

BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknologi transportasi digital berperan penting dalam meningkatkan efektivitas, efisiensi, dan kualitas layanan transportasi umum. Melalui aplikasi digital, masyarakat dapat menikmati transportasi yang lebih efisien, nyaman, aman, dan canggih, sementara operator dan pemerintah dapat meningkatkan manajemen layanan. Sistem transportasi umum di negara-negara maju seperti Jerman dan Jepang telah membuktikan bahwa mereka dapat menawarkan kenyamanan, efisiensi, dan keterjangkauan, yang esensial bagi masyarakat untuk memfasilitasi mobilitas harian (Siti Hamidah, 2025). Namun, masalah utama yang dihadapi transportasi umum di Indonesia tetap berkaitan dengan keterbatasan infrastruktur, ketidaknyamanan, dan kurangnya minat masyarakat untuk beralih dari kendaraan pribadi ke transportasi umum. Khususnya, TransJakarta diperkenalkan pada tahun 2004 di ibu kota untuk mengatasi kemacetan lalu lintas di Jakarta akibat peningkatan signifikan kendaraan pribadi dan layanan transportasi umum tradisional yang tidak memadai (Haryono et al., 2018).

Untuk mengatasi masalah serupa, pemerintah meluncurkan beberapa program angkutan massal, termasuk Trans Jatim Bus yang diluncurkan pada tahun 2022 (Rukmana, 2025). Layanan bus ini menghubungkan wilayah dan kota-kota di Jawa Timur, dengan tujuan meningkatkan mobilitas masyarakat untuk berbagai keperluan seperti pendidikan, pekerjaan, dan perjalanan. Inisiatif ini telah meningkatkan aksesibilitas dan kemudahan transportasi di wilayah Jawa Timur. Selain itu, peningkatan angkutan umum difokuskan pada wilayah-wilayah strategis yang memiliki dampak signifikan terhadap perekonomian nasional, khususnya wilayah Gerbangkertosusila, yang meliputi kota Gresik, Bangkalan, Mojokerto, Surabaya, Sidoarjo, dan Lamongan (Widasari & Rosdiana, 2024). Untuk meningkatkan kenyamanan penumpang, aplikasi Trans Jatim Ajaib disediakan, yang memberikan akses mudah ke informasi mengenai layanan transportasi ini. Aplikasi ini mencakup

berbagai fitur seperti pelacakan bus secara real-time, jadwal keberangkatan, rincian tarif, serta informasi koridor dan rute, membantu masyarakat dalam perencanaan perjalanan.

Dibalik meningkatnya popularitas layanan TransJatim, baik aplikasi TransJatim Ajaib maupun operasional busnya, menghadapi berbagai permasalahan yang banyak dikeluhkan oleh pengguna. Berdasarkan ulasan di Google Play Store, sejumlah pengguna melaporkan kendala teknis seperti informasi jadwal yang tidak akurat, pembaruan lokasi bus secara real-time yang terlambat, serta data kapasitas penumpang yang tidak sesuai dengan kondisi aktual di lapangan. Ketidaksesuaian antara informasi di aplikasi dan situasi nyata ini sering menimbulkan kebingungan dan ketidakpuasan. Salah satu ulasan menyebutkan bahwa “antara aplikasi dan aktualnya tidak sinkron; di aplikasi tertulis masih terisi satu penumpang, namun kenyataannya bus sudah penuh dan hanya tersisa tempat berdiri”, yang menunjukkan bahwa permasalahan tidak hanya terdapat pada sistem aplikasi, tetapi juga mencerminkan kekurangan dalam manajemen layanan TransJatim secara keseluruhan (Perhubungan, 2022).

Selain itu, terdapat komentar positif mengenai kemudahan pengguna, harga yang terjangkau serta efektivitas layanan transportasi digital dalam menghubungkan berbagai wilayah. Beberapa ulasan pengguna juga menilai bahwa “aplikasi ini merupakan layanan transportasi yang canggih dan terkoneksi antar lini, murah meriah, serta sangat bermanfaat bagi masyarakat menengah ke bawah”, dengan harapan agar layanan ini terus bertransformasi menuju sistem yang lebih baik dan maju (Perhubungan, 2022). Selain di platform *Google Play Store*, masyarakat juga aktif menyampaikan pendapat, kritik, dan pengalaman mereka mengenai layanan TransJatim melalui media sosial *X* dan TikTok.

Penggunaan data dari tiga platform yang berbeda, yaitu *Google Play Store*, Media Sosial *X* dan TikTok, dinilai mampu meningkatkan validitas dan kedalaman analisis dalam penelitian berbasis opini publik. *Google Play Store* memberikan umpan balik yang lebih terarah karena berasal dari pengguna aplikasi yang memiliki pengalaman langsung, sedangkan Media Sosial *X* dan TikTok menyajikan opini yang lebih spontan, luas, dan dinamis terhadap layanan TransJatim (Ningrum et al., 2025). Kombinasi

ketiga sumber ini tidak hanya memperkaya variasi konteks ulasan, tetapi juga memungkinkan peneliti untuk menangkap beragam perspektif pengguna dari sisi fungsional maupun emosional. Integrasi data multi-platform ini sejalan dengan tren penelitian terkini di bidang analisis opini publik digital, di mana pendekatan lintas sumber digunakan untuk memperoleh gambaran sentimen yang lebih representatif dan mendalam terhadap persepsi masyarakat terhadap layanan digital publik.

Platform media sosial menjadi sumber data penting karena memungkinkan pengguna mengekspresikan opini secara spontan dan real-time, termasuk saat mengalami kendala atau memuji layanan TransJatim. Media sosial X dan TikTok yang berisi ulasan, keluhan, maupun dukungan terhadap TransJatim mencerminkan persepsi publik yang lebih luas dan beragam dibandingkan dengan ulasan di toko aplikasi. Berdasarkan data resmi dari DataReportal (Oktober 2025), media sosial X memiliki sekitar 557 juta pengguna, sedangkan TikTok memiliki sekitar 1,99 miliar pengguna aktif bulanan (DataReportal, n.d.). Jumlah pengguna yang besar ini menunjukkan bahwa kedua platform tersebut merupakan ruang digital yang aktif dan representatif untuk menelusuri opini masyarakat. Namun, tantangan muncul karena ulasan pengguna di kedua platform ini bersifat tidak terstruktur dan terdiri dari campuran sentimen positif, negatif, serta netral yang sering kali menggunakan bahasa informal atau slang regional, sehingga sulit dipahami secara menyeluruh.

Menurut data dari *Google Play Store* per November 2024, aplikasi TransJatim Ajaib telah melampaui 100.000 unduhan dan memiliki sekitar 1.123 ulasan pengguna (Perhubungan, 2022). Hal ini menunjukkan bahwa aplikasi tersebut telah menarik minat yang signifikan dari pengguna angkutan umum di Jawa Timur. Sementara di Media Sosial X dan TikTok, terdapat ratusan unggahan yang menyinggung layanan TransJatim dengan berbagai nada sentimen positif, negatif, dan netral terkait kenyamanan, ketepatan waktu, hingga interaksi petugas. Data tersebut menjadi sumber informasi potensial untuk memahami kepuasan dan harapan pengguna terhadap layanan transportasi publik digital di Jawa Timur. Selain itu, berdasarkan pengumuman resmi dari Dinas Perhubungan Jawa Timur, TransJatim berhasil melayani 1.603.554 penumpang dari Januari hingga Maret 2025 di seluruh rute operasional. Setiap tahun,

jumlah total penumpang pada 2024 mencapai 4.715.809 orang yang tersebar di lima koridor utama (Azmi, 2025). Informasi ini menunjukkan peningkatan yang signifikan dalam penggunaan transportasi umum digital di wilayah Jawa Timur. Namun, transportasi di Jawa Timur masih didominasi oleh kendaraan pribadi. Menurut data dari Badan Pusat Statistik (BPS) Provinsi Jawa Timur pada tahun 2024, jumlah kendaraan bermotor di provinsi ini telah melebihi 26 juta unit, meliputi sepeda motor, mobil, bus, dan truk (Badan Pusat Statistik, 2025).

Mengingat volume unduhan aplikasi yang signifikan dan jumlah penumpang yang sangat besar, menganalisis ulasan pengguna TransJatim, baik yang berasal dari *Google Play Store* maupun Media Sosial *X* dan TikTok, menjadi sangat relevan. Ulasan-ulasan tersebut memberikan wawasan langsung mengenai tingkat kepuasan, keluhan, serta kebutuhan pengguna terhadap layanan transportasi digital TransJatim. Data dari *Google Play Store* mencerminkan pengalaman pengguna yang telah menginstal dan menggunakan aplikasi, sementara data dari Media Sosial *X* dan TikTok menggambarkan persepsi publik yang lebih luas dan *real-time* terhadap kinerja layanan di lapangan. Dengan memanfaatkan ketiga sumber data tersebut, penelitian ini dapat menghasilkan analisis yang lebih komprehensif dalam menggambarkan opini masyarakat. Oleh karena itu, studi ini penting untuk memahami sentimen pengguna dan mengidentifikasi topik utama yang muncul dalam ulasan dengan menerapkan teknik analisis data modern seperti *Logistic Regression* untuk klasifikasi sentimen dan *Non-Negative Matrix Factorization* (NMF) untuk pemodelan topik.

Metode Logistic Regression (LR) dinilai efisien untuk analisis sentimen karena mampu memetakan data teks ke dalam kelas positif, negatif, dan netral melalui fungsi sigmoid yang menghasilkan probabilitas secara linier dan akurat (Assiroj et al., 2023). Dalam penelitian Priati Assiroj, Asep Kurnia, dan Sirojul Alam (2023) yang berjudul “*The Performance of Naïve Bayes, Support Vector Machine, and Logistic Regression on Indonesia Immigration Sentiment Analysis*” ditemukan bahwa metode LR memberikan akurasi tertinggi sebesar 77%, dibandingkan dengan Support Vector Machine (SVM) sebesar 76% dan Naïve Bayes (NB) sebesar 70% (Assiroj et al., 2023). Perbandingan tersebut menunjukkan bahwa Logistic Regression lebih unggul

dalam mengklasifikasikan polaritas sentimen karena model ini sederhana, tidak memerlukan asumsi distribusi data seperti NB, serta mampu menyesuaikan bobot fitur dengan lebih baik dibandingkan SVM yang sensitif terhadap parameter dan skala data. Dalam konteks analisis sentimen terhadap ulasan pengguna TransJatim di Google Play Store, Media Sosial X, dan TikTok, penggunaan Logistic Regression menjadi relevan karena model ini mampu memberikan hasil klasifikasi yang stabil meskipun data bersifat heterogen dan tidak terdistribusi normal.

Metode pemodelan topik seperti *Non-Negative Matrix Factorization* (NMF) digunakan untuk melengkapi analisis dengan mengekstraksi tema-tema utama dari kumpulan komentar pengguna, sehingga penelitian ini tidak hanya mengukur tingkat kepuasan tetapi juga mengidentifikasi aspek layanan yang perlu ditingkatkan berdasarkan topik yang paling sering dibicarakan. Penggunaan metode NMF dinilai lebih unggul dibandingkan dengan *Latent Dirichlet Allocation* (LDA) karena menghasilkan topik yang lebih koheren dan waktu eksekusi yang lebih singkat. Berdasarkan penelitian yang dilakukan oleh Shini George dan V. Srividhya (2021) yang berjudul “*Comparasion of LDA and NMF Topic Modeling Techniques for Restaurant Reviews*”, model NMF memperoleh nilai *Mean Coherence Score* tertinggi sebesar 0.729, sedangkan LDA hanya mencapai 0.469. Selain itu, NMF juga memiliki waktu eksekusi yang lebih efisien, menunjukkan kemampuannya dalam mengidentifikasi informasi tersembunyi dengan lebih cepat dan akurat (Srividhya & Shini, 2020). Oleh karena itu, NMF dipilih karena memberikan hasil topik yang lebih stabil, bermakna, dan relevan terhadap kumpulan data komentar pengguna.

Sebagian besar penelitian sebelumnya hanya berfokus pada salah satu pendekatan, yaitu analisis sentimen atau pemodelan topik secara terpisah, sehingga menghasilkan gambaran yang terbatas. Dengan mengombinasikan analisis sentimen menggunakan algoritma *Logistic Regression* dan pemodelan topik *Non-Negative Matrix Factorization* (NMF) terhadap ulasan pengguna yang diperoleh dari *Google Play Store*, Media Sosial X, dan TikTok, penelitian ini berupaya menilai tingkat kepuasan sekaligus mengidentifikasi tema-tema utama yang muncul dalam keluhan maupun apresiasi pengguna. Hasil dari penelitian ini diharapkan dapat menjadi dasar bagi

perumusan kebijakan transportasi umum yang lebih tepat sasaran dan berbasis data, sehingga mampu meningkatkan kualitas layanan TransJatim sesuai kebutuhan serta harapan masyarakat.

1.2 Perumusan Masalah

Pada bagian ini, peneliti akan menjelaskan permasalahan ulasan pada pengguna TransJatim, metode yang diusulkan untuk mengatasi masalah tersebut, serta pertanyaan penelitian yang akan dijawab.

1.2.1 Permasalahan

Bersumber dari latar belakang diatas, penelitian ini berfokus pada bagaimana memahami persepsi dan kepuasan masyarakat terhadap layanan transportasi digital TransJatim melalui analisis ulasan pengguna yang berasal dari tiga sumber utama, yaitu *Google Play Store*, Media Sosial X dan TikTok. Meskipun aplikasi TransJatim Ajaib telah banyak digunakan oleh masyarakat Jawa Timur sebagai sarana pendukung transportasi publik, ulasan pengguna menunjukkan berbagai keluhan seperti bug teknis, ketidakakuratan jadwal, serta keterlambatan pembaruan posisi bus secara real-time. Ulasan-ulasan tersebut bersifat tidak terstruktur, beragam, dan sering menggunakan bahasa informal, baik dalam bentuk komentar di toko aplikasi maupun cuitan di media sosial, sehingga sulit dianalisis secara manual. Kondisi ini menyebabkan banyak masukan penting dari pengguna berpotensi terabaikan dan menghambat peningkatan kualitas layanan transportasi digital.

1.2.2 Metode Usulan

Penelitian ini menggunakan algoritma *Logistic Regression* untuk analisis sentimen dan *Non-Negative Matrix Factorization* (NMF) untuk pemodelan topik. *Logistic Regression* digunakan untuk mengklasifikasikan ulasan pengguna aplikasi TransJatim Ajaib karena mampu memetakan data teks ke dalam kelas positif, negatif, dan netral melalui fungsi sigmoid yang menghasilkan probabilitas secara linier dan akurat. Sementara itu, NMF digunakan untuk mengidentifikasi tema-tema utama yang muncul dalam ulasan negatif dengan menganalisis distribusi kata yang sering muncul bersama.

1.2.3 Pertanyaan Penelitian

Dengan merujuk kepada masalah yang telah dijelaskan sebelumnya, dan metode yang diusulkan, maka didapatkan pertanyaan bahwa:

1. Bagaimana performa algoritma *Logistic Regression* dalam klasifikasi sentimen ulasan pengguna TransJatim berdasarkan hasil *Confusion Matrix*, yang mencakup nilai akurasi, presisi, *recall*, dan *F1-score*?
2. Bagaimana perbandingan hasil akurasi, presisi, *recall*, dan *F1-score* model *Logistic Regression* ketika data latih menggunakan metode SMOTE dan tanpa SMOTE ?
3. Bagaimana pengaruh variasi jumlah topik terhadap nilai coherence dan kualitas interpretasi topik pada pemodelan topik menggunakan Non-negative Matrix Factorization untuk mengidentifikasi tema utama pada ulasan berlabel negatif?
4. Bagaimana perbandingan nilai *coherence score* (c_v) antara model yang tidak menggunakan custom stopwords dan model yang menggunakan custom stopwords pada pemodelan topik menggunakan metode *Non-negative Matrix Factorization* (NMF)?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan sebelumnya, penelitian ini bertujuan untuk mengatasi permasalahan analisis ulasan pengguna TransJatim yang tidak terstruktur dan sulit diinterpretasikan secara manual. Secara spesifik, tujuan penelitian ini adalah sebagai berikut:

1. Untuk mengevaluasi performa algoritma *Logistic Regression* dalam mengklasifikasikan sentimen dengan menggunakan hasil *Confusion Matrix* yang mencakup nilai akurasi, presisi, *recall*, dan *F1-score*, guna menilai sejauh mana model mampu memprediksi sentimen pengguna secara akurat.
2. Untuk membandingkan performa model *Logistic Regression* pada kondisi data tidak seimbang (tanpa SMOTE) dan kondisi data yang telah diseimbangkan menggunakan metode SMOTE, melalui evaluasi akurasi, presisi, *recall*, dan *F1-score*.

3. Untuk mengidentifikasi topik-topik keluhan dominan yang muncul dari ulasan pengguna aplikasi TransJatim Ajaib, khususnya pada ulasan negatif, melalui pemodelan topik menggunakan *Non-Negative Matrix Factorization* (NMF), sehingga dapat memetakan tema-tema utama yang perlu diperbaiki dalam layanan transportasi digital dan menjadikan teknologi informasi sebagai *partner strategis* dalam peningkatan kualitas layanan.
4. Untuk menilai kualitas hasil pemodelan topik yang dihasilkan melalui metode *Non-Negative Matrix Factorization* (NMF) dengan menggunakan ukuran *Coherence Score*, sehingga dapat diketahui tingkat koherensi makna antar-topik serta sejauh mana topik-topik tersebut mampu merepresentasikan tema utama yang terkandung dalam kumpulan ulasan pengguna secara efektif.

1.3.2 Manfaat Penulisan

1. Bagi Penulis

Penelitian ini bermanfaat untuk memberikan pemahaman mendalam kepada penulis dalam menganalisis sentimen dan memodelkan topik ulasan pengguna aplikasi TransJatim Ajaib menggunakan algoritma *Logistic Regression* dan *Non-Negative Matrix Factorization* (NMF).

2. Bagi Pembaca

Penelitian ini bermanfaat sebagai bahan literatur dan sumber informasi mengenai tingkat kepuasan dan keluhan utama pengguna transportasi publik digital dengan menggunakan algoritma *Logistic Regression* dan *Non-Negative Matrix Factorization* (NMF), serta dapat menjadi rujukan bagi pengelola layanan maupun penelitian selanjutnya dalam upaya meningkatkan kualitas transportasi berbasis aplikasi.

1.4 Batasan Masalah

Batasan penelitian ini adalah sebagai berikut :

1. Sumber data yang digunakan pada penelitian ini dari *Goggle Play Store*, Media Sosial X dan TikTok dengan jumlah data 2.656 ulasan.
2. Data yang digunakan merupakan ulasan dari pengguna TransJatim.

3. Klasifikasi sentimen pada penelitian ini dibagi menjadi tiga yaitu, sentimen positif, negatif, dan netral.
4. Pemodelan topik hanya berfokus pada sentimen negatif.
5. Algoritma yang digunakan adalah *Logistic Regression* dan *Non-Negative Matrix Factorization* (NMF), tanpa perbandingan dengan algoritma lainnya.

1.5 Sistematika Penulisan Proposal Skripsi

Sistematika penulisan skripsi ini disusun dengan mengikuti aturan penulisan skripsi pada program studi S1 Sistem Informasi Universitas Trunodjoyo Madura. Berikut sistematika penulisan yang digunakan terdiri dari lima bab sebagai berikut.

BAB I PENDAHULUAN

Bab ini menjelaskan terkait latar belakang, permasalahan penelitian, metode usulan, pertanyaan penelitian, tujuan dan manfaat penelitian serta batasan masalah yang dilakukan untuk analisis sentimen dan pemodelan topik ulasan pengguna TransJatim, serta sistematika penulisan skripsi.

BAB II KAJIAN PUSTAKA

Bab ini mencakup teori penunjang dari berbagai literatur yang akan digunakan sebagai dasar dalam penelitian ini. Bab ini juga mencantumkan beberapa penelitian terdahulu yang relevan dengan penelitian ini.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan terkait tahapan-tahapan analisis penelitian, kerangka kerja penelitian dan teknik-teknik yang diterapkan dalam penelitian ini. Meliputi proses scraping data, labelling data, preprocessing data.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil dari penerapan metode analisis sentimen *Logistic Regression* dan pemodelan topik *Non Negative Matrix Factorization* dalam menganalisis ulasan terkait pengguna Transjatim, serta pembahasan atas temuan-temuan yang diperoleh. Analisis dilakukan mulai dari proses preprocessing data teks, perhitungan ekstraksi fitur TF UDF, pembagian data, penyeimbangan data, analisis sentimen, pemodelan topik, evaluasi hasil, hingga visualisasi.

BAB V PENUTUP

Bab ini menjelaskan kesimpulan dari hasil penelitian yang telah dilakukan terkait analisis sentimen *Logistic Regression* dan pemodelan topik *Non Negative Matrix Factorization* dalam menganalisis ulasan terkait pegguan Transjatim., serta saran untuk pengembangan dan penelitian lanjutan. Kesimpulan disusun berdasarkan hasil analisis topik yang diperoleh dari proses analisis sentimen dengan algoritma *Logistic Regression* serta pemodelan topik dengan *Non Negative Matrix Factorization*. Selain itu, bab ini juga memberikan masukan terkait potensi perbaikan metode maupun penggunaan pendekatan lain yang dapat diterapkan dalam penelitian sejenis di masa mendatang.

DAFTAR PUSTAKA

Pada referensi berisi kumpulan rujukan jurnal, prosiding dari penelitian terdahulu yang digunakan untuk menyusun skripsi ini.