahwa setiap wilayah memiliki karakteristik lalu lintas yang berbeda. Perbedaan tersebut terlihat dari nilai rata-rata kecepatan (mean), standar deviasi (std), nilai minimum, maksimum, serta distribusi kuartil pada masing-masing wilayah. Sebagai contoh, wilayah Kota Administratif Jakarta Pusat memiliki rata-rata kecepatan sekitar 27 km/jam dengan nilai maksimum mencapai 57,93 km/jam, sedangkan beberapa wilayah lain menunjukkan rata-rata kecepatan yang lebih rendah. Variasi ini menunjukkan bahwa kondisi lalu lintas antar wilayah tidak homogen dan memiliki pola yang berbeda-beda.

Wilayah	Jumlah Data	Rata-rata	Std Deviasi	Kecepatan Min	Kecepatan Max
JAKARTA BARAT	71	21.47	6.91	12.15	33.35
JAKARTA PUSAT	347	26.54	8.15	11.0	57.93
JAKARTA SELATAN	147	27.36	7.04	9.32	47.05
JAKARTA TIMUR	48	34.36	5.94	21.42	42.15

Gambar 4. 4 EDA Dataset

Pada gambar 4.4 diatas, dilakukan proses pembentukan fitur menggunakan metode agregasi statistik berdasarkan atribut wilayah dan ruas_jalan. Proses agregasi dilakukan menggunakan fungsi groupby() untuk menghasilkan representasi setiap ruas jalan dalam bentuk statistik kecepatan. Feature engineering pada penelitian ini menghasilkan beberapa fitur utama.

1) rata_kecepatan (Mean)

Fitur ini menunjukkan rata-rata kecepatan kendaraan pada suatu ruas jalan. Nilai ini digunakan untuk merepresentasikan kondisi umum lalu lintas pada ruas jalan tersebut.

2) median_kecepatan
   Fitur median digunakan untuk mengetahui nilai tengah distribusi kecepatan dan lebih stabil terhadap pengaruh outlier dibandingkan nilai rata-rata.
3) min_kecepatan
   Fitur ini menunjukkan nilai kecepatan minimum yang pernah terjadi pada suatu ruas jalan, sehingga dapat merepresentasikan kondisi lalu lintas terpadat.
4) max_kecepatan
   Fitur ini menunjukkan nilai kecepatan maksimum yang terjadi pada ruas jalan dan digunakan untuk menggambarkan kondisi lalu lintas paling lancar.
5) std_kecepatan (Standar Deviasi)
   Fitur standar deviasi digunakan untuk mengukur tingkat variasi kecepatan kendaraan. Nilai standar deviasi yang tinggi menunjukkan bahwa kondisi lalu lintas pada ruas jalan cenderung fluktuatif.
6) range_kecepatan
   Fitur ini merupakan selisih antara nilai kecepatan maksimum dan minimum. Range digunakan untuk melihat rentang perubahan kecepatan pada suatu ruas jalan.
7) iqr_kecepatan (Interquartile Range)
   Fitur IQR diperoleh dari selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1). Fitur ini digunakan untuk mengukur sebaran data tanpa terlalu dipengaruhi oleh outlier.
8) cv_kecepatan (Coefficient of Variation)
   Fitur coefficient of variation digunakan untuk mengukur tingkat variasi relatif kecepatan terhadap rata-rata kecepatan. Nilai CV yang tinggi menunjukkan bahwa kondisi lalu lintas pada ruas jalan cenderung tidak stabil atau fluktuatif.

Melalui proses feature engineering ini, data mentah kecepatan lalu lintas berhasil diubah menjadi representasi numerik yang lebih informatif dan sesuai untuk proses clustering menggunakan *algoritma K-Means*. Fitur-fitur tersebut diharapkan mampu menggambarkan karakteristik kondisi lalu lintas pada setiap ruas jalan secara lebih akurat.

3. Normalisasi Data

Setelah proses feature engineering selesai dilakukan, tahap selanjutnya adalah normalisasi data menggunakan metode StandardScaler. Proses normalisasi bertujuan untuk menyamakan skala antar fitur agar seluruh variabel memiliki kontribusi yang seimbang dalam proses clustering menggunakan algoritma K-Means. Tahap ini penting dilakukan karena algoritma K-Means menggunakan perhitungan jarak berbasis Euclidean Distance yang sangat sensitif terhadap perbedaan skala data. Apabila terdapat fitur dengan rentang nilai yang jauh lebih besar dibandingkan fitur lainnya, maka fitur tersebut dapat mendominasi proses pengelompokan.

a) rata_kecepatan
b) median_kecepatan
c) min_kecepatan
d) max_kecepatan
e) std_kecepatan
f) range_kecepatan
g) iqr_kecepatan
h) cv_kecepatan

Metode StandardScaler bekerja dengan cara mengubah distribusi data sehingga setiap fitur memiliki.

1) rata-rata (mean) mendekati 0
2) standar deviasi (standard deviation) mendekati 1

Secara matematis, proses standardisasi dilakukan menggunakan persamaan berikut.

$$z = \frac{x - \mu}{\sigma}$$

Keterangan.

a) Z = nilai hasil standardisasi
b) X = nilai data asli
c) μ = rata-rata data
d) σ = standar deviasi data

Hasil proses standardisasi menunjukkan bahwa seluruh fitur berhasil dinormalisasi dengan baik. Berdasarkan hasil perhitungan, nilai rata-rata setiap fitur berada sangat dekat dengan 0, seperti.

rata_kecepatan = 2.22×10^{-16}

median_kecepatan = -3.08×10^{-16}

cv_kecepatan = -2.72×10^{-16}

Nilai rata-rata hasil standardisasi tidak menghasilkan angka 0 secara mutlak, melainkan mendekati 0 dalam bentuk bilangan sangat kecil seperti 2.22×10

$^{-16}$. Nilai tersebut muncul akibat keterbatasan presisi perhitungan floating point pada komputer dan secara praktis tetap dianggap sebagai 0. Hasil ini menunjukkan bahwa proses standardisasi menggunakan StandardScaler telah berhasil dilakukan dengan baik.
Selain itu, nilai standar deviasi seluruh fitur berada mendekati 1, yaitu sekitar 1.0067. Hasil tersebut menunjukkan bahwa proses standardisasi berhasil membuat distribusi data menjadi lebih seragam sehingga tidak terdapat fitur yang mendominasi proses clustering.

Tahap normalisasi ini menghasilkan dataset baru dalam bentuk data terstandardisasi (X_scaled) yang kemudian digunakan sebagai input utama dalam proses penentuan jumlah cluster menggunakan metode Elbow, evaluasi Silhouette Score, dan proses clustering menggunakan algoritma K-Means. Sebagai hasil akhir, data yang telah distandardisasi memiliki rentang distribusi yang lebih seimbang dan siap digunakan dalam proses analisis clustering untuk mengidentifikasi segmentasi ruas jalan berdasarkan karakteristik kecepatan lalu lintas di DKI Jakarta.

4.1.3 Hasil Elbow Method

Pada tahap modeling, sebelum proses clustering dilakukan, terlebih dahulu dilakukan penentuan jumlah cluster optimal menggunakan metode *Elbow Method*. Metode ini digunakan untuk menentukan jumlah cluster terbaik dengan menganalisis nilai inertia atau *Sum of Squared Errors (SSE)* pada beberapa jumlah cluster yang berbeda. Inertia merupakan total jarak kuadrat antara setiap data dengan centroid cluster terdekat. Semakin kecil nilai inertia, maka data dalam cluster semakin rapat dan homogen.

Dalam penelitian ini, pengujian dilakukan dengan menggunakan jumlah kluster berkisar dari k=1 hingga k=10. Perhitungan dilakukan menggunakan algoritma *KMeans* dengan parameter *random_state = 42* untuk memastikan konsistensi hasil pengelompokan dan *n_init = 20* untuk menginisialisasi pusat kluster sebanyak 20 kali, sehingga hasil pengelompokan menjadi lebih stabil dan mencapai konvergensi optimal.

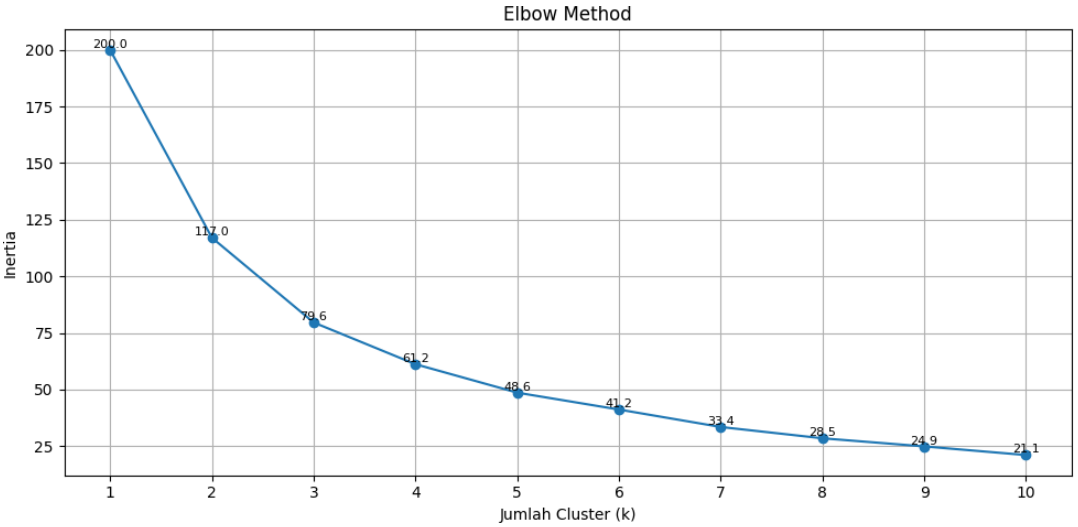
Hasil perhitungan inersia kemudian divisualisasikan dalam bentuk grafik Metode Siku. Berdasarkan grafik ini, dapat dilihat bahwa nilai inersia menurun sangat signifikan pada awal peningkatan jumlah kluster. Nilai inersia untuk setiap kluster adalah sebagai berikut.

Tabel 4. 2 Grafik Elbow

k=1	206.0
k=2	117.0
k=3	79.609
k=4	61.8
k=5	48.8
k=6	41.7
k=7	33.44
k=8	28.45
k=9	24.87
k=10	21.05

Penurunan drastis dari k=1 hingga k=3 *Inertia* menurun dari 206.0 menjadi 79.609, yakni sebesar 61.4% pengurangan. Penurunan yang sangat besar ini menunjukkan adanya pengelompokan yang sangat efektif pada tahap awal, di mana penambahan setiap *cluster* baru memberikan kontribusi signifikan dalam mengurangi jarak data terhadap *centroid*. Penurunan moderat dari k=3 hingga k=6 Inertia menurun dari 79.609 menjadi 41.153, sebesar 48.3% pengurangan. Penurunan masih signifikan tetapi sudah mulai melambat jika dibandingkan dengan tahap pertama. Ini menunjukkan bahwa efektivitas penambahan cluster baru mulai berkurang.

Penurunan gradual dari k=6 hingga k=10 *Inertia* menurun dari 41.153 menjadi 21.057, sebesar 48.8% pengurangan total. Meskipun persentasenya serupa dengan tahap sebelumnya, nilai pengurangan inertia per langkah (per cluster tambahan) menjadi semakin kecil. Sebagai contoh, penambahan cluster dari k=9 ke k=10 hanya mengurangi inertia sebesar 3.814 (dari 24.871 menjadi 21.057). Titik *elbow* yang menunjukkan perubahan kecepatan penurunan inertia terlihat jelas pada k=3 hingga k=4. Pada rentang ini, terjadi transisi dari penurunan yang sangat dramatis menjadi penurunan yang lebih gradual. Hal ini mengindikasikan bahwa setelah k=3, penambahan *cluster* baru tidak lagi memberikan perbaikan yang signifikan dalam hal pemisahan data.



Gambar 4. 5 Grafik Elbow Method

Berdasarkan visualisasi grafik *Elbow* Method pada Gambar 4.5, dapat dilihat dengan jelas bahwa kurva menunjukkan pola yang khas dari *elbow* point, yaitu:

a. Kurva curam di k=1 hingga k=3 Garis tersebut turun dengan kemiringan yang sangat tinggi (slope, -43.2 per *cluster*), menunjukkan efektivitas tinggi dalam mengurangi *inertia*.
b. Perubahan kecepatan di k=3 hingga k=4 Kemiringan kurva mulai berkurang (slope, -18.4 per *cluster*), menandai awal dari *elbow* point atau siku kurva.
c. Kurva relatif datar dari k=4 hingga k=10 Setelah k=4, kurva terus menurun tetapi dengan kemiringan yang jauh lebih lambat (slope, -5.2 per *cluster*), hampir linear dan mendekati asimptotik.

Dalam metode *Elbow*, titik di mana perubahan kecepatan penurunan inertia mulai melandai disebut sebagai elbow point, dan ini dianggap sebagai indikasi jumlah cluster optimal. Berdasarkan analisis visual dan numerik, elbow point berada pada rentang k=3 hingga k=4. Baik k=3 maupun k=4 memiliki potensi untuk menjadi jumlah cluster optimal dalam penelitian ini.

Hasil dari metode *Elbow* ini kemudian digunakan sebagai dasar awal *(candidate values)* dalam menentukan jumlah cluster yang akan dievaluasi lebih lanjut menggunakan metode Silhouette Score. Kombinasi kedua metode ini memungkinkan pemilihan jumlah cluster yang tidak hanya optimal secara matematis tetapi juga interpretabel secara praktis dalam konteks analisis kemacetan lalu lintas.

4.1.4 Hasil Silhouette Score

Setelah dilakukan penentuan jumlah *cluster* menggunakan metode *Elbow*, tahap selanjutnya adalah evaluasi kualitas cluster menggunakan metode *Silhouette Score.* Evaluasi ini bertujuan untuk mengukur seberapa baik setiap data (ruas jalan) berada pada *cluster* yang terbentuk dibandingkan dengan cluster lainnya. *Silhouette Score* digunakan karena mampu mengevaluasi tingkat kepadatan atau *cohesion* (kesamaan internal dalam *cluster*) dan pemisahan atau separation (perbedaan eksternal antar *cluster*) secara bersamaan.

Pada penelitian ini, pengujian dilakukan terhadap jumlah cluster k=2 hingga k=10. Proses evaluasi dilakukan menggunakan algoritma *KMeans* pada data yang telah melalui proses standardisasi *(X_scaled)* dengan fitur input, rata_kecepatan, std_kecepatan, iqr_kecepatan, dan jumlah_periode. Untuk setiap nilai k, model *KMeans* dijalankan menggunakan parameter *random_state* = 42 dan n_init = 20 untuk menghasilkan proses *clustering* yang lebih stabil dan konsisten. Selanjutnya, nilai *Silhouette Score* dihitung menggunakan fungsi *silhouette_score()* dari pustaka *Scikit-learn.*

Hasil evaluasi menunjukkan bahwa nilai *silhouette score* mengalami perubahan pada setiap jumlah cluster yang diuji. Nilai *silhouette score* yang diperoleh adalah sebagai berikut.

Tabel 4. 3 silhouette Score

Jumlah Cluster (k)	Nilai Silhouette Score
2	0.414703
3	0.415005
4	0.389434
5	0.36817
6	0.354135
7	0.397977
8	0.390897
9	0.392491
10	0.385689

Analisis terhadap hasil evaluasi Silhouette Score mengungkapkan beberapa temuan penting

Silhouette Score Tertinggi Berdasarkan tabel di atas, Silhouette Score tertinggi diperoleh untuk k=3, dengan nilai 0,415005, disusul sangat dekat oleh k=2 dengan nilai 0,414703. Perbedaan antara kedua nilai ini sangat kecil (0,000302), yang menunjukkan bahwa kedua opsi tersebut memberikan kualitas pengelompokan yang hampir setara dalam hal kepadatan internal dan pemisahan antar *cluster*.

Pola Perubahan *Silhouette Score* berdasarkan kurva.

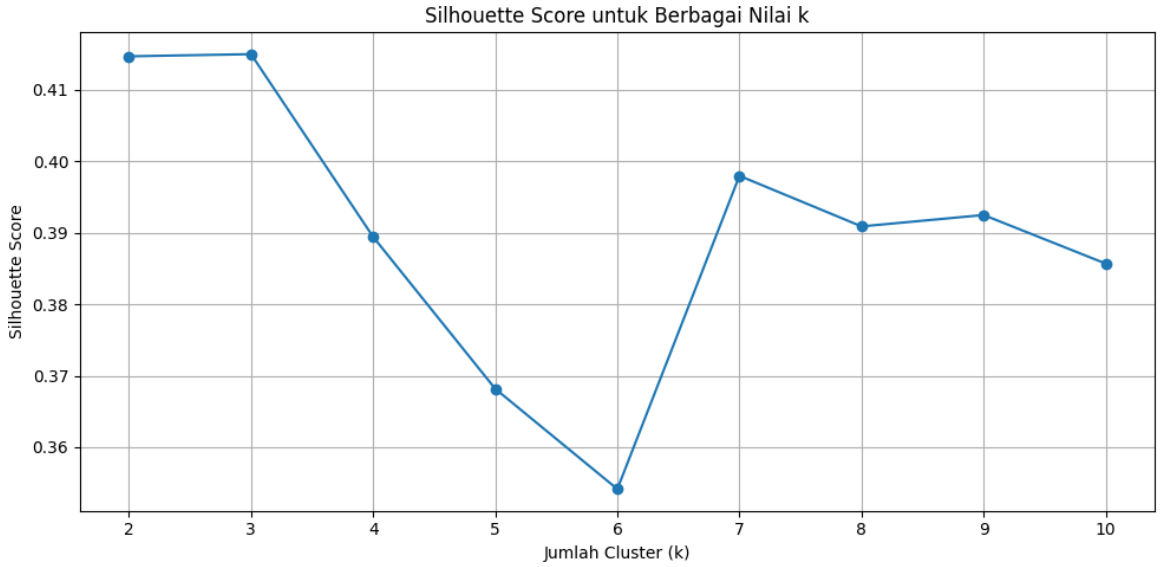
a. Peningkatan dari k=2 ke k=3 Terjadi peningkatan sebesar 0.000302 (0.02%), menunjukkan perbaikan kualitas cluster yang sangat minimal.
b. Penurunan signifikan dari k=3 ke k=6 *Silhouette Score* menurun dari 0.415005 menjadi 0.354135, sebesar 14.7%. Penurunan ini menunjukkan bahwa penambahan cluster melebihi k=3 menghasilkan kualitas pengelompokan yang menurun.
c. Peningkatan kembali di k=7 *Silhouette Score* naik menjadi 0.397977, menunjukkan adanya nilai lokal yang lebih tinggi. Namun, nilai ini masih di bawah nilai k=3.
d. Fluktuasi dari k=7 ke k=10 Nilai berfluktuasi dalam rentang 0.386-0.397, menunjukkan ketidakstabilan dalam kualitas cluster di jumlah cluster yang lebih tinggi.

Interpretasi Kualitas *Cluster* Secara umum, nilai *silhouette score* pada penelitian ini berada dalam rentang 0.35 hingga 0.42. Menurut standar interpretasi *Silhouette Score,* rentang nilai ini mengindikasikan struktur cluster dalam kategori sedang hingga baik. Hal ini berarti.

a. Sebagian besar data telah berada pada *cluster* yang sesuai.
b. Terdapat pemisahan yang cukup baik antara cluster.
c. Namun, masih terdapat kemungkinan overlap atau ketidakjelasan batas pada beberapa data.

Analisis Perbandingan k=2 vs k=3 Meskipun k=2 dan k=3 memberikan *Silhouette Score* yang hampir identik, terdapat perbedaan penting dalam hal interpretabilitas.

a. k=2 Menghasilkan dua kategori cluster, misalnya *"Lancar"* dan *"Tidak Lancar"* atau *"Padat"* dan *"Tidak Padat"*. Pembagian ini terlalu sederhana untuk menggambarkan tingkat kemacetan yang kompleks di DKI Jakarta.
b. k=3 Menghasilkan tiga kategori cluster yang lebih informatif, yaitu *"Lancar", "Sedang",* dan *"Macet*". Pembagian ini lebih sesuai dengan kondisi lapangan dan lebih mudah dipahami oleh praktisi transportasi.



Gambar 4. 6 Grafik Silhouette Score untuk Berbagai Nilai k

Berdasarkan pertimbangan holistik terhadap nilai Silhouette Score yang hampir setara antara k=2 dan k=3, serta mempertimbangkan aspek praktis dan interpretabilitas hasil, *penelitian ini memilih k=3 sebagai jumlah cluster optimal.* Pemilihan ini didasarkan pada:

1. Kualitas *Cluster* yang Sebanding Nilai *Silhouette Score* untuk k=3 (0.415005) hanya 0.02% lebih tinggi dari k=2 (0.414703), menunjukkan kualitas yang sebanding.

b. *Elbow* Method Mendukung k=3 Titik *elbow* yang jelas terlihat pada k=3 dalam grafik Elbow Method memperkuat pemilihan jumlah cluster ini.

c. Interpretabilitas Praktis: Pembagian ke dalam tiga kategori *(Lancar, Sedang, Macet)* lebih sesuai dengan konteks analisis kemacetan lalu lintas di DKI Jakarta dan lebih mudah dipahami oleh pengambil keputusan.

d. Relevansi terhadap Tujuan Penelitian: Tujuan penelitian adalah untuk mengelompokkan ruas jalan berdasarkan karakteristik kecepatan agar dapat memberikan gambaran yang lebih terstruktur tentang tingkat kemacetan di berbagai lokasi. Pembagian tiga kategori lebih informatif dalam mencapai tujuan ini.

Dengan demikian, pemilihan jumlah cluster tidak hanya mempertimbangkan nilai evaluasi matematis semata, tetapi juga relevansi hasil terhadap konteks praktis dan kebutuhan analisis segmentasi ruas jalan di DKI Jakarta.

4.1.5 Hasil Clustering

Setelah dilakukan proses penentuan jumlah cluster dan evaluasi model menggunakan metode *Elbow* Method serta Silhouette Score, tahap selanjutnya

adalah implementasi clustering menggunakan algoritma *KMeans*. Pada penelitian ini, jumlah cluster yang digunakan adalah sebanyak 3 cluster (k=3) yang merepresentasikan kategori kondisi lalu lintas, yaitu cluster lancar, sedang, dan macet. Pemilihan jumlah cluster tersebut dilakukan untuk mempermudah interpretasi segmentasi ruas jalan berdasarkan karakteristik kecepatan lalu lintas di DKI Jakarta.

Berdasarkan hasil evaluasi model akhir dengan menggunakan data terstandarisasi dan parameter *KMeans* yang optimal (*random_state* = 42, n_init = 20), diperoleh nilai *Silhouette Score* sebesar 0.415005. Nilai tersebut menunjukkan bahwa hasil clustering memiliki struktur *cluster* dalam kategori baik (strong), di mana sebagian besar data telah berada pada cluster yang sesuai dengan pemisahan yang cukup memuaskan antar *cluster*. Meskipun masih terdapat kemungkinan *overlap* antar cluster pada beberapa ruas jalan, nilai *Silhouette Score* ini menunjukkan bahwa model clustering yang dibangun telah mampu membentuk segmentasi ruas jalan berdasarkan pola kecepatan lalu lintas dengan cukup efektif.

Hasil *clustering* menghasilkan tiga kelompok utama dengan karakteristik yang berbeda. Ringkasan hasil clustering dapat dilihat berdasarkan jumlah ruas jalan dan nilai statistik rata-rata pada masing-masing *cluster*.

Tabel 4. 4 Ringkasan Hasil Clustering

Kategori	Jumlah Ruas	Rata-rata (km/h)	Min	Max	Std Dev	CV
Lancar	7	33.43	27.92	39.97	7.85	0.235
Sedang	25	28.86	13.50	42.14	2.57	0.087
Macet	18	23.91	12.89	31.82	4.18	0.177
Total	50	28.73	12.89	42.14	4.86	0.169

Analisis terhadap tabel di atas mengungkapkan beberapa insight penting *cluster* lancar (n=7 ruas jalan)Cluster lancar merupakan kategori dengan rata-rata kecepatan tertinggi, yaitu 33.43 km/jam. Cluster ini mencakup 7 ruas jalan (14%) dari total 50 ruas jalan yang dianalisis. Karakteristik utama cluster ini adalah:

a. Range kecepatan: 27.92 km/jam (minimum) hingga 39.97 km/jam (maksimum)
b. Standar deviasi: 7.85 km/h, menunjukkan variasi kecepatan yang tinggi
c. *Coefficient of Variation CV* 0.235, mengindikasikan bahwa meskipun rata-rata tinggi, terdapat fluktuasi kecepatan yang signifikan pada beberapa periode.
d. Interpretasi Ruas jalan dalam cluster ini memiliki kondisi lalu lintas yang relatif lancar, namun dengan variasi kecepatan yang cukup tinggi. Hal ini

menunjukkan bahwa meskipun arus lalu lintas secara umum lancar, terdapat periode tertentu di mana kecepatan menurun (mungkin pada jam sibuk).

Contoh ruas jalan dalam cluster ini mencakup Jalan Salemba Raya 27.92 km, Jalan Kramat Raya 30.34 km, dan Jalan Hayam Wuruk 39.97 km, yang merupakan ruas jalan utama di wilayah Jakarta Pusat *cluster* Sedang n=25 ruas jalan *cluster* sedang merupakan kategori terbesar dengan 25 ruas jalan 50% dari total sampel. Kategori ini memiliki rata-rata kecepatan 28.86 km/jam, yang berada di antara cluster macet dan lancar. Karakteristik utama cluster ini adalah.

a. Range kecepatan 13.50 km/jam minimum hingga 42.14 km/jam maksimum
b. Standar deviasi 2.57 km/h, merupakan nilai terkecil di antara ketiga cluster
c. *CV* 0.087, nilai terendah menunjukkan kondisi lalu lintas yang sangat stabil dan konsisten
d. Interpretasi *Cluster* sedang menunjukkan kondisi lalu lintas yang stabil dan dapat diprediksi. Fluktuasi kecepatan kecil menunjukkan bahwa pola lalu lintas pada ruas jalan ini konsisten sepanjang periode observasi.

Range kecepatan yang lebar 13.50 hingga 42.14 km/h menunjukkan bahwa meskipun sebagian besar data menunjukkan kecepatan sedang, terdapat beberapa outlier atau periode dengan kondisi ekstrem. Namun, nilai *CV* yang rendah 0.087 menunjukkan bahwa *outlier* ini tidak mendominasi data.

Ruas jalan dalam cluster ini tersebar di berbagai wilayah Jakarta, mencakup Jalan Gunung Sahari 22.56 km/h, Jalan D.I. Panjaitan 27.88 km/h, dan Jalan Balikpapan 42.14 km/h. *cluster* macet (n=18 ruas jalan). *Cluster* macet merupakan kategori dengan rata-rata kecepatan terendah, yaitu 23.91 km/jam, *cluster* ini mencakup 18 ruas jalan 36% dari total sampel dan merepresentasikan kondisi lalu lintas yang paling padat. Karakteristik utama adalah.

a. Range kecepatan 12.89 km/jam minimum hingga 31.82 km/jam maksimum
b. Standar deviasi: 4.18 km/h, menunjukkan variasi moderat
c. *CV* 0.177, mengindikasikan beberapa fluktuasi tetapi lebih stabil dari cluster lancar
d. Interpretasi Ruas jalan dalam cluster ini mengalami kemacetan yang konsisten dengan kecepatan rata-rata di bawah 24 km/h. Nilai *CV* 0.177 menunjukkan bahwa meskipun terdapat fluktuasi harian, ruas jalan ini secara umum selalu padat.

Kecepatan minimum 12.89 km/h Jalan Pintu Besar Selatan menunjukkan kondisi yang sangat padat pada periode tertentu atau terus-menerus. Ruas jalan dalam cluster ini mencakup Jalan Majapahit 17.44 km/h, Jalan Kyai Caringin 19.63

km/h, dan Jalan Sisimanggaraja 21.72 km/h, yang semuanya terletak di wilayah Jakarta Pusat dan Jakarta Selatan.

4.1.2 Karakteristik *Cluster*

Hasil clustering menggunakan algoritma *KMeans* menghasilkan tiga kategori ruas jalan, yaitu cluster lancar, sedang, dan macet berdasarkan karakteristik kecepatan lalu lintas. Perbedaan signifikan antar cluster terlihat jelas dari nilai rata-rata kecepatan.

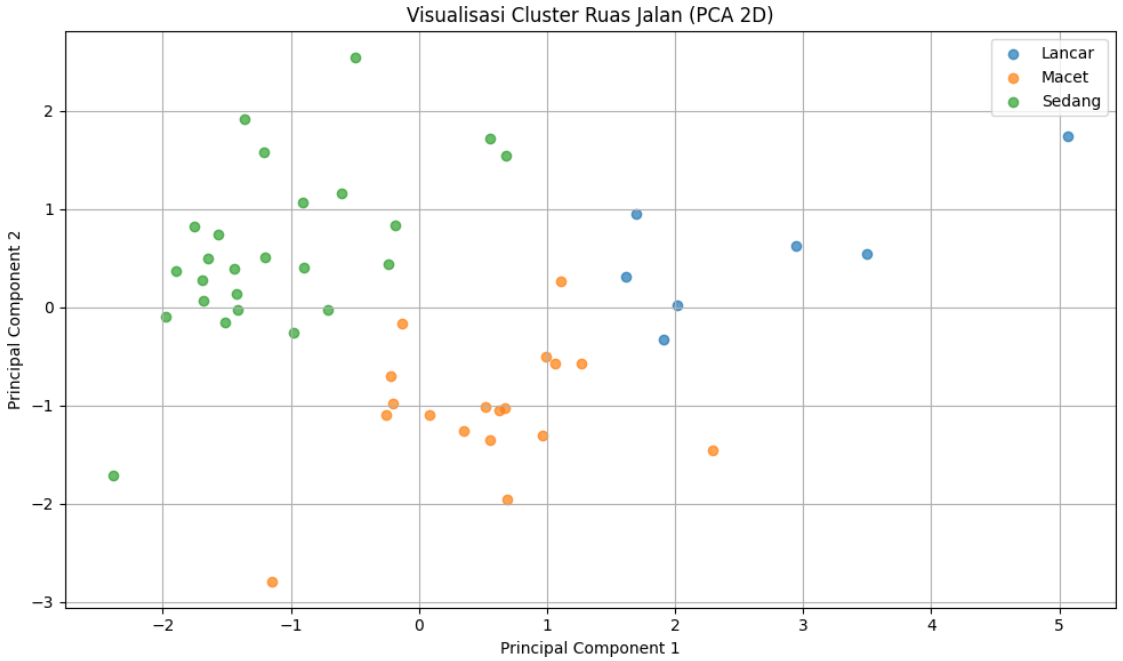
1) Perbedaan Lancar vs Macet: 33.43 - 23.91 = 9.52 km/h (39.8% perbedaan)
2) Perbedaan Lancar vs Sedang: 33.43 - 28.86 = 4.57 km/h (15.8% perbedaan)
3) Perbedaan Sedang vs Macet: 28.86 - 23.91 = 4.95 km/h (20.7% perbedaan)

Perbedaan rata-rata kecepatan yang signifikan (terutama antara lancar dan macet) memvalidasi bahwa algoritma *KMeans* berhasil mengidentifikasi segmentasi yang bermakna dalam data. *Diferensiasi* 9.52 km/h antara *cluster* lancar dan macet adalah perbedaan yang cukup substansial dalam konteks lalu lintas perkotaan.

4.1.6 Visualisasi

Visualisasi hasil clustering dilakukan untuk memperjelas pola pengelompokan ruas jalan berdasarkan karakteristik kecepatan lalu lintas. Pada penelitian ini digunakan tiga bentuk visualisasi, yaitu visualisasi PCA 2D, visualisasi rata-rata kecepatan per kategori cluster, dan visualisasi scatter berdasarkan fitur asli. Ketiga visualisasi tersebut digunakan karena masing-masing memiliki fungsi yang berbeda dalam membantu memahami hasil segmentasi ruas jalan.

1) Visualisasi cluster 2D dengan PCA
   Visualisasi PCA 2D digunakan untuk menampilkan hasil clustering dalam bentuk dua dimensi. Hal ini diperlukan karena proses clustering dilakukan menggunakan beberapa fitur sekaligus, seperti rata-rata kecepatan, median kecepatan, standar deviasi, range, IQR, dan CV. Dengan menggunakan PCA, fitur-fitur tersebut diringkas menjadi dua komponen utama, yaitu Principal Component 1 dan Principal Component 2, sehingga sebaran cluster dapat ditampilkan secara visual.

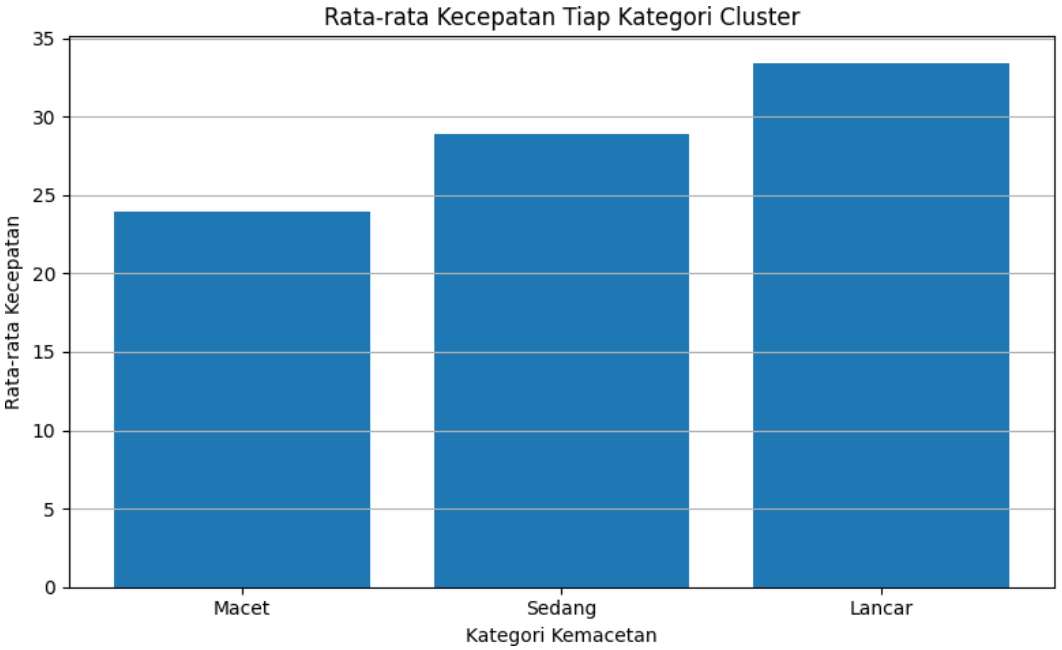


Gambar 4. 7 PCA visualization

Berdasarkan Gambar 4.6, terlihat bahwa cluster macet, sedang, dan lancar memiliki pola sebaran yang berbeda. Cluster macet cenderung berada pada area sisi kiri dan bawah grafik, cluster sedang banyak tersebar pada area tengah hingga atas, sedangkan cluster lancar cenderung berada pada sisi kanan grafik. Pola ini menunjukkan bahwa hasil clustering memiliki perbedaan karakteristik antar kelompok, meskipun masih terdapat beberapa titik yang berdekatan antar cluster. Hal tersebut sesuai dengan nilai silhouette score model akhir yang berada pada kategori sedang.

2) Visualisasi Kategori Cluster

Visualisasi rata-rata kecepatan per kategori cluster dibuat untuk melihat perbedaan tingkat kecepatan pada setiap kelompok hasil clustering. Grafik ini penting karena rata-rata kecepatan merupakan indikator utama dalam penelitian ini untuk merepresentasikan kondisi lalu lintas pada ruas jalan.

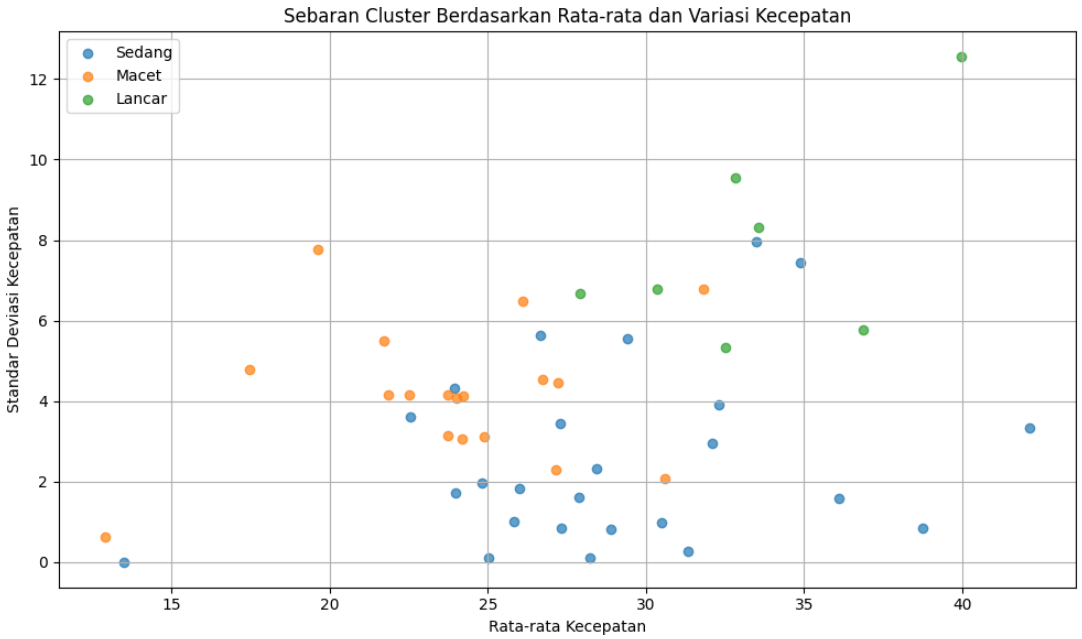


Gambar 4. 8 Visualization Categori Clsuter

Berdasarkan gambar 4.7, cluster macet memiliki rata-rata kecepatan paling rendah, yaitu sekitar 20,99 km/jam. Cluster sedang memiliki rata-rata kecepatan sekitar 29,14 km/jam, sedangkan cluster lancar memiliki rata-rata kecepatan tertinggi, yaitu sekitar 32,08 km/jam. Perbedaan tersebut menunjukkan bahwa hasil clustering sudah sesuai dengan tujuan penelitian, yaitu mengelompokkan ruas jalan berdasarkan karakteristik kecepatan lalu lintas. Semakin tinggi rata-rata kecepatan suatu cluster, maka kondisi lalu lintas pada cluster tersebut cenderung lebih lancar.

3) Visualilasi Scarter

Visualisasi scatter berdasarkan fitur asli digunakan untuk melihat hubungan antara rata-rata kecepatan dan standar deviasi kecepatan pada setiap ruas jalan. Visualisasi ini dibuat agar hasil clustering tidak hanya dilihat dari data hasil reduksi PCA, tetapi juga dari fitur asli yang lebih mudah diinterpretasikan dalam konteks lalu lintas.



Gambar 4. 9 Visulization Scater

Pada grafik ini, sumbu X menunjukkan rata-rata kecepatan, sedangkan sumbu Y menunjukkan standar deviasi kecepatan. Cluster macet cenderung berada pada area rata-rata kecepatan yang rendah, cluster sedang berada pada rentang rata-rata kecepatan menengah dengan variasi yang relatif rendah, sedangkan cluster lancar berada pada rentang rata-rata kecepatan yang lebih tinggi dan memiliki variasi kecepatan yang lebih besar pada beberapa ruas jalan.

Visualisasi ini menunjukkan bahwa pengelompokan tidak hanya dipengaruhi oleh tinggi atau rendahnya rata-rata kecepatan, tetapi juga oleh tingkat variasi kecepatan pada ruas jalan. Dengan demikian, ruas jalan yang memiliki kecepatan tinggi tetapi sangat fluktuatif dapat memiliki karakteristik berbeda dibandingkan ruas jalan dengan kecepatan menengah namun stabil.

Secara keseluruhan, ketiga visualisasi tersebut saling melengkapi dalam menjelaskan hasil clustering. Visualisasi PCA membantu melihat sebaran cluster secara umum, visualisasi rata-rata kecepatan membantu membandingkan kategori kemacetan, sedangkan visualisasi fitur asli membantu memahami hubungan antara kecepatan rata-rata dan variasi kecepatan pada setiap ruas jalan.

4.2 Pembahasan

4.2.1 Interpretasi Cluster

Hasil pengelompokan K-Means dengan k=3 menghasilkan tiga kategori ruas jalan, yang masing-masing mewakili kondisi lalu lintas yang berbeda di DKI Jakarta.

a) *Cluster* lancar (7 ruas, 14%) menunjukkan arus lalu lintas yang lancar dengan kecepatan rata-rata 33,43 km/jam. Kategori ini mencerminkan ruas jalan dengan kapasitas yang memadai untuk menampung volume kendaraan yang ada. Simpangan baku sebesar 7,85 km/jam (*CV* = 0,234) menunjukkan fluktuasi kecepatan yang signifikan selama periode tertentu, yang mengindikasikan bahwa meskipun kondisi rata-ratanya lancar, segmen jalan ini tetap rentan terhadap perubahan beban lalu lintas.
b) *Cluster* Sedang (25 ruas, 50%) merupakan kategori terbesar dengan rata-rata kecepatan 28.86 km/jam. Fitur paling membedakan cluster ini adalah stabilitas yang tinggi dengan standar deviasi terendah (2.57 km/h, *CV* = 0.087). Hal ini menunjukkan bahwa meskipun kecepatan berada pada level menengah (kemacetan ringan), pola lalu lintas pada ruas-ruas ini sangat konsisten dan predictable sepanjang periode observasi. Karakteristik ini sangat berharga untuk perencanaan dan manajemen lalu lintas jangka panjang.
c) *Cluster* Macet (18 ruas, 36%) menunjukkan kondisi kemacetan konsisten dengan rata-rata kecepatan 23.91 km/jam (terendah di antara ketiga kategori). Ruas-ruas dalam kategori ini mengalami hambatan dalam pergerakan kendaraan dengan kecepatan yang jauh di bawah kondisi optimal. Nilai *CV* 0.177 menunjukkan bahwa kemacetan pada ruas-ruas ini bersifat sistematis, bukan anomali temporer.

Perbedaan rata-rata kecepatan antar *cluster* sangat signifikan: 9.52 km/h (39.8%) antara Lancar dan Macet, memvalidasi bahwa k=3 menghasilkan segmentasi yang bermakna dengan diferensiasi substantive.

4.2.2 Analisis Pola

Analisis distribusi *cluster* menunjukkan pola spasial yang sejalan dengan teori lalu lintas perkotaan. Kluster kemacetan lalu lintas terkonsentrasi di Jakarta Pusat dan Jakarta Selatan (pusat kota), sejalan dengan temuan yang mengidentifikasi bahwa ruas jalan dengan karakteristik lalu lintas serupa cenderung membentuk kluster berdasarkan lokasi geografis dan fungsi jalan. Konsentrasi kemacetan lalu lintas di pusat kota mencerminkan kepadatan aktivitas ekonomi

yang tinggi, pusat-pusat lapangan kerja, serta pola mobilitas perkotaan yang kompleks.

Sebaliknya, *Cluster* Sedang tersebar merata di seluruh DKI Jakarta, mencakup Jakarta Pusat, Selatan, Timur, Barat, dan Utara. Pola distribusi ini menunjukkan bahwa kondisi "cukup lancar dengan variasi terbatas" adalah kondisi umum yang ditemukan di berbagai lokasi DKI Jakarta, mengindikasikan bahwa manajemen lalu lintas telah mencapai level operasional yang stabil di area-area tersebut. Analisis terhadap nilai standar deviasi dan Coefficient of Variation mengungkapkan pola temporal yang penting.

Tabel 4. 5 Pola Temporal variabelitas kecepatan

Cluster	**Std Dev**	***CV***	**Pola Temporal**
Lancar	7.85	0.234	Fluktuasi besar, sensitif beban
Sedang	2.57	0.087	Stabil & konsisten ← OPTIMAL
Macet	4.18	0.177	Kemacetan sistematis

Berdasarkan table 4.1 di atas ditemukan *cluster* Sedang menunjukkan variabilitas terendah *CV* = 0.087, mengindikasikan predictability tertinggi dalam pola lalu lintas. Sebaliknya, Cluster Lancar menunjukkan fluktuasi besar *CV* = 0.234, menunjukkan bahwa kecepatan tinggi tidak selalu berkorelasi dengan stabilitas operasional.

Hal ini menunjukkan bahwa karakteristik stabilitas temporal lebih berharga daripada kecepatan rata-rata untuk keperluan perencanaan dan prediksi. kapasitas struktural Jalan Analisis nilai maksimum kecepatan per cluster mengungkapkan batas struktural.

a. *Cluster* Macet Max = 31.82 km/h (tetap dalam kategori macet meskipun kondisi optimal)
b. *Cluster* Sedang Max = 42.14 km/h (dapat mencapai lancar pada kondisi optimal)
c. *Cluster* Lancar Max = 39.97 km/h (tetap lancar dalam kondisi terbaik)

Temuan ini menunjukkan bahwa kapasitas jalan memiliki batas struktural yang jelas. Ruas-ruas dalam *Cluster* Macet tidak dapat mencapai kecepatan tinggi bahkan dalam skenario ideal, mengindikasikan bahwa permasalahan kemacetan pada ruas-ruas tersebut bersifat struktural dan memerlukan intervensi fisik (perbaikan geometrik, perluasan kapasitas) bukan hanya manajemen operasional.

4.2.3 Alasan Pemilihan Nilai K

Berdasarkan hasil evaluasi clustering, dari kedua metode evaluasi tersebut, nilai K di tentukan berdasarkan 2 pertimbangan berikut ini. Yaitu pertimbangankuantitatif dan pertimbangan praktis implementasi.

1) Pertimbangan Kuantitatif
   a) *Silhouette Score* yang optimal nilai *Silhouette Score* untuk k=3 0.415 hanya 0.02% lebih tinggi dari k=2 0.414), menunjukkan kualitas cluster yang setara. Namun, elbow point yang jelas terlihat pada k=3 dalam kurva *elbow method*, di mana inertia menurun 61.4% dari k=1 ke k=3 206.0 → 79.609, menunjukkan efektivitas tinggi dalam pengelompokan.
   b) Stabilitas Evaluasi Dari k=3 ke k=4, Silhouette Score menurun dari 0.415 menjadi 0.389 6.1% penurunan, mengindikasikan bahwa penambahan cluster ke-4 tidak meningkatkan kualitas, malah menguranginya.
   c) Pola *Inertia* yang Jela: Penurunan *inertia* per step menunjukkan perubahan slope yang signifikan pada k=3, mengindikasikan transisi dari penurunan dramatis menjadi *gradual*, yang merupakan karakteristik *elbow* point.
2) Pertimbangan Praktis dan Implementasi
   a) Kemudahan Interprasi Cluster

      Pemilihan K=3 menghasilkan tiga kelompok kondisi lalu lintas yang mudah dipahami, yaitu Lancar, Sedang, dan Macet. Pembagian ini memungkinkan hasil clustering diinterpretasikan berdasarkan tingkat kecepatan lalu lintas pada masing-masing ruas jalan.

   b) Kesesuaian dengan Karakteristik Data

      K=3 mampu membedakan ruas jalan berdasarkan pola kecepatan lalu lintas menjadi tiga kondisi utama tanpa menghasilkan pembagian cluster yang terlalu banyak. Jika menggunakan K=6, segmentasi menjadi lebih rinci tetapi lebih kompleks dan sulit diinterpretasikan.

   c) Relevansi terhadap Tujuan Penelitian

      Tujuan penelitian adalah menganalisis segmentasi ruas jalan berdasarkan karakteristik kecepatan lalu lintas. Dengan K=3, hasil clustering dapat menggambarkan perbedaan kondisi ruas jalan secara lebih sederhana dan jelas melalui kategori Lancar, Sedang, dan Macet.

BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan studi segmentasi ruas jalan di DKI Jakarta berdasarkan karakteristik kecepatan lalu lintas dengan menggunakan algoritma K-Means, serta evaluasi yang dilakukan dengan metode Elbow dan Silhouette Score, dapat ditarik beberapa kesimpulan.

1. Segmentasi ruas jalan di DKI Jakarta dilakukan dengan menggunakan algoritma *KMeans* dengan karakteristik kecepatan lalu lintas sebagai variabel, yaitu kecepatan rata-rata, simpangan baku, koefisien variasi (CV), dan kecepatan maksimum pada setiap ruas jalan. Jumlah klaster yang optimal ditentukan dengan membandingkan metode Elbow dan *Silhouette Score*. Hasil evaluasi menunjukkan bahwa K=3 adalah jumlah klaster yang paling optimal, dengan penurunan nilai inersia yang signifikan sebesar 61,4% dari k=1 ke k=3 (206,0 menjadi 79,609) dan Silhouette Score sebesar 0,415. Meskipun perbedaan Skor *Silhouette* antara k=3 (0,415) dan k=2 (0,414) sangat kecil, titik siku yang terlihat jelas pada k=3 dan penurunan Skor *Silhouette* pada k=4 (0,389) memperkuat bahwa k=3 memberikan struktur pengelompokan yang paling stabil dan tidak mengalami *overfitting* pada jumlah *cluster.*
2. Hasil klasterisasi menghasilkan tiga kategori ruas jalan yang mencerminkan kondisi lalu lintas berbeda di DKI Jakarta, yaitu *Cluster* Lancar (7 ruas, 14%) dengan kecepatan rata-rata 33,43 km/jam namun fluktuasi tinggi (*CV* = 0,234), *Cluster* Sedang (25 ruas, 50%) dengan kecepatan rata-rata 28,86 km/jam dan stabilitas tertinggi (*CV* = 0,087), serta *Cluster* Macet (18 ruas, 36%) dengan kecepatan rata-rata terendah yaitu 23,91 km/jam dan pola kemacetan yang bersifat sistematis (*CV* = 0,177). Selisih kecepatan rata-rata antara "*Cluster* Lancar dan *Cluster* Macet sebesar 9,52 km/jam (39,8%)" menunjukkan bahwa segmentasi yang dihasilkan memiliki diferensiasi yang substantif dan bermakna secara operasional, bukan sekadar perbedaan statistik semata
3. Analisis pola distribusi klaster menunjukkan bahwa kemacetan lalu lintas terkonsentrasi di wilayah Jakarta Pusat dan Jakarta Selatan yang merupakan pusat aktivitas ekonomi dan perkantoran, sementara Cluster Sedang tersebar merata di seluruh wilayah DKI Jakarta. Pola ini menunjukkan bahwa kondisi arus lalu lintas menengah dengan variasi

terbatas merupakan kondisi umum yang dijumpai di berbagai lokasi, sedangkan kemacetan lebih bersifat spesifik lokasi dan berkaitan erat dengan fungsi kawasan. Selain itu, analisis nilai kecepatan maksimum pada setiap klaster mengungkapkan adanya batas struktural pada ruas-ruas Cluster Macet, di mana kecepatan maksimum yang dapat dicapai hanya sebesar 31,82 km/jam meskipun dalam kondisi lalu lintas optimal. Hal ini mengindikasikan bahwa permasalahan kemacetan pada ruas-ruas tersebut bersifat struktural, bukan semata-mata akibat fluktuasi volume kendaraan, sehingga memerlukan intervensi fisik seperti perbaikan geometrik atau perluasan kapasitas jalan, dan tidak cukup diselesaikan hanya melalui manajemen operasional lalu lintas.

4. Pemilihan K=3 tidak hanya didasarkan pada pertimbangan kuantitatif berupa nilai inertia dan Silhouette Score, tetapi juga pertimbangan praktis implementasi. Kategori Lancar–Sedang–Macet sejalan dengan standar yang lazim digunakan dalam bidang transportasi, sehingga memudahkan interpretasi dan komunikasi hasil penelitian kepada para pemangku kepentingan. Selain itu, tiga kategori tersebut memungkinkan perumusan strategi intervensi yang berbeda untuk setiap kondisi ruas jalan, sehingga hasil segmentasi ini bersifat actionable dan relevan dengan tujuan penelitian, yaitu menganalisis segmentasi ruas jalan berdasarkan karakteristik kecepatan lalu lintas di DKI Jakarta.

5.2 Saran

Adapun saran untuk penelitian selanjutnya adalah berikut.

1. Penelitian selanjutnya disarankan untuk membandingkan metode *Elbow* dan *Silhouette Score* dengan metode evaluasi *cluster* lain, seperti *Davies-Bouldin Index* atau Gap *Statistic*, mengingat selisih *Silhouette Score* antara k=2 dan k=3 pada penelitian ini tergolong sangat tipis (0,414 berbanding 0,415). Perbandingan antar metode evaluasi diharapkan dapat memperkuat validitas penentuan jumlah klaster optimal secara lebih komprehensif.
2. Pada penelitian lanjutan diharapkan dapat mengeksplorasi variabel tambahan di luar karakteristik kecepatan, seperti volume kendaraan, tingkat kepadatan *(density),* lebar ruas jalan, atau fungsi kawasan, agar segmentasi ruas jalan tidak hanya mencerminkan pola kecepatan, tetapi juga faktor-faktor penyebab kemacetan secara lebih menyeluruh.
3. Pada penelitian selanjutnya diharapkan dapat mengeksplorasi algoritma pembelajaran mesin lain, seperti DBSCAN atau Hierarchical

Clustering, untuk mengevaluasi apakah metode tersebut mampu menangkap pola spasial dan temporal data kecepatan lalu lintas dengan lebih baik, khususnya pada ruas jalan dengan karakteristik distribusi yang tidak sepenuhnya berbentuk bulat (non-spherical), yang menjadi salah satu keterbatasan algoritma KMeans.

4. Pengambilan data kecepatan lalu lintas pada penelitian ini disarankan untuk diperluas cakupan periode waktunya pada penelitian selanjutnya, misalnya dengan membedakan pola pada jam sibuk *(peak hour)* dan jam normal *(off-peak),* maupun perbedaan hari kerja dan akhir pekan. Hal ini penting karena karakteristik lalu lintas di perkotaan cenderung bersifat dinamis, sehingga segmentasi yang mempertimbangkan dimensi waktu diharapkan dapat menghasilkan gambaran kondisi lalu lintas yang lebih representatif dan aplikatif dalam pengambilan kebijakan.

5. Ruas jalan pada *Cluster* Macet, khususnya di wilayah Jakarta Pusat dan Jakarta Selatan, menunjukkan indikasi keterbatasan kapasitas struktural karena kecepatan maksimumnya tetap rendah bahkan pada kondisi optimal. Oleh karena itu, disarankan kepada instansi terkait, seperti Dinas Perhubungan DKI Jakarta, untuk menindaklanjuti hasil segmentasi ini dengan kajian teknis lebih lanjut guna merumuskan intervensi yang tepat sasaran, baik berupa perbaikan geometrik jalan, penambahan kapasitas, maupun rekayasa lalu lintas, alih-alih hanya mengandalkan manajemen operasional semata.

DAFTAR PUSTAKA

Abdurrasyid, Agtriadi, H. B., Wijayanti, R. R., Ragillia, A., & Rifai, M. F. (2025). Data-Driven Classification and Spatial Clustering of Stunting Cases in Rural Indonesia. *2025 International Conference on Advanced Technologies in Energy and Informatic (ICATEI)*, 349–354. https://doi.org/10.1109/ICATEI67676.2025.11404831

Akbar, S. C. D., Defit, S., & Hendrik, B. (2025). Segmentasi Tunggakan Pelanggan Menggunakan Algoritma K-Means Cluster pada Perusahaan Air Minum Daerah. *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, *5*(2), 349–355. https://doi.org/10.55382/jurnalpustakaai.v5i2.1215

Amal, Andi Alfiah, Fitriani Yessy, & Cahyaningtyas Rizqia. (n.d.). *PENGGUNAAN WIFI ID INSTITUT TEKNOLOGI PLN MENGGUNAKAN METODE KLASTERISASI K-MEANS*.

Cui, L. (2019). Complex industrial automation data stream mining algorithm based on random Internet of robotic things. *Automatika*, *60*(5), 570–579. https://doi.org/10.1080/00051144.2019.1683287

Etemad, S., Mosayebi, R., Khodavirdian, T. A., Dastan, E., Telmadarreh, A. S., Jafari, M., & Rafiei, S. (2023). *Clustering of Urban Traffic Patterns by K-Means and Dynamic Time Warping: Case Study*. http://arxiv.org/abs/2309.09830

Faktor, A., Mempengaruhi, Y., Angkutan, P., Beralih, U., Transportasi, K., Go-Jek, O., Sahputra, M. I., Triayudi, A., & Rubhasy, A. (2022). Metode K-Means Clustering. *Jurnal Teknologi Informasi Dan Komunikasi)*, *6*(1), 2022. https://doi.org/10.35870/jti

filki.widyatami. (2025). *Analisis Kemacetan Lalu Lintas di Ruas Jalan Kota*.

Gupta, S., Kumar, A., & Kumar, A. (2025). Analysis of Traffic Flow Congestion by Integrating Supervised Machine Learning with K-mean Clustering. *SN Computer Science*, *6*(3), 255. https://doi.org/10.1007/s42979-025-03761-4

Hafidz Ardana, C., Aldita, A., Aisyah, A., Khoyum, A., Faisal, M., Informatika, T., Sains, F., Teknologi, D., Maulana, U., & Ibrahim, M. (2024). Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-means Clustering. In *Jurnal Informatika Sunan Kalijaga)* (Vol. 9, Number 1).

Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM), 3(1), 46–56. https://doi.org/10.58602/jima-ilkom.v3i1.26

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, *622*, 178–210. https://doi.org/https://doi.org/10.1016/j.ins.2022.11.139

Istiqomah, N., Dianita Utami, W., & Yuliati, D. (2025). Penerapan K-Means Clustering untuk Analisis Kondisi Lalu Lintas di Jalan Ir. H. Soekarno Surabaya. *Jurnal Keselamatan Transportasi Jalan (Indonesian Journal of Road Safety)*, *12*(2), 78–88. https://doi.org/10.46447/ktj.v12i2.725

Kai Jiang, Yujie Long, Zejian Li, Zhilan Zhou, & Junwei Liang. (2025). Time Period Analysis of Traffic Flow around Scenic Spots in Small Towns Based on K-Means Clustering. *Academic Journal of Computing & Information Science*, *8*(5). https://doi.org/10.25236/ajcis.2025.080509

Kilang Permatasari, -Cahyaning, & Kristiyanto, H. (2025). *DAMPAK TRANSPORTASI SISTEM LIGH RAIL TRANSITS TERHADAP KEMACETAN LALU LINTAS DAN EMISI CARBON DI JAKARTA* (Vol. 7, Number 1). https://jurnal.ucy.ac.id/index.php/CivETech/issue/archive

Li, L. (2024a). Smart City Traffic Data Analysis and Prediction Based on Weighted K-means Clustering Algorithm. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 15, Number 6). www.ijacsa.thesai.org

Li, L. (2024b). Smart City Traffic Data Analysis and Prediction Based on Weighted K-means Clustering Algorithm. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 15, Number 6). www.ijacsa.thesai.org

M Khatami, & HB Agtriadi. (n.d.). *PENGELOMPOKAN PENILAIAN KEY PERFORMANCE INDICATOR INSTITUT TEKNOLOGI PLN MENGGUNAKAN ALGORITMA K-MEANS*.

Hafidz Ardana, C., Aldita, A., Aisyah, A., Khoyum, A., Faisal, M., Informatika, T., Sains, F., Teknologi, D., Maulana, U., & Ibrahim, M. (2024). Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-means Clustering. In Jurnal Informatika Sunan Kalijaga) (Vol. 9, Number 1).

Nugraha, Fadlan Triana, Max Teja Ajie Cipta Widiyanto, & Satrio Yudho. (2024). *Analisis Clustering Metode K-Means Berdasarkan Karakteristik Pemakaian Listrik PLN Untuk Pemasangan Wifi Iconet*.

Oce Ovalina Yumame, Sapto Pramono, & Ika Devy Pramudiana. (2025). Analisis Dampak Transportasi Rel terhadap Penurunan Tingkat Kemacetan di Jakarta. *JURNAL PENDIDIKAN DAN ILMU SOSIAL (JUPENDIS)*, *3*(1), 265–283. https://doi.org/10.54066/jupendis.v3i1.2903

Safitri, Nailul Muna, Dine Tiara Kusuma, & Dwina Kuswardani. (2024). *PENDEKATAN METODE K-MEANS UNTUK CLUSTERISASI WAKTU PENGGUNAAN ENERGI LISTRIK PADA PELANGGAN RUMAH TANGGA*.

Satria Ardi Perdana, Sara Famayla Florentin, & Agus Santoso. (2022). ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN K-MEANS CLUSTERING STUDI KASUS APLIKASI ALFAGIFT. *Sebatik*, *26*(2), 420–427. https://doi.org/10.46984/sebatik.v26i2.2134

Septia Dini Arifta, Tanaya Anindita Irawan, Naufal Ricko Maulana Pasha, Trimono, & Muhammad Nasrudin. (n.d.). *ANALISIS SEGMENTASI PENUMPANG TRANSJAKARTA BERDASARKAN POLA WAKTU, RUTE, DAN JUMLAH PEMBAYARAN MENGGUNAKAN PCA DAN K- MEANS CLUSTERING*.

Shimaoka, A. M., Ferreira, R. C., & Goldman, A. (2024). The evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks. *SBC Reviews on Computer Science*, *4*(1), 28–43. https://doi.org/10.5753/reviews.2024.3757

Sohail, A. M., Khattak, K. S., & Khan, Z. H. (2024). Data-driven insights: Unravelling traffic dynamics with k-means clustering and vehicle type differentiation. *Information System and Smart City*, *4*(1), 1737. https://doi.org/10.59400/issc1737

Voutama, A. (2022). Sistem Antrian Cucian Mobil Berbasis Website Menggunakan Konsep CRM dan Penerapan UML. *Komputika : Jurnal Sistem Komputer*, *11*, 102–111. https://doi.org/10.34010/komputika.v11i1.4677

DAFTAR RIWAYAT HIDUP



Data Pribadi

Nama mahasiswa : Dea Enjelika

Program studi : S1 Teknik Informatika

NIM : 202231048

Tempat, tanggal lahir : Bengkulu, 27 Juli 2004

Jenis kelamin : Perempuan

Agama : Islam

No HP : 081368193107

Email : Dea2231048@itpln.ac.id

Alamat rumah : JL TELAGA DEWA ASRI NO. 09

Pendidikan

SD : SD Negeri 59 Bengkulu, tahun 2016

SMP : SMP Negeri 14 Bengkulu, tahun lulus 2019

SMA : SMA Negeri 2 Bengkulu, tahun lulus 2022

Demikianlah daftar riwayat hidup ini dibuat dengan sebenarnya.

LAMPIRAN

LEMBAR PENGESAHAN PEMBIMBING

Lampiran 1 Lembar Bimbingan Skripsi

INSTITUT TEKNOLOGI-PLN

LEMBAR BIMBINGAN SKRIPSI

Nama Mahasiswa : Dea Enjelika

NIM : 202231048

Program Studi : Teknik Informatika

Nama Dosen Pembimbing : Pritasari Palupiningsih S.Kom., M.Kom

Judul Skripsi : Analisis segmentasi ruas jalan berdasarkan karakteristik kecepatan lalu lintas kecepatan lalu lintas di DKI Jakarta menggunakan K-Means Clustering.

No	Tanggal	Materi Bimbingan	Tanda Tangan
1	20 Februari 2026	Konsultasi awal dan pengajuan topik penelitian	
2	4 Maret 2026	Persetujuan Elbow Method, Silhouette Score	
3	5 Maret 2026	Penyusunan dan pembahasan BAB I	
4	1 April 2026	Penyusunan BAB II dan BAB III	
5	9 April 2026	Seminar proposal dan saran perbaikan	
6	13 Mei 2026	Pembahasan data penelitian dan kesiapan dataset	
7	30 Mei 2026	Review hasil preprocessing dan validasi dataset	
8	20 June 2026	Pembahasan kesiapan data	
9	27 June 2026	Evaluasi hasil dan analisis performa model	
10	14 Juli 2026	Revisi BAB I–V	
11	21 Juli 2026	Revisi lanjutan: perbaikan abstrak, kerangka penelitian, dan dasar perhitungan manual	
12	23 Juli 2026	Finalisasi skripsi dan persetujuan pengumpulan	

Lampiran 2 Pemohonan Mengikuti Sidang Skripsi

INSTITUT TEKNOLOGI PLN

PERMOHONAN MENGIKUTI SIDANG TUGAS AKHIR SARJANA[*)]

Nama mahasiswa : Dea Enjelika
Program studi : Teknik Informatika
Fakultas/sekolah : Telematika Energi
No HP/Email : 081368193107/ dea2231048@itpln.ac.id

Mengajukan permohonan mengikuti sidang tugas akhir pada Semester 8 Tahun Akademik 2025/2026, dimana judul tugas akhir dan dosen pembimbingnya sebagai berikut:

Judul tugas akhir : Analisis Segmentasi ruas jalan berdasarkan karakteristik Kecepatan lalu lintas di DKI Jakarta Menggunakan K-Means Clustering
Nama Pembimbing : Pritasari Palupiningsih S.Kom, M.Kom

Sebagai bahan pertimbangan disampaikan data-data pendukung sebagai berikut:

No	Persyaratan	Memenuhi syarat **)	
1	Telah menyelesaikan naskah tugas akhir (foto copy lembar bimbingan, dan surat persetujuan Dosen pembimbing)	Ya	Tidak
2	Persetujuan Dosen pembimbing pertama	Ya	Tidak
3	Memiliki surat bebas plagiasi	Ya	Tidak
4	Memiliki Sertifikat Toefl	Ya	Tidak
5	Memiliki Sertfikat soft skill/hard skill (minimal 5)	Ya	Tidak
6	Memiliki Sertifikat seminar (minimal 5)	Ya	Tidak
7	Memiliki Sertifikat MOS	Ya	Tidak
8	Memiliki Sertifikat Uji Kompetensi	Ya	Tidak
9	Lulus mata kuliah dengan nilai minimum C	Ya	Tidak
10	Tidak mempunyai tunggakan keuangan	Ya	Tidak
11	Tidak mempunyai pinjaman buku perpustakaan	Ya	Tidak
12	Tidak mempunyai pinjaman alat-alat laboratorium	Ya	Tidak

*) Pilih salah satu
**) Diisi oleh Kepala Program Studi,

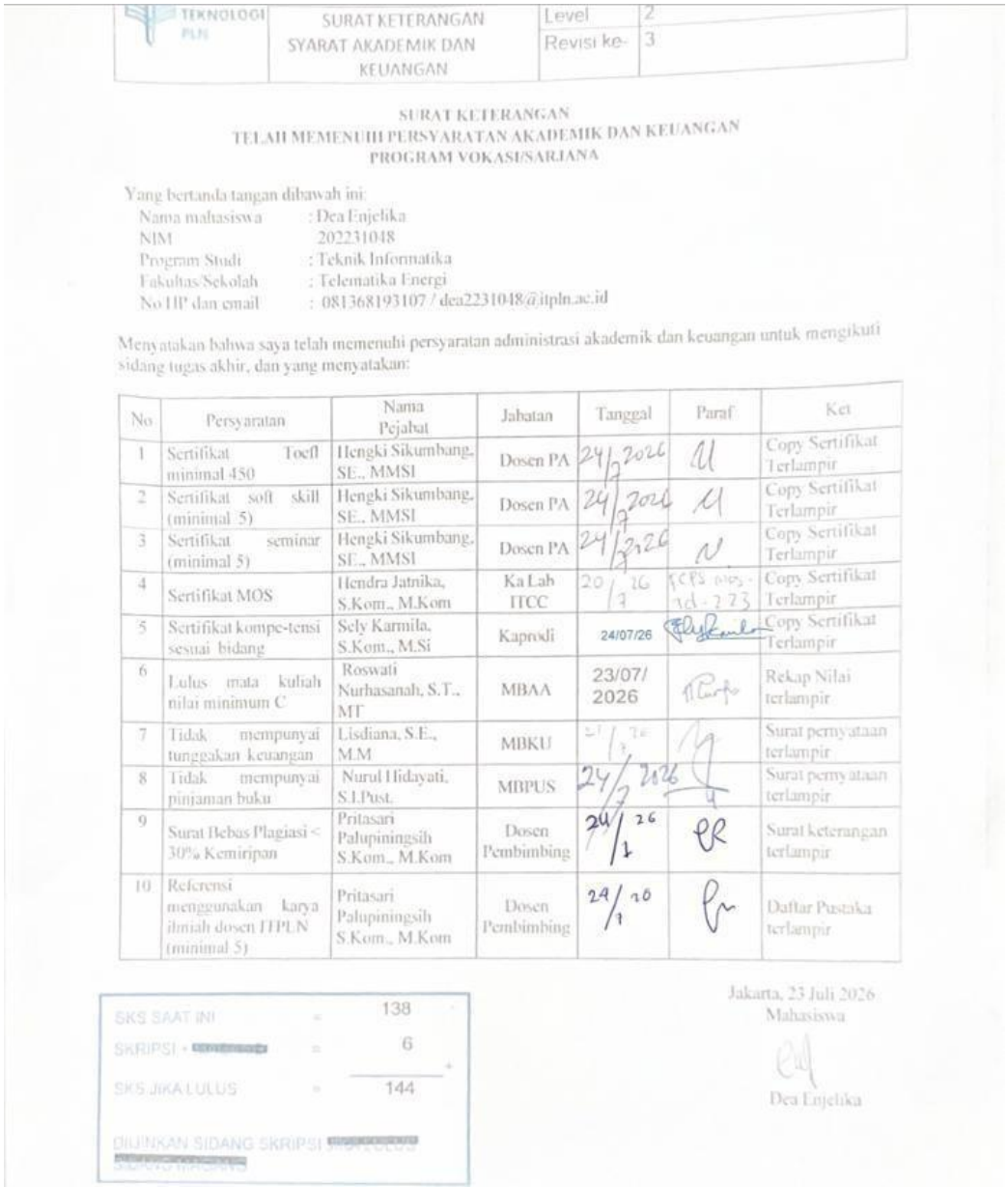
Disetujui/ ~~Tidak disetujui~~ mengikuti Sidang

Jakarta, 21 Juli 2026

Diperiksa oleh:

Kepala Program Studi	Dosen Pembimbing Akademik	Mahasiswa
Sely Karmila, S.Kom., M.Si	Hengki Sihombing, SE., MMSI	Dea Enjelika

Lampiran 3 Surat Keterangan Persyaratan Akademik dan Keuangan



TEKNOLOGI PLN	SURAT KETERANGAN SYARAT AKADEMIK DAN KEUANGAN	Level	2
		Revisi ke-	3

SURAT KETERANGAN
TELAH MEMENUHI PERSYARATAN AKADEMIK DAN KEUANGAN
PROGRAM VOKASI/SARJANA

Yang bertanda tangan dibawah ini:

Nama mahasiswa	: Dea Enjelika
NIM	202231048
Program Studi	: Teknik Informatika
Fakultas/Sekolah	: Telematika Energi
No HP dan email	: 081368193107 / dea2231048@itpln.ac.id

Menyatakan bahwa saya telah memenuhi persyaratan administrasi akademik dan keuangan untuk mengikuti sidang tugas akhir, dan yang menyatakan:

No	Persyaratan	Nama Pejabat	Jabatan	Tanggal	Paraf	Ket
1	Sertifikat Toefl minimal 450	Hengki Sikumbang, SE., MMSI	Dosen PA	24/7/2026		Copy Sertifikat Terlampir
2	Sertifikat soft skill (minimal 5)	Hengki Sikumbang, SE., MMSI	Dosen PA	24/7/2026		Copy Sertifikat Terlampir
3	Sertifikat seminar (minimal 5)	Hengki Sikumbang, SE., MMSI	Dosen PA	24/7/26		Copy Sertifikat Terlampir
4	Sertifikat MOS	Hendra Jatnika, S.Kom., M.Kom	Ka Lab ITCC	20/7/26		Copy Sertifikat Terlampir
5	Sertifikat kompe-tensi sesuai bidang	Sely Karmila, S.Kom., M.Si	Kaprodi	24/07/26		Copy Sertifikat Terlampir
6	Lulus mata kuliah nilai minimum C	Roswati Nurhasanah, S.T., MT	MBAA	23/07/2026		Rekap Nilai terlampir
7	Tidak mempunyai tunggakan keuangan	Lisdiana, S.E., M.M	MBKU	21/7/26		Surat pernyataan terlampir
8	Tidak mempunyai pinjaman buku	Nurul Hidayati, S.I.Pust.	MBPUS	24/7/2026		Surat pernyataan terlampir
9	Surat Bebas Plagiasi < 30% Kemiripan	Pritasari Palupiningsih S.Kom., M.Kom	Dosen Pembimbing	24/1/26		Surat keterangan terlampir
10	Referensi menggunakan karya ilmiah dosen ITPLN (minimal 5)	Pritasari Palupiningsih S.Kom., M.Kom	Dosen Pembimbing	29/1/20		Daftar Pustaka terlampir

SKS SAAT INI	=	138
SKRIPSI + [redacted]	=	6
SKS JIKA LULUS	=	144

DIIJINKAN SIDANG SKRIPSI [redacted]

Jakarta, 23 Juli 2026
Mahasiswa

Dea Enjelika

Lampiran 4 Surat Keterangan Bebas Administrasi

SURAT KETERANGAN
BEBAS ADMINISTRASI

Yang bertanda tangan dibawah ini menerangkan :

Nama	:	Dea Enjelika
NIM	:	202231048
Jenjang / Program Studi	:	S1 / Teknik Informatika

Telah menyelesaikan Administrasi Keuangan dan Administrasi Perpustakaan IT-PLN.

Surat keterangan ini dikeluarkan sebagai syarat untuk mengikuti **sidang skripsi.**

Jakarta, 21 Juli 2026

Yang Menerangkan,

M – BKU

Lisdiana, S.E., M.M

M - Perpustakaan

Nurul Hidayati, S.I.Pust.

Lampiran 5 Surat Keterangan Bebas Pinjam Alat Laboratorium

TEKNOLOGI PLN	ALAT-ALAT LABORATORIUM	Revisi ke-	3

SURAT KETERANGAN BEBAS
PINJAM ALAT LABORATORIUM

Yang bertanda tangan dibawah ini:

Nama mahasiswa	: Dea Enjelika
NIM	: 202231048
Program studi	: Teknik Informatika
Fakultas/Sekolah	: Telematika Energi/Institut Teknologi PLN Jakarta
No HP dan email	: 081368193107 dan Dea2231048@itpln.ac.id

Menyatakan bahwa saya tidak mempunyai pinjaman alat-alat laboratorium di lingkungan Institut Teknologi PLN, dan yang menyatakan:

No	Persyaratan	Nama Pejabat	Jabatan	Tanggal	Tanda tangan
1	Lab. Software Engineering	Agus Mulyanto, MT., M.SC	Kepala Laboratorium	23/7/26	
2	Lab. Intelligent Computing	Rosida Nur Aziza, ST., M.Eng.Sc	Kepala Laboratorium	23/7/26	
3	Lab. Embedded System	Karina Djunaidi, ST., M.TI	Kepala Laboratorium	23/7	
4	Lab. Computer Network	Karina Djunaidi, ST., M.TI	Kepala Laboratorium		
5	Lab. Information Retrieval	Dine Tiara Kusuma., ST., M.Kom	Kepala Laboratorium	23/7	
6	Lab. Multimedia	Yasni Djamain S.Kom., M.Kom	Kepala Laboratorium	24/07/26	
7	Lab. Software Architecture & Quality	Abdurrasyid, S.Kom., MMSI	Kepala Laboratorium	24/7/26	
8	Lab. Virtual Reality	Dr. Darma Rusjdi., ST., M.Kom	Kepala Laboratorium	23/7-26	

Jakarta, 10 Juli 2026
Mahasiswa

Dea Enjelika

baeabd1d55464069_Turnitin

ORIGINALITY REPORT

15 %

SIMILARITY INDEX

PRIMARY SOURCES

	ejurnal.ubharajaya.ac.id Internet	Arrahman, Mhd. Furqan. "SEGMENTASI KEAKTIFAN
2	repository.itpln.ac.id Internet	
3	journal.smartpublisher.id Internet	
4	dspace.uii.ac.id Internet	
5	ejurnal.methodist.ac.id Internet	
6	digilib.ulm.ac.id Internet	
7	Isra Yanuar Giu, Juang Akbardin, Erna Krisnanto. "ANALISIS HUBUNGAN FUNDAMENTAL LALU LINTAS BERBASIS SMP DENGAN MODEL GREENSHIELDS", JURNAL SIMETRIK, 2026 Crossref	
8	eprints.amikom.ac.id Internet	

Lampiran 6 Ringkasan Perbaikan Naskah Tugas Akhir

INSTITUT TEKNOLOGI PLN	**INSTITUT TEKNOLOGI PLN**	Kode	FO-PD-A03-2-003-03-011
		Tanggal	10 Februari 2025
	PERBAIKAN NASKAH TUGAS AKHIR	Level	2
		Revisi ke-	3

RINGKASAN PERBAIKAN NASKAH TUGAS AKHIR **Ruang : 106**

Tanggal ujian sidang	:	05 Agustus 2026
Nama mahasiswa	:	DEA ENJELIKA
NIM	:	202231048
Program studi	:	S1 Teknik Informatika
Fakultas/Sekolah	:	Telematika Energi
Judul tugas akhir	:	ANALISIS SEGMENTASI RUAS JALAN BERDASARKAN KARAKTERISTIK KECEPATAN LALU LINTAS DI DKI JAKARTA MENGGUNAKAN K-MEANS CLUSTERING

Saran perbaikan naskah tugas akhir adalah sebagai berikut:

Kerjakan sesuai saran dan masukkan penguji

Jangka waktu untuk perbaikan naskah tugas akhir diatas harus diselesikan paling lambat satu minggu setelah tanggal ujian yakni pada tanggal 12 Agustus

Apabila dalam jangka waktu tersebut mahasiswa tersebut diatas tidak dapat menyelesaikan **REVISI** tugas akhirnya, maka mahsiswa harus kembali mengulang sidang.

Mahassiswa	**Dosen pembimbing**	**Ketua Penguji**
(DEA ENJELIKA)	(Pritasari Palupiningsih, S.Kom, M.Kom)	(Dr. Puji Catur Siswipraptini, ST., MTI)

Tugas akhir telah selesai diperbaiki pada tanggal .....

Mahassiswa	**Dosen pembimbing**	**Ketua Penguji**
(DEA ENJELIKA)	(Pritasari Palupiningsih, S.Kom, M.Kom)	(Dr. Puji Catur Siswipraptini, ST., MTI)

Lampiran 7 Perbaikan Naskah Tugas Akhir Ketua Penguji

INSTITUT TEKNOLOGI PLN	INSTITUT TEKNOLOGI PLN	Kode	FO-PD-A03-2-003-03-012
		Tanggal	10 Februari 2025
	PERBAIKAN NASKAH TUGAS AKHIR	Level	2
		Revisi ke-	3

PERBAIKAN NASKAH TUGAS AKHIR **Ruang : 106**

Tanggal ujian sidang : 05 Agustus 2026

Nama mahasiswa : DEA ENJELIKA

NIM : 202231048

Program studi : S1 Teknik Informatika

Fakultas/Sekolah : Telematika Energi

Judul tugas akhir : ANALISIS SEGMENTASI RUAS JALAN BERDASARKAN KARAKTERISTIK KECEPATAN LALU LINTAS DI DKI JAKARTA MENGGUNAKAN K-MEANS CLUSTERING

Saran perbaikan naskah tugas akhir adalah sebagai berikut:

1. Perbaiki format penulisan: numbering dll.

Jangka waktu untuk perbaikan naskah tugas akhir diatas harus diselesikan paling lambat satu minggu setelah tanggal ujian yakni pada tanggal .....

Apabila dalam jangka waktu tersebut mahasiswa tersebut diatas tidak dapat menyelesaikan **REVISI** tugas akhirnya, maka mahsiswa harus kembali mengulang sidang.

Mahasiswa **Ketua Penguji**

(DEA ENJELIKA) (Dr. Puji Catur Siswipraptini, ST., MTI)

Tugas akhir telah selesai diperbaiki pada tanggal .....

Mahasiswa **Ketua Penguji**

(DEA ENJELIKA) (Dr. Puji Catur Siswipraptini, ST., MTI)

Lampiran 8 Perbaikan Naskah Tugas Akhir Anggota Penguji 1

INSTITUT TEKNOLOGI PLN	INSTITUT TEKNOLOGI PLN	Kode	FO-PD-A03-2-003-03-012
		Tanggal	10 Februari 2025
	PERBAIKAN NASKAH TUGAS AKHIR	Level	2
		Revisi ke-	3

PERBAIKAN NASKAH TUGAS AKHIR **Ruang : 106**

Tanggal ujian sidang	:	05 Agustus 2026
Nama mahasiswa	:	DEA ENJELIKA
NIM	:	202231048
Program studi	:	S1 Teknik Informatika
Fakultas/Sekolah	:	Telematika Energi
Judul tugas akhir	:	ANALISIS SEGMENTASI RUAS JALAN BERDASARKAN KARAKTERISTIK KECEPATAN LALU LINTAS DI DKI JAKARTA MENGGUNAKAN K-MEANS CLUSTERING

Saran perbaikan naskah tugas akhir adalah sebagai berikut:

1. Format Penulisan (Font, Spasi, margin, Paragraf, Penomoran OK
2. Pertajam masalah di latar belakang OK
3. Gambar 3.1 Perbaiki dan tambah penjelasan
4. Data collection OK
5. Perbaiki tabel 3.1, 3.2 OK
6. Tampilkan data sebelum dan sesudah dilakukan proses Feature Engineering

Jangka waktu untuk perbaikan naskah tugas akhir diatas harus diselesikan paling lambat satu minggu setelah tanggal ujian yakni pada tanggal 12 Agustus

Apabila dalam jangka waktu tersebut mahasiswa tersebut diatas tidak dapat menyelesaikan **REVISI** tugas akhirnya, maka mahsiswa harus kembali mengulang sidang.

Mahasiswa	**Anggota Penguji 1**
(DEA ENJELIKA)	(Karina Djunaidi, ST., M.TI)

Tugas akhir telah selesai diperbaiki pada tanggal 19 Agustus

Mahasiswa	**Anggota Penguji 1**
(DEA ENJELIKA)	(Karina Djunaidi, ST., M.TI)

Lampiran 9 Perbaikan Naskah Tugas Akhir Anggota Penguji 2

INSTITUT TEKNOLOGI PLN	**INSTITUT TEKNOLOGI PLN**	Kode	FO-PD-A03-2-003-03-012
		Tanggal	10 Februari 2025
	PERBAIKAN NASKAH TUGAS AKHIR	Level	2
		Revisi ke-	3

PERBAIKAN NASKAH TUGAS AKHIR **Ruang : 106**

Tanggal ujian sidang	:	05 Agustus 2026
Nama mahasiswa	:	DEA ENJELIKA
NIM	:	202231048
Program studi	:	S1 Teknik Informatika
Fakultas/Sekolah	:	Telematika Energi
Judul tugas akhir	:	ANALISIS SEGMENTASI RUAS JALAN BERDASARKAN KARAKTERISTIK KECEPATAN LALU LINTAS DI DKI JAKARTA MENGGUNAKAN K-MEANS CLUSTERING

Saran perbaikan naskah tugas akhir adalah sebagai berikut:

1. Hilangkan teori UML di bab 2
2. Tambahkan narasi terkait jumlah data yang digunakan pada bab 3.
3. Ikuti arahan tim penguji.

Jangka waktu untuk perbaikan naskah tugas akhir diatas harus diselesikan paling lambat satu minggu setelah tanggal ujian yakni pada tanggal .....

Apabila dalam jangka waktu tersebut mahasiswa tersebut diatas tidak dapat menyelesaikan **REVISI** tugas akhirnya, maka mahsiswa harus kembali mengulang sidang.

Mahasiswa **Anggota Penguji 2/ Pemb.**

(DEA ENJELIKA) (Pritasari Palupiningsih, S.Kom, M.Kom)

Tugas akhir telah selesai diperbaiki pada tanggal .....

Mahasiswa **Anggota Penguji 2/ Pemb.**

(DEA ENJELIKA) (Pritasari Palupiningsih, S.Kom, M.Kom)