Journal Software, Hardware and Information Technology SHiFT

ISSN 2776-8961 (online)
2808-3385 (print)
Vol. 5 No. 2 Tahun 2025
doi: https://doi.org/10.24252/shift.v5i2.205

Analisa Pola Penyebaran Pengguna Layanan Transjakarta dengan Metode K-Means Clustering

Reynaldi [1], Raihan Jamal Faiz Djarot [2*], Mochamad Wahyudi [3], Sumanto [4], Ade Surya Budiman [5]

[1,2,3,4,5] Jurusan Teknik Informatika, Universitas Bina Sarana Informatika, Indonesia
[1] 15220563@bsi.ac.id, [2] 15220561@bsi.ac.id, [3] wahyudi@bsi.ac.id, [4] sumanto@bsi.ac.id, [5] adeaum@bsi.ac.id

Diajukan: 15 Juni 2025 | Direvisi: 21 Juni 2025 | Diterima: 23 Juni 2025 | Diterbitkan: 30 Juni 2025

Abstrak

Penelitian ini menganalisis pola penyebaran spasial pengguna layanan Transjakarta di wilayah Jakarta menggunakan algoritma *K-Means Clustering*. Dataset yang digunakan bersumber dari platform *Kaggle* dan terdiri atas 189.501 data transaksi penumpang, mencakup informasi lokasi *tap-in* dan *tap-out*, waktu perjalanan, serta data pengguna. Proses penelitian melibatkan pengumpulan data, prapemrosesan untuk menghapus nilai kosong, penerapan algoritma *K-Means Clustering*, serta penentuan jumlah klaster optimal menggunakan metode *elbow*. Berdasarkan hasil perhitungan, jumlah klaster optimal ditetapkan sebanyak empat (K=4). Visualisasi menggunakan *scatter plot* menunjukkan pola distribusi pengguna berdasarkan koordinat geografis dan waktu penggunaan layanan. Setiap klaster merepresentasikan kelompok pengguna dengan karakteristik perjalanan yang serupa. Analisis ini menghasilkan segmentasi yang menggambarkan variasi pola mobilitas penumpang Transjakarta, serta menunjukkan bagaimana aktivitas perjalanan tersebar dalam dimensi ruang dan waktu di wilayah perkotaan Jakarta.

Kata Kunci: K-Means Clustering, Transjakarta, Transportasi Umum, Data Mining

Abstract

This study analyzes the spatial distribution patterns of Transjakarta service users in Jakarta using the K-Means Clustering algorithm. The dataset, obtained from the Kaggle platform, consists of 189,501 passenger transaction records, including tap-in and tap-out locations, travel times, and user-related information. The research process involves data collection, preprocessing to remove missing values, application of the K-Means Clustering algorithm, and determination of the optimal number of clusters using the elbow method. Based on the analysis, the optimal number of clusters is identified as four (K=4). A scatter plot visualization presents user distribution patterns based on geographic coordinates and service usage times. Each cluster represents a group of users with similar travel characteristics. This analysis results in a segmentation that reflects variations in Transjakarta passenger mobility patterns and illustrates how travel activity is distributed across spatial and temporal dimensions within the urban area of Jakarta.

Keywords: *K-Means Clustering, Transjakarta, Public Transportation, Data Mining*

This work is an open access article and licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0). Copyright (C) Author's.

1. PENDAHULUAN

Transportasi umum memiliki peran krusial dalam menunjang kehidupan masyarakat di suatu wilayah. Fungsinya tidak hanya sebagai pilihan alternatif, tetapi telah menjadi kebutuhan utama yang mendukung aktivitas harian penduduk [1]. Angkutan umum menjadi alternatif bagi masyarakat Jakarta yang padat penduduk. Terlebih lagi, perkembangan teknologi mendorong terciptanya solusi untuk mengatasi permasalahan akibat kepadatan penduduk yang memicu kemacetan di perkotaan [2]. Bus Transjakarta adalah salah satu transportasi darat yang mampu membawa penumpang dalam jumlah besar karena dilengkapi dengan tempat duduk yang nyaman, interior yang layak, dan tarif yang terjangkau [3]. Transjakarta merupakan salah satu moda

ISSN 2776-8961 (online)
2808-3385 (print)
Vol. 5 No. 2 Tahun 2025
doi: https://doi.org/10.24252/shift.v5i2.205

transportasi publik yang disediakan oleh Pemerintah Provinsi DKI Jakarta dengan tujuan mempermudah mobilitas masyarakat serta mengurangi tingkat kemacetan di wilayah ibu kota [4].

Salah satu permasalahan yang sering terjadi adalah kurangnya perhatian Bus Transjakarta terhadap kenyamanan dan ketertiban penumpang. Masih banyak penumpang yang memaksakan diri masuk ke dalam bus meskipun kapasitas sudah melebihi batas yang dianjurkan. Kondisi ini menyebabkan penumpang berdesakan dan menurunkan tingkat kenyamanan penumpang lainnya [3]. Oleh karena itu, dibutuhkan sistem transportasi yang aman, nyaman, tepat waktu, serta terintegrasi, guna mendukung perkembangan kota, meningkatkan kepuasan pengguna transportasi, dan mengurangi tingkat kemacetan di Jakarta [5]. Dalam beberapa tahun terakhir, transportasi umum di wilayah perkotaan semakin mendapat perhatian untuk mendukung keberlanjutan dan meningkatkan kualitas hidup masyarakat kota. Efisiensi dalam menghubungkan sumber daya ke berbagai tujuan serta kemudahan mobilitas massal dapat memberikan dampak positif bagi kinerja ekonomi dan lingkungan suatu kota. Transjakarta menjadi salah satu layanan transportasi umum yang berkualitas, dengan menawarkan kenyamanan dan keamanan bagi warga Jakarta, sehingga menarik minat banyak masyarakat untuk menggunakannya [6]. Transportasi umum diharapkan mampu mengatasi berbagai permasalahan perkotaan seperti kemacetan lalu lintas dan polusi udara akibat emisi kendaraan bermotor. Tingginya mobilitas penduduk kota dalam menjalankan aktivitas seperti sekolah, bekerja, dan memenuhi kebutuhan sehari-hari menyebabkan meningkatnya kebutuhan waktu tempuh. Jika masyarakat lebih memilih menggunakan kendaraan pribadi dibandingkan transportasi umum, maka tingkat kemacetan akan semakin tinggi. Beberapa penelitian menunjukkan bahwa pengendara dapat menghabiskan waktu antara 45 menit hingga 1 jam untuk bepergian dari rumah ke tempat tujuan, dengan persentase waktu perjalanan mencapai sekitar 40% hingga 60% dari total waktu aktivitas harian mereka [7].

Dalam upaya mengoptimalkan layanan Transjakarta, penelitian ini menerapkan metode *K-Means Clustering*, yaitu salah satu teknik *Clustering non-hierarki* yang bekerja dengan membagi data ke dalam beberapa kelompok (*cluster*). Metode ini mengelompokkan data berdasarkan kesamaan karakteristik, sehingga data dengan ciri serupa ditempatkan dalam satu *cluster*, sedangkan data dengan karakteristik yang berbeda dimasukkan ke dalam *cluster* yang lain[8]. Penelitian terdahulu menjelaskan bahwa *K-Means Clustering* mampu melakukan pengkategorian pada sebuah data untuk menyajikan pengetahuan baru. Pengetahuan yang disajikan dapat digunakan untuk memberikan pengelompokan pada sebuah data [9]. *K-Means Clustering* adalah salah satu jenis algoritma *unsupervised learning* yang bekerja dengan mengelompokkan data berdasarkan kesamaan jarak antar data. Algoritma ini membagi sekumpulan data ke dalam beberapa klaster sesuai tingkat kemiripannya. *K-Means Clustering* banyak digunakan dalam berbagai bidang, seperti pemetaan inventaris dan deteksi potensi tanah longsor. Namun, algoritma ini hanya dapat diterapkan pada data yang memiliki atribut berupa nilai numerik [10]. Beberapa penelitian terdahulu telah mengkaji penggunaan algoritma *K-Means Clustering*, Penelitian tersebut bertujuan untuk mengelompokkan tingkat persebaran kasus *COVID-19* di berbagai provinsi di Indonesia ke dalam kategori tinggi, sedang, dan rendah. Berdasarkan studi-studi sebelumnya mengenai penyebaran kasus *COVID-19* di Kabupaten Bogor, tingkat risiko tinggi, sedang, dan rendah dapat dianalisis melalui sejumlah faktor seperti kepadatan penduduk, jumlah kasus positif, serta jumlah kematian akibat *COVID-19* [11]. Dalam penelitian yang lain juga menggunakan teknik clustering serupa, algoritma *K-Means Clustering* diterapkan untuk mengelompokkan data pengguna narkoba, dengan memanfaatkan metode *elbow* guna menentukan jumlah klaster yang paling optimal [12].

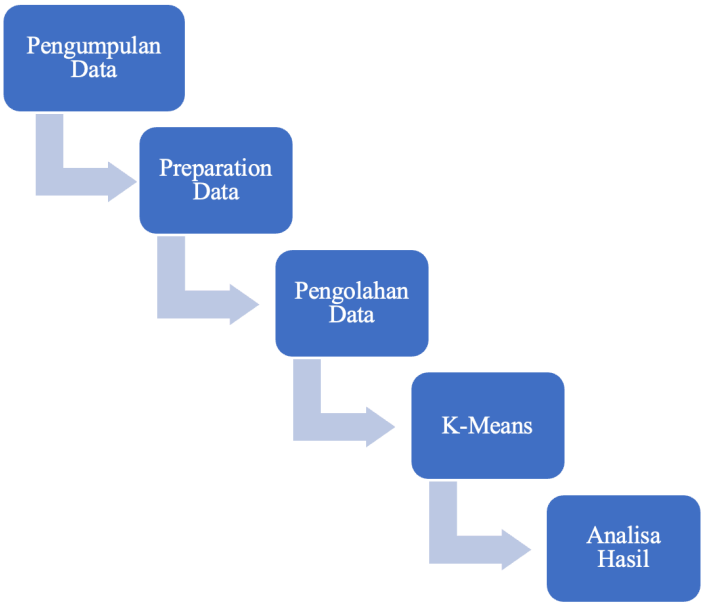
Berdasarkan permasalahan lonjakan penumpang di halte Transjakarta, diperlukan pendekatan yang tepat untuk mengenali pola keramaian penumpang di halte tertentu. Salah satu

Journal Software, Hardware and Information Technology SHiFT
ISSN 2776-8961 (online)
2808-3385 (print)
Vol. 5 No. 2 Tahun 2025
doi: https://doi.org/10.24252/shift.v5i2.205

metode yang dapat digunakan yaitu *K-Means Clustering*, Metode ini dapat mengelompokkan data yang mirip ke dalam satu kelompok, sehingga kita bisa melihat pola keramaian dan menentukan kapan serta di mana layanan perlu ditingkatkan.

2. METODE PENELITIAN/ALGORITMA

Tahapan penelitian merupakan sebuah metode yang digunakan untuk mendeskripsikan langkah-langkah yang dilakukan mulai dari identifikasi masalah sampai mendapatkan sebuah solusi dari permasalahan yang ada [13]. Metodologi penelitian ini mengadaptasi sebagian tahapan dari *CRISP-DM*, dengan fokus pada tahapan pemrosesan data, modeling menggunakan *K-Means Clustering*, dan evaluasi hasil. Tahapan *business understanding* dan *deployment* tidak dibahas karena ruang lingkup penelitian ini hanya pada aspek teknis. Tahapan penelitian yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Proses ini mencakup pengambilan data dari sumber eksternal, yaitu *Kaggle*, yang menyediakan dataset Pengguna Transjakarta. Pengambilan data dilakukan mengikuti tahapan metodologi penelitian untuk memperoleh informasi yang diperlukan [1]. data tersebut dibuat dari tahun 2023 yang berfokus kepada data penumpang Transjakarta di Jabodetabek.

2.2 Preparation Data

Pada tahap ini dilakukan penyiapan data yang akan digunakan dalam analisis klasterisasi. Proses *preparation data* dalam penelitian ini berfokus pada pembersihan data, yang bertujuan untuk mengatasi nilai kosong atau *missing value*. Pembersihan ini dilakukan untuk menjaga kualitas data dan validitas hasil analisis, sehingga dapat meningkatkan ketepatan dalam pengelompokan menggunakan algoritma *K-Means Clustering* [14].

2.3. Pengolahan Data

Data yang telah diperoleh pada tahap pengumpulan dan pembersihan, selanjutnya diolah ke dalam algoritma *K-Means Clustering* sebagai bagian dari proses analisis data [13].

Journal Software, Hardware and Information Technology SHiFT
ISSN 2776-8961 (online)
2808-3385 (print)
Vol. 5 No. 2 Tahun 2025
doi: https://doi.org/10.24252/shift.v5i2.205

2.4. K-Means

Langkah berikutnya adalah menerapkan model machine learning dengan algoritma *K-Means Clustering*, yang termasuk dalam jenis *unsupervised learning*. *K-Means Clustering* merupakan salah satu metode *nonhierarki* untuk pengelompokan data, yang memungkinkan data dipartisi menjadi dua kelompok atau lebih. Metode ini akan memisahkan data ke dalam kelompok-kelompok, di mana data dengan karakteristik yang serupa dikelompokkan bersama, sedangkan data dengan karakteristik yang berbeda akan dikelompokkan kedalam kelompok lain yang serupa[15]. Hasil dievaluasi menggunakan metode *elbow* guna menentukan jumlah klaster yang paling sesuai atau optimal [16].

2.5. Analisa Hasil

Analisis dilakukan terhadap hasil yang diperoleh dari data yang telah melalui proses pengelompokan menggunakan algoritma *K-Means Clustering*. Tujuan dari analisis ini adalah untuk mengidentifikasi pola atau kecenderungan tertentu yang berkaitan dengan permasalahan pada sistem transportasi umum di Transjakarta. Dengan memahami hasil pengelompokan tersebut, diharapkan dapat diperoleh wawasan yang lebih mendalam mengenai distribusi permasalahan, sehingga dapat menjadi dasar pertimbangan dalam perencanaan atau pengambilan keputusan yang lebih efektif.

3. HASIL PENELITIAN DAN PEMBAHASAN

Data yang diolah dalam penelitian ini merupakan data pembayaran penumpang Transjakarta melalui pembayaran data digital, total dataset tersebut berjumlah 189.501, adapun atribut yang dimiliki yaitu transID, payCard, payCardBank, payCardName, payCardsex, payCardBirthdate, corridorID, corridorName, direction, tapInStopsName, tapInStopsLat, tapInStopLon, stopStartSeq, tapOutStops, tapOutStopsName, tapOutStopsLat, stopEndSeq, tapOutTime, payAmount. Contoh data Transjakarta yang digunakan dapat dilihat pada Gambar 2.

	transID	payCardID	payCardBank	payCardName	payCardSex	payCardBirthDate	corridorID	corridorName
0	VRPJ892P3M98RA	3561407960318444	dki	Dr. Janet Nashiruddin, M.Ak	M	2010	4	Pulo Gadung 2 - Tosari
1	ZWCH834I6M26HS	347728053419394	emoney	Balamantri Rahayu	M	2002	JAK.28	Kp. Rambutan - Taman Wiladatika
2	YRLD835V6L82GO	377105453850671	emoney	Dian Mustofa	F	1993	B13	Bekasi Barat - Blok M
3	ZZBX143N6N83HQ	4486493302356581	dki	Cut Janet Suryatmi, M.Ak	M	1980	8K	Batusari - Grogol
4	EWEG491A2W45DR	30139379978125	bni	dr. Mulyanto Pudjiastuti, M.Ak	F	1997	NaN	NaN
...	...	...	...	...	...	...	...	...
8859	ILEO434C6Y65NP	6011311045426843	dki	Yulia Najmudin	F	1983	JAK.48A	Stasiun Tebet - Rusun Karet Tengsin
8860	BOCU509F4I66GP	6011359814736421	dki	Latif Narpati	F	1999	JAK.12	Tanah Abang - Kebayoran Lama via Pos Pengumben
8861	UEID963M7X88QZ	346712638899859	emoney	Lala Maryati	F	1971	B14	Bekasi Barat - Kuningan

Gambar 2. Data Penumpang Transjakarta

Journal Software, Hardware and Information Technology SHiFT
ISSN 2776-8961 (online)
2808-3385 (print)
Vol. 5 No. 2 Tahun 2025
doi: https://doi.org/10.24252/shift.v5i2.205

Berdasarkan Dataset yang ada, selanjutnya data tersebut akan diolah menggunakan *Python* yang ada di *Google Colab* untuk menampilkan visualisasi. Gambar ini juga menunjukan data yang berkaitan dengan transaksi penumpang dalam sistem transportasi, mencakup beragam informasi mengenai pengguna kartu pembayaran. Kolom transID berfungsi sebagai identitas khusus untuk setiap transaksi *tap-in*. Kolom payCardID dan payCardBank menunjukan nomor kartu dan nama lembaga yang mengeluarkan kartu pembayaran, dengan contoh dki, *emoney*, atau bni. Pemilik kartu dicatat di bagian payCardName, sedangkan bagian jenis kelamin pengguna terdapat pada payCardSex dan juga tahun lahir di payCardBirthDate. Selain itu, kolom corridorID menentukan nomor koridor bus Transjakarta yang digunakan, sedangkan corridorName mengartikulasikan rute koridor, dicontohkan dengan Pulo Gadung 2 - Tosari. Dari penjelasan ini dapat memahami bahwa data tidak hanya merangkum detail transaksi, tetapi juga menunjukan data pengguna dan rute yang mereka pilih.

Jumlah Missing Values Setelah Pembersihan:

	0
transID	0
payCardID	0
payCardBank	0
payCardName	0
payCardSex	0
payCardBirthDate	0
corridorID	0
corridorName	0
direction	0
tapInStops	0
tapInStopsName	0
tapInStopsLat	0
tapInStopsLon	0
stopStartSeq	0
tapInTime	0
tapOutStops	0
tapOutStopsName	0
tapOutStopsLat	0
tapOutStopsLon	0
stopEndSeq	0
tapOutTime	0
payAmount	0

Jumlah Missing Values Sebelum Pembersihan:

	0
transID	0
payCardID	0
payCardBank	0
payCardName	0
payCardSex	0
payCardBirthDate	0
corridorID	6980
corridorName	13528
direction	0
tapInStops	7241
tapInStopsName	0
tapInStopsLat	0
tapInStopsLon	0
stopStartSeq	0
tapInTime	0
tapOutStops	12369
tapOutStopsName	6720
tapOutStopsLat	6720
tapOutStopsLon	6720
stopEndSeq	6720
tapOutTime	6720
payAmount	3718

Gambar 3. Pembersihan Data

Gambar tersebut menampilkan jumlah nilai kosong atau *missing value* pada masing-masing atribut sebelum dan sesudah dilakukan proses pembersihan data. Atribut corridorID, corridorName, tapInStops, tapOutStops, tapOutStopsName, tapOutStopsLat, tapOutStopsLon, stopsEndSeq, tapOutTime, dan payAmount sebelumnya memiliki banyak nilai kosong yang dapat mengganggu proses analisis. Setelah dilakukan pembersihan, jumlah data menjadi 166.395 baris dengan semua atribut yang sudah bersih dari *missing value*. Proses ini penting untuk memastikan kualitas data sebelum dilanjutkan ke tahap berikutnya.

Journal Software, Hardware and Information Technology SHiFT
ISSN 2776-8961 (online)
2808-3385 (print)
Vol. 5 No. 2 Tahun 2025
doi: https://doi.org/10.24252/shift.v5i2.205

Pola Penyebaran Pengguna Layanan Transjakarta (Tap-In Terbanyak):

tapInStopsName	tapInHour	tapInCount
Penjaringan	17	205
BKN	6	155
Penjaringan	6	150
Penjaringan	16	131
BNN LRT	17	118
Rusun Penjaringan	6	112
BNN LRT	6	107
Cibubur Junction	17	99
BNN LRT	16	95
Cibubur Junction	16	94
Garuda Taman Mini	17	94
Tegalan	6	91
Indosiar	6	88
Penjaringan	19	87
Penjaringan	8	86
Duren Tiga	17	85
Buperta Cibubur	6	83
Penjaringan	7	82
Penjaringan	20	80
BNN Arah Barat	6	79

Gambar 4. Tap-in berdasarkan waktu

Gambar diatas menampilkan 20 nama halte dan jam dengan jumlah *tap-in* terbanyak, memberikan gambaran yang jelas tentang titik-titik keramaian dan waktu puncaknya. Dari data tersebut, terlihat bahwa halte Penjaringan mendominasi dengan frekuensi *tap-in* yang sangat tinggi, terutama pada sore hari tepatnya pukul 17:00 dengan 205 *tap-in*, serta pagi hari pada pukul 06:00 dengan 150 *tap-in*. Halte penting lainnya yang menunjukkan aktivitas tinggi di jam sibuk pagi adalah BKN pukul 06.00, BNN LRT pagi hari pukul 06.00, BNN LRT sore hari pukul 17:00, serta Rusun Penjaringan pagi hari pukul 06:00. Pola ini menunjukkan bahwa halte-halte tersebut berperan sebagai titik keberangkatan atau transit utama, terutama selama jam sibuk komuter di pagi dan sore hari, mencerminkan pergerakan pengguna dari area residensial atau titik kumpul menuju pusat aktivitas atau sebaliknya. Informasi ini sangat berharga untuk memahami distribusi dan kepadatan pengguna Transjakarta di berbagai lokasi dan waktu.

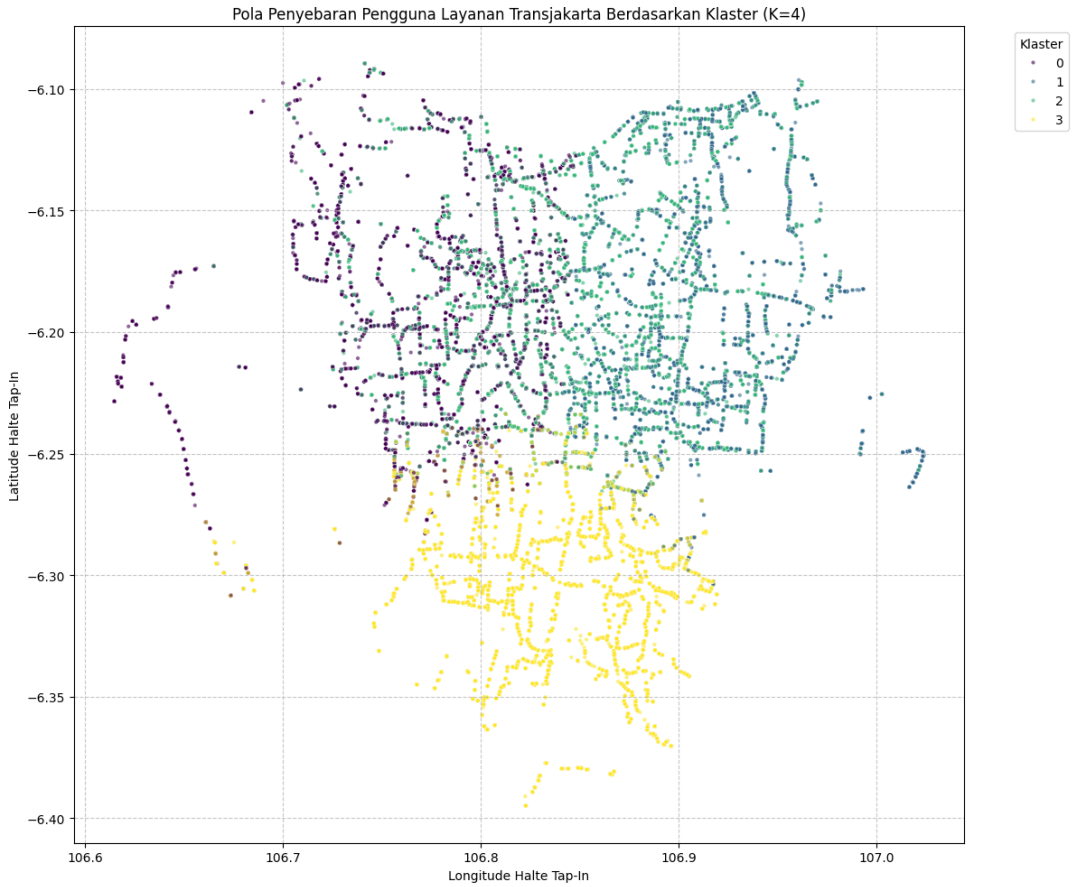
Pola Penyebaran Pengguna Layanan Transjakarta (Tap-Out Terbanyak):

tapOutStopsName	tapOutHour	tapOutCount
Penjaringan	6	201
BKN	18	175
Penjaringan	7	171
Penjaringan	18	144
BKN	7	132
BKN	6	131
Penjaringan	9	129
BKN	19	126
Penjaringan	8	124
Penjaringan	19	108
Penjaringan	10	106
Rusun Penjaringan	18	93
St. MRT Fatmawati	18	93
Puri Beta 2	7	90
Puri Beta 2	6	85
BKN	17	83
Jelambar	6	79
Summarecon Bekasi	18	77
Tirtayasa	18	77
Rusun Penjaringan	19	77

Gambar 4. Tap out berdasarkan waktu

Journal Software, Hardware and Information Technology SHiFT
ISSN 2776-8961 (online)
2808-3385 (print)
Vol. 5 No. 2 Tahun 2025
doi: https://doi.org/10.24252/shift.v5i2.205

Gambar di atas menampilkan 20 nama halte dan jam dengan jumlah *tap-out* terbanyak, memberikan gambaran yang jelas tentang titik-titik tujuan akhir dan waktu puncaknya. Dari data tersebut, terlihat bahwa halte Penjaringan juga mendominasi dengan frekuensi *tap-out* yang sangat tinggi, terutama pada pagi hari tepatnya pukul 06:00 dengan 201 *tap-out*, serta pukul 07:00 dengan 171 *tap-out*. Selain itu, halte BKN menunjukkan aktivitas *tap-out* tinggi pada sore hari pukul 18:00 dengan 175 *tap-out*, dan juga pada pagi hari pukul 06:00 dengan 131 *tap-out*. Halte penting lainnya yang menunjukkan aktivitas tinggi di jam sibuk pada sore hari yaitu St. MRT Fatmawati pukul 18:00 dan Rusun Penjaringan pukul 18:00. Pola ini menunjukkan bahwa halte-halte tersebut berperan sebagai titik tujuan utama, terutama selama jam sibuk komuter di pagi dan sore hari, mencerminkan pergerakan pengguna dari pusat aktivitas menuju area residensial atau titik kumpul lainnya. Informasi ini sangat berharga untuk memahami distribusi dan kepadatan pengguna Transjakarta di berbagai lokasi dan waktu.



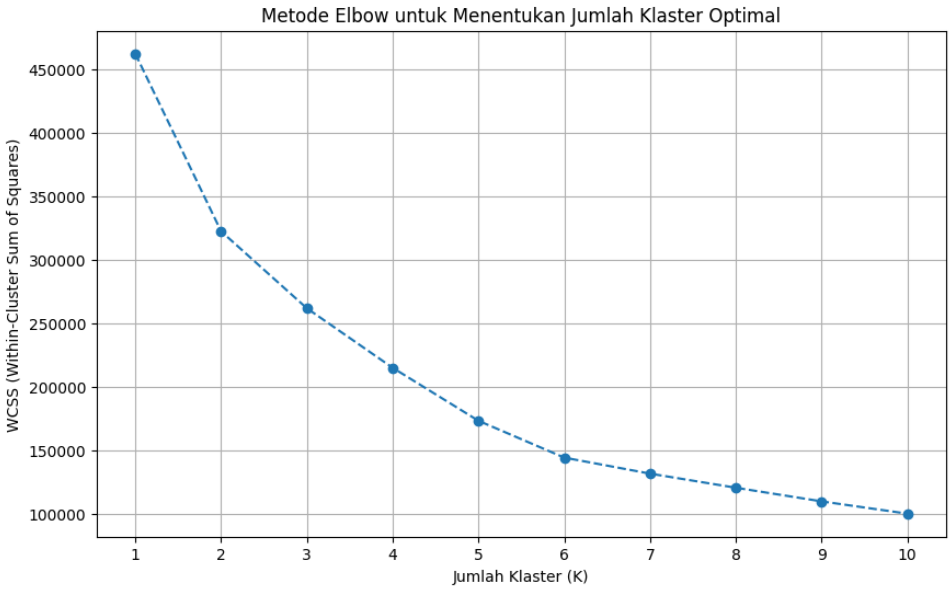
Gambar 5. Scatter Plot

Scatter plot merupakan bentuk visualisasi data yang menampilkan hubungan antara dua variabel numerik melalui pemetaan titik-titik pada sumbu horizontal (x) dan vertikal (y) [17]. Visualisasi *scatter plot* pada Gambar 5 memberikan gambaran yang jelas mengenai hasil klasterisasi pengguna layanan Transjakarta berdasarkan koordinat geografis halte *tap-in* dan waktu penggunaan (*tap-in hour*). Setiap titik pada plot mewakili satu transaksi penumpang, dengan warna berbeda menunjukkan klaster hasil pengelompokan *K-Means Clustering* sebanyak empat klaster. Visualisasi ini tidak hanya menampilkan distribusi spasial pengguna, tetapi juga mengungkapkan pola mobilitas yang terkonsentrasi di wilayah-wilayah tertentu seperti Jakarta Barat dan Jakarta Pusat. Implikasi dari hasil klasterisasi ini sangat penting bagi pengelola layanan

Journal SHiFT Software, Hardware and Information Technology
ISSN 2776-8961 (online)
2808-3385 (print)
Vol. 5 No. 2 Tahun 2025
doi: https://doi.org/10.24252/shift.v5i2.205

Transjakarta, karena dapat digunakan sebagai dasar untuk mengidentifikasi area dengan tingkat permintaan tinggi, serta membantu dalam pengambilan keputusan seperti penambahan armada, penyesuaian jadwal, atau perbaikan infrastruktur di halte-halte strategis.

Jika dibandingkan dengan penelitian serupa yang dilakukan oleh Adi Supriyatna [18], terdapat perbedaan signifikan dalam pendekatan dan hasil yang diperoleh. Penelitian Adi Supriyatna berfokus pada pengelompokan koridor Transjakarta berdasarkan total jumlah penumpang per tahun dengan menggunakan algoritma *K-Means Clustering* dan menghasilkan tiga klaster yaitu tinggi, sedang, dan rendah. Fitur yang digunakan terbatas pada data kuantitatif tahunan per koridor, serta dilakukan menggunakan *Microsoft Excel* dan *RapidMiner*. Sementara itu, penelitian ini menggunakan data transaksi individual yang jauh lebih granular, serta mempertimbangkan dimensi spasial dan temporal dalam pembentukan klaster. Dengan pendekatan tersebut, hasil klasterisasi dalam penelitian ini memiliki ketajaman analisis yang lebih tinggi karena mampu menunjukkan pola keramaian pengguna secara langsung pada lokasi dan waktu tertentu. Oleh karena itu, penelitian ini memberikan kontribusi lebih lanjut dalam mendukung perencanaan layanan transportasi publik berbasis data *real-time* dan geospasial.



Gambar 6. Grafik Elbow

Metode *elbow* adalah metode yang digunakan untuk menentukan jumlah klaster yang optimal dalam analisis klaster dengan algoritma *K-Means Clustering* [19]. Gambar 6 menunjukkan grafik *elbow* yang digunakan untuk menentukan jumlah klaster optimal dalam proses klasterisasi menggunakan algoritma *K-Means Clustering*. Dari visualisasi tersebut, terlihat bahwa titik siku (*elbow*) muncul pada K = 4, yang menandakan bahwa penggunaan empat klaster memberikan keseimbangan terbaik antara kompleksitas model dan kualitas klasterisasi. Nilai inersia menurun tajam hingga titik K = 4 dan kemudian mulai melandai, menunjukkan bahwa penambahan klaster setelah titik tersebut hanya memberikan sedikit peningkatan kualitas dan cenderung mengarah ke *overfitting*.

Interpretasi dari hasil ini menunjukkan bahwa empat klaster cukup mampu merepresentasikan variasi pola penggunaan layanan Transjakarta berdasarkan fitur spasial dan temporal. Pembentukan empat klaster ini membantu menyederhanakan data pengguna ke dalam

ISSN 2776-8961 (online)
2808-3385 (print)
Vol. 5 No. 2 Tahun 2025
doi: https://doi.org/10.24252/shift.v5i2.205

kelompok-kelompok dengan karakteristik yang homogen secara geografis dan waktu, sehingga sangat berguna untuk pengambilan keputusan operasional.

Jika dibandingkan dengan hasil penelitian oleh Saputra [1], yang juga menggunakan algoritma *K-Means Clustering* namun dengan pendekatan berbasis *RFM (Recency, Frequency, Monetary)*, terdapat beberapa perbedaan signifikan. Dalam penelitian tersebut, jumlah klaster optimal juga ditemukan pada K = 4, namun data yang digunakan bersifat transaksi individual pelanggan dan berfokus pada perilaku pelanggan dari sisi bisnis, seperti frekuensi dan nilai pengeluaran. Penelitiannya juga menggunakan metrik evaluasi tambahan seperti *Silhouette Score* dan *Davies-Bouldin Index*, di mana *K-Means Clustering* menunjukkan performa lebih baik dibandingkan *DBSCAN* dalam membedakan karakteristik pelanggan. Sementara itu, dalam penelitian ini, pemilihan jumlah klaster tidak hanya didasarkan pada perubahan inersia, tetapi juga dipadukan dengan analisis visualisasi *scatter plot* dan distribusi geografis pengguna layanan.

Perbedaan mendasar terletak pada fokus analisis, penelitian Saputa lebih menekankan pada segmentasi pelanggan berbasis perilaku finansial, sementara penelitian ini lebih menekankan pada analisis spasial dan temporal dalam konteks layanan publik. Dengan demikian, temuan dalam penelitian ini memberikan kontribusi yang lebih aplikatif dalam konteks operasional transportasi massal dan pengelolaan rute serta jadwal layanan.

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis pola penyebaran spasial dan temporal pengguna layanan Transjakarta menggunakan algoritma *K-Means Clustering*. Dengan memanfaatkan 189.501 data transaksi penumpang dari platform *Kaggle*, penelitian ini berhasil mengidentifikasi karakteristik mobilitas pengguna berdasarkan lokasi *tap-in* dan waktu perjalanan. Hasil visualisasi *scatter plot* menunjukkan adanya konsentrasi aktivitas penumpang di beberapa titik utama, seperti wilayah Jakarta Barat dan Jakarta Pusat, khususnya pada jam-jam sibuk pagi dan sore hari. Sementara itu, analisis grafik *elbow* menunjukkan bahwa jumlah klaster optimal adalah empat (K=4), yang menggambarkan segmentasi pengguna berdasarkan kesamaan pola perjalanan.

Temuan ini memberikan implikasi strategis bagi pengelola layanan Transjakarta, khususnya dalam perencanaan operasional, seperti penyesuaian jumlah armada, waktu layanan, dan pengembangan fasilitas halte pada area dengan kepadatan tinggi. Selain itu, dibandingkan dengan penelitian sebelumnya, pendekatan berbasis spasial dan temporal dalam jurnal ini menghasilkan segmentasi yang lebih presisi karena menggunakan data transaksi individual secara real-time. Untuk kedepannya, metode klasterisasi ini dapat dikembangkan lebih lanjut dengan mengintegrasikan algoritma prediksi serta variabel eksternal seperti cuaca, hari libur, atau event khusus guna mendukung pengambilan keputusan yang lebih adaptif dan berbasis data.

DAFTAR PUSTAKA

[1] A. Saputra and R. Yusuf, “Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1346–1361, 2024, doi: 10.57152/malcom.v4i4.1516.

[2] I. Iwandini, A. Triayudi, and G. Soepriyono, “Analisa Sentimen Pengguna Transportasi Jakarta Terhadap Transjakarta Menggunakan Metode Naives Bayes dan K-Nearest Neighbor,” *J. Inf. Syst. Res.*, vol. 4, no. 2, pp. 543–550, 2023, doi: 10.47065/josh.v4i2.2937.

[3] A. Purnama, F. Fauziah, and N. D. Nathasia, "Smart Counter Pada Kapasitas Bus Transjakarta Menggunakan Sensor Infrared Berbasis Arduino Uno Atmega328," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 7, no. 1, pp. 175–185, 2022, doi: 10.29100/jipi.v7i1.2623.

[4] M. Fathur, "BUS TRANSJAKARTA DI WILAYAH JAKARTA," vol. 7, no. Sukarto 2006, pp. 7612–7617, 2024.

[5] Andri Tio Sitorus, "ANALISIS KINERJA OPERASIONAL BUS RAPID TRANSIT TRANSJAKARTA KORIDOR 9C (PINANG RANTI – BUNDARAN SENAYAN) (Transjakarta Corridor 9C Rapid Transit Bus Operational," vol. 4, no. 2, pp. 205–212, 2024.

[6] M. Hidayat, "Perancangan Dashboard Untuk Visualisasi Data Jumlah Penumpang Transjakarta," *Insantek*, vol. 4, no. 1, pp. 32–36, 2023, doi: 10.31294/insantek.v4i1.2222.

[7] Y. Rachmadina, M. Isradi, J. Prasetijo, A. K. Negara Dalimunte, and A. Mufhidin, "Analysis of the Choice of Commuter Line Electric Rail Train (Krl) Modes and Transjakarta Buses for the Bekasi City - East Jakarta Route," *Eng. Technol. J.*, vol. 8, no. 08, pp. 2655–2664, 2023, doi: 10.47191/etj/v8i8.23.

[8] D. Tri Cahaya, D. Puspita, and R. Syahri, "Penerapan Metode K-Means Clustering Untuk Pengelompokan Potensi Padi Di Kota Pagar Alam," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 2187–2193, 2024, doi: 10.36040/jati.v8i2.9432.

[9] M. Djaka Permana, A. Lia Hananto, E. Novalia, B. Huda, and T. Paryono, "Klasterisasi Data Jamaah Umrah pada Tanurmutmainah Tour Menggunakan Algoritma K-Means," *J. KomtekInfo*, vol. 10, pp. 15–20, 2023, doi: 10.35134/komtekinfo.v10i1.332.

[10] M. Bagus Herlambang and L. Theresia, "Pemetaan Kota/Kabupaten Endemis Demam Berdarah Dengue Dengan Analisis Data Science Menggunakan Algoritma Clustering," *Multimed. Jar.*, vol. 8, no. 1, 2023.

[11] M. Iqbal and M. Nurul Huda, "Implementasi Algoritma K-Means Clustering dengan Jarak Euclidean dalam Mengelompokkan Daerah Penyebaran COVID-19 di Kabupaten Bogor," vol. 2, no. 1, pp. 47–56, 2023, [Online]. Available: http://jurnal.fmipa.unmul.ac.id/index.php/basis

[12] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.

[13] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 20–28, 2023, doi: 10.58602/jaiti.v1i1.22.

[14] J. Elektronika and D. A. N. Komputer, "Mengoptimalkan Proses Pembersihan Data dalam Analisis Big Data Menggunakan Pipeline Berbasis AI," vol. 17, no. 2, pp. 657–666, 2024.

[15] T. B. Pamungkas, S. Maesaroh, and P. Ardiansyah, "Implementasi Data Mining Pada Stok Penggunaan Barang Di Gmf Aeroasia Menggunakan Algoritma K-Means Clustering," *J. Ilm. Sains dan Teknol.*, vol. 7, no. 2, pp. 112–123, 2023, doi: 10.47080/saintek.v7i2.2697.

[16] A. Rajsya and A. Rachman, "Rancang Bangun Penerapan Metode Elbow Pada K-Means Untuk Clustering Data Persediaan Barang," vol. 1, no. 4, pp. 395–403, 2024.

ISSN 2776-8961 (online)
2808-3385 (print)
Vol. 5 No. 2 Tahun 2025
doi: https://doi.org/10.24252/shift.v5i2.205

[17] G. Joelio, L. Nugrah, and M. Saputra, "Penerapan Metode Service Quality Dalam Analisis Persepsi Konsumen Pada Pelayanan Pemesanan Barang E-Commerce Di Universitas Prima Indonesia," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 40–46, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2311.

[18] A. Supriyatna, I. Carolina, S. Janti, and A. Haidir, "Clustering Koridor Transjakarta Berdasarkan Jumlah Penumpang Dengan Algoritma K-Means," *J. Sains Komput. Inform. (J-SAKTI*, vol. 4, no. 2, pp. 682–693, 2020.

[19] B. G. Aji, D. C. A. Sondawa, M. R. Gifari, and S. Wijayanto, "Penerapan Algoritma K-Means Untuk Clustering Harga Rumah Di Bandung," *J. Ilm. Inform. Glob.*, vol. 14, no. 2, pp. 17–23, 2023, doi: 10.36982/jiig.v14i2.3189.