



BERITA ACARA SIDANG PENDADARAN TUGAS AKHIR

S/UBL/FTI/1675/VIII/26

Pada hari ini, Senin 03 Agustus 2026 telah dilaksanakan Ujian Sidang Pendadaran Tugas Akhir sebagai berikut:

Judul: IMPLEMENTASI DATA MINING DENGAN ALGORITMA KMEANS UNTUK
SEGMENTASI POLA PENUMPANG TRANSPORTASI PUBLIK DI DKI JAKARTA

Nama : Rio Adriyanto
NIM : 2211501925
Dosen Pembimbing : Joko Christian Chandra, S.Kom., M.Kom.

Berdasarkan penilaian pada Presentasi + Demo, Penulisan, Penguasaan Materi, Penguasaan Program maka Mahasiswa tersebut di atas dinyatakan:

LULUS

dengan nilai angka : **66** huruf : **B-**

Mahasiswa tersebut di atas wajib menyerahkan hasil perbaikan tulisan Tugas Akhir dalam bentuk terjilid sesuai dengan Panduan Perbaikan Tugas Akhir, selambat-lambatnya Senin 17 Agustus 2026.

Panitia Penguji:

- 1 Ketua Sejati Waluyo, S.Kom., M.Kom.
- 2 Anggota Painem, S.Kom., M.Kom.
- 3 Moderator Joko Christian Chandra, S.Kom., M.Kom.

Keterangan:

Nilai Huruf: A:85-100 A-:80-84,99 B+:75-79,99 B:70-74,99 B-:65-69,99 C:60-64,99 D:40-59,99
E-:0-39,99



UNIVERSITAS BUDI LUHUR
FAKULTAS TEKNOLOGI INFORMASI

Kartu Bimbingan Tugas Akhir

NIM: 2211501925

Nama: Rio Adriyanto

Pembimbing: Joko Christian Chandra, S.Kom., M.Kom.

No.	Tanggal	Materi
1	07-04-2026	konsep dan tema TA
2	16-04-2026	Draft Bab 1
3	28-04-2026	perbaikan bab 1 dan draft bab 2
4	12-05-2026	perbaikan bab 2 dan draft bab 3
5	27-05-2026	Perbaikan bab 3 dan metodologi
6	09-06-2026	perbaikan diagram dan layout, draft bab 4
7	13-07-2026	perbaikan bab 4 dan draft bab 5
8	22-07-2026	perbaikan keseluruhan bab



LEMBAR PENGESAHAN

Nama : Rio Adriyanto
Nomor Induk Mahasiswa : 2211501925
Program Studi : Teknik Informatika
Bidang Peminatan : Cyber Security
Jenjang Studi : Strata 1
Judul : IMPLEMENTASI DATA MINING DENGAN ALGORITMA KMEANS
UNTUK SEGMENTASI POLA PENUMPANG TRANSPORTASI PUBLIK
DI DKI JAKARTA



Laporan Tugas Akhir ini telah disetujui, disahkan dan direkam secara elektronik sehingga tidak memerlukan tanda tangan tim penguji.

Jakarta, Senin 03 Agustus 2026

Tim Penguji:

Ketua : Sejati Waluyo, S.Kom., M.Kom.
Anggota : Painem, S.Kom., M.Kom.
Pembimbing : Joko Christian Chandra, S.Kom., M.Kom.
Ketua Program Studi : Dr. Indra, S.Kom., M.T.I.

**IMPLEMENTASI DATA MINING DENGAN ALGORITMA K-
MEANS UNTUK SEGMENTASI POLA PENUMPANG
TRANSPORTASI PUBLIK DI DKI JAKARTA**

TUGAS AKHIR



Oleh:
RIO ADRIYANTO
NIM: 2211501925

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR**

**JAKARTA
2026**

ABSTRAK

IMPLEMENTASI DATA MINING DENGAN ALGORITMA K-MEANS UNTUK SEGMENTASI POLA PENUMPANG TRANSPORTASI PUBLIK DI DKI JAKARTA

Oleh: Rio Adriyanto (2211501925)

Sistem transportasi publik di DKI Jakarta melayani jutaan penumpang setiap hari dengan tingkat kepadatan yang berbeda antar-hari, sehingga pengelola membutuhkan gambaran objektif mengenai kelompok hari berdasarkan kepadatan penumpang. Penelitian ini menerapkan algoritma K-Means Clustering untuk menyegmentasi 365 hari sepanjang satu tahun terakhir (Mei 2025–April 2026) berdasarkan jumlah penumpang harian, sekaligus menganalisis karakteristik tiap segmen yang terbentuk. Data diproses mengikuti tahapan Knowledge Discovery in Database, meliputi seleksi data, pra-pemrosesan, transformasi dengan StandardScaler, penentuan jumlah cluster menggunakan Elbow Method dan Silhouette Score, serta implementasi K-Means. Seluruh tahapan diwujudkan dalam aplikasi antarmuka grafis berbasis Streamlit dengan pola Input-Proses-Output yang menerima dataset melalui fitur unggah. Hasil pengujian menetapkan jumlah cluster $k=3$ dengan Silhouette Score 0,7784 yang tergolong kuat, membentuk tiga segmen kepadatan, yaitu Rendah (121 hari), Sedang (233 hari), dan Tinggi (11 hari). Mayoritas hari berada pada kepadatan sedang, sedangkan hari berkepadatan sangat tinggi relatif jarang, sehingga hasil segmentasi dapat menjadi dasar bagi pengelola dalam mengoptimalkan alokasi armada dan frekuensi perjalanan secara lebih tepat sasaran.

Kata kunci: Data Mining, K-Means Clustering, Segmentasi, Transportasi Publik, Silhouette Score.

DAFTAR ISI

ABSTRAK.....	iv
PERNYATAAN TIDAK PLAGIAT.....	v
SURAT PERNYATAAN PENGGUNAAN AI GENERATIF DALAM PENULISAN KARYA ILMIAH	vi
KATA PENGANTAR	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR ALGORITMA	xii
DAFTAR SIMBOL	xiii
DAFTAR ISI.....	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Perumusan Masalah	1
1.3. Batasan Masalah	2
1.4. Tujuan	2
1.5. Manfaat	2
1.6. Sistematika Penulisan	2
BAB II LANDASAN TEORI.....	4
2.1. Data Mining	4
2.2. Knowledge Discovery in Database (KDD)	4
2.3. Clustering.....	4
2.4. Jenis Metode Clustering.....	4
2.5. Algoritma K-Means	4
2.6. Centroid	7
2.7. Ukuran Jarak.....	7
2.7.1. Euclidean Distance	7
2.7.2. Manhattan Distance	7
2.7.3. Cosine Similarity	8
2.8. Normalisasi Data.....	8
2.8.1. Mean (Rata-rata).....	8
2.8.2. StandardScaler (Z-Score).....	8
2.8.3. MinMaxScaler	9

2.9. Penentuan Jumlah Cluster.....	9
2.9.1. SSE, WCSS, dan Inertia	9
2.9.2. Elbow Method.....	9
2.9.3. Silhouette Score	10
2.9.4. Kombinasi Elbow Method dan Silhouette Score.....	11
2.10. Perangkat Lunak Pendukung	12
2.11. Metode Penelitian Kuantitatif.....	12
2.12. Knowledge Discovery in Database sebagai Metodologi Penelitian	12
2.13. Pengujian Perangkat Lunak (Black Box Testing).....	13
2.14. Evaluasi Model Clustering.....	13
2.15. Penelitian Terdahulu	13
2.16. Analisis Research Gap	15
BAB III METODOLOGI PENELITIAN	16
3.1. Data Penelitian.....	16
3.2. Variabel Penelitian.....	16
3.3. Kerangka Berpikir.....	17
3.4. Pemilihan Metode	17
3.5. Metode Penelitian	18
3.5.1. Alur Metode Penelitian.....	18
3.5.2. Tahapan Data Mining dan Pemetaannya pada Aplikasi	18
3.5.3. Tahapan Pra-pemrosesan Data.....	19
3.6. Rancangan Pengujian.....	20
3.7. Arsitektur dan Struktur Program.....	21
3.8. Struktur Folder, Modul, dan Fungsi Program.....	24
3.9. Rancangan Menu	25
3.10. Rancangan Antarmuka (Wireframe).....	25
3.10.1. Rancangan Layar Beranda	25
3.10.2. Rancangan Layar Input Dataset.....	26
3.10.3. Rancangan Layar Seleksi Data	26
3.10.4. Rancangan Layar Pre-processing.....	27
3.10.5. Rancangan Layar Transformasi	27
3.10.6. Rancangan Layar Data Mining	28
3.10.7. Rancangan Layar Interpretasi	28
3.10.8. Rancangan Layar Evaluasi Model	29

BAB IV HASIL DAN PEMBAHASAN	30
4.1. Lingkungan Percobaan	30
4.1.1. Spesifikasi Perangkat Keras.....	30
4.1.2. Spesifikasi Perangkat Lunak.....	30
4.1.3. Diagram Deployment.....	30
4.2. Implementasi Metode	31
4.2.1. Seleksi Data (Import Dataset).....	31
4.2.2. Pra-pemrosesan Data	31
4.2.3. Transformasi (Normalisasi)	32
4.2.4. Penentuan Jumlah Cluster.....	33
4.2.5. Ilustrasi Perhitungan K-Means	33
4.2.6. Hasil Clustering	34
4.3. Flowchart	35
4.3.1. Flowchart Import Dataset	35
4.3.2. Flowchart Seleksi Data	36
4.3.3. Flowchart Pre-processing	36
4.3.4. Flowchart Transformasi.....	37
4.3.5. Flowchart Penentuan Jumlah Cluster	37
4.3.6. Flowchart Algoritma K-Means.....	38
4.3.7. Flowchart Interpretasi Hasil.....	38
4.3.8. Flowchart Evaluasi Model	39
4.3.9. Flowchart Pengujian Black Box	39
4.3.10. Flowchart Keseluruhan Sistem	40
4.4. Algoritma	40
4.5. Pengujian	41
4.5.1. Lingkungan Pengujian	41
4.5.2. Tahapan Pengujian.....	41
4.5.3. Skenario dan Hasil Pengujian Black Box	42
4.5.4. Pengujian Silhouette Score dan Elbow Method	42
4.5.5. Interpretasi dan Kesimpulan Pengujian	42
4.6. Tampilan Layar Aplikasi	42
4.6.1. Tampilan Layar Beranda	43
4.6.2. Tampilan Layar Input Dataset	44
4.6.3. Tampilan Layar Seleksi Data.....	45

4.6.4. Tampilan Layar Pre-processing	46
4.6.5. Tampilan Layar Transformasi	47
4.6.6. Tampilan Layar Data Mining	48
4.6.7. Tampilan Layar Elbow Method dan Silhouette Score.....	49
4.6.8. Tampilan Layar Hasil Clustering.....	50
4.6.9. Tampilan Layar Interpretasi.....	51
4.6.10. Tampilan Layar Evaluasi Model.....	52
4.6.11. Tampilan Layar Ilustrasi Algoritma	53
4.7. Analisis Hasil Segmentasi	53
4.7.1. Pola Hari dan Bulan.....	53
4.7.2. Perbandingan Weekday dan Weekend	54
4.7.3. Karakteristik dan Insight Tiap Cluster.....	54
4.7.4. Alasan Pemilihan k=3 dan Interpretasi Metrik	54
BAB V PENUTUP	55
5.1. Kesimpulan	55
5.2. Saran	55
DAFTAR PUSTAKA	56

DAFTAR PUSTAKA

- Afira, N., & Wijayanto, A. W. (2021). Analisis Cluster dengan Metode Partitioning dan Hierarki pada Data Informasi Kemiskinan Provinsi di Indonesia Tahun 2019. *Komputika : Jurnal Sistem Komputer*, 10(2), 101–109. <https://doi.org/10.34010/komputika.v10i2.4317>
- Chong, B. (2021). K-means clustering algorithm: a brief review. *Academic Journal of Computing & Information Science*, 4(5), 37–40. <https://doi.org/10.25236/ajcis.2021.040506>
- Daffauzan, M. F. (2023). Analisa Penumpang Bus Transjakarta dengan Metode K-Means. *Journal Computer Science and Information Systems : J-Cosys*, 3(2), 61–64. <https://doi.org/10.53514/jco.v3i2.395>
- Fahmi, M. N. (2023). Implementasi Mechine Learning menggunakan Python Library : Scikit-Learn (Supervised dan Unsupervised Learning). *Sains Data Jurnal Studi Matematika dan Teknologi*, 1(2), 87–96. <https://doi.org/10.52620/sainsdata.v1i2.31>
- Iqbal, M., Syaripuddin, S., & Huda, M. N. (2023). Implementasi Algoritma K-Means Clustering dengan Jarak Euclidean dalam Mengelompokkan Daerah Penyebaran COVID-19 di Kabupaten Bogor. *Basis : Jurnal Ilmiah Matematika*, 2(1), 47–56. <https://doi.org/10.30872/basis.v2i1.1019>
- MacQUEEN, J. (1967). SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 3439–3447.
- Maori, N. A., & Evanita, E. (2023). Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 14(2), 277–288. <https://doi.org/10.24176/simet.v14i2.9630>
- Pangestu, M. S., & Fitriani, M. A. (2022). Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means. *Sainteks*, 19(2), 141. <https://doi.org/10.30595/sainteks.v19i2.14495>
- Pemerintah Provinsi DKI Jakarta. (2024). *Jumlah Penumpang Angkutan Umum yang Terlayani per Hari*. Portal Satudata Jakarta. https://satudata.jakarta.go.id/open-data/detail?kategori=dataset&page_url=jumlah-penumpang-angkutan-umum-yang-terlayani-perhari&data_no=1
- Prisuna, B. F. (2021). Pengaruh Penggunaan Aplikasi Google Meet terhadap Hasil Belajar. *Jurnal Penelitian Ilmu Pendidikan*, 14(2), 137–147. <https://doi.org/10.21831/jpipfip.v14i2.39160>
- Ramadhan, A. R., & Diana, A. (2025). PENERAPAN ALGORITMA K-MEANS UNTUK KLASTERISASI JUMLAH PENUMPANG ANGKUTAN UMUM DI JAKARTA. *Infotek: Jurnal Informatika dan Teknologi*, 8(1), 228–238. <https://doi.org/10.29408/jit.v8i1.28195>
- Reynaldi, R., Djarot, R. J. F., Wahyudi, M., Sumanto, S., & Budiman, A. S. (2025). Analisa Pola Penyebaran Pengguna Layanan Transjakarta dengan Metode K-Means Clustering. *Journal Software, Hardware and Information Technology*,

- 5(2), 128–138. <https://doi.org/10.24252/shift.v5i2.205>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C), 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Sakinah, F. A., Aditiawan, F. P., & Lina Nurlaili, A. (2024). Pengujian Pada Aplikasi Manajemen Aset Menggunakan Black Box Testing. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 2814–2823. <https://doi.org/10.36040/jati.v8i3.9524>
- Saputra, A., & Yusuf, R. (2024). Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM: Comparison of the DBSCAN and K-MEANS Algorithms in Segmenting Customers Using Public Transportation of Transjakarta Usin. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1346–1361.
- Setyaningtyas, S., Nugroho, B. I., & Arif, Z. (2022). Tinjauan Pustaka Sistematis pada Data Mining: Studi Kasus Algoritma K-Means Clustering. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, 10(2), 52–61.
- Shutaywi, M., & Kachouie, N. N. (2021). Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy*, 23(6), 1–17. <https://doi.org/10.3390/e23060759>
- Silangen, R. T., & Matdoan, M. Y. (2022). Klasifikasi Hasil Seleksi Kompetensi Dasar CPNS Menggunakan Metode Decision Tree. *Inferensi*, 5(2), 69. <https://doi.org/10.12962/j27213862.v5i2.12353>
- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Supriyatna, A., Carolina, I., Janti, S., & Haidir, A. (2020). Clustering Koridor Transjakarta Berdasarkan Jumlah Penumpang Dengan Algoritma K-Means. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 4(2), 682–693.
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, 336(1). <https://doi.org/10.1088/1757-899X/336/1/012017>
- Whendasromo, R. G., & Joseph, J. (2022). Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada Kinerja Algoritma K-NN. *JURIKOM (Jurnal Riset Komputer)*, 9(4), 872. <https://doi.org/10.30865/jurikom.v9i4.4526>
- Widiarti Ningsih, T., & Ningsih, S. (2025). Metode Clustering Dengan Model Rfm Dan Algoritma K-Means Untuk Segmentasi Pelanggan Transjakarta. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(5), 7898–7905. <https://doi.org/10.36040/jati.v9i5.14947>