

# Penerapan Algoritma K-Means untuk Pengelompokan Halte Transjakarta Berdasarkan Aktivitas Harian Penumpang

Esteria Rohanauli Sidauruk, Anisa Fitriyani, Akmal Faiz Abdillah, M Syamsuddin Wisnubroto\*, Fajri Farid

Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera, Jalan Terusan Ryacudu, Lampung, Indonesia

## ARTICLE INFO

### Keywords:

*Clustering, Halte, Transjakarta, Activity, Peta*

**Received:** November 03, 2025

**Revised:** December 03, 2025

**Accepted:** May 18, 2026

### \*Corresponding author:

E-mail: [syamsuddin.wisnubroto@sd.itera.ac.id](mailto:syamsuddin.wisnubroto@sd.itera.ac.id)

**DOI:** [10.37253/telcomatics.v1i1.11462](https://doi.org/10.37253/telcomatics.v1i1.11462)

## ABSTRACT

Ketidakeimbangan aktivitas di halte bus Transjakarta, yang menyebabkan kemacetan dan penumpukan penumpang pada titik-titik tertentu, menjadi tantangan penting dalam peningkatan efisiensi layanan transportasi publik perkotaan. Penelitian ini bertujuan untuk mengelompokkan halte bus Transjakarta berdasarkan pola aktivitas penumpang harian (pagi, siang, sore, dan malam) menggunakan algoritma *K-Means Clustering* pada dataset transaksi Transjakarta. Setelah tahap pra-pemrosesan data dan penentuan jumlah kluster optimal dengan metode *Elbow*, diperoleh dua kluster ( $K=2$ ) yang menunjukkan perbedaan signifikan. Kluster 0 merepresentasikan halte dengan tingkat aktivitas rendah yang menandakan pemanfaatan minimal, sedangkan Kluster 1 mencakup halte dengan volume aktivitas tinggi, terutama pada periode sore hari. Visualisasi peta interaktif menunjukkan distribusi geografis yang jelas: halte dengan aktivitas sibuk (berwarna merah) terkonsentrasi di pusat kota Jakarta, sementara halte dengan aktivitas rendah (berwarna biru) tersebar di wilayah pinggiran. Hasil pengelompokan ini dapat dimanfaatkan oleh manajemen Transjakarta dan Dinas Perhubungan DKI Jakarta sebagai dasar dalam pengambilan keputusan strategis, seperti pengalokasian armada dan penyesuaian jadwal keberangkatan yang lebih adaptif terhadap tingkat permintaan di Kluster 1, serta evaluasi efektivitas operasional halte di Kluster 0. Evaluasi model menghasilkan nilai *Silhouette Score* sebesar 0.7205 yang menandakan pemisahan kluster yang baik, dan *Davies-Bouldin Index* sebesar 0.8763, yang mengindikasikan kluster yang cukup kompak dan terpisah. Penelitian ini memberikan kontribusi praktis bagi perencanaan transportasi publik berbasis data dalam mendukung kebijakan distribusi layanan yang lebih merata di seluruh jaringan Transjakarta.

## I. PENDAHULUAN

Transportasi umum memiliki peran vital dalam kehidupan masyarakat perkotaan, tidak hanya sebagai sarana mobilitas, tetapi juga sebagai infrastruktur penunjang aktivitas ekonomi dan sosial masyarakat. Di tengah pertumbuhan populasi Jakarta yang terus meningkat, keberadaan sistem transportasi publik yang efisien menjadi semakin krusial untuk mengatasi permasalahan kemacetan dan menekan tingkat polusi udara yang disebabkan oleh tingginya penggunaan kendaraan pribadi [1]. Transjakarta, yang mulai beroperasi pada tahun 2004, merupakan sistem *Bus Rapid Transit* (BRT) dengan jalur terpanjang di dunia, membentang sepanjang 251,2 kilometer melalui 14 koridor dan 287 halte yang tersebar di seluruh wilayah Jakarta. Dengan armada sebanyak 1.347 unit bus, Transjakarta berperan sebagai tulang punggung mobilitas masyarakat ibu kota [2].

Peningkatan jumlah pengguna Transjakarta secara signifikan telah memunculkan sejumlah tantangan operasional, seperti penumpukan penumpang pada halte tertentu, antrian panjang, serta kondisi bus yang melebihi kapasitas. Fenomena ini menurunkan tingkat kenyamanan penumpang dan keselamatan penumpang serta berdampak

pada efisiensi operasional sistem secara keseluruhan [3]. Ketidakeimbangan tingkat aktivitas di antara halte-halte Transjakarta mengindikasikan perlunya analisis mendalam mengenai pola aktivitas penumpang agar dapat dirancang strategi operasional yang lebih efektif dan adaptif terhadap kondisi lapangan [4].

Pemahaman terhadap karakteristik dan pola aktivitas di setiap halte merupakan aspek penting dalam meningkatkan kualitas layanan transportasi publik. Setiap halte memiliki karakteristik yang berbeda berdasarkan volume penumpang, waktu puncak aktivitas, dan pola pergerakan penumpang pengguna layanan. Dengan mengelompokkan halte berdasarkan kesamaan karakteristik pola aktivitasnya, pengelola Transjakarta dapat mengoptimalkan alokasi sumber daya seperti jumlah armada, frekuensi keberangkatan, serta kapasitas bus pada rute tertentu [5]. Selain itu, hasil pengelompokan halte dapat dimanfaatkan untuk perencanaan pengembangan infrastruktur, peningkatan fasilitas halte, dan penyusunan jadwal operasional yang lebih responsif terhadap kebutuhan pengguna [6].

Dalam konteks tersebut, analisis berbasis data menjadi pendekatan strategis untuk memahami pola aktivitas halte [7]. Metode *clustering* atau pengelompokan data merupakan salah

satu teknik dalam *data mining* yang bertujuan mengelompokkan objek berdasarkan kesamaan karakteristiknya [8]. Algoritma *K-Means Clustering* merupakan salah satu metode *clustering* yang paling populer dan banyak digunakan karena kesederhanaan implementasinya, efisiensi komputasi, serta kemampuannya dalam menangani dataset berukuran besar. Algoritma ini bekerja dengan membagi data ke dalam sejumlah kelompok berdasarkan jarak terdekat ke *centroid*, sehingga objek dengan karakteristik serupa tergabung dalam satu kelompok yang sama. Penerapan *K-Means Clustering* telah terbukti efektif dalam berbagai studi kasus, seperti segmentasi pelanggan, pemetaan wilayah endemis penyakit, dan analisis pola penggunaan layanan [9].

Sejumlah penelitian terdahulu juga telah menerapkan *K-Means Clustering* dalam konteks transportasi Transjakarta. Saputra dan Yusuf [10] menerapkan pendekatan ini dengan metode *Recency, Frequency*, dan *Monetary* untuk segmentasi pelanggan Transjakarta dan menghasilkan nilai *Silhouette Score* sebesar 0,714917 dan *Davies-Bouldin Index* sebesar 0,365776, yang menunjukkan kualitas pengelompokan yang baik. Sementara itu, Reynaldi dkk [11] menganalisis pola penyebaran spasial pengguna Transjakarta menggunakan *K-Means Clustering* berbasis koordinat geografis dan waktu penggunaan layanan, menghasilkan empat kluster yang merepresentasikan variasi spasial-temporal penggunaan layanan. Kedua penelitian ini memperlihatkan efektivitas *K-Means Clustering* dalam analisis data transportasi, tetapi belum mengkaji dimensi aktivitas halte berdasarkan intensitas temporal penumpang.

Meskipun berbagai penelitian terdahulu telah memanfaatkan *K-Means Clustering*, konteks analisis transportasi publik, termasuk penelitian Matseliukh dan Lytvyn [11] yang menerapkan *K-Means* untuk mengorganisasikan transportasi penumpang di *smart city*, sebagian besar penelitian tersebut masih berfokus pada aspek spasial, pola pergerakan, atau segmentasi pengguna tanpa mengintegrasikan secara komprehensif dinamika temporal aktivitas halte. Penelitian-penelitian tersebut belum menguraikan bagaimana variasi aktivitas harian termasuk periode pagi, siang, sore, dan malam secara langsung memengaruhi pola operasional halte. Dengan demikian, masih terdapat kesenjangan dalam literatur mengenai pemetaan halte berdasarkan intensitas aktivitas temporal dengan menggunakan pendekatan *clustering*, terutama dalam konteks sistem BRT seperti Transjakarta, yang memiliki variasi penggunaan yang sangat dipengaruhi oleh ritme mobilitas masyarakat perkotaan.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan menerapkan algoritma *K-Means Clustering* untuk mengelompokkan halte Transjakarta berdasarkan aktivitas operasional harian (pagi, siang, sore, dan malam) menggunakan data transaksi penumpang yang mencakup informasi mengenai lokasi *tap-in/tap-out*, waktu penggunaan, dan volume penumpang di setiap halte. Pendekatan ini mengintegrasikan dimensi temporal dan spasial dalam satu model analisis, sehingga menghasilkan representasi yang lebih komprehensif terhadap dinamika aktivitas halte. Kebaruan penelitian ini terletak pada penerapan *K-Means* untuk pemetaan halte berdasarkan pola aktivitas temporal

penumpang, yang belum pernah dibahas dalam studi terdahulu. Hasil pengelompokan diharapkan dapat digunakan oleh manajemen Transjakarta dan Dinas Perhubungan DKI Jakarta untuk mendukung pengambilan keputusan strategis, termasuk optimalisasi armada, penyesuaian jadwal keberangkatan, dan peningkatan efisiensi operasional berbasis analisis data.

## II. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan *Design Science Research (DSR)* dengan fokus pada pengembangan artefak teknologi untuk menyelesaikan masalah praktis. Implementasi mengikuti metodologi iteratif dengan tahapan *problem identification, design & development, demonstration, dan evaluation*.

### A. Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *unsupervised learning*, khususnya penerapan algoritma *K-Means clustering*. Tujuan dari penelitian ini adalah untuk mengelompokkan halte Transjakarta berdasarkan kesamaan pola aktivitas harian yang ditinjau dari waktu penggunaannya (pagi, siang, sore, dan malam). Dengan metode ini, halte-halte yang memiliki pola penggunaan serupa dapat tergabung dalam satu *cluster*, sehingga mempermudah analisis pola pergerakan penumpang dan pengambilan kebijakan operasional transportasi publik.

### B. Sumber dan Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari Transjakarta Open Data Portal (Transjakarta, 2024) yang menyediakan rekaman aktivitas harian penumpang pada setiap halte. Dataset ini mencakup informasi waktu *tap in* dan *tap out*, nama halte keberangkatan (*tapInStopsName*) dan tujuan (*tapOutStopsName*), serta koordinat geografis halte (*latitude* dan *longitude*). Selain itu, data ini merepresentasikan pola aktivitas pengguna Transjakarta pada berbagai rentang waktu dalam sehari sehingga dapat digunakan sebagai dasar dalam proses pengelompokan menggunakan algoritma *K-Means*.

### C. Tahapan Penelitian

Tahapan penelitian dilakukan secara sistematis melalui empat tahapan utama, yaitu pra-pemrosesan data, penentuan jumlah cluster optimal, pembentukan model kluster akhir, dan visualisasi hasil. Setiap fase dijelaskan sebagai berikut:

#### 1. Pra-pemrosesan Data

Tahap pra-pemrosesan data merupakan langkah awal untuk memastikan data siap digunakan dalam proses pengelompokan. Pertama, dilakukan pemuatan data dari file *dfTransjakarta-1.xlsx* menggunakan library *pandas* di Python. Selanjutnya, dilakukan proses pembersihan data (*data cleaning*) dengan cara mengkonversi kolom waktu seperti *tapInTime* dan *tapOutTime* ke format *datetime*, serta menghapus data yang memiliki nilai kosong pada kolom penting seperti nama halte atau waktu [13]. Setelah itu, dilakukan restrukturisasi data aktivitas dengan cara menggabungkan aktivitas *tap in* dan *tap out* ke dalam satu

*DataFrame* yang berisi nama halte beserta waktu aktivitasnya. Dari kolom waktu tersebut, diekstraksi nilai jam dan dikategorikan menjadi empat kelompok waktu, yaitu pagi (06.00–09.00), siang (10.00–15.00), sore (16.00–19.00), dan malam (20.00–05.00). Selanjutnya, dilakukan agregasi aktivitas menggunakan *pivot table* untuk menghitung jumlah aktivitas per kategori waktu berdasarkan nama halte. Sebelum digunakan dalam model, data dinormalisasi dengan *MinMaxScaler* agar seluruh fitur berada dalam rentang 0–1, sehingga setiap variabel memiliki bobot yang seimbang dalam proses *clustering*.

## 2. Penentuan Jumlah Cluster Optimal

Tahap ini bertujuan untuk menentukan jumlah *cluster* ( $K$ ) yang paling sesuai agar hasil pengelompokan menghasilkan kelompok halte dengan karakteristik yang jelas. Proses dilakukan dengan mencoba berbagai nilai  $K$  dari 2 hingga 10, kemudian dilakukan evaluasi menggunakan dua metrik, yaitu metode *Elbow* dan *Silhouette Score* [14]. Metode *Elbow* digunakan untuk melihat perubahan nilai *inertia*, yakni jumlah kuadrat jarak antara setiap titik data dan pusat cluster. Sementara itu, *Silhouette Score* digunakan untuk menilai seberapa baik data telah terkelompok berdasarkan jarak antar cluster. Kedua hasil evaluasi divisualisasikan dalam grafik, yang menunjukkan hubungan antara jumlah cluster dengan nilai *inertia* dan *Silhouette Score*. Nilai  $K$  optimal ditentukan berdasarkan titik siku (*elbow point*) pada grafik inersia serta skor *Silhouette* tertinggi, karena pada titik inilah model dianggap paling efisien dalam membedakan kelompok data.

## 3. Pembentukan Model Kluster

Setelah jumlah kluster optimal ditentukan, misalnya  $K = 2$ , langkah selanjutnya adalah membentuk model akhir menggunakan algoritma *K-Means*. Algoritma *K-Means* bekerja secara iteratif dengan menentukan  $K$  *centroid* awal, kemudian menghitung jarak dari setiap data halte bus ke setiap *centroid* untuk mengelompokkan data ke dalam kluster terdekat. *Centroid* kemudian diperbarui berdasarkan rata-rata data di setiap kluster, dan proses ini diulang hingga posisi *centroid* stabil atau tidak berubah secara signifikan, menghasilkan pengelompokan konvergen akhir. Model dilatih menggunakan data yang telah dinormalisasi pada tahap sebelumnya, dan setiap halte bus Transjakarta kemudian diberi label kluster berdasarkan hasil pengelompokan. Selanjutnya, analisis profil kluster dilakukan untuk memahami karakteristik setiap kelompok halte bus. Analisis ini dilakukan dengan menghitung rata-rata jumlah aktivitas halte bus di setiap kategori waktu untuk setiap kluster. Hasil analisis divisualisasikan dalam bentuk diagram batang untuk lebih jelas menunjukkan perbedaan pola aktivitas antar kluster. Untuk memberikan pemahaman yang lebih konkret, beberapa halte bus dari setiap *cluster* diambil, yang mewakili pola aktivitas tertentu, seperti halte bus dengan aktivitas tinggi dan aktivitas rendah.

## 4. Metrik Evaluasi

Tahapan ini dilakukan untuk menilai sejauh mana kualitas hasil pengelompokan yang dihasilkan oleh algoritma *K-Means*. Evaluasi dilakukan menggunakan dua metrik utama,

yaitu *Silhouette Score* dan *Davies-Bouldin Index*. *Silhouette Score* digunakan untuk mengukur seberapa baik setiap data dalam suatu kluster memiliki kesamaan dengan klusternya sendiri dibandingkan dengan kluster lain, dengan nilai yang lebih tinggi menunjukkan pemisahan kluster yang lebih baik. Sementara itu, *Davies-Bouldin Index* menilai tingkat kemiripan antar-kluster, di mana nilai yang lebih rendah mengindikasikan bahwa kluster memiliki jarak yang cukup jauh satu sama lain dan lebih kompak di dalamnya. Kedua metrik ini digunakan secara bersamaan untuk memastikan bahwa hasil pengelompokan yang diperoleh tidak hanya terpisah dengan baik, tetapi juga memiliki konsistensi internal yang kuat. Evaluasi ini menjadi langkah penting sebelum dilakukan interpretasi dan visualisasi hasil *clustering* agar model yang dibangun benar-benar merepresentasikan pola aktivitas halte Transjakarta secara akurat [15]. Selain mendeskripsikan fungsi utamanya, kedua metrik tersebut dihitung menggunakan pendekatan matematis standar dalam evaluasi klusterisasi. *Silhouette Score* dihitung berdasarkan selisih antara rata-rata jarak titik terhadap klusternya sendiri dan kluster terdekat, yang dirumuskan sebagai:

$$s(i) = \frac{(b - a)}{\max(a, b)} \quad (1)$$

di mana  $a$  adalah jarak rata-rata titik terhadap anggota klusternya dan  $b$  adalah jarak rata-rata terhadap kluster terdekat lainnya. Sementara itu, *Davies-Bouldin Index* dihitung melalui perbandingan antara dispersi intrakluster dan jarak antarkluster, dengan formula:

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left( \frac{s_i + s_j}{d_{ij}} \right) \quad (2)$$

di mana  $s_i$  merupakan dispersi rata-rata dalam kluster ke- $i$  dan  $d_{ij}$  adalah jarak antara centroid kluster ke- $i$  dan ke- $j$ .

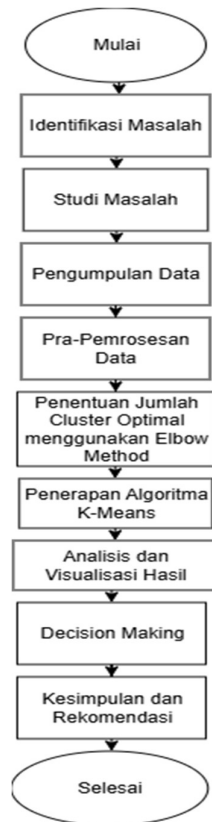
## 5. Visualisasi Peta Persebaran Cluster

Tahap terakhir adalah pembuatan visualisasi peta interaktif yang menggambarkan persebaran *cluster* halte Transjakarta secara geografis. Proses ini dilakukan dengan menggabungkan hasil *clustering* dengan data koordinat halte (*latitude* dan *longitude*). Visualisasi dibuat menggunakan library *Folium*, di mana setiap halte ditandai dengan warna berbeda sesuai dengan *cluster*-nya, misalnya biru untuk *cluster* dengan aktivitas rendah dan merah untuk *cluster* dengan aktivitas tinggi. Peta interaktif ini membantu dalam memahami distribusi spasial pola aktivitas halte di seluruh wilayah Jakarta, serta dapat digunakan sebagai alat bantu dalam perencanaan operasional atau kebijakan transportasi.

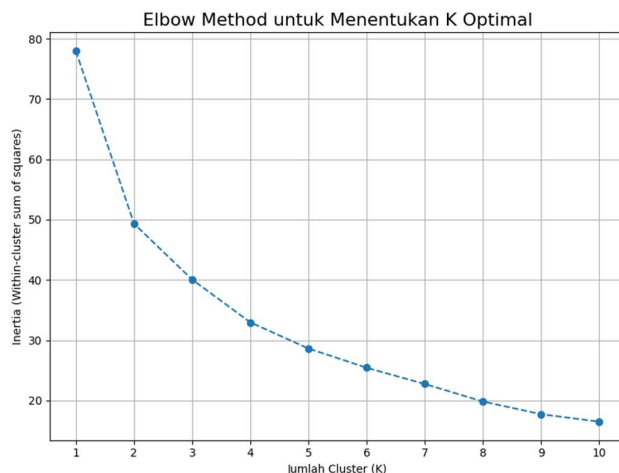
## 6. Alur Penelitian

Alur penelitian ini diawali dengan identifikasi masalah mengenai ketidakseimbangan aktivitas halte Transjakarta, kemudian dilanjutkan dengan studi literatur untuk menentukan metode analisis yang tepat menggunakan algoritma *K-Means Clustering*. Setelah pengumpulan data dari dataset Transjakarta, dilakukan pra-pemrosesan data berupa pembersihan, pengelompokan waktu aktivitas, dan

normalisasi. Selanjutnya ditentukan jumlah *cluster* optimal melalui metode *Elbow* dan *Silhouette Score*, lalu diterapkan algoritma *K-Means* untuk mengelompokkan halte berdasarkan pola aktivitas harian. Hasil pengelompokan dianalisis dan divisualisasikan untuk memudahkan interpretasi, kemudian dimanfaatkan dalam tahap *decision making* guna mendukung kebijakan operasional Transjakarta. Penelitian diakhiri dengan kesimpulan dan rekomendasi sebagai dasar peningkatan efisiensi pelayanan transportasi publik. Secara keseluruhan alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian



Gambar 2. Penentuan K-Optimal

### III. HASIL DAN PEMBAHASAN

#### A. Pra-pemrosesan Data

Sebelum dilakukan penentuan jumlah *cluster* optimal menggunakan metode *elbow*, dilakukannya pra-pemrosesan data. Tabel 1 menampilkan hasil 5 data teratas setelah dilakukannya pra-pemrosesan.

Tabel 1. Data setelah Pra-pemrosesan

nama_titik_henti	aktivitas_m alam	aktivitas_ pagi	aktivitas_ siang	aktivitas_ sore
18 Office Park	0.046729	0.070352	0.000000	0.000000
ABA	0.000000	0.005025	0.020408	0.000000
ACC Simatupang	0.000000	0.005025	0.000000	0.005747
ACE Hardware	0.028037	0.000000	0.000000	0.028736
AKR Tower	0.000000	0.000000	0.020408	0.000000

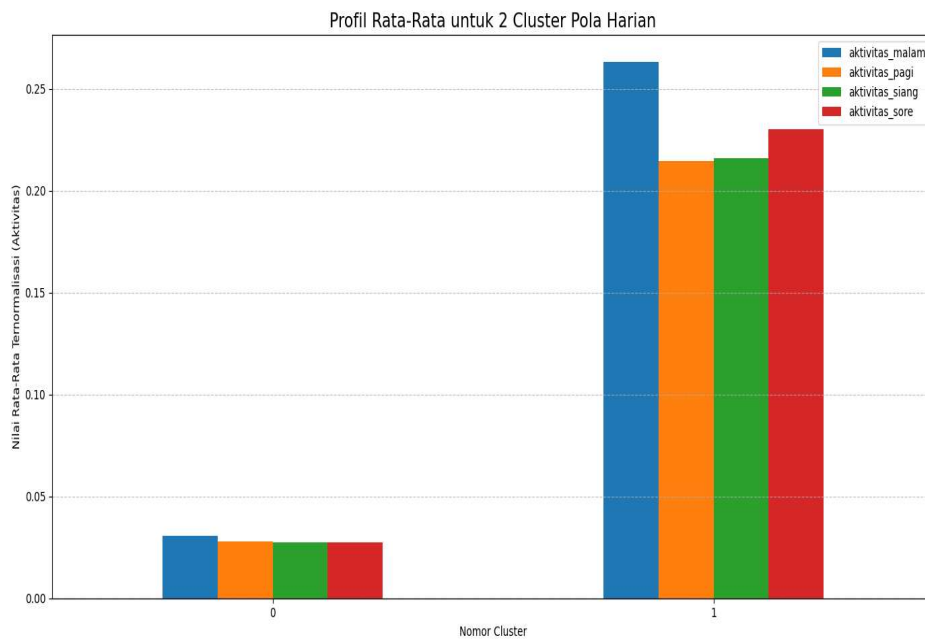
#### B. Hasil Penentuan Jumlah Cluster Optimal

Sebelum melakukan proses pengelompokan menggunakan algoritma *K-Means*, langkah awal yang dilakukan adalah menentukan jumlah *cluster* ( $K$ ) yang paling optimal. Dalam penelitian ini digunakan metode *Elbow Method* untuk mengetahui nilai  $K$  terbaik. Metode ini bekerja dengan cara menghitung nilai *inertia* atau *within-cluster sum of squares (WCSS)* untuk setiap nilai  $K$  yang diuji, kemudian hasilnya diplotkan dalam bentuk grafik. Berikut merupakan hasil visualisasi dari *Elbow Method* yang ditunjukkan pada Gambar 2.

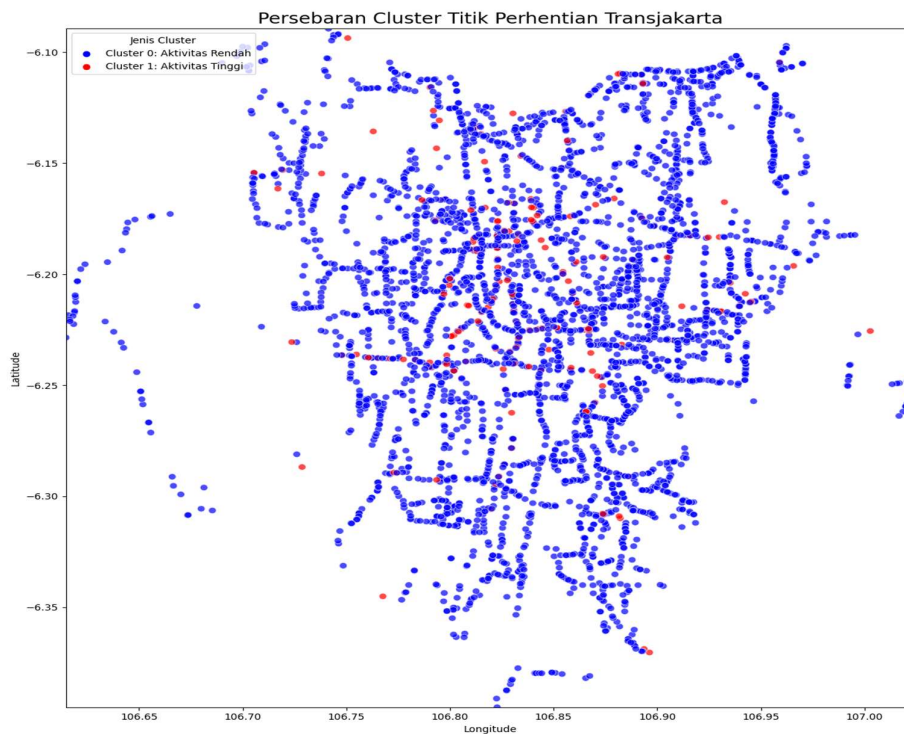
Berdasarkan grafik pada Gambar 2, terlihat bahwa penurunan nilai *inertia* mulai melambat setelah nilai  $K = 2$ . Titik tersebut membentuk “siku” (*elbow*), yang menandakan bahwa penambahan jumlah *cluster* setelah nilai tersebut tidak memberikan peningkatan signifikan terhadap penurunan *inertia*. Oleh karena itu, jumlah *cluster* optimal yang digunakan dalam penelitian ini adalah 2 *cluster*, yang terdiri dari *cluster* 0 dan *cluster* 1.

#### C. Visualisasi Hasil

Proses penelitian diakhiri dengan pembuatan visualisasi yang mudah dipahami. Setelah menggunakan metode *K-Means* untuk mengelompokkan halte bus Transjakarta berdasarkan pola aktivitas harian, data kemudian ditampilkan dalam dua format. Pertama, sebagai diagram batang untuk menunjukkan dengan jelas perbedaan rata-rata aktivitas antara halte bus yang ramai dan sepi. Kedua, data pengelompokan ini digabungkan dengan lokasi GPS halte bus untuk membuat peta interaktif. Pada peta ini, setiap halte bus ditandai dengan warna yang sesuai yaitu biru untuk halte bus yang sepi dan merah untuk halte bus yang ramai. Kedua visualisasi ini bertujuan agar semua orang dapat melihat dan memahami perbedaan tingkat aktivitas halte bus, baik dari segi jumlah maupun lokasi, sehingga pihak Transjakarta dapat merencanakan operasional dan rute dengan lebih baik.



Gambar 3. Visualisasi Hasil Evaluasi Clustering



Gambar 4. Peta Interaktif Clustering

Berdasarkan Gambar 3 yaitu diagram batang "Profil Rata-Rata untuk 2 Klaster Pola Harian" menunjukkan hasil pengelompokan *K-Means*, yang membagi halte Transjakarta menjadi dua klaster dengan profil penggunaan yang sangat berbeda. Klaster 0 mewakili halte dengan penggunaan minimal, ditandai dengan tingkat aktivitas yang sangat rendah dan stabil di seluruh segmen waktu harian (malam (bagan

warna biru), pagi (bagan warna oranye), siang (bagan warna hijau), dan sore (bagan warna merah)), dengan nilai normalisasi sekitar 0,03. Di sisi lain, Klaster 1 menunjukkan aktivitas yang tinggi dan dominan di seluruh jaringan, jauh melebihi Klaster 0. Klaster utama ini mengalami puncak aktivitas tertinggi pada malam hari (sekitar 0,26) dan sore hari (sekitar 0,23). Perbedaan yang jelas ini menunjukkan adanya

ketidakseimbangan operasional, yang secara jelas memisahkan halte yang sibuk dan strategis (Klaster 1) dari halte yang kurang populer (Klaster 0), sehingga memberikan informasi penting untuk panduan alokasi sumber daya.

Tabel 2. Metrik Evaluasi

Silhouette Score	0.7205
Davies-Bouldin Index	0.8763

Berdasarkan Tabel 2, dua metrik utama yang digunakan untuk mengevaluasi kualitas pengelompokan yaitu *Silhouette Score* dan *Davies-Bouldin Index*. Nilai *Silhouette Score* sebesar 0.7205 menunjukkan bahwa hasil clustering memiliki kualitas pemisahan yang sangat baik, dengan tingkat kemiripan yang tinggi dalam satu klaster dan jarak yang cukup jauh antar klaster. Sementara itu, *Davies-Bouldin Index* dihitung sebagai rata-rata ukuran kesamaan antara setiap klaster dan klaster yang paling mirip dengannya, di mana nilai yang lebih rendah menunjukkan performa *clustering* yang lebih baik. Nilai *Davies-Bouldin Index* sebesar 0.8763 mengindikasikan tingkat tumpang tindih antar klaster yang rendah dan struktur klaster yang terbentuk dengan jelas. Secara keseluruhan, kedua metrik ini memperkuat validitas pemilihan  $K = 2$  sebagai jumlah klaster optimal karena mampu menghasilkan pemisahan data yang baik serta konsistensi internal yang tinggi pada setiap klaster.

Gambar 4 memvisualisasikan hasil geospasial dari pengelompokan *K-Means*, yang membagi semua halte bus Transjakarta menjadi dua kelompok berdasarkan tingkat aktivitas: Klaster 0 (Aktivitas Rendah) ditunjukkan dengan titik-titik biru, dan Klaster 1 (Aktivitas Tinggi) ditunjukkan dengan titik-titik merah. Secara visual, terjadi polarisasi geografis yang jelas, dengan halte bus biru (Aktivitas Rendah) mendominasi area periferal dan rute yang kurang padat, membentuk jaringan yang luas di seluruh Jabodetabek. Sebaliknya, halte bus merah (Aktivitas Tinggi) tampak terkonsentrasi padat di pusat Jakarta (terutama Jakarta Pusat dan Jakarta Selatan), yang mengonfirmasi bahwa halte bus yang paling vital dan sibuk sebagian besar terletak di distrik bisnis, area pemerintah, dan pusat transit utama di dalam lingkaran kota. Distribusi ini memberikan wawasan penting bahwa upaya pengoptimalan sumber daya dan peningkatan layanan harus difokuskan pada area merah yang padat, sementara area biru mungkin memerlukan evaluasi ulang efisiensi rute.

#### IV. KESIMPULAN

Penerapan algoritma *K-Means Clustering* terhadap data aktivitas harian penumpang berhasil mengelompokkan halte-halte Transjakarta ke dalam dua klaster utama dengan karakteristik berbeda secara signifikan. Berdasarkan hasil analisis Metode *Elbow*, jumlah klaster optimal ditetapkan pada  $K = 2$ , yang membagi halte menjadi Klaster 0 dengan tingkat aktivitas rendah dan Klaster 1 dengan tingkat aktivitas tinggi. Temuan ini menunjukkan adanya perbedaan mencolok dalam pola mobilitas penumpang antar wilayah operasional Transjakarta..

Klaster 0 merepresentasikan halte dengan tingkat aktivitas yang rendah dan relatif stabil, yang ditandai dengan nilai

ternormalisasi sekitar 0,03 dan umumnya berada di wilayah pinggiran kota dengan pemanfaatan rendah. Sebaliknya, Klaster 1 mencakup halte dengan volume penumpang tinggi, terutama pada sore hingga malam hari, yang banyak terkonsentrasi di wilayah pusat dan selatan Jakarta sebagai kawasan dengan intensitas aktivitas perkotaan tinggi. Perbedaan spasial ini menegaskan hubungan erat antara tingkat aktivitas penumpang dengan fungsi wilayah dan densitas mobilitas di sekitar halte. Evaluasi kuantitatif Evaluasi kuantitatif menunjukkan bahwa kualitas hasil klasterisasi tergolong baik, dengan nilai *Silhouette Score* sebesar 0,7205 yang mengindikasikan pemisahan klaster yang kuat dan kohesi internal yang tinggi, serta *Davies-Bouldin Index* sebesar 0,8763 yang menunjukkan tingkat kemiripan antar klaster yang rendah. Kombinasi kedua metrik tersebut menegaskan bahwa model klasterisasi dengan  $K=2$  memberikan hasil yang valid dan representatif dalam menggambarkan segmentasi aktivitas penumpang Transjakarta. Hasil penelitian ini dapat menjadi dasar pengambilan kebijakan dan manajemen Transjakarta untuk merumuskan strategi optimalisasi layanan. Fokus peningkatan armada, penyesuaian frekuensi perjalanan, dan penyediaan fasilitas tambahan perlu diarahkan pada halte-halte di Klaster 1 untuk mengantisipasi lonjakan penumpang pada jam sibuk. Sementara itu, halte di Klaster 0 perlu dievaluasi dari sisi efisiensi rute dan pemanfaatan fasilitas agar alokasi sumber daya dapat lebih proporsional serta mendukung pemerataan layanan transportasi publik di seluruh wilayah Jakarta.

Meskipun penelitian ini berhasil mengidentifikasi dua kelompok aktivitas halte Transjakarta dengan kualitas klasterisasi yang baik, penelitian ini memiliki beberapa keterbatasan, terutama terkait periode data yang digunakan dan jumlah variabel yang dianalisis. Untuk penelitian selanjutnya, disarankan untuk menggunakan data multi-tahun agar pola musiman dapat diamati, menguji algoritma klasterisasi lain seperti *DBSCAN* atau *Hierarchical Clustering* untuk memperoleh perbandingan hasil, serta menambahkan variabel kontekstual seperti kepadatan penduduk, karakteristik wilayah, atau integrasi antar moda untuk memperkaya interpretasi klaster. Pendekatan ini diharapkan dapat meningkatkan robustness dan kedalaman analisis pada studi-studi berikutnya.

#### V. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Sains Data ITERA atas dukungan fasilitas komputasi, Badan Meteorologi Klimatologi dan Geofisika (BMKG), Badan Nasional Penanggulangan Bencana (BNPB), dan Badan Informasi Geospasial (BIG) atas akses data untuk penelitian, serta komunitas *open source* atas *tools* yang digunakan dalam implementasi sistem ini.

#### VI. DAFTAR PUSTAKA

- [1] Liu, X., Zhang, Y., He, Y., dkk. (2025). Nonlinear effects of multilevel factors on public transport commuting: A machine-learning framework. *Transportation Research Part D: Transport and Environment*. <https://doi.org/10.1016/j.trd.2025.104724>

- [2] Tang, Y., Alhadlaq, A., Bagabaldo, A. R., & Gonzalez, M. C. (2025). Designing transit routes based on vehicle routing behavior determined through location-based services data. *EPJ Data Science*, 14(1), 45. <https://doi.org/10.1140/epjds/s13688-025-00559-5>
- [3] Wang, D. (2025). A spatio-temporal node-place-ridership model for classifying public transit stations using K-Means. *Journal of Transport and Land Use*, 18(1), 447–466. <https://doi.org/10.5198/jtlu.2025.2573>
- [4] Siswanto, J., Hendry, H., Rahardja, U., Sembiring, I., & Lisangan, E. A. (2025). Predicting TransJakarta Passengers with LSTM-BiLSTM (case study / model development). *APTISI Transactions on Technopreneurship*, 7(1), 84–96. <https://doi.org/10.34306/att.v7i1.440>
- [5] Saputra, A., & Yusuf, R. (2024). Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1346–1361. <https://doi.org/10.57152/malcom.v4i4.1516>
- [6] Kristianto, A. (2021). Analisa Performa K-Means dan DBSCAN dalam Clustering Public Transportation. *Jurnal Elektronika dan Komputer (Elkom)*, 14(2), 368–372. <https://doi.org/10.51903/elkom.v14i2.551>
- [7] Liperda, R. I., & Fatahayu, N. R. (2025). A Simulation-based Study of Bus Interarrival and Passenger Flow at a TransJakarta Transfer Point. *International Journal / Institutional Repository (Mercu Buana)*. <https://doi.org/10.22441/ijiem.v4i3.20721>
- [8] Sinaga, A. O., & Passarella, R. (2025). Analysis of Passenger and Bus Numbers Using Unsupervised Learning (K-Means, FCM, dll.). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(3), 1029–1036. <https://doi.org/10.57152/malcom.v5i3.2005>
- [9] Nurlaela, S., & William, A. (2023). TransJakarta Service Evaluation in Controlling COVID-19 Transmission Using Twitter Sentiment Analysis. *Journal of Regional and City Planning (JPWK)*, 34(2), 156–174. <https://doi.org/10.5614/jpwk.2023.34.2.2>
- [10] Shafiq, M., Rocha, H., Couto, A., & Ferreira, S. (2024). A Clustering Approach for Analyzing Access to Public Transport. *Sustainability*, 16(16), 6944. <https://doi.org/10.3390/su16166944>
- [11] Matseliukh, Y., & Lytvyn, V. (2025). Application of the K-means clustering method in organizing passenger transportation in a smart city. *Innovative Technologies and Scientific Solutions for Industries*, 1(31), 83–101. <https://doi.org/10.30837/2522-9818.2025.1.083>
- [12] Ikasari, D., Permana, R. A., & Widiastuti. (2024). Prediction in the Number of Passengers on TransJakarta Public Transportation Using the Single Exponential Smoothing Method. *International Research Journal of Advanced Engineering and Science (IRJAES)*, 9(2), 197–202. Available online at: <https://irjaes.com/wp-content/uploads/2024/06/IRJAES-V9N2P197Y24.pdf>
- [13] Shafiq, M., Rocha, H., Couto, A., & Ferreira, S. (2024). A Clustering Approach for Analyzing Access to Public Transportation and Destinations. *Sustainability*, 16(16), 6944. <https://doi.org/10.3390/su16166944>
- [14] Liu, X., Huang, Z., & Jian, W. (2025). Nonlinear effects of multilevel factors on public transport commuting in China's cities. *Transportation Research Part D: Transport and Environment*, 143, Article 104724. <https://doi.org/10.1016/j.trd.2025.104724>
- [15] Wang, D., Wu, J., & Tao, Z. (2025). A spatio-temporal node-place-ridership model for classifying metro station areas: The case of Shenzhen, China. *Journal of Transport and Land Use*, 18(1), 447–466. <https://doi.org/10.5198/jtlu.2025.2573>