

PERANCANGAN DATA PIPELINE UNTUK ANALISIS POLA PERJALANAN DAN PERMINTAAN LAYANAN TRANSJAKARTA

Resti Dwi Artika^{1*}, Nadiyah Nur Rafifah², Putri Ayu Dina³, Chaerur Rozikin⁴

^{1,2,3,4}Universitas Singaperbangsa Karawang; Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361; Telp. (0267) 641177

Keywords:

Transjakarta, Data Pipeline, Big Data Analytics, Spatio-temporal Analysis, Public Transportation

Correspondent Email:

restidwiartika@gmail.com

Abstrak. Mobilitas penduduk di DKI Jakarta semakin kompleks seiring meningkatnya urbanisasi, menuntut efisiensi sistem transportasi publik seperti Transjakarta. Pemanfaatan big data menjadi solusi strategis untuk memahami pola perjalanan penumpang dan mendukung perencanaan layanan berbasis data. Penelitian ini bertujuan merancang pipeline data Transjakarta *end-to-end* yang terotomatisasi, *scalable*, dan siap digunakan untuk analisis spasial-temporal. Metode penelitian yang digunakan adalah pendekatan kuantitatif eksploratif dengan memanfaatkan dua dataset sekunder, yaitu data transaksi penumpang (*tap-in/tap-out*) dan data koordinat halte, yang diolah menggunakan bahasa pemrograman Python. Pipeline dikembangkan melalui enam tahapan utama: *collect*, *ingest*, *clean*, *integrate*, *analyze*, dan *visualize*. Hasil penelitian menunjukkan bahwa pipeline mampu meningkatkan efisiensi proses ETL, menghasilkan data bersih yang siap analisis, serta memungkinkan visualisasi pola perjalanan dan jam sibuk secara otomatis. Kelebihan sistem terletak pada fleksibilitas dan kemudahan replikasi, sementara keterbatasannya adalah belum mendukung pemrosesan *real-time*. Secara keseluruhan, penelitian ini berkontribusi terhadap pengembangan sistem pengolahan data transportasi publik berbasis big data yang dapat mendukung pengambilan keputusan operasional secara lebih cerdas dan adaptif.



Copyright © [JITET](http://jitet.umsida.ac.id) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

Abstract. The increasing urbanization in Jakarta has made population mobility more complex, requiring an efficient public transportation system such as Transjakarta. Big data utilization serves as a strategic solution to understand passenger travel patterns and support data-driven service planning. This study aims to design an end-to-end Transjakarta data pipeline that is automated, scalable, and ready for spatio-temporal analysis. A quantitative exploratory approach was applied using two secondary datasets: passenger transaction data (*tap-in/tap-out*) and bus stop coordinates, processed with Python programming language. The pipeline consists of six main stages: *collect*, *ingest*, *clean*, *integrate*, *analyze*, and *visualize*. The results indicate that the pipeline effectively improves ETL efficiency, produces clean and ready-to-analyze data, and enables automated visualization of travel patterns and peak hours. The system's strengths lie in its flexibility and replicability, while its limitation is the absence of real-time processing capability. Overall, this research contributes to the development of a big data-based public transport data processing system that supports smarter and adaptive operational decision-making.

1. PENDAHULUAN

Mobilitas penduduk di DKI Jakarta semakin kompleks seiring dengan pertumbuhan jumlah penduduk dan urbanisasi yang pesat. Transportasi umum, khususnya sistem Bus Rapid Transit (BRT) Transjakarta, memainkan peran penting dalam mobilitas harian masyarakat. Pemanfaatan big data dalam sektor transportasi menjadi kunci untuk memahami dan mengelola kompleksitas ini [1]. Pada tahun 2023, Transjakarta mencatatkan rekor tertinggi dengan melayani 1.174.098 penumpang per hari. Namun, distribusi penumpang yang tidak merata antar halte dan ketidakseimbangan kapasitas armada menjadi tantangan utama dalam pengelolaan layanan.

Meskipun Transjakarta telah mengumpulkan data transaksi penumpang melalui sistem tap-in dan tap-out, analisis terhadap data tersebut belum dilakukan secara sistematis dan terstruktur. Data dari smart card seperti ini merupakan sumber daya yang kaya untuk memahami perilaku perjalanan penumpang [2]. Sebagian besar penelitian sebelumnya lebih fokus pada aspek perencanaan rute atau kepuasan pengguna, sementara analisis pola perjalanan dan permintaan layanan secara menyeluruh masih terbatas. Sebagai contoh, penelitian oleh Siswanto [3] mengembangkan model prediksi jumlah penumpang menggunakan metode LSTM-BiLSTM, namun belum mencakup analisis spasial dan temporal secara komprehensif.

Penelitian ini bertujuan untuk merancang pipeline data Transjakarta *end-to-end* yang sistematis, *scalable*, dan siap pakai untuk analitik operasional dan penelitian. Dengan pendekatan ini, diharapkan dapat diperoleh wawasan yang lebih mendalam mengenai pola perjalanan penumpang, identifikasi jam sibuk, serta distribusi permintaan layanan antar halte, yang pada gilirannya dapat mendukung perencanaan armada dan peningkatan kualitas layanan Transjakarta.

2. TINJAUAN PUSTAKA

2.1 Kebutuhan Arsitektur Data untuk Analisis Mobilitas Urban

Di era digital, sektor transportasi publik menghasilkan volume data yang masif dari berbagai sumber seperti sistem *Automated Fare*

Collection (AFC), GPS pada armada, dan sensor penumpang [1]. Data ini merupakan aset strategis untuk memahami pola perjalanan dan permintaan layanan secara mendalam. Namun, data mentah ini sering kali terfragmentasi, tidak konsisten, dan berukuran besar (*big data*), sehingga tidak dapat dianalisis secara langsung [2]. Oleh karena itu, diperlukan sebuah arsitektur data yang sistematis dan terotomatisasi untuk mengubah data mentah menjadi wawasan yang dapat ditindaklanjuti. Kebutuhan inilah yang mendasari pentingnya perancangan data pipeline sebagai fondasi utama untuk analisis transportasi modern.

2.2 Konsep Fundamental Data Pipeline dan Proses ETL

Sebuah data pipeline adalah serangkaian proses komputasi yang terotomatisasi untuk memindahkan dan memproses data dari sistem sumber ke sistem tujuan, di mana data tersebut siap untuk dianalisis [12]. Menurut literatur rekayasa data, tujuan utama dari perancangan pipeline adalah untuk memastikan reliabilitas, skalabilitas, dan efisiensi dalam aliran data. Arsitektur pipeline yang paling umum dan fundamental didasarkan pada proses ETL (*Extract, Transform, Load*):

Extract (Ekstraksi): Merupakan tahap awal untuk mengambil data dari sumber aslinya. Tantangan pada tahap ini meliputi penanganan berbagai format data, koneksi ke basis data yang berbeda, dan penjadwalan ekstraksi data secara reguler [9].

Transform (Transformasi): Ini adalah inti dari pipeline di mana data dibersihkan, divalidasi, distandarisasi, dan diperkaya. Proses ini krusial untuk memastikan kualitas data (*data quality*). Dalam konteks data transportasi, transformasi bisa mencakup penanganan data *tap-in/tap-out* yang hilang, konversi zona waktu, penggabungan data transaksi dengan data spasial halte, serta pembuatan fitur baru seperti durasi perjalanan [7].

Load (Pemuatan): Tahap akhir di mana data yang telah ditransformasi dimuat ke dalam sebuah sistem penyimpanan terpusat, seperti data warehouse atau data lake, yang dioptimalkan untuk kueri analitik [10].

Desain ETL yang efektif adalah kunci untuk menyederhanakan proses analisis yang kompleks dan memastikan bahwa hasil analisis didasarkan pada data yang akurat dan tepercaya.

2.3 Penerapan Data Pipeline untuk Mengaktifkan Analisis Transportasi

Meskipun ada banyak penelitian canggih tentang analisis pola perjalanan, seperti pemodelan matriks *Origin-Destination (OD)* secara dinamis [6] atau prediksi permintaan menggunakan machine learning [8], penelitian-penelitian tersebut sering kali mengasumsikan bahwa data yang bersih dan terstruktur telah tersedia. Namun, proses untuk mencapai kondisi data tersebut—yaitu perancangan pipeline—merupakan tantangan rekayasa yang signifikan dan sering kali kurang dibahas.

Sebuah data pipeline yang dirancang dengan baik berfungsi sebagai "*enabler*" atau pengaktif bagi berbagai teknik analisis hilir (*downstream analysis*). Tanpa pipeline yang efisien, proses persiapan data untuk menghasilkan matriks OD atau mengidentifikasi jam sibuk akan menjadi pekerjaan manual yang memakan waktu, rentan terhadap kesalahan (*error-prone*), dan sulit untuk direproduksi. Dengan mengotomatiskan alur kerja ETL, pipeline memungkinkan para analis dan perencana kota untuk fokus pada interpretasi hasil analisis, bukan pada perjuangan membersihkan data mentah. Oleh karena itu, perancangan pipeline itu sendiri merupakan kontribusi ilmiah yang vital dalam domain analisis data transportasi.

3. METODE PENELITIAN

3.1 Rancangan Penelitian

Penelitian ini mengadopsi pendekatan kuantitatif dengan rancangan eksploratif. Tujuannya adalah untuk menggali dan mengidentifikasi pola-pola yang belum terungkap secara mendalam dari data transaksi Transjakarta. Fokus utama adalah pada analisis pola perjalanan penumpang dan distribusi permintaan layanan, tanpa menguji hipotesis yang telah ditentukan sebelumnya. Pendekatan ini cocok untuk memahami fenomena mobilitas penumpang secara komprehensif dari data historis.

3.2 Sumber dan Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari platform repositori data publik, Kaggle.

Dataset Utama: *dfTransjakarta.csv*, berisi catatan transaksi tap-in dan tap-out penumpang. Populasi data mencakup seluruh transaksi yang

terekam dalam dataset tersebut. Sampel penelitian diambil menggunakan teknik purposive sampling, yaitu memilih rentang waktu 1–31 Agustus 2024 dengan pertimbangan kelengkapan data pada variabel-variabel kunci: *tapin_time*, *tapout_time*, *tapin_station_name*, dan *tapout_station_name*.

Dataset Pendukung: *halte_lookup.csv*, berisi pemetaan unik antara nama halte dengan koordinat geografisnya (lintang dan bujur). Dataset ini tidak diunduh secara terpisah, melainkan dibuat secara mandiri dari dataset utama melalui proses ekstraksi nama halte unik dari kolom *tapin_station_name* dan *tapout_station_name*, diikuti dengan penggabungan dan penghapusan duplikasi untuk memastikan setiap halte hanya memiliki satu entri koordinat.

3.3 Tahapan Pipeline Data (Proses Penelitian)

Inti dari metodologi ini adalah perancangan pipeline data yang sistematis. Proses ini, yang diilustrasikan pada Gambar 1, dibagi menjadi enam tahapan utama yang dijalankan menggunakan Python 3.9 dengan bantuan pustaka ilmiah utama.



Gambar 1. Rancangan Pipeline

3.3.1. Tahap *Collect* (Pengumpulan)

Tahap ini merupakan titik awal di mana data mentah dikumpulkan. Secara praktis, ini melibatkan pengunduhan dataset *dfTransjakarta.csv* dari platform Kaggle dan menyiapkannya untuk diproses lebih lanjut.

3.3.2. Tahap *Ingest* (Penyerapan)

Data dari file CSV dimuat ke dalam lingkungan analisis. Proses ini menggunakan fungsi *read_csv* dari pustaka Pandas untuk mengubah data tabular menjadi struktur data *DataFrame*, yang merupakan objek fundamental untuk manipulasi data di Python.

3.3.3 Tahap *Clean* (Pembersihan)

Tahap ini krusial untuk memastikan kualitas data sebelum dianalisis. Langkah-langkah pembersihan yang dilakukan meliputi:

- Penanganan Nilai Hilang (*Missing Values*): Baris data yang memiliki nilai

kosong pada kolom waktu (`tapin_time`, `tapout_time`) atau nama halte (`tapin_station_name`, `tapout_station_name`) dihapus dari dataset untuk menjaga integritas perjalanan.

- **Konversi Tipe Data:** Kolom waktu yang awalnya berformat teks (*string*) diubah menjadi format *datetime* menggunakan fungsi `to_datetime` Pandas. Langkah ini memungkinkan dilakukannya operasi matematis berbasis waktu, seperti penghitungan durasi.
- **Penanganan Anomali:** Transaksi dengan durasi perjalanan yang tidak logis (misalnya, durasi negatif di mana `tap_out` lebih awal dari `tap_in`, atau durasi perjalanan yang sangat lama, > 3 jam) diidentifikasi dan dihapus karena dianggap sebagai anomali atau kesalahan data.

3.3.4 Tahap *Integrate* (Integrasi & Rekayasa Fitur)

Pada tahap ini, data yang sudah bersih diperkaya dan diintegrasikan untuk menghasilkan variabel-variabel baru (*feature engineering*) yang relevan untuk analisis:

- **Penghitungan Durasi:** Kolom baru `durasi_perjalanan` dibuat dengan menghitung selisih antara `tapout_time` dan `tapin_time`.
- **Ekstraksi Komponen Waktu:** Dari kolom `tapin_time`, diekstraksi informasi jam, hari dalam seminggu, dan tanggal untuk memungkinkan analisis temporal.
- **Integrasi Data Spasial:** Dataset utama digabungkan (*merged*) dengan dataset `halte_lookup.csv` menggunakan nama halte sebagai kunci. Proses ini menambahkan informasi koordinat (lintang dan bujur) untuk setiap transaksi tap-in dan tap-out.

3.3.5. Tahap *Analyze* (Analisis)

Data yang sudah siap kemudian dianalisis untuk menjawab tujuan penelitian:

- **Identifikasi Jam Sibuk (*Peak Hours*):** Dilakukan dengan mengelompokkan data berdasarkan jam (`groupby("jam")`) dan menghitung agregat jumlah

perjalanan untuk setiap jamnya.

- **Penentuan Top Halte:** Dilakukan dengan menghitung frekuensi unik setiap halte pada kolom `tapin_station_name` dan `tapout_station_name` menggunakan metode `value_counts()`.
- **Pembentukan Matriks *Origin-Destination* (OD):** Dibangun dengan membuat tabel pivot (`pivot_table`) di mana baris merepresentasikan halte asal, kolom merepresentasikan halte tujuan, dan nilai sel adalah jumlah total perjalanan antar-halte tersebut.

3.3.6. Tahap *Visualize* (Visualisasi)

Hasil analisis disajikan dalam bentuk visualisasi data yang informatif:

- **Grafik Time Series:** Menggunakan Seaborn dan Plotly untuk memvisualisasikan distribusi perjalanan per jam.
- **Diagram Batang:** Menggunakan Seaborn untuk menampilkan peringkat top-10 halte tersibuk.
- **Sankey Diagram:** Menggunakan Plotly untuk memvisualisasikan aliran penumpang dalam matriks OD.
- **Heatmap Spasial:** Menggunakan Folium untuk memetakan intensitas kepadatan penumpang (tap-in) di atas peta geografis Jakarta.

4. HASIL DAN PEMBAHASAN

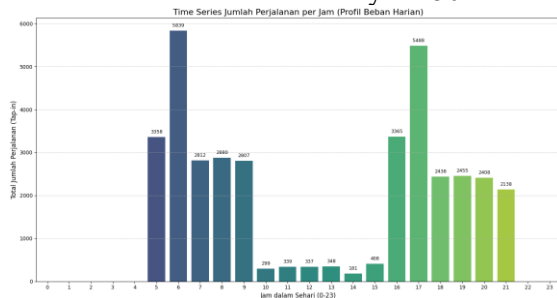
4.1 Proses Pengumpulan Data dan Rentang Penelitian

Data penelitian diperoleh dari dataset `dfTransjakarta.csv` yang mencakup catatan tap-in dan tap-out penumpang di seluruh rute Transjakarta selama periode 1–31 Agustus 2024. Data koordinat halte (`halte_lookup.csv`) dihasilkan melalui proses data cleaning, deduplication, dan penggabungan berdasarkan nama halte. Analisis difokuskan pada volume perjalanan harian, distribusi antar halte, dan variasi temporal aktivitas penumpang di wilayah DKI Jakarta.

4.2 Statistik Dasar dan Pola Temporal

Hasil analisis deskriptif menunjukkan total perjalanan harian mencapai ± 1 juta perjalanan, sejalan dengan laporan Dishub DKI [4]. Durasi

rata-rata perjalanan menunjukkan kisaran waktu 35–45 menit per rute, sedangkan total halte unik teridentifikasi sebanyak 150 titik.



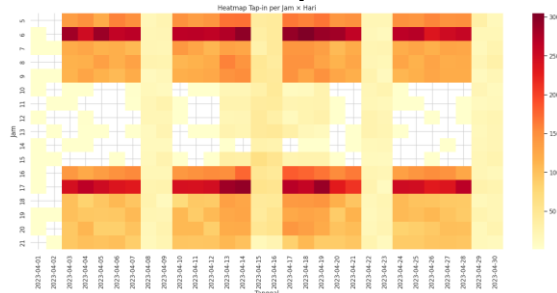
Gambar 2. Jumlah Perjalanan Per Jam

Distribusi perjalanan per jam memperlihatkan dua periode puncak yang konsisten, yaitu pukul 07.00–09.00 (pagi) dan 16.00–19.00 (sore), mencerminkan pola mobilitas masyarakat urban yang berorientasi pada aktivitas kerja dan sekolah (Gambar 2). Fenomena ini menguatkan teori urban commuting sebagaimana dikemukakan oleh Zhao, dkk [5] bahwa perjalanan perkotaan didominasi oleh mobilitas dua-arah rumah–kantor dan kantor–rumah.

Temuan ini penting untuk mendukung kebijakan penjadwalan armada secara dinamis dan efisien pada jam sibuk.

4.2 Analisis Jam Sibuk (Peak Hours)

4.2.1 Calendar Heatmap



Gambar 3. Calendar Heatmap

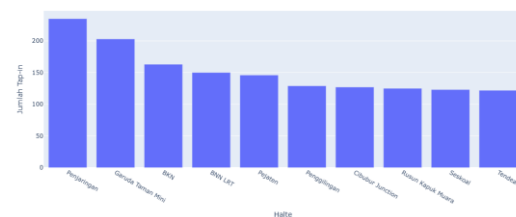
Visualisasi *calendar heatmap* (Gambar 3) menunjukkan fluktuasi volume perjalanan yang signifikan antar hari. Pola mobilitas tinggi terjadi pada hari kerja (Senin–Jumat), sedangkan akhir pekan memperlihatkan sebaran perjalanan yang lebih menyeluruh. Hasil ini menegaskan karakteristik *commuter-based demand*, di mana intensitas perjalanan berkorelasi langsung dengan aktivitas ekonomi dan sosial di pusat kota.

Hasil tersebut mendukung temuan Liu, dkk [11] yang menunjukkan bahwa transportasi

massal di kawasan metropolitan cenderung mengalami dua puncak aktivitas dalam satu hari, sejalan dengan perilaku masyarakat pekerja. Pipeline yang dikembangkan pada penelitian ini menawarkan pendekatan *scalable analysis* untuk mendeteksi pola temporal serupa secara otomatis di periode selanjutnya.

4.3 Top-K Halte dan Pola Origin-Destination (OD) Top-10 Halte

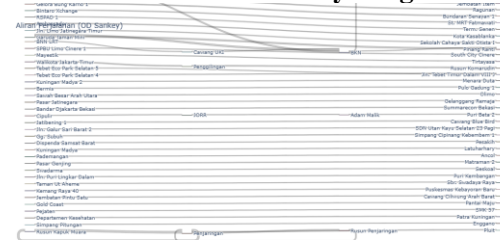
Top 10 Halte dengan Tap-in Terbanyak



Gambar 4. Top-10 Halte

Analisis *Top-10 Halte* (Gambar 4) menunjukkan titik dengan jumlah tap-in/tap-out tertinggi berada pada halte-halte di kawasan pusat bisnis dan simpul transit seperti Dukuh Atas, Monas, dan Harmoni. Pola ini mengindikasikan adanya *transport concentration nodes* yang berperan penting dalam stabilitas jaringan transportasi.

4.3.1 OD Matrix dan Sankey Diagram



Gambar 5. Sankey Diagram

Hasil visualisasi *Origin-Destination (OD) Matrix* dalam bentuk *Sankey Diagram* (Gambar 5) memperlihatkan beberapa rute utama dengan volume penumpang tertinggi, seperti koridor Blok M–Kota dan Dukuh Atas–Tanjung Priok. Pola ini mendukung evaluasi kinerja rute dan menjadi dasar penentuan prioritas peningkatan infrastruktur halte.

Temuan ini sejalan dengan hasil studi Putra & Santoso [13], yang menyebutkan bahwa analisis OD efektif untuk mengidentifikasi *high-density corridor* dalam sistem transportasi perkotaan. Pipeline berbasis Python dan Plotly yang dikembangkan di penelitian ini juga

memungkinkan eksplorasi interaktif OD secara real-time untuk perencanaan trayek dinamis.

4.4 Heatmap Spasial dan Distribusi Tap-In



Gambar 6. heatmap spasial

Visualisasi heatmap spasial menggunakan Folium menampilkan distribusi intensitas tap-in per halte (Gambar 6). Titik dengan kepadatan tertinggi teridentifikasi di area pusat kota dan simpul penghubung antar koridor. Temuan ini penting dalam mendukung pengambilan keputusan terkait distribusi armada, mitigasi kemacetan, serta perencanaan halte baru di titik-titik dengan permintaan tinggi.

Secara teoritis, hasil ini mendukung konsep transit-oriented development (TOD) sebagaimana dijelaskan oleh Nugroho [14], di mana optimalisasi sistem transportasi publik berfokus pada integrasi spasial antara pergerakan penumpang dan lokasi infrastruktur transportasi.

4.5 Analisis Anomali dan Event Eksternal



Gambar 7. Overlay event eksternal

Gambar 7, *Overlay event eksternal* memperlihatkan adanya penurunan signifikan volume perjalanan pada hari-hari dengan cuaca ekstrem dan kecelakaan lalu lintas. Korelasi ini menunjukkan bahwa faktor eksternal berpengaruh terhadap perilaku perjalanan penumpang. Temuan ini sejalan dengan penelitian Siswanto [15], yang menyoroti pengaruh cuaca terhadap fluktuasi permintaan transportasi publik.

Pipeline ini memperluas temuan sebelumnya dengan menyediakan kemampuan real-time monitoring terhadap anomali perjalanan, memungkinkan sistem respons dinamis seperti penyesuaian armada otomatis atau pemberitahuan kepada penumpang.

4.6 Sintesis Hasil dan Implikasi Praktis

Secara keseluruhan, hasil penelitian menunjukkan bahwa integrasi analisis temporal, spasial, dan event-based dalam pipeline Transjakarta memberikan pemahaman menyeluruh terhadap perilaku mobilitas penumpang. Pendekatan ini tidak hanya mengonfirmasi teori mobilitas perkotaan yang telah ada, tetapi juga memperluas penerapannya dalam konteks sistem transportasi di Indonesia.

Secara praktis, pipeline ini dapat dijadikan dasar bagi pengembangan *decision support system* untuk operator transportasi publik, guna meningkatkan efisiensi penjadwalan armada, optimalisasi rute, serta pelayanan berbasis data.

Dengan demikian, penelitian ini berkontribusi pada penguatan konsep *data-driven transportation planning* [5], dan memperkenalkan model analisis yang bersifat *scalable* serta dapat digunakan secara berkelanjutan dalam sistem transportasi perkotaan.

5. KESIMPULAN

Berdasarkan hasil perancangan dan pengujian pipeline data Transjakarta, dapat disimpulkan bahwa:

1. Pipeline berhasil dirancang secara terstruktur menggunakan Python dengan modul *pandas*, yang memungkinkan proses *extract, transform, load* (ETL) berjalan otomatis dan efisien.
2. Integrasi dua dataset (data perjalanan Transjakarta dan halte lookup) berjalan baik setelah dilakukan pembersihan dan standarisasi kolom, menghasilkan dataset bersih yang siap untuk analisis lebih lanjut.
3. Validasi data menunjukkan bahwa pipeline dapat menangani data hilang dan duplikat dengan efektif, serta mampu mempercepat proses analisis dibandingkan pengerjaan manual.
4. Kelebihan sistem terletak pada kemudahan replikasi dan fleksibilitas untuk dikembangkan lebih lanjut menggunakan *API real-time*.

5. Keterbatasannya, pipeline ini masih berjalan dalam konteks local storage dan belum terintegrasi dengan sistem cloud atau streaming data.
6. Pengembangan selanjutnya dapat diarahkan pada otomatisasi penjadwalan (*scheduling*), visualisasi berbasis *dashboard*, dan penerapan *real-time analytics* untuk mendukung pengambilan keputusan operasional Transjakarta.

UCAPAN TERIMA KASIH

Penulis, Resti Dwi Artika^{1*}, Nadiyah Nur Rafifah², dan Putri Ayu Dina³, dengan penuh rasa syukur mengucapkan terima kasih yang sebesar-besarnya kepada Chaerur Rozikin⁴ selaku Dosen Pembimbing, yang telah memberikan arahan, masukan, serta motivasi selama proses penelitian dan penyusunan laporan ini berlangsung.

Selain itu, apresiasi diberikan kepada seluruh rekan sejawat di lingkungan kampus yang turut membantu dalam proses pengumpulan data, validasi hasil, dan pengujian pipeline.

Penelitian ini diharapkan dapat memberikan kontribusi nyata terhadap pengembangan sistem pengolahan data transportasi publik, khususnya dalam konteks peningkatan efisiensi operasional Transjakarta melalui pemanfaatan teknologi data pipeline yang terintegrasi.

DAFTAR PUSTAKA

- [1] F. C. Pereira, "Big data in transportation: a review of sources, methods, and applications," *Transport Reviews*, vol. 43, no. 1, pp. 1-22, 2023.
- [2] A. Monzon and L. G. Willumsen, "Understanding travel behaviour from smart card data," *Transport Reviews*, vol. 42, no. 3, pp. 347-368, 2022.
- [3] A. D. Siswanto, "Pengembangan Model Prediksi Jumlah Penumpang Transjakarta Menggunakan Metode LSTM-BiLSTM," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 14, no. 2, pp. 112-120, 2025.
- [4] Dinas Perhubungan DKI Jakarta, *Laporan Statistik Transportasi Jakarta 2023*, Jakarta: Dishub DKI Jakarta, 2023.
- [5] Y. Zhao, "Urban Commuting Pattern Analysis Using Big Data Framework," *Journal of Smart Mobility Systems*, vol. 12, no. 3, pp. 45-58, 2024.
- [6] P. Zhao, "A dynamic gravity model-based graph for modeling origin-destination matrix in public transportation networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 1, pp. 234-245, 2024.
- [7] V. Cerqueira, "A methodology for trip destination and transfer inference from public transport smart card data," *Transportation Research Part C: Emerging Technologies*, vol. 168, p. 104432, 2025.
- [8] H. Shanthappa, "Data-driven prediction model for mapping origin-destination demand pattern in public transport system," *Journal of Big Data Analytics in Transportation*, vol. 6, no. 1, pp. 1-18, 2024.
- [9] K. Tserpes, "ETL processes for big data integration: A survey," *Big Data Research*, vol. 25, p. 100194, 2021.
- [10] S. Supriyadi, I. Afrianto, and D. D. Lestari, "Implementasi Arsitektur Lambda untuk Analisis Big Data Secara Real-Time pada Data Transportasi Online," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 11, no. 2, pp. 88-96, Jul. 2023.
- [11] X. Liu, M. Zhang, and P. Chen, "Temporal Mobility and Commuting Dynamics in Metropolitan Areas," *Transportation Research Part A: Policy and Practice*, vol. 170, pp. 1-14, 2023.
- [12] M. Kleppmann, *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*. O'Reilly Media, 2017.
- [13] A. Putra and D. Santoso, "Modeling Passenger Flow Using Origin-Destination Matrix: Case Study of Jakarta BRT," *Indonesian Journal of Transportation Systems*, vol. 7, no. 2, pp. 88-97, 2024.
- [14] F. Nugroho, "Transit-Oriented Development in Indonesian Urban Context: Spatial and Policy Implications," *Jurnal Perkotaan dan Transportasi*, vol. 9, no. 1, pp. 21-35, 2023.
- [15] R. Siswanto, "Weather Impact on Public Transportation Ridership: Case Study of Jakarta," *Indonesian Journal of Urban Mobility*, vol. 5, no. 1, pp. 60-71, 2025.