JORAPI : Journal of Research and Publication Innovation
Vol. 4, No. 1, Januari 2026 **ISSN : 2985-4768**
Halaman : 1-8

ANALISIS SENTIMEN ULASAN PUBLIK PADA X MENGGUNAKAN NAIVE BAYES UNTUK MENINGKATKAN KUALITAS LAYANAN TRANSJAKARTA

Daffa Alif Ruriyanto [1], Fauzan Abhip Raya [2], and Reffiano Siswoyo[3]

Program Studi Sistem Informasi, Universitas Pamulang, Tangerang Selatan, Indonesia, 15310
e-mail: [1] notdaffa@gmail.com, [2] fauzanabhipraya14@gmail.com, [3] reffzhii@gmail.com

Abstract

As the largest public transportation system in Jakarta, TransJakarta frequently becomes a topic of discussion on Twitter due to various user experiences ranging from satisfaction to service-related complaints. The abundance of public opinions provides an opportunity to apply machine learning–based sentiment analysis as a faster, more objective, and measurable method for evaluating service quality. This study employs the Naive Bayes algorithm to classify tweet sentiments into positive, negative, and neutral categories. Data were collected through a crawling process using keywords related to TransJakarta, then processed through several stages including tokenization, text cleaning, stopword removal, and stemming to produce analysis-ready data. Several service aspects discussed most frequently include bus arrival punctuality, travel comfort, and fleet conditions. The analysis results indicate that the Naive Bayes model can accurately identify public sentiment patterns in near real-time, enabling the detection of opinion trends across different service dimensions. These findings demonstrate that Naive Bayes–based sentiment analysis can serve as an effective tool for monitoring public perception and provide a strong foundation for TransJakarta in formulating strategies to improve service quality.

Keywords: Sentiment Analysis, TransJakarta, Twitter, Naive Bayes, Service Quality

Abstrak

Sebagai moda transportasi publik terbesar di Jakarta, TransJakarta menjadi topik yang banyak diperbincangkan di Twitter karena berbagai pengalaman pengguna, mulai dari kepuasan hingga keluhan terhadap layanan. Banyaknya opini publik ini memberikan peluang untuk menerapkan analisis sentimen berbasis machine learning sebagai metode evaluasi kualitas layanan yang lebih cepat, objektif, dan terukur. Penelitian ini menggunakan algoritma Naive Bayes untuk mengklasifikasikan sentimen tweet ke dalam kategori positif, negatif, dan netral. Data dikumpulkan melalui proses crawling menggunakan kata kunci yang berkaitan dengan TransJakarta, kemudian diproses melalui tahapan tokenisasi, pembersihan teks, penghapusan stopword, dan stemming untuk menghasilkan data yang siap dianalisis. Beberapa aspek layanan yang paling sering dibahas antara lain ketepatan waktu kedatangan bus, kenyamanan perjalanan, serta kondisi armada. Hasil analisis menunjukkan bahwa model Naive Bayes mampu mengidentifikasi pola sentimen publik secara akurat dan mendekati real-time, sehingga memungkinkan deteksi kecenderungan opini terhadap berbagai aspek layanan. Temuan ini membuktikan bahwa analisis sentimen berbasis Naive Bayes dapat menjadi alat yang efektif untuk memantau persepsi masyarakat dan memberikan dasar yang kuat bagi pihak TransJakarta dalam merumuskan strategi peningkatan kualitas layanan.

Kata Kunci: Analisis Sentimen, TransJakarta, Twitter, Naive Bayes, Kualitas Layanan

1. PENDAHULUAN

Transportasi umum memiliki peranan penting sebagai sarana mobilitas masyarakat, khususnya di wilayah perkotaan dengan tingkat kepadatan penduduk yang tinggi seperti Jakarta. Ketersediaan transportasi yang memadai tidak hanya membantu kelancaran aktivitas masyarakat, tetapi juga berkontribusi pada pengurangan kemacetan serta peningkatan kualitas lingkungan. Transportasi umum menjadi pilihan masyarakat di Jakarta yang padat penduduk. Menurut laporan Dinas Lingkungan Hidup Pemprov DKI Jakarta, Transjakarta merupakan salah satu transportasi umum yang memiliki jumlah penumpang harian terbanyak dibandingkan dengan transportasi umum lainnya [1]. Dengan jumlah pengguna yang besar, TransJakarta menjadi moda transportasi utama bagi masyarakat dalam menjalankan aktivitas sehari-hari. Banyaknya interaksi pengguna dengan layanan tersebut memungkinkan munculnya berbagai pendapat, pengalaman, kritik, dan saran terkait kualitas pelayanan. Hal tersebut membuat para pengguna Transjakarta dapat memberikan pendapat, kritik maupun saran dari pengalaman yang dialami oleh pengguna moda transportasi umum Transjakarta [2].

Perkembangan teknologi digital membuat masyarakat kini lebih mudah menyampaikan opini secara terbuka melalui platform media sosial. Salah satu platform yang paling aktif digunakan untuk menyuarakan pendapat adalah media sosial X (sebelumnya Twitter). Media sosial ini memiliki karakteristik komunikasi yang cepat, terbuka, dan berbasis teks, sehingga sangat relevan untuk mengumpulkan opini publik. Media sosial X adalah media sosial terkenal yang berguna sebagai wadah komunikasi di masyarakat [3]. Setiap harinya, pengguna aktif twitter dapat menulis lebih dari 400 juta tweet terkenal [4]. Para pengguna media sosial X dapat berbagi hal dalam bentuk tulisan status atau tweet seperti mengungkapkan kebahagian, informasi, inspirasi, dan mendiskusikan topik tertentu [5]. Semua pendapat, saran, dan kritik yang diperoleh dari pengguna sosial media yaitu media sosial X dapat disebut sebagai sentimen [6]. Oleh karena itu, Twitter/X menjadi sumber data yang sangat potensial dalam menggambarkan persepsi masyarakat terhadap pelayanan TransJakarta.

Untuk mengolah opini publik tersebut, diperlukan metode analisis yang mampu mengklasifikasikan kecenderungan atau polaritas emosi dalam teks, yang dikenal sebagai analisis sentimen. Teknik ini berfungsi mengidentifikasi apakah suatu ulasan bersifat positif, negatif, atau netral. Analisis sentimen yaitu metode dalam menentukan sentimen untuk mengklasifikasikan polaritas teks untuk dimasukan kedalam dokumen atau kalimat [7] untuk menentukan kategori seperti sentimen positif, negatif atau netral. Saat ini, analisis sentimen banyak digunakan oleh para peneliti sebagai bagian dari penelitian ilmu komputer [8]. Dengan analisis sentimen, instansi layanan publik seperti TransJakarta dapat mengetahui persepsi masyarakat secara real-time dan berbasis data, sehingga bisa digunakan sebagai bahan evaluasi peningkatan kualitas layanan.

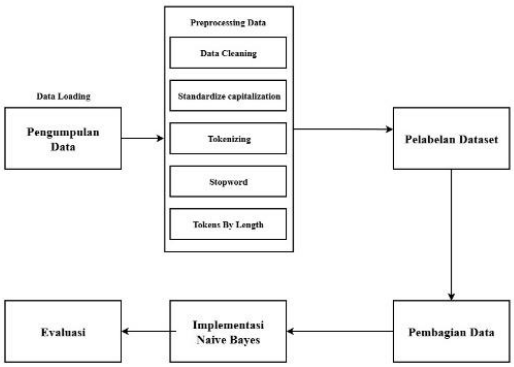
Dalam penelitian analisis sentimen, salah satu algoritma yang paling banyak digunakan adalah Naïve Bayes. Algoritma ini memiliki keunggulan karena perhitungannya yang sederhana, cepat, dan efektif dalam memproses data teks berukuran besar. Naïve Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma mengunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabe kelas [9]. Karena sifatnya yang efisien, Naïve Bayes sangat sesuai digunakan untuk menganalisis sentimen publik terhadap layanan TransJakarta yang tersebar melimpah di media sosial X.

Berdasarkan latar belakang tersebut, penelitian ini dilakukan untuk menganalisis sentimen masyarakat terhadap layanan TransJakarta melalui unggahan pengguna di media sosial X menggunakan algoritma Naïve Bayes. Hasil analisis ini diharapkan dapat memberikan gambaran kualitas layanan berdasarkan opini publik serta menjadi masukan dalam upaya peningkatan pelayanan TransJakarta di masa depan.

2. METODE

Penelitian ini dilakukan melalui beberapa tahapan yang terstruktur untuk menghasilkan analisis sentimen yang akurat terhadap ulasan publik mengenai

layanan TransJakarta. Adapun tahapan penelitian yang dilakukan adalah sebagai berikut:



Gambar 1. Tahapan Proses Penelitian

a) Pengumpulan Data

Data ulasan publik dikumpulkan dari platform Twitter menggunakan metode web scraping dengan kata kunci "TransJakarta". Data yang diperoleh meliputi nama akun, teks tweet dan metadata pendukung lainnya. Seluruh tweet yang berbahasa Indonesia menjadi fokus utama agar relevan dengan konteks pengguna TransJakarta.

b) Preprocessing Data

Analisis Tahap preprocessing dilakukan untuk meningkatkan kualitas data teks sebelum masuk pada proses analisis dan pemodelan. Langkah ini bertujuan mengurangi noise, menyamakan format penulisan, serta menyiapkan representasi data yang sesuai untuk algoritma klasifikasi. Adapun tahapan preprocessing yang digunakan dalam penelitian ini meliputi:

- Data Cleaning

Proses data cleaning dilakukan untuk menghilangkan elemen-elemen yang tidak memberikan kontribusi terhadap analisis sentimen. Elemen tersebut mencakup URL, mention (@username), hashtag, emoji, angka, serta tanda baca yang tidak relevan. Penghapusan elemen tersebut bertujuan untuk menghasilkan teks yang lebih bersih dan terfokus pada konten semantik.

- Standardisasi Kapitalisasi

Seluruh teks diubah ke huruf kecil (lowercase) guna mengurangi kompleksitas analisis dan menghindari perbedaan makna akibat variasi format kapitalisasi. Dengan demikian, kata yang sama dalam bentuk berbeda seperti "Bus", "BUS", dan "bus" dapat diperlakukan sebagai satu entitas

- Tokenisasi

Tokenisasi dilakukan dengan memecah teks menjadi unit-unit kata (tokens). Pemecahan ini memungkinkan algoritma untuk memproses teks pada tingkat kata sehingga memudahkan analisis frekuensi dan pola dalam kalimat.

- Stopword

Tahap berikutnya adalah menghilangkan kata-kata umum (stopwords) yang tidak memiliki kontribusi signifikan terhadap penentuan sentimen. Stopwords yang digunakan merujuk pada daftar kata Bahasa Indonesia yang umum dan bersifat non-informatif, seperti "yang", "di", "dan", "ke", serta kata penghubung lainnya.

Tahap *preprocessing* secara keseluruhan menghasilkan representasi teks yang lebih terstruktur dan siap untuk dilakukan ekstraksi fitur pada tahap berikutnya.

c) Pelabelan Dataset

Pelabelan dataset merupakan tahap penting dalam membangun model klasifikasi yang berbasis supervised learning. Data yang telah melalui proses preprocessing kemudian diberi label sentimen berdasarkan konteks dan makna isi tweet. Proses pelabelan dilakukan secara manual untuk memastikan kesesuaian interpretasi terhadap konten yang dianalisis.

Tweet dikelompokkan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Tweet dengan keluhan terkait keterlambatan bus, kepadatan penumpang, dan kualitas fasilitas dikategorikan sebagai sentimen negatif. Tweet berisi apresiasi atau penilaian baik terhadap layanan masuk ke kategori positif. Sementara itu, tweet yang bersifat informatif, tidak menunjukkan emosi

tertentu, atau netral dalam penyampaian dimasukkan ke kategori netral.

Proses pelabelan ini menghasilkan dataset berlabel yang kemudian digunakan sebagai dasar untuk pelatihan dan pengujian model Naive Bayes.

d) Pembagian Data

Pelabelan Dataset yang telah dilabeli kemudian dibagi menjadi dua subset, yaitu data pelatihan (*training set*) dan data pengujian (*testing set*). Pembagian dilakukan dengan proporsi 80% untuk data pelatihan dan 20% untuk data pengujian. Proporsi ini dipilih untuk memastikan model memperoleh data yang cukup untuk belajar sekaligus menyediakan dataset terpisah untuk mengukur kemampuan generalisasi model. Proses pembagian dilakukan secara acak agar distribusi kelas pada kedua subset tetap seimbang.

e) Implementasi Metode Naive Bayes

Model klasifikasi sentimen dibangun menggunakan algoritma *Multinomial Naive Bayes*, yang umum digunakan untuk pemrosesan teks karena kemampuannya dalam menangani fitur berupa frekuensi kata. Sebelum pelatihan model, teks direpresentasikan menggunakan metode ekstraksi fitur *Term Frequency–Inverse Document Frequency (TF-IDF)* untuk memberikan bobot yang lebih proporsional terhadap kata-kata yang relevan dalam konteks sentimen.

Model dilatih menggunakan *training set* untuk mempelajari distribusi probabilitas antar fitur dan kelas sentimen. Selanjutnya, model digunakan untuk memprediksi sentimen pada *testing set*.

f) Evaluasi

Pada tahap evaluasi, penelitian ini melakukan analisis terhadap distribusi kelas sentimen untuk memastikan bahwa dataset yang digunakan berada dalam kondisi yang representatif untuk proses pelatihan model. Evaluasi distribusi kelas dilakukan menggunakan fitur *SimpleDistribution* untuk memeriksa proporsi masing-masing kategori sentimen, yaitu positif, negatif, dan netral.

Analisis distribusi kelas ini bertujuan untuk mengidentifikasi potensi ketidakseimbangan kelas (*class imbalance*) yang dapat mempengaruhi kinerja algoritma *Naive Bayes*. Ketidakseimbangan kelas berpotensi menyebabkan model cenderung memprediksi kelas mayoritas dan mengabaikan karakteristik kelas minoritas. Oleh karena itu, proses evaluasi ini dilakukan sebelum tahap pelatihan model untuk memastikan bahwa struktur label pada dataset telah dipahami dengan baik dan dapat dipertimbangkan dalam proses pemodelan selanjutnya.

3. PEMBAHASAN

Pembahasan ini menguraikan proses analisis sentimen yang dilakukan mulai dari tahap pengambilan data, pengolahan data awal, hingga pembentukan model menggunakan RapidMiner. Alur ini menjelaskan bagaimana data mentah dari Twitter diproses hingga menghasilkan klasifikasi sentimen yang dapat dianalisis lebih lanjut.

a. Pengambilan Data menggunakan Google colab

Tahap awal penelitian dimulai dengan proses pengumpulan data ulasan publik terkait TransJakarta melalui platform Twitter. Pengambilan data dilakukan menggunakan Google Colab karena platform tersebut mendukung eksekusi library Python seperti tweet harvest . Google Colab juga mempermudah pengolahan data awal karena menyediakan lingkungan komputasi berbasis cloud.

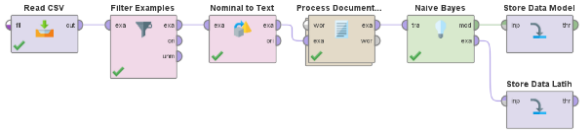
Data tweet dikumpulkan dengan kata kunci yang relevan, kemudian disimpan dalam format CSV. Dalam penelitian ini, jumlah data yang berhasil dikumpulkan adalah **249 tweet**, yang selanjutnya digunakan sebagai dataset analisis sentimen. Setiap tweet memuat atribut teks dan label sentimen yang diperlukan untuk proses pelatihan model.

Gambar 2. Alur Pengumpulan Tweet Untuk Dataset

b. Pemrosesan Data Menggunakan RapidMiner

Setelah data diperoleh, seluruh proses *preprocessing* dan pembangunan model dilakukan menggunakan RapidMiner. Platform ini dipilih karena menyediakan alur kerja visual yang mudah diatur serta mendukung operator text mining dan algoritma klasifikasi.

Alir proses yang digunakan meliputi beberapa tahapan sebagai berikut:



Gambar 3. Proses Pembuatan Data Latih dan Model Naïve Bayes

c. Preprocessing Data

Setelah data dimasukkan ke RapidMiner, tahap selanjutnya adalah mempersiapkan teks agar dapat digunakan dalam model analisis sentimen. Proses ini mencakup:

Gambar 4. Alur Proses Preprocessing Teks pada RapidMiner

- Data Cleaning

Setelah Tahap cleansing dilakukan dengan dua langkah utama, yaitu menghilangkan data yang bersifat duplikat serta membersihkan berbagai karakter atau simbol yang tidak relevan, seperti emoji, tanda baca (.?@#,!-/), dan lainnya. Pada penelitian ini, proses pembersihan tweet dilakukan menggunakan operator *subprocess* dan *remove duplicated*. Tujuan dari tahapan ini adalah memastikan bahwa data yang masuk ke proses analisis sentimen berada dalam kondisi yang bersih sehingga hasil yang diperoleh lebih akurat.

- Standarize Capitalization

Proses *standardize capitalize* merupakan tahapan normalisasi awal yang bertujuan menyeragamkan seluruh karakter huruf pada tweet menjadi huruf kecil (lowercase). Langkah ini dilakukan untuk mencegah sistem memperlakukan kata yang sama sebagai entitas berbeda hanya karena perbedaan kapitalisasi. Dengan demikian, konsistensi data semakin terjaga dan memudahkan proses analisis lebih lanjut. Hasil perubahan bentuk huruf ini dapat diamati melalui perbandingan data sebelum dan sesudah normalisasi pada tabel berikut.

- Tokenizing

Tokenizing memecah teks tweet menjadi unit-unit kata atau token sehingga setiap kata dapat diproses secara terpisah. Tahap ini sangat penting karena model analisis sentimen bekerja berdasarkan representasi token, bukan kalimat utuh. Dengan adanya tokenisasi, pola kata, frekuensi kemunculan, dan konteks dapat diidentifikasi lebih akurat. Perubahan struktur data akibat proses tokenisasi ini ditunjukkan pada tabel before–after yang disajikan setelah penjelasan ini.

Tabel 1. Tokenizing

no	Isi tweet	tokens
1	siapapun yang nempelin permen karet di tiang pembatas tije gini sumpah ga jelas bgt jadi orang	nempelin,permen, karet,tiang,pembatas, tije,gini,sumpah, ga,bgt,orang
2	pahlawan jalur 13 nih yang berhasil ngedorong pinky jadi ga nuutupin jalan makasih ya bapak bapak hebat	pahlawan,jalur,13, nih,berhasil,ngedorong, pinky,ga,nuutupin, jalan,makasih,ya, hebat
3	setelah 30mnt menunggu bs msk bus busny sangat optimal pak sering2 cek cek koridor	30mnt,menunggu, bs,msk,bus,busny, optimal,serang2,cek, cek,koridor

- Stopword

Tahap stopword removal menghilangkan kata-kata yang bersifat umum dan tidak memberikan kontribusi signifikan pada penentuan sentimen, seperti “dan”, “yang”, atau “di”. Penghapusan kata-kata ini membantu mengurangi noise dan membuat analisis lebih fokus pada kata bermakna yang benar-benar mencerminkan opini pengguna.

Pengaruh dari proses ini dapat dilihat melalui perbandingan data sebelum dan setelah stopword dihapus pada tabel berikutnya.

Tabel 2. Stopword

no	Isi_tweet	Clean_tokens
1	siapapun yang nempelin permen karet di tiang pembatas tije gini sumpah ga jelas bgt jadi orang	nempelin,permen, karet,tiang,pembatas, tije,gini,sumpah,ga, bgt,orang
2	pahlawan jalur 13 nih yang berhasil ngedorong pinky jadi ga nuutupin jalan makasih ya bapak bapak hebat	pahlawan,jalur,13, nih,berhasil,ngedorong, pinky,ga,nuutupin,jalan, makasih,ya,hebat
3	setelah 30mnt menunggu bs msk bus busny sangat optimal pak sering2 cek cek koridor	30mnt,menunggu, bs,msk,bus,busny, optimal,sering2, cek,cek,koridor

- Filter Tokens by Length

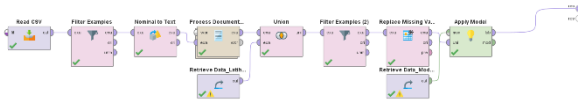
Penyaringan token berdasarkan panjang karakter dilakukan untuk membuang token yang terlalu pendek atau tidak bermakna, seperti satu huruf atau sisa simbol pembersihan. Token semacam ini biasanya tidak berkontribusi terhadap makna keseluruhan dan dapat menurunkan kinerja model. Perubahan kualitas data setelah proses penyaringan ini dapat dilihat secara jelas pada tabel before–after yang disertakan pada bagian berikut.

Tabel 3. Tokens by Length

no	Isi_tweet	after tokens
1	siapapun yang nempelin permen karet di tiang pembatas tije gini sumpah ga jelas bgt jadi orang	nempelin,permen, karet,tiang,pembatas, tije,gini,sumpah, ga,bgt,orang
2	pahlawan jalur 13 nih yang berhasil ngedorong pinky jadi ga nuutupin jalan makasih ya bapak bapak hebat	pahlawan,jalur,13, nih,berhasil,ngedorong, pinky,ga,nuutupin, jalan,makasih,ya, hebat
3	setelah 30mnt menunggu bs msk bus busny sangat optimal pak sering2 cek cek koridor	30mnt,menunggu, bs,msk,bus,busny, optimal,serang2,cek, cek,koridor

Seluruh proses ini dilakukan dalam satu rangkaian *text preprocessing pipeline* sehingga menghasilkan representasi teks yang rapi, terstruktur, dan siap diproses oleh algoritma klasifikasi.

d. Pembentukan Model Analisis Sentimen



Gambar 5. Tahapan Model Building Naive Bayes untuk Analisis Sentimen

Tahap berikutnya adalah pembangunan model sentimen menggunakan algoritma Multinomial Naive Bayes. Algoritma ini dipilih karena efektif untuk analisis teks dan memiliki performa baik pada data berdimensi tinggi seperti TF-IDF.

Model dilatih menggunakan dataset yang telah diproses sebelumnya. Proses pelatihan dilakukan secara otomatis oleh sistem, di mana model mempelajari pola kemunculan kata yang berkaitan dengan sentimen positif, negatif, maupun netral.

Model akhir kemudian disimpan agar dapat digunakan dalam tahap prediksi atau evaluasi lebih lanjut.

- Pembagian Data

Pada tahap pembentukan model, pembagian data menghasilkan distribusi yang cukup proporsional antara data latih dan data uji. Berdasarkan dataset tweet yang telah melalui proses *preprocessing*, terlihat bahwa jumlah tweet bernada negatif lebih dominan dibandingkan sentimen netral maupun positif. Ketimpangan ini otomatis tercermin pada kedua subset data sehingga model dilatih dengan komposisi yang mencerminkan kondisi opini publik yang sebenarnya.

Melalui pembagian tersebut, terlihat bahwa karakteristik bahasa pengguna Twitter terkait layanan TransJakarta relatif konsisten antara data

latih dan data uji. Hal ini penting karena memastikan bahwa performa model yang dievaluasi tidak bias akibat ketidakseimbangan distribusi saat proses *splitting*. Dengan demikian, hasil evaluasi yang diperoleh dapat dianggap merepresentasikan kemampuan model ketika diterapkan pada data publik di luar sampel penelitian.

- Implementasi Metode Naïve Bayes

Pada proses pelatihan model, algoritma Naive Bayes menunjukkan kemampuan untuk membedakan pola kata yang paling sering muncul pada tiap kategori sentimen. Dari hasil pelatihan terlihat bahwa tweet bernada negatif memuat kata-kata yang sangat repetitif terkait isu layanan, seperti keterlambatan bus, kepadatan penumpang, serta keluhan fasilitas. Frekuensi kemunculan kata-kata tersebut membentuk probabilitas yang kuat sehingga model lebih mudah mengidentifikasi tweet negatif.

Sebaliknya, tweet positif jauh lebih sedikit dan kata-katanya cenderung lebih beragam, sehingga probabilitasnya lebih tersebar. Kondisi ini membuat model lebih menantang dalam mengenali sentimen positif. Namun secara keseluruhan, pola probabilistik yang muncul selama proses pelatihan menunjukkan bahwa Naive Bayes cukup responsif terhadap struktur bahasa yang pendek dan langsung, seperti tweet.

Temuan ini sejalan dengan karakteristik data Twitter, dimana pengguna cenderung mengekspresikan keluhan dengan lebih eksplisit dibanding pujian. Hal tersebut memberi dampak pada kekuatan model dalam menangkap sisi negatif layanan TransJakarta secara lebih tajam.

- Evaluasi Model

Evaluasi model dilakukan dengan meninjau distribusi prediksi sentimen pada data uji. Berdasarkan hasil klasifikasi, kategori sentimen negatif menjadi kelompok dengan jumlah prediksi terbanyak, yaitu sebanyak 159 tweet. Dominasi tweet negatif ini menunjukkan bahwa percakapan publik mengenai layanan TransJakarta di Twitter lebih banyak berfokus pada keluhan. Sebagian besar keluhan berkaitan dengan isu operasional seperti keterlambatan bus, kepadatan penumpang, dan keluhan fasilitas yang dianggap belum optimal. Pola ini mengindikasikan bahwa aspek-aspek tersebut merupakan sumber ketidakpuasan yang paling sering dialami pengguna.

Kategori sentimen positif berada pada posisi kedua dengan jumlah 51 tweet. Meskipun jumlahnya tidak sebesar sentimen negatif, keberadaan tweet positif menunjukkan bahwa masih terdapat pengguna yang memberikan apresiasi terhadap aspek tertentu dari layanan TransJakarta. Umpan balik positif biasanya muncul terkait ketepatan waktu pada kondisi tertentu, kenyamanan armada baru, atau pengalaman layanan yang dinilai memuaskan.

Sementara itu, kategori sentimen netral menjadi kelompok yang paling sedikit terklasifikasi, yakni sebanyak 39 tweet. Tweet kategori ini umumnya berisi informasi atau komentar yang tidak secara eksplisit menunjukkan kepuasan maupun ketidakpuasan, misalnya laporan kondisi lalu lintas, pengamatan situasi halte, atau pernyataan yang sifatnya informatif tanpa nada emosional.

Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa model Naive Bayes berhasil menggambarkan kecenderungan persepsi publik secara jelas. Dominasi sentimen negatif menandakan bahwa terdapat beberapa aspek layanan TransJakarta yang masih menjadi perhatian utama pengguna, sementara keberadaan sentimen positif dan netral memberikan gambaran yang lebih seimbang mengenai variasi opini yang muncul di platform Twitter.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan metode Naive Bayes mampu mengidentifikasi kecenderungan sentimen publik terhadap layanan TransJakarta secara terstruktur dan mudah direproduksi. Distribusi sentimen yang dihasilkan mengindikasikan bahwa sentimen negatif menempati proporsi terbesar, diikuti oleh sentimen positif, sementara sentimen netral muncul sebagai kategori paling kecil. Pola tersebut mencerminkan bahwa persepsi publik masih didominasi oleh keluhan atau pengalaman kurang memuaskan, terutama terkait aspek operasional harian. Keunggulan pendekatan ini terletak pada kemampuannya menangani data teks dalam skala besar dengan proses komputasi yang efisien.

Namun demikian, model masih memiliki keterbatasan dalam menangkap nuansa bahasa informal, sarkasme, dan konteks spesifik yang sering muncul pada percakapan di Twitter. Ke depan, penelitian dapat dikembangkan dengan memperluas korpus data, mengintegrasikan model berbasis deep learning, atau menambahkan analisis temporal untuk memahami dinamika perubahan sentimen. Secara keseluruhan, hasil penelitian ini memberikan gambaran yang dapat menjadi dasar bagi perumusan strategi peningkatan kualitas layanan TransJakarta berbasis persepsi publik.

DAFTAR PUSTAKA

[1] Cindy Mutia Annur, "No Title," Ini Rute Bus Transjakarta dengan Jumlah Penumpang Terbanyak pada 2021. Accessed: Dec. 12, 2025. [Online]. Available: https://databoks.katadata.co.id/transportasi-logistik/statistik/9d26bc18b58a174/ini-rute-bus-transjakarta-dengan-jumlah-penumpang-terbanyak-pada-2021

[2] K. F. Ramdhania, D. F. Hidayat, and R. Salkiawati, "Implementasi Metode Naïve Bayes dan Support Vector Machine (SVM) untuk Menganalisis Sentimen Pengguna Twitter terhadap Transjakarta (Implementation of Naïve Bayes and SVM Methods to Analyze Twitter User Sentiment on Transjakarta)," *J. Mat. dan Pendidik. Mat.*, vol. 9, no. 1, pp. 1–14, 2024.

[3] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.

[4] O. P. Zusrotun, A. C. Murti, and R. Fiati, "Sentimen Analisis Belajar Online Di Twitter Menggunakan Naïve Bayes Jurnal Nasional Pendidikan Teknik Informatika : Janapati | 311," *J. Nas. Pendidik. Tek. Inform. JANAPATI* |, vol. 11, pp. 310–320, 2022.

[5] I. B. G. Sarasvananda, D. Selivan, M. L. Radhitya, and I. N. T. A. Putra, "Analisis Sentimen Pada Pembelajaran Daring Di Indonesia Melalui Twitter Menggunakan Naïve Bayes Classifier," *SINTECH (Science Inf. Technol. J.*, vol. 5, no. 2, pp. 227–233, 2022, doi: 10.31598/sintechjournal.v5i2.1241.

[6] H. Zhafran Muflih and F. Noor Hasan, "KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Terhadap Pelayanan TransJakarta Berdasarkan Tweets Menggunakan Metode Naïve Bayes Classifier," *Media Online)*, vol. 4, no. 6, pp. 3044–3052, 2024, doi: 10.30865/klik.v4i6.1927.

[7] N. Safitri, R. Alfiran, D. Tamitiadini, D. Dewi, W. W. A. Febriani, *Analisis Sentimen: Metode Alternatif Penelitian Big Data*. Universitas Brawijaya Press, 2021.

[8] 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.

[9] E. Martantoh and N. Yanih, "Implementasi Metode Naïve Bayes Untuk Klasifikasi Karakteristik Kepribadiaan Siswa Di Sekolah MTS Darussa'adah Menggunakan Php Mysql," *J. Teknol. Sist. Inf.*, vol. 3, no. 2, pp. 166–175, 2022, doi: 10.35957/jtsi.v3i2.2896.