

BAB I

PENDAHULUAN

1.1 Latar Belakang

TransJakarta merupakan salah satu layanan transportasi perkotaan utama di DKI Jakarta yang berperan penting dalam mendukung mobilitas masyarakat. Keberadaan TransJakarta membantu pengguna melakukan perjalanan untuk bekerja, belajar, mengakses layanan publik, maupun menjalankan aktivitas sosial dan ekonomi[1]. Dalam konteks kota besar seperti Jakarta, TransJakarta tidak hanya berfungsi sebagai sarana perpindahan, tetapi juga menjadi bagian dari sistem mobilitas perkotaan yang mendukung aksesibilitas, produktivitas, dan keterhubungan antar wilayah [2].

Kebutuhan terhadap layanan TransJakarta semakin penting seiring dengan tingginya aktivitas masyarakat di wilayah perkotaan. Pada Januari 2025, jumlah penduduk Indonesia tercatat sekitar 285 juta jiwa, dengan 59,5% penduduk tinggal di wilayah perkotaan [3]. Kondisi tersebut menunjukkan bahwa sebagian besar masyarakat berada pada kawasan dengan intensitas mobilitas tinggi. Sebagai pusat pemerintahan, bisnis, pendidikan, dan aktivitas ekonomi nasional, DKI Jakarta membutuhkan layanan transportasi yang mampu melayani pergerakan masyarakat dalam jumlah besar secara berkelanjutan [4].

DKI Jakarta merupakan wilayah dengan aktivitas harian yang padat dan tingkat kebutuhan mobilitas yang tinggi. Dalam konteks tersebut, TransJakarta menjadi salah satu layanan yang banyak digunakan masyarakat karena memiliki jaringan layanan yang mencakup berbagai wilayah Jakarta [5]. Layanan TransJakarta tidak hanya terbatas pada koridor BRT, tetapi juga mencakup layanan pengumpan dan Mikrotrans yang membantu memperluas jangkauan perjalanan pengguna. Keberadaan layanan tersebut menunjukkan bahwa TransJakarta memiliki peran penting dalam mendukung perjalanan harian masyarakat Jakarta [6].

Integrasi layanan TransJakarta juga didukung melalui ekosistem JakLingko [7]. Sistem ini membantu pengguna dalam mengakses layanan perjalanan dan

pembayaran secara lebih praktis. Melalui integrasi tersebut, pengguna dapat merencanakan perjalanan dan melakukan pembayaran dengan lebih mudah, sehingga layanan TransJakarta dapat digunakan secara lebih efisien dalam aktivitas sehari-hari. Integrasi layanan ini juga memperkuat posisi TransJakarta sebagai bagian penting dari sistem mobilitas perkotaan di Jakarta [8].

Tingginya penggunaan TransJakarta dapat dilihat dari jumlah penumpang yang terus menunjukkan skala layanan yang besar. Badan Pusat Statistik Provinsi DKI Jakarta mencatat bahwa pada September 2025 jumlah penumpang TransJakarta mencapai 36.437.848 orang [9]. Selain itu, TransJakarta melayani sekitar 156 juta pelanggan pada periode Januari hingga Mei 2025 [10]. Transjakarta juga mencatat 413 juta pelanggan sepanjang tahun 2025 [11]. Data tersebut menunjukkan bahwa TransJakarta memiliki peran besar dalam mendukung aktivitas perjalanan masyarakat Jakarta.

Tingginya intensitas penggunaan TransJakarta turut diiringi oleh tingginya intensitas perbincangan publik mengenai layanan tersebut di media sosial. Salah satu contohnya adalah insiden keributan antara seorang pramudi TransJakarta koridor 8 dengan pengemudi mobil listrik yang terjadi di Jalan Panjang, Jakarta Barat, pada awal Maret 2026. Video kejadian tersebut menyebar luas di media sosial dan memicu beragam respons dari masyarakat, mulai dari kritik terhadap sikap petugas di lapangan hingga sorotan terhadap standar pelayanan dan etika kerja pramudi TransJakarta secara umum. Pihak TransJakarta sendiri turut memberikan klarifikasi resmi atas insiden tersebut dan menyatakan akan melakukan investigasi serta pembinaan internal. Fenomena ini menunjukkan bahwa peristiwa operasional sekecil apa pun yang melibatkan TransJakarta berpotensi memicu perbincangan luas di platform seperti Twitter/X, sehingga media sosial menjadi sumber data yang relevan untuk memahami opini publik secara real-time terhadap kualitas layanan TransJakarta, baik dari sisi armada, petugas, maupun aspek operasional lainnya.

Meskipun jumlah pengguna TransJakarta cukup tinggi, kondisi tersebut belum tentu menunjukkan bahwa seluruh pengguna memiliki penilaian yang sama terhadap layanan yang diberikan [12]. Pengalaman pengguna dapat dipengaruhi

oleh berbagai faktor, seperti ketepatan waktu kedatangan bus, kepadatan penumpang, kenyamanan halte, kondisi armada, aksesibilitas rute, sistem pembayaran, keamanan, serta respons penyedia layanan terhadap pertanyaan atau keluhan pengguna. Oleh karena itu, analisis terhadap opini pengguna menjadi penting untuk melihat bagaimana layanan TransJakarta dibicarakan oleh masyarakat [13].

Kualitas layanan TransJakarta menjadi aspek penting karena berhubungan dengan kepuasan dan loyalitas pengguna. Beberapa penelitian mengenai TransJakarta menunjukkan bahwa kualitas layanan, fasilitas, kepercayaan, dan nilai yang dirasakan pengguna dapat memengaruhi kepuasan terhadap layanan. Hal tersebut menunjukkan bahwa pengalaman pengguna tidak hanya ditentukan oleh ketersediaan armada, tetapi juga oleh aspek pelayanan yang dirasakan selama menggunakan TransJakarta. Dengan demikian, opini pengguna dapat menjadi salah satu sumber informasi untuk memahami bagian layanan yang masih perlu diperhatikan [14].

Analisis terhadap opini pengguna TransJakarta dapat dilakukan melalui data media sosial. Saat ini, pengguna tidak hanya menyampaikan pengalaman melalui kanal resmi, tetapi juga melalui media sosial. Media sosial memungkinkan pengguna menyampaikan keluhan, pertanyaan, kritik, saran, maupun apresiasi secara cepat dan terbuka. Pada Januari 2025, Indonesia memiliki sekitar 212 juta pengguna internet atau setara dengan 74,6% dari total populasi, sedangkan jumlah identitas pengguna media sosial mencapai sekitar 143 juta atau setara dengan 50,2% dari total populasi [3].

Salah satu platform media sosial yang relevan untuk digunakan dalam penelitian ini adalah *Twitter/X*. Platform ini memiliki karakteristik percakapan yang singkat, terbuka, aktual, dan sering digunakan untuk membahas isu layanan secara *real-time* [3]. Selain itu, fitur seperti *mention*, *reply*, *hashtag*, dan pencarian berbasis kata kunci memungkinkan pengumpulan data berdasarkan topik tertentu atau akun resmi penyedia layanan. Pada awal tahun 2025, *Twitter/X* tercatat

memiliki sekitar 25,2 juta pengguna di Indonesia, sehingga platform ini masih relevan digunakan sebagai sumber data opini pengguna [15].

Dalam konteks TransJakarta, percakapan di *Twitter/X* dapat mencerminkan berbagai bentuk respons pengguna terhadap layanan yang digunakan. Percakapan tersebut dapat berupa keluhan terhadap keterlambatan bus, kepadatan halte, antrean, gangguan operasional, kendala pembayaran, perubahan rute, maupun apresiasi terhadap pelayanan dan informasi perjalanan [16]. Akun resmi @PT_Transjakarta juga menjadi salah satu ruang interaksi antara penyedia layanan dan pengguna, sehingga *tweet*, *mention*, dan balasan yang berkaitan dengan akun tersebut dapat dimanfaatkan sebagai sumber data penelitian [17].

Penelitian sebelumnya menunjukkan bahwa data *Twitter/X* dapat digunakan untuk menganalisis opini pengguna terhadap layanan TransJakarta. Analisis sentimen pada data *Twitter/X* dapat membantu mengidentifikasi kecenderungan opini pengguna terhadap layanan, sedangkan pemodelan topik dapat membantu menemukan isu atau tema pembahasan yang sering muncul [18]. Pendekatan berbasis media sosial ini menjadi relevan karena data dapat dikumpulkan dari percakapan publik yang terjadi secara aktual dan dapat digunakan untuk melihat respons pengguna terhadap layanan TransJakarta [19].

Namun, data percakapan di *Twitter/X* memiliki karakteristik yang tidak terstruktur. *Tweet* umumnya ditulis dalam kalimat pendek, menggunakan bahasa informal, singkatan, *slang*, emoji, dan terkadang mengandung sarkasme [20]. Selain itu, jumlah data yang besar membuat proses analisis manual menjadi tidak efisien karena membutuhkan waktu lama dan berisiko menimbulkan subjektivitas. Kondisi tersebut mendorong perlunya pendekatan *Natural Language Processing* (NLP) untuk membantu mengolah data teks secara lebih sistematis [21].

Analisis sentimen merupakan salah satu pendekatan dalam NLP yang dapat digunakan untuk mengidentifikasi kecenderungan opini dalam teks. Melalui analisis sentimen, *tweet* pengguna dapat diklasifikasikan ke dalam kategori tertentu, seperti negatif, netral, atau positif [22]. Dalam penelitian ini, analisis sentimen digunakan untuk mengklasifikasikan opini pengguna terhadap layanan

TransJakarta berdasarkan percakapan pada *Twitter/X*. Hasil klasifikasi tersebut dapat menunjukkan kecenderungan sentimen pengguna terhadap layanan yang dibahas dalam data penelitian [23].

Selain analisis sentimen, pemodelan topik juga diperlukan untuk mengetahui tema pembahasan yang muncul dalam percakapan pengguna. Analisis sentimen dapat menunjukkan kecenderungan opini, sedangkan pemodelan topik dapat membantu mengidentifikasi aspek layanan yang paling sering dibicarakan [20]. Dengan demikian, penelitian ini tidak hanya mengetahui apakah opini pengguna terhadap TransJakarta cenderung negatif, netral, atau positif, tetapi juga dapat melihat topik yang muncul pada masing-masing sentimen, seperti rute, halte, waktu tunggu, kepadatan, pembayaran, fasilitas, dan informasi operasional [24].

Pada beberapa penelitian sebelumnya, analisis sentimen dan pemodelan topik banyak dilakukan menggunakan metode *machine learning* konvensional. Namun, metode tersebut memiliki keterbatasan dalam memahami konteks kalimat secara mendalam, terutama pada data media sosial yang bersifat informal. Perkembangan model berbasis *transformer* memberikan pendekatan yang lebih kuat dalam pemrosesan teks karena mampu memahami hubungan antar kata secara kontekstual. BERT merupakan model representasi bahasa berbasis *transformer* yang melakukan *pre-training* secara *bidirectional*, sehingga mampu memahami konteks kata berdasarkan sisi kiri dan kanan dalam kalimat [25].

Dalam penelitian ini, model yang digunakan untuk analisis sentimen adalah IndoBERT. IndoBERT merupakan model bahasa *pre-trained* untuk bahasa Indonesia yang diperkenalkan dalam penelitian IndoLEM. Model ini dikembangkan untuk menjawab kebutuhan pemrosesan bahasa Indonesia yang sebelumnya masih memiliki keterbatasan sumber daya NLP. Hasil penelitian IndoLEM menunjukkan bahwa IndoBERT mencapai performa *state-of-the-art* pada sebagian besar tugas NLP bahasa Indonesia yang diuji [26].

Pemilihan IndoBERT dalam penelitian ini didasarkan pada kemampuannya dalam memahami konteks kalimat berbahasa Indonesia secara lebih baik dibandingkan pendekatan berbasis kata kunci atau metode konvensional. Hal ini

penting karena data *Twitter/X* sering memuat ungkapan informal, kalimat pendek, dan konteks yang tidak selalu eksplisit. Selain itu, IndoLEM juga mencakup tugas NLP yang berkaitan dengan data *tweet*, sehingga penggunaan model berbasis bahasa Indonesia relevan untuk pemrosesan teks media sosial.

Sebelum dilakukan *fine-tuning* IndoBERT, penelitian ini menggunakan proses *auto labeling* untuk memberikan label sentimen awal pada data. Pelabelan otomatis dilakukan menggunakan model Indonesian RoBERTa *sentiment classifier* yang tersedia pada platform Hugging Face [27]. Model ini digunakan karena data penelitian berupa *tweet* berbahasa Indonesia dan kelas sentimen yang digunakan terdiri dari negatif, netral, dan positif. Dengan menggunakan *auto labeling*, proses pemberian label dapat dilakukan secara konsisten terhadap seluruh data sebelum digunakan dalam pelatihan model IndoBERT.

Sementara itu, pemodelan topik dalam penelitian ini menggunakan BERTopic. BERTopic merupakan metode pemodelan topik yang memanfaatkan *embedding* dari model berbasis *transformer*, melakukan *clustering* terhadap dokumen, dan menghasilkan representasi topik menggunakan pendekatan *class-based TF-IDF* [28]. Pendekatan ini relevan untuk data teks media sosial karena mampu memanfaatkan representasi semantik dari teks, sehingga topik yang dihasilkan tidak hanya bergantung pada frekuensi kata, tetapi juga mempertimbangkan kedekatan makna antar dokumen.

Berdasarkan uraian tersebut, penelitian ini berfokus pada analisis sentimen dan pemodelan topik terhadap layanan TransJakarta di *Twitter/X* menggunakan IndoBERT dan BERTopic. Penelitian ini menggunakan data *tweet* berbahasa Indonesia yang berkaitan dengan layanan TransJakarta. Analisis sentimen dilakukan untuk mengetahui kecenderungan opini pengguna ke dalam kelas negatif, netral, dan positif, sedangkan pemodelan topik dilakukan untuk mengidentifikasi isu utama yang muncul pada masing-masing kelompok sentimen. Hasil penelitian ini diharapkan dapat memberikan gambaran berbasis data mengenai opini pengguna terhadap layanan TransJakarta serta menjadi masukan bagi pihak terkait dalam mengevaluasi aspek layanan yang banyak dibicarakan masyarakat [29].

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, ada beberapa permasalahan yang ditemui.

1. Bagaimana sentimen pengguna Twitter/X terhadap layanan TransJakarta berdasarkan data percakapan di media sosial?
2. Bagaimana hasil analisis sentimen menggunakan model IndoBERT terhadap tweet berbahasa Indonesia yang membahas layanan TransJakarta?
3. Bagaimana hasil pemodelan topik menggunakan BERTopic dalam mengidentifikasi isu atau tema utama yang muncul pada masing-masing kelompok sentimen terhadap layanan TransJakarta?

1.3 Batasan Masalah

Ada beberapa batasan agar penelitian ini tidak keluar dari konteks.

1. Objek penelitian dibatasi pada layanan TransJakarta.
2. Penelitian ini hanya menggunakan data teks dari media sosial Twitter/X yang berkaitan dengan layanan TransJakarta.
3. Data yang digunakan berupa tweet, mention, dan balasan yang berkaitan dengan kata kunci TransJakarta atau akun resmi @PT_Transjakarta.
4. Data *tweet* yang digunakan hanya berbahasa Indonesia.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Penelitian ini memiliki tujuan untuk:

1. Menganalisis sentimen pengguna Twitter/X terhadap layanan TransJakarta berdasarkan data percakapan di media sosial.
2. Mengetahui hasil analisis sentimen menggunakan model IndoBERT terhadap tweet berbahasa Indonesia yang membahas layanan TransJakarta.
3. Membandingkan performa model IndoBERT hasil fine-tuning dengan model Indonesian RoBERTa yang digunakan pada tahap auto

labeling, guna mengetahui sejauh mana peningkatan keandalan klasifikasi sentimen setelah dilakukan proses fine-tuning.

4. Mengidentifikasi isu atau tema utama yang muncul pada masing-masing kelompok sentimen terhadap layanan TransJakarta menggunakan BERTopic.

1.4.2 Manfaat Penelitian

- Manfaat Teoritis
 - a. Penelitian ini diharapkan dapat menambah referensi dalam bidang *text mining*, *natural language processing*, analisis sentimen, dan pemodelan topik pada data teks berbahasa Indonesia.
 - b. Penelitian ini dapat menjadi rujukan dalam penerapan model *transformer*, khususnya IndoBERT, untuk analisis sentimen pada data media sosial Twitter/X.
 - c. Penelitian ini dapat memberikan kontribusi akademik dalam penggunaan kombinasi IndoBERT dan BERTopic untuk menganalisis kecenderungan sentimen serta isu pembahasan utama pada data opini pengguna berbasis media sosial.
- Manfaat Praktis
 - a. Bagi TransJakarta, hasil penelitian ini dapat menjadi bahan masukan untuk memahami kecenderungan sentimen pengguna terhadap layanan yang diberikan berdasarkan percakapan pada Twitter/X.
 - b. Bagi pihak pengelola layanan TransJakarta, hasil pemodelan topik dapat membantu mengidentifikasi isu layanan yang banyak dibicarakan pengguna, seperti rute, halte, waktu tunggu, kepadatan, pembayaran, fasilitas, dan informasi operasional.
 - c. Bagi pemerintah atau pihak terkait dalam pengelolaan transportasi di DKI Jakarta, hasil penelitian ini dapat menjadi bahan pendukung dalam mengevaluasi aspek

layanan TransJakarta yang sering menjadi perhatian masyarakat.

- d. Bagi peneliti selanjutnya, penelitian ini dapat menjadi referensi dalam pengembangan penelitian analisis sentimen dan pemodelan topik menggunakan *auto labeling*, IndoBERT, dan BERTopic pada data media sosial berbahasa Indonesia.

1.5 Sistematika Penulisan

Penulisan skripsi ini disusun ke dalam lima bab dengan sistematika penulisan sebagai berikut.

BAB I PENDAHULUAN

Bab ini berisi penjelasan mengenai latar belakang penelitian yang membahas layanan TransJakarta serta pemanfaatan data percakapan pengguna pada media sosial Twitter/X. Selain itu, bab ini juga memuat rumusan masalah yang berkaitan dengan analisis sentimen dan pemodelan topik terhadap layanan TransJakarta, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi teori-teori yang menjadi dasar dalam penelitian, khususnya yang berkaitan dengan analisis sentimen, pemodelan topik, pre-processing data teks, auto labeling, framework KDD, algoritma IndoBERT, RoBERTa, BERTopic, serta tools yang digunakan. Selain itu, bab ini juga memuat penelitian terdahulu yang relevan berdasarkan hasil literature review sebagai acuan dalam mendukung penelitian.

BAB III METODE PENELITIAN

Bab ini menjelaskan metode penelitian yang digunakan, yaitu Knowledge Discovery in Database (KDD). Objek penelitian yang digunakan adalah data tweet dari media sosial Twitter/X yang membahas layanan TransJakarta. Pada analisis sentimen, variabel independen yang digunakan adalah data tweet yang telah melalui proses data cleaning, sedangkan variabel dependennya adalah klasifikasi sentimen negatif, netral, dan positif. Pelabelan awal dilakukan menggunakan Indonesian

RoBERTa, kemudian data digunakan untuk proses fine-tuning IndoBERT. Pada pemodelan topik, variabel independen yang digunakan adalah tweet yang telah memiliki label sentimen, sedangkan variabel dependennya adalah hasil topik yang diperoleh menggunakan BERTopic. Selain itu, bab ini juga menjelaskan teknik pengumpulan data dan teknik analisis data yang digunakan dalam penelitian.

BAB IV ANALISIS DAN HASIL PENELITIAN

Bab ini berisi proses analisis dan hasil penelitian berdasarkan tahapan dalam framework KDD, yaitu data selection, data cleaning, data transformation, data mining, serta interpretation/evaluation. Pada bab ini dijelaskan hasil auto labeling menggunakan Indonesian RoBERTa, proses fine-tuning dan evaluasi IndoBERT, distribusi sentimen pengguna Twitter/X terhadap layanan TransJakarta, serta hasil pemodelan topik menggunakan BERTopic pada masing-masing kelompok sentimen.

BAB V SIMPULAN DAN SARAN

Bab ini berisi simpulan dari hasil penelitian berdasarkan rumusan masalah yang telah dijelaskan pada Bab I. Selain itu, bab ini juga memuat limitasi penelitian serta saran yang dapat digunakan sebagai acuan untuk penelitian selanjutnya, khususnya dalam pengembangan analisis sentimen dan pemodelan topik pada data media sosial berbahasa Indonesia.