

**ANALISIS POLA PERJALANAN PENUMPANG
TRANSJAKARTA MENGGUNAKAN ALGORITMA
K-MEANS DAN APRIORI**



SKRIPSI

Hotmauli Kristiani Siboro

00000063975

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**ANALISIS POLA PERJALANAN PENUMPANG
TRANSJAKARTA MENGGUNAKAN ALGORITMA
K-MEANS DAN APRIORI**



Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Komputer (S.Kom.)

Hotmauli Kristiani Siboro

00000063975

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Hotmauli Kristiani Siboro

Nomor Induk Mahasiswa : 00000063975

Program Studi : Sistem Informasi

Skripsi dengan judul:

ANALISIS POLA PERJALANAN PENUMPANG TRANSJAKARTA MENGUNAKAN ALGORITMA K-MEANS DAN APRIORI

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 4 Juni 2025



Hotmauli Kristiani Siboro

HALAMAN PENGESAHAN

Skripsi dengan judul

ANALISIS POLA PERJALANAN PENUMPANG TRANSJAKARTA MENGUNAKAN ALGORITMA K-MEANS DAN APRIORI

Oleh

Nama : Hotmauli Kristiani Siboro
NIM : 00000063975
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Rabu, 11 Juni 2025
Pukul 10.00 s.d 12.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut.

Ketua Sidang



24/6
JW

Jansen Wiratama, S.Kom., M.Kom.
0409019301/074887

Penguji



Dr. Erick Fernando, S.Kom., M.S.I.
1029118501/083109

Pembimbing



24/6 '25

Monika Evelin Johan, S.Kom., M.M.S.I.
0327059501/071381

Ketua Program Studi Sistem Informasi



25/6 '25

Ririn Ikana Desanti, S.Kom., M.Kom.
0313058001/051314

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Hotmauli Kristiani Siboro

NIM : 00000063975

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah : Analisis Pola Perjalanan Penumpang Transjakarta Menggunakan Algoritma K-Means Dan Apriori

Menyatakan dengan sesungguhnya bahwa saya bersedia* (pilih salah satu):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 4 Juni 2025

(Hotmauli Kristiani Siboro)

* Pilih salah satu

** Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

Dengan penuh rasa syukur kepada Tuhan Yang Maha Esa, penulis dapat menyelesaikan penelitian ini dengan baik. Ucapan terima kasih yang tulus juga penulis sampaikan kepada dosen pembimbing, keluarga, dan teman-teman atas segala dukungan, dan semangat yang telah diberikan selama proses penyusunan penelitian ini.

Mengucapkan terima kasih kepada:

1. Bapak Dr. Ir. Andrey Andoko, M.Sc, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc, selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Ibu Monika Evelin Johan, S.Kom., M.M.S.I., sebagai Dosen Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
6. Teman-teman saya yang telah menemani dan mendukung saya dalam menyelesaikan tugas akhir ini.

Semoga karya ilmiah ini dapat menjadi referensi awal dalam penerapan data mining untuk menganalisis pola perjalanan pada transportasi publik. Selain itu, hasilnya juga dapat memberikan gambaran bagi penelitian selanjutnya dalam mengembangkan model analisis mobilitas penumpang yang lebih komprehensif.

Tangerang, 4 Juni 2025



Hotmauli Kristiani Siboro

ANALISIS POLA PERJALANAN PENUMPANG TRANSJAKARTA MENGGUNAKAN ALGORITMA K-MEANS DAN APRIORI

Hotmauli Kristiani Siboro

ABSTRAK

Transportasi publik memainkan peran krusial dalam mendukung mobilitas masyarakat, terutama di kota besar seperti Jakarta yang kerap menghadapi kemacetan. Transjakarta, sebagai sistem Bus Rapid Transit (BRT) terbesar di Indonesia, menjadi tulang punggung transportasi publik ibu kota. Salah satu tantangan dalam pengelolaan sistem ini adalah memahami pola perjalanan penumpang dan hubungan antarhalte yang sering digunakan secara bersamaan. Penelitian ini dilakukan untuk mengidentifikasi pola perjalanan penumpang Transjakarta dan menganalisis asosiasi antarhalte guna mendukung pengambilan keputusan dalam perencanaan transportasi yang lebih efisien.

Penelitian ini menggunakan pendekatan *framework* CRISP-DM dengan data perjalanan penumpang Transjakarta yang diperoleh dari platform Kaggle. Dua metode utama yang digunakan adalah algoritma K-Means untuk melakukan clustering penumpang berdasarkan karakteristik perjalanan, serta algoritma Apriori untuk menemukan aturan asosiasi antarhalte. Penentuan jumlah cluster pada K-Means dilakukan menggunakan metode Elbow dan Davies-Bouldin Index, sementara evaluasi hasil cluster menggunakan Calinski-Harabasz Index. Proses analisis dilakukan menggunakan bahasa pemrograman Python di lingkungan Jupyter Notebook.

Hasil penelitian menunjukkan bahwa penumpang dapat dikelompokkan ke dalam empat cluster dengan karakteristik perjalanan yang berbeda, seperti frekuensi, durasi, dan jarak tempuh. Evaluasi menghasilkan nilai Calinski-Harabasz Index sebesar 96.335 yang menunjukkan pemisahan cluster yang baik. Algoritma Apriori berhasil menemukan pola asosiasi kuat antarhalte, terutama yang melibatkan halte Penjarangan dengan confidence sempurna dan nilai lift tinggi, yang menunjukkan keterkaitan perjalanan signifikan. Meskipun penelitian ini bersifat eksploratif dan menggunakan data historis, temuan yang dihasilkan dapat memberikan insight awal dalam segmentasi penumpang serta perencanaan pengelolaan rute dan armada secara lebih strategis.

Kata kunci: Apriori, K-Means, Segmentasi Penumpang, Transjakarta

ANALYSIS OF TRANSJAKARTA PASSENGER TRAVEL PATTERNS USING K-MEANS AND APRIORI ALGORITHMS

Hotmauli Kristiani Siboro

ABSTRACT (English)

Public transportation plays a crucial role in supporting urban mobility, especially in major cities like Jakarta, which frequently faces severe traffic congestion. Transjakarta, as the largest Bus Rapid Transit (BRT) system in Indonesia, serves as the backbone of public transportation in the capital. One of the key challenges in managing such a system is understanding passenger travel patterns and identifying associations between bus stops that are frequently used together. This study aims to identify the travel patterns of Transjakarta passengers and analyze associations between stops to support more efficient transportation planning and decision-making.

This research adopts the CRISP-DM framework approach using Transjakarta passenger trip data obtained from the Kaggle platform. Two main algorithms are applied: K-Means for clustering passengers based on travel characteristics, and Apriori for discovering association rules between bus stops. The optimal number of clusters in K-Means is determined using the Elbow Method and Davies-Bouldin Index, while the clustering results are evaluated using the Calinski-Harabasz Index. The entire analysis is conducted using the Python programming language in a Jupyter Notebook environment.

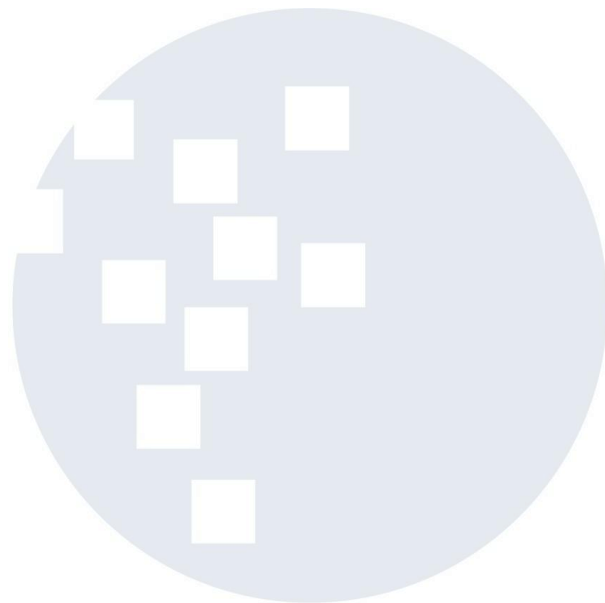
The results show that passengers can be grouped into four distinct clusters based on characteristics such as trip frequency, duration, and distance. The clustering evaluation yielded a Calinski-Harabasz Index score of 96.335, indicating a well-separated clustering structure. The Apriori algorithm successfully uncovered strong association patterns between certain stops—particularly involving the Penjaringan stop—with perfect confidence and high lift values, indicating significant travel correlations. Although this research is exploratory and based on historical data, the findings offer valuable initial insights into passenger segmentation and strategic planning for route and fleet management.

Keywords: *Apriori, K-Means, Passenger Segmentation, Transjakarta*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
KATA PENGANTAR.....	v
ABSTRAK	vi
<i>ABSTRACT (English)</i>	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR RUMUS	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah	5
1.4 Tujuan dan Manfaat Penelitian	6
1.4.1 Tujuan Penelitian.....	6
1.4.2 Manfaat Penelitian	6
1.5 Sistematika Penulisan	7
BAB II LANDASAN TEORI	8
2.1 Penelitian Terdahulu	8
2.2 Teori tentang Topik Skripsi	15
2.2.1 Transportasi Publik.....	15
2.2.2 Transjakarta.....	16
2.3 Teori tentang Algoritma yang digunakan	17
4.3.1 Data Mining	17
4.3.2 K-Means	18
4.3.3 Elbow Method	19
4.3.4 Davies-Bouldin Index.....	20
4.3.5 Calinski-Harabasz Index	21

4.3.6	Apriori.....	21
2.4	Teori tentang tools/software yang digunakan	23
2.4.1	Jupyter Notebook.....	23
2.4.2	Python	24
BAB III	METODOLOGI PENELITIAN	25
3.1	Gambaran Umum Objek Penelitian.....	25
3.2	Metode Penelitian.....	25
3.2.1	Alur Penelitian.....	25
3.2.2	Metode Data Mining	26
3.3	Teknik Pengumpulan Data	33
3.4	Variabel Penelitian.....	33
3.4.1	Variabel Independen.....	33
3.4.1	Variabel Dependen.....	34
3.5	Teknik Analisis Data.....	34
BAB IV	ANALISIS DAN HASIL PENELITIAN	35
4.1	Business Understanding	35
4.2	Data Understanding	36
4.3	Data Preparation	44
4.3.1	Data Cleaning	44
4.3.2	Data Type Conversion	45
4.3.3	Feature Engineering.....	46
4.3.4	Data Encoding	48
4.4	Modeling	50
4.4.1	Pre-processing Data untuk K-Means	50
4.4.2	Penerapan K-Means	52
4.4.3	Penerapan Apriori	59
4.5	Evaluation	64
4.6	Hasil dan Diskusi.....	66
BAB V	SIMPULAN DAN SARAN	70
5.1	Simpulan.....	70
5.2	Saran.....	71
DAFTAR PUSTAKA		72

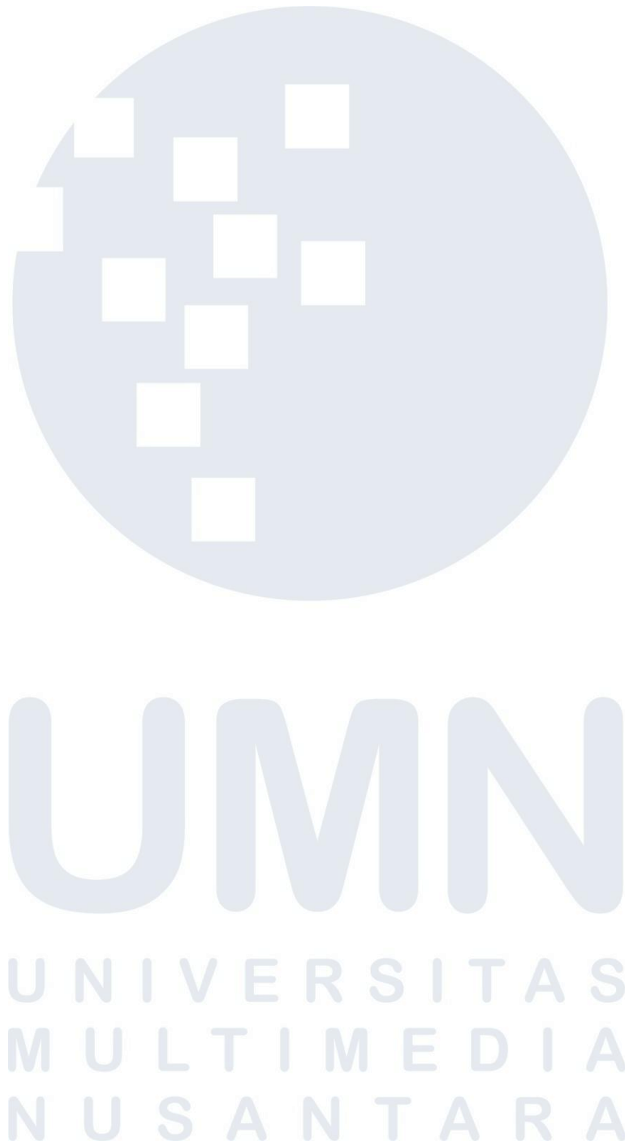


UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	8
Tabel 3.1 Perbandingan Metode Data Mining	26
Tabel 4.1 Informasi Dataset Perjalanan Penumpang Transjakarta	37
Tabel 4.2 Nilai Rata-Rata <i>Cluster</i>	58



DAFTAR GAMBAR

Gambar 1.1 Persentase Penggunaan Transportasi Publik di Jakarta pada Tahun 2023	2
Gambar 1.2 Jumlah Penumpang Transjakarta pada Tahun 2023 dan 2024	3
Gambar 3.1 Alur Penelitian CRISP DM	28
Gambar 4.1 Dataset Perjalanan Penumpang Transjakarta	36
Gambar 4.2 Informasi Dataset Perjalanan Penumpang Transjakarta	37
Gambar 4.3 Top 10 Koridor dengan Perjalanan Terbanyak	38
Gambar 4.4 Top 10 Halte Tap In	39
Gambar 4.5 Distribusi Waktu Tap In	40
Gambar 4.6 Heatmap Distribusi Waktu Tap In pada 10 Halte Tap In Terbanyak	41
Gambar 4.7 Top 10 Halte Tap Out	42
Gambar 4.8 Distribusi Waktu Tap Out	43
Gambar 4.9 Heatmap Distribusi Waktu Tap Out pada 10 Halte Tap Out Terbanyak	44
Gambar 4.10 Hasil Proses Data Cleaning	45
Gambar 4.11 Hasil Proses Konversi Tipe Data	46
Gambar 4.12 Membuat Kolom tripDuration	46
Gambar 4.13 Membuat Kolom timeOfDay	47
Gambar 4.14 One-Hot Encoding Kolom timeOfDay	47
Gambar 4.15 Membuat Kolom countTrips	47
Gambar 4.16 Membuat Kolom tripDistance	48
Gambar 4.17 Tampilan Fitur-Fitur Baru pada Dataset	48
Gambar 4.18 Mengubah Kolom Kategorik Menjadi Numerik	49
Gambar 4.19 Mengubah Kolom Boolean Menjadi 0 Dan 1	49
Gambar 4.20 Hasil Proses Data Encoding	50
Gambar 4.21 Memilih Fitur untuk <i>Clustering</i>	51
Gambar 4.22 Proses Normalisasi Data	52
Gambar 4.23 Menentukan Jumlah <i>Cluster</i> dengan <i>Elbow Method</i>	53
Gambar 4.24 Evaluasi Jumlah <i>Cluster</i> dengan DBI	54
Gambar 4.25 Perbandingan antara <i>Elbow Method</i> dan <i>Davies-Bouldin Index</i> (DBI)	55
Gambar 4.26 Penerapan K-Means dengan $k = 4$	56
Gambar 4.27 Visualisasi Hasil <i>Clustering</i>	59
Gambar 4.28 Mengembalikan Nilai Numerik Menjadi Nilai Kategorik	60
Gambar 4.29 Pembuatan Fitur Route dan Pengelompokan Transaksi Harian Berdasarkan Kartu dan Tanggal	61
Gambar 4.30 Proses Encoding Transaksi dengan TransactionEncoder	61
Gambar 4.31 <i>Frequent Route Itemsets</i>	62
Gambar 4.32 <i>Association Rules</i> Antar Halte Awal	63
Gambar 4.33 Visualisasi <i>Association Rules</i> Antar Halte Awal	63
Gambar 4.34 Evaluasi K-Means Menggunakan CH Index	64

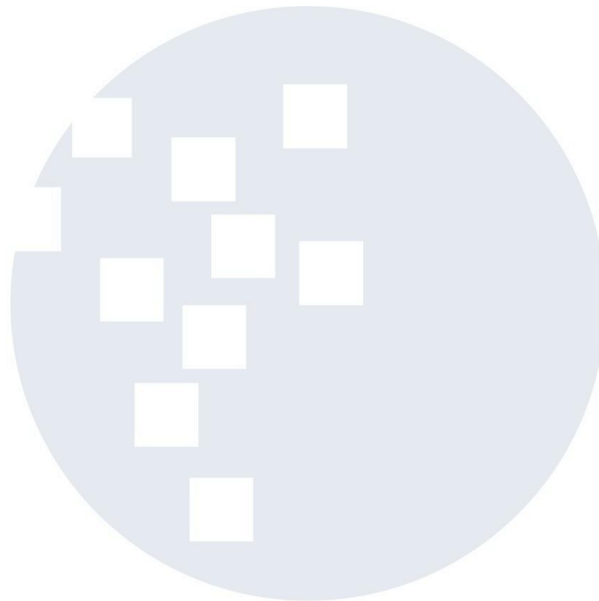
DAFTAR RUMUS

Rumus 2.1 Jarak <i>Euclidean</i> antara dua titik dalam ruang dua dimensi	18
Rumus 2.2 Jarak <i>Euclidean</i> antara dua titik dalam ruang berdimensi n	19
Rumus 2.3 Menghitung <i>centroid</i> suatu <i>cluster</i>	19
Rumus 2.4 Mengitung nilai <i>Within-Cluster Sum of Squares</i> (WCSS).....	20
Rumus 2.5 Menghitung nilai <i>Davies-Bouldin Index</i> (DBI)	20
Rumus 2.6 Menghitung nilai <i>Calinski-Harabasz Index</i>	21
Rumus 2.7 Menghitung nilai <i>support</i> sebuah item	22
Rumus 2.8 Menghitung nilai <i>support</i> dua item.....	22
Rumus 2.9 Menghitung nilai <i>confidence</i>	22
Rumus 2.10 Menghitung nilai lift.....	22



DAFTAR LAMPIRAN

Lampiran A Turnitin Similarity Report	78
Lampiran B Form Konsultasi Bimbingan.....	87



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA