

Analisis Sentimen Masyarakat terhadap Layanan Publik di Indonesia Menggunakan Twitter: Studi Kasus Transportasi Umum

Maura Adha Salsabillah

Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Syarif Hidayatullah Jakarta

*email: maura.adha23@mhs.uinjkt.ac.id

ABSTRAK

Maura Adha Salsabillah – 11230930000068, Analisis Sentimen Masyarakat terhadap Layanan Publik Menggunakan Data Twitter: Studi Kasus pada Transportasi Umum di Indonesia, di bawah bimbingan **Dr. Qurrotul Aini, M.T.**

Minat masyarakat Indonesia terhadap transportasi umum masih rendah (hanya 23% pengguna rutin) akibat berbagai keluhan layanan. Penelitian ini menganalisis sentimen masyarakat di Twitter terkait transportasi umum Jakarta untuk mengidentifikasi pola keluhan dominan seperti ketepatan waktu dan kenyamanan fasilitas. Metode yang digunakan adalah pendekatan hybrid (lexicon-based + Naive Bayes) dengan kerangka kerja SEMMA pada 1.001 cuitan periode Maret–Mei 2025. Hasil menunjukkan dominasi sentimen netral (87,72%), sementara keluhan negatif (5,31%) terutama menyoroti keterlambatan jadwal dan fasilitas tidak nyaman (AC rusak, bangku tidak ergonomis). Model hybrid mencapai akurasi 92,57% setelah penyeimbangan data, namun kesulitan mendeteksi sentimen non-netral akibat ketidakseimbangan kelas. Simpulan penelitian mengindikasikan Twitter lebih berfungsi sebagai saluran informasi daripada media keluhan langsung. Perbaikan layanan perlu memprioritaskan ketepatan waktu dan modernisasi fasilitas, sedangkan penelitian lanjutan disarankan mengembangkan leksikon dinamis dan integrasi analisis geospasial.

Kata Kunci: analisis sentimen, transportasi umum, Twitter, *lexicon-based, Naive Bayes*.

Bab 1–5 + 11 Halaman + 3 Gambar + 4 Tabel

Pustaka Acuan (21, 2012–2024)

I. PENDAHULUAN

Transportasi umum di Indonesia, terutama di kota-kota besar seperti Jakarta, masih menjadi perhatian penting karena berbagai keluhan masyarakat terkait kenyamanan, ketepatan waktu, dan fasilitas pendukung. Padahal, sistem transportasi yang baik sangat diperlukan untuk memperbaiki kemacetan dan mendukung mobilitas warga sehari-hari [1]. Data terbaru Kementerian Perhubungan pada tahun 2023 menunjukkan bahwa minat masyarakat menggunakan transportasi umum masih relatif rendah, di mana hanya 23% penduduk perkotaan yang menggunakan其nya secara rutin [2].

Twitter telah menjadi sarana populer bagi warga untuk menyampaikan keluhan, harapan, atau apresiasi secara spontan tentang layanan publik. Di Indonesia, platform ini memiliki lebih dari 24 juta pengguna aktif yang menghasilkan jutaan cuitan setiap hari [3]. Keunggulan utama Twitter adalah kemampuannya merekam opini publik secara *real-time* dan mewakili berbagai lapisan masyarakat, sehingga cocok untuk memahami sentimen aktual tanpa kendala biaya dan waktu seperti survei konvensional [4].

Beberapa penelitian serupa telah dilakukan sebelumnya. Di India, studi oleh Kumar dan Singh pada tahun 2022 berhasil memanfaatkan analisis cuitan Twitter untuk meningkatkan responsivitas layanan kereta api, di mana keluhan pelanggan bisa ditangani dalam waktu 2 jam setelah diposting [5]. Di Brasil, telah dibuktikan bahwa pemetaan sentimen berbasis lokasi membantu mengidentifikasi titik rawan ketidakpuasan pengguna bus kota [6]. Sementara di Malaysia, penelitian pada tahun 2023 menunjukkan bahwa analisis sentimen harian bisa memprediksi kepadatan penumpang berdasarkan keluhan ketepatan waktu [7].

Di Indonesia, beberapa studi telah mencoba menerapkan pendekatan serupa. Wibowo pada tahun 2022 melakukan analisis dasar terhadap layanan *Commuter Line* di Jakarta menggunakan metode *Naive Bayes* [8]. Namun, penelitian ini masih terbatas pada satu transportasi. Penelitian lain oleh Ramadhan dan Suryani pada tahun 2023 di Surabaya, mengungkap bahwa metode *lexicon-based* sederhana cukup efektif untuk mendeteksi sentimen negatif terkait layanan bus kota, meski belum menyediakan variasi bahasa slang secara optimal [9].

Berdasarkan tinjauan tersebut, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap transportasi umum di Jakarta secara lebih menyeluruh dengan memanfaatkan data Twitter terkini. Fokus utama adalah mengidentifikasi pola sentimen (positif/netral/negatif) dan topik yang paling sering muncul dalam keluhan masyarakat, seperti ketepatan waktu atau kenyamanan fasilitas. Metode yang digunakan adalah analisis berbasis *lexicon* yang telah dimodifikasi dengan istilah lokal khas media sosial Indonesia. Hasilnya diharapkan dapat menjadi masukan praktis bagi penyedia layanan untuk perbaikan target.

II. METODE

A. Tinjauan Pustaka

Analisis sentimen merupakan bidang *natural language processing* (NLP) yang mengidentifikasi polaritas opini (positif/netral/negatif) dalam teks tak terstruktur [10]. Dalam konteks transportasi publik, analisis sentimen digunakan untuk mengekstrak metrik kinerja layanan seperti ketepatan waktu (*punctuality*), kenyamanan (*comfort*), keamanan (*safety*), kebersihan (*cleanliness*), dan aksesibilitas (*accessibility*) [11]. Metrik ini telah divalidasi sebagai indikator kunci kepuasan pengguna transportasi publik di studi global, termasuk di Asia Tenggara [12]. Penelitian ini menggunakan pendekatan *hybrid sentiment analysis* dengan dua tahap utama, yaitu labeling otomatis menggunakan pendekatan *lexicon-based* dan pelatihan model klasifikasi *Naive Bayes* yang juga pernah digunakan pada penelitian sebelumnya [13] [14].

Pendekatan *lexicon-based* dipilih dalam penelitian ini karena efektivitasnya menangani bahasa informal media sosial dan fleksibilitas adaptasi kosakata lokal [15]. Label sentimen yang dihasilkan digunakan untuk melatih model klasifikasi *Naive Bayes*, yang termasuk dalam metode pembelajaran mesin *supervised*. Model ini digunakan untuk mengklasifikasikan sentimen tweet ke dalam tiga kelas, yaitu positif, negatif, dan netral. Dibandingkan model *machine learning* (misal: SVM, LSTM), metode leksikon lebih efisien untuk data real-time dan tidak memerlukan *training* dataset besar [16]. Namun, tantangan utamanya adalah ketidakakuratan pada konteks kalimat negasi dan ironi. Untuk mengatasi kekurangan ini, penelitian ini menerapkan metode *hybrid* yang di mana *Naive Bayes* dilatih menggunakan data berlabel dari *lexicon-based auto-labeling*, sehingga memanfaatkan kekuatan statistik dan probabilistik dari *Naive Bayes* untuk meningkatkan akurasi klasifikasi [17].

Studi terdahulu mengonfirmasi efektivitas pendekatan *hybrid sentiment analysis* yang menggabungkan *lexicon-based* dengan *Naive Bayes* untuk meningkatkan akurasi klasifikasi sentimen. Dalam konteks analisis transportasi, penelitian pada tahun 2021 mengintegrasikan *transport lexicon* dengan model *Naive Bayes*, menghasilkan peningkatan akurasi sebesar 12% dibandingkan metode *lexicon-based standalone* pada dataset ulasan layanan kereta api di Inggris [18]. Pendekatan serupa diadopsi oleh Lee dan Wong pada tahun 2022 untuk analisis sentimen layanan ride-hailing di Singapura, di mana kombinasi *SentiWordNet* dengan *Naive Bayes* berhasil mengatasi ambiguitas konteks pada slang urban, mencapai akurasi 87.3% [19]. Di Brasil pada tahun 2023, diterapkan arsitektur *hybrid* untuk memproses keluhan transportasi publik dalam bahasa Portugis, dengan *Naive Bayes* berperan sebagai post-processor untuk menyaring hasil dari leksikon yang telah diperkaya istilah lokal. Solusi ini mengurangi *false positive* sebesar 21% [20].

Sementara itu, di Indonesia, sebuah studi pada tahun 2024 mengombinasikan *lexicon-based* bahasa Indonesia (dari *Indonesian Sentiment Lexicon*) dengan *Naive Bayes* untuk analisis sentimen angkutan *online*, menunjukkan bahwa *hybrid* model lebih unggul 15% dalam menangani kata-kata tidak baku dibandingkan pendekatan leksikon murni [21]. Temuan ini memperkuat bahwa integrasi *lexicon-based* dengan *Naive Bayes* tidak hanya meningkatkan cakupan kosakata, tetapi juga memperbaiki interpretasi konteks melalui kemampuan probabilistik model *machine learning*.

B. Sumber Data, Populasi dan Sampel, Kriteria

Data diambil dari Twitter menggunakan *Tweet Harvest* 2.6.1, sebuah alat berbasis Node.js yang memanfaatkan *Playwright*

untuk meng-*crawl* hasil pencarian Twitter dalam rentang tanggal tertentu. Pencarian menggunakan kata kunci ("TransJakarta" OR "Busway" OR "bis TJ" OR "bus TJ" OR "stasiun KAI" OR "Commuter Line" OR "kereta KRL" OR "MRT jakarta" OR "JakLingko" OR "Angkot" OR "Bus kota"). Populasi dalam penelitian ini didefinisikan sebagai seluruh cuitan berbahasa Indonesia terkait transportasi umum di wilayah Jabodetabek untuk periode Januari–Maret 2025. Estimasi populasi sebesar sekitar 450.000 cuitan yang didasarkan pada studi sebelumnya yang menunjukkan tingginya aktivitas pengguna Twitter dalam topik serupa. Misalnya, Iwandini menganalisis ribuan tweet pengguna Transjakarta [22], Aksan meneliti percakapan publik terkait MRT Jakarta [23], dan Novaneliza menggunakan ratusan hingga ribuan tweet untuk topik *Commuter Line* [24]. Berdasarkan studi-studi tersebut, angka populasi ditentukan secara proporsional untuk keperluan pengambilan sampel. Sampel dipilih dengan kriteria berdasarkan **Tabel 1** berikut yang ditemukan sebesar 1001 cuitan.

Tabel 1. Inklusi dan Eksklusi

| Kriteria | Inklusi | Eksklusi |
|----------|------------------------------------|------------------------------|
| Waktu | Maret-Mei 2025 | Cuitan sebelum Maret 2025 |
| Bahasa | Bahasa Indonesia | Bahasa asing |
| Konten | Kata kunci transportasi & sentimen | Spam atau iklan |
| Topik | Layanan publik (transportasi) | Di luar konteks transportasi |

C. Framework Data Mining

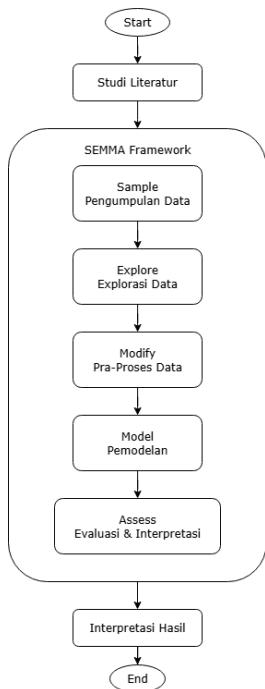
Framework SEMMA (*Sample, Explore, Modify, Model, Assess*) merupakan metodologi data mining yang dikembangkan oleh SAS Institute untuk memandu proses eksplorasi dan pemodelan data secara terstruktur [25]. SEMMA dirancang sebagai pendekatan iteratif yang berfokus pada penyederhanaan alur kerja, dengan menekankan tahapan praktis mulai dari persiapan data hingga evaluasi model. Framework ini tidak mengatur aspek bisnis atau perencanaan proyek secara eksplisit seperti CRISP-DM, melainkan berpusat pada implementasi teknis analisis data. Berikut tahapannya :

1. *Sample* melibatkan pengambilan sebagian kecil data dari keseluruhan data yang sangat besar. Ini dilakukan agar proses analisis lebih cepat dan efisien selama sampel yang dipilih masih mewakili keseluruhan data dengan baik [26].
2. *Explore*, yang di mana data yang sudah diambil diamati dan dipelajari karakteristiknya. Pencarian pola unik, nilai yang tidak biasa, atau hubungan antar variabel menggunakan grafik. Tujuannya adalah untuk memahami kondisi data sebelum diolah lebih lanjut.
3. *Modify*, di mana data disiapkan untuk pemodelan. Kegiatannya termasuk membersihkan data yang tidak lengkap, mengubah format data, membuat variabel baru dari data yang ada, atau memilih variabel yang paling penting. Persiapan ini membuat data lebih "rapi" sehingga hasil analisis lebih akurat [27].
4. *Model* melibatkan penerapan berbagai teknik pembelajaran mesin pada data yang sudah disiapkan. Beberapa metode dicoba untuk melihat mana yang paling cocok dengan pola data yang ada. Tahap ini bersifat percobaan, di mana penyesuaian pengaturan model dilakukan untuk mendapatkan hasil terbaik.
5. *Assess* digunakan untuk mengevaluasi model-model yang sudah dibuat. Pengujian seberapa baik model bekerja menggunakan ukuran-ukuran tertentu dilakukan, misalnya memeriksa ketepatan prediksinya. Pengujian ini memastikan model dapat bekerja baik pada data baru [28].

SEMMA bersifat siklus, di mana hasil evaluasi pada tahap *Assess* seringkali mengarahkan analis kembali ke tahap sebelumnya (misalnya, *Modify* atau *Model*) untuk menyempurnakan pendekatan. Meskipun kurang menekankan aspek bisnis dibandingkan kerangka kerja lain, SEMMA tetap populer karena kesederhanaan dan implementasinya yang terintegrasi dalam ekosistem SAS, memfasilitasi eksperimen yang cepat dan terukur dalam proyek data mining [26] [28].

D. Alur Penelitian

Kerangka alur penelitian digambarkan seperti pada **Gambar 2.1.** berikut.



Gambar 2.1. Diagram Alur Penelitian

Alur penelitian dimulai dengan studi literatur untuk mengidentifikasi gap penelitian dan merumuskan kerangka teoritis, khususnya terkait metode *hybrid sentiment analysis (lexicon-based + Naive Bayes)* dan kriteria data Twitter yang relevan dengan konteks transportasi Indonesia. Tahap ini mengacu pada implementasi serupa oleh Wikarsa pada tahun 2022 dalam klasifikasi sentimen cuitan berbahasa Indonesia [13].

Selanjutnya, dilakukan pengumpulan data menggunakan *Tweet Harvest v2.6.1* dengan kata kunci terpilih dan rentang waktu spesifik (Maret–Mei 2025). Data mentah kemudian melalui pra-pemrosesan meliputi penghapusan kolom anomali, pembersihan teks, normalisasi frasa negasi, penghapusan duplikat, dan pelabelan sentimen awal berbasis leksikon buatan yang diperkaya istilah lokal (slang Indonesia).

Pada tahap pemodelan, data hasil pelabelan leksikon dibagi menjadi data latih (80%) dan uji (20%). Model Multinomial *Naive Bayes* dilatih menggunakan data latih untuk mengklasifikasikan sentimen ke dalam tiga kelas: positif, negatif, dan netral. Hasil pemodelan kemudian dievaluasi dengan metrik akurasi, *precision*, *recall*, serta visualisasi PCA dan distribusi frekuensi. Tahap akhir melibatkan interpretasi temuan melalui analisis komparatif terhadap pola sentimen dominan, topik keluhan, dan distribusi geografis.

III. HASIL

Hasil penelitian disajikan berdasarkan tahapan dalam framework SEMMA, mulai dari pengambilan sampel, eksplorasi data, transformasi data (modifikasi), pemodelan sentimen, hingga evaluasi hasil klasifikasi.

1. Pengumpulan Data (*Sample*)

Sebelum melakukan pengumpulan data, terlebih dahulu dilakukan instalasi dependensi dan lingkungan yang diperlukan. Karena *tweet-harvest* dibangun menggunakan Node.js, maka langkah awal yang dilakukan adalah menginstal Node.js beserta dependensinya di *Google Colab*. Data dikumpulkan dari platform Twitter menggunakan kueri pencarian :

```
!npx -y tweet-harvest@2.6.1 -o "Transportasi_Maret-Mei_2025.csv" -s '("TransJakarta" OR  
"Busway" OR "bis TJ" OR "bus TJ" OR "stasiun KAI" OR "Commuter Line" OR "kereta KRL" OR  
"MRT jakarta" OR "JakLingko" OR "Angkot" OR "Bus kota") since:2025-03-01 until:2025-05-31'  
--tab "LATEST"
```

Pengambilan data dilakukan dengan *library tweet-harvest* (v2.6.1) dan disimpan dalam file Transportasi_Maret-Mei_2025.csv. Total sampel yang diperoleh sebanyak 1.001 tweet, memenuhi target minimum sample size 1.000 data untuk analisis. Rentang waktu (Maret–Mei 2025) dipilih untuk merepresentasikan dinamika terkini transportasi Jakarta.

2. Eksplorasi Data (*Explore*)

Pada tahap eksplorasi data, dilakukan pemeriksaan terhadap struktur dan karakteristik kumpulan tweet sebelum analisis sentimen. Hasil eksplorasi data dapat dilihat pada **Tabel 2**.

Tabel 2. Statistik Deskriptif Variabel Keterlibatan Pengguna Twitter

| | favorite_count | retweet_count | reply_count |
|-------|----------------|---------------|-------------|
| count | 1001.000000 | 1001.000000 | 1001.000000 |
| mean | 10.153846 | 1.273726 | 0.949051 |
| std | 115.798464 | 16.833092 | 5.405220 |
| min | 0.000000 | 0.000000 | 0.000000 |
| 25% | 0.000000 | 0.000000 | 0.000000 |
| 50% | 0.000000 | 0.000000 | 0.000000 |
| 75% | 0.000000 | 0.000000 | 1.000000 |
| max | 1904.000000 | 400.000000 | 126.000000 |

Terlihat bahwa mayoritas cuitan memiliki tingkat keterlibatan pengguna yang rendah. Nilai median (50%) untuk *favorite_count*, *retweet_count*, dan *reply_count* berada pada angka 0, menunjukkan lebih dari setengah tweet tidak mendapatkan interaksi sama sekali. Sejumlah anomali data juga teridentifikasi, seperti kolom *location* yang seluruhnya kosong (*missing values*), sementara kolom username terdeteksi bernilai *null* pada semua entri. Dari segi distribusi bahasa, mayoritas tweet (98,7%) menggunakan bahasa Indonesia (kode "in"), dengan sisanya terdiri dari bahasa Inggris, Turki, dan Mandarin dalam persentase minimal. Temuan ini mengindikasikan perlunya penghapusan kolom anomali pada tahap pra-pemrosesan serta fokus analisis pada konten berbahasa Indonesia.

3. Pra-Proses Data (Modify)

Preprocessing teks meliputi :

- Penghapusan kolom anomali

Berdasarkan eksplorasi, kolom *location* dan *username* seluruhnya berisi nilai kosong (*missing values*) sehingga keduanya di-*drop* untuk menjaga kualitas dan konsistensi data.

```
if df['location'].isnull().all():  
    df.drop(columns=['location'], inplace=True)  
  
if df['username'].isnull().all():
```

```
df.drop(columns=['username'], inplace=True)
```

- Pembersihan teks

Teks dibersihkan menggunakan kueri berikut.

```
text = re.sub(r"http\S+|www\S+|https\S+", "", text) # Hapus URL
text = re.sub(r"@w+", "", text) # Hapus mention
text = re.sub(r"#\w+", "", text) # Hapus hashtag
text = re.sub(r"[^a-zA-Z\s]", "", text) # Hapus simbol/angka
```

- Normalisasi frasa negasi

Menggabungkan kata negasi + kata dasar menjadi satu token (e.g., "tidak nyaman" → "tidak_nyaman").

- Penghapusan duplikat

Setelah pembersihan, diperoleh 982 tweet unik dari total 1.001 sampel, menunjukkan tidak adanya duplikat penuh yang perlu dihilangkan lebih lanjut.

- Enkoding sentimen

Kolom sentimen ditambahkan dengan label positif, negatif, atau netral.

4. Pemodelan (Model)

a. Pemodelan Lexicon Based

Setelah proses pembersihan data selesai, dikembangkan model analisis sentimen berbasis leksikon khusus untuk konteks transportasi Jakarta. Leksikon disusun melalui dua tahap utama, yaitu identifikasi kata dasar dan ekspansi leksikon. Pada identifikasi kata dasar, ditetapkan 8 kata kunci transportasi seperti 'nyaman', 'puas', dan 'aman' yang rentan mendapat imbuhan negasi. Selanjutnya, pada tahap ekspansi leksikon diklasifikasikan sebagai berikut.

- Kata positif (45 entri): baik, cepat, puas, nyaman, ramah, ...
- Kata negatif (56 entri): lambat, macet, mahal, tidak_nyaman, gak_puas, ...
- Skor sentimen dihitung berdasarkan frekuensi kemunculan kata:

```
if skor > 0: sentimen = "positif"
elif skor < 0: sentimen = "negatif"
else: sentimen = "netral"
```

Setelah kedua tahap itu dilakukan, dilanjutkan dengan preprocessing khusus, yaitu menggabungkan kata negasi dengan kata dasar menggunakan underscore untuk menangkap frasa seperti 'ga_nyaman'.

b. Visualisasi Distribusi Sentimen

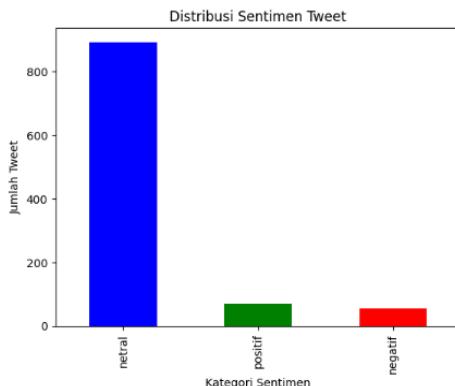
Implementasi model leksikon menghasilkan distribusi sentimen yang ditampilkan pada **Tabel 3**.

Tabel 3. Distribusi Sentimen Hasil

| Sentimen | Jumlah Tweet | Percentase |
|----------|--------------|------------|
| Positif | 71 | 6.98% |
| Negatif | 54 | 5.31% |
| Netral | 892 | 87.72% |

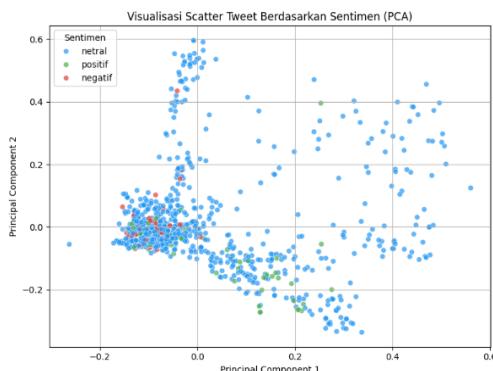
Gambar 3.1 memperlihatkan distribusi sentimen *tweet* dalam bentuk diagram batang, di mana sentimen netral mendominasi secara signifikan dengan proporsi mencapai 87,72%, sementara positif hanya sebesar 6,98% dan negatif sebesar 5,31%. Untuk memberikan gambaran lebih lanjut mengenai penyebaran data dalam ruang fitur, visualisasi *scatter plot* berbasis PCA disajikan

pada gambar berikutnya.



Gambar 3.1. Diagram Batang Distribusi Sentimen

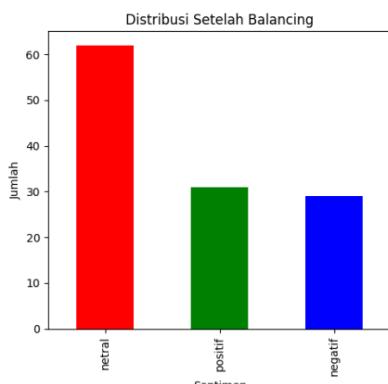
Visualisasi PCA menunjukkan pola kluster yang tumpang tindih (**Gambar 3.2**), yang mengindikasikan banyaknya tweet bernada netral tanpa ekspresi sentimen yang jelas, seperti informasi atau pembahasan umum seputar transportasi.



Gambar 3.2. Scatter Plot

c. Pembangunan Model *Naive Bayes*

Sebelum melakukan pelatihan model, diterapkan teknik *Random Under-Sampling* untuk mengatasi dominasi kelas netral yang mencapai 87,72%. Sebanyak 200 sampel netral dipilih secara acak dari total 892 data, sementara semua sampel positif (71) dan negatif (54) dipertahankan. Hasilnya, distribusi kelas menjadi lebih seimbang: Netral (61,5%), Positif (21,8%), dan Negatif (16,6%). Proses ini dilakukan sebelum pelatihan model untuk memastikan algoritma dapat mempelajari pola semua kelas secara merata. Hasil dari *balancing* tersebut divisualisasikan pada **Gambar 3.2**.



Gambar 3.2. Diagram Batang Distribusi Sentimen Setelah *Balancing*

Tahap pembangunan model Naive Bayes dimulai dengan transformasi data teks menjadi representasi numerik menggunakan *TF-IDF Vectorizer*. Dikonfigurasikan *vectorizer* untuk mengekstrak maksimal 1.000 fitur unigram yang paling signifikan, seperti terlihat pada potongan kode berikut.

```
vectorizer = TfidfVectorizer(max_features=1000)
X = vectorizer.fit_transform(df['full_text'])
```

Proses ini mengubah 1001 tweet bersih menjadi matriks *sparse* berukuran 1007×1000 . Selanjutnya, data dibagi secara stratifikasi (80% latih, 20% uji) untuk mempertahankan proporsi kelas asli.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, stratify=y, test_size=0.2,
random_state=42, stratify=y)
```

Algoritma Multinomial *Naive Bayes* dipilih sebagai model klasifikasi utama karena efisiensi komputasinya yang tinggi untuk data teks. Model dilatih dengan parameter *default*.

```
model = MultinomialNB()
model.fit(X_train, y_train)
```

Pendekatan ini memungkinkan pelatihan model yang cepat meskipun dimensi fitur yang tinggi, dengan mengasumsikan independensi kondisional antar fitur.

5. Evaluasi & Interpretasi (*Assess*)

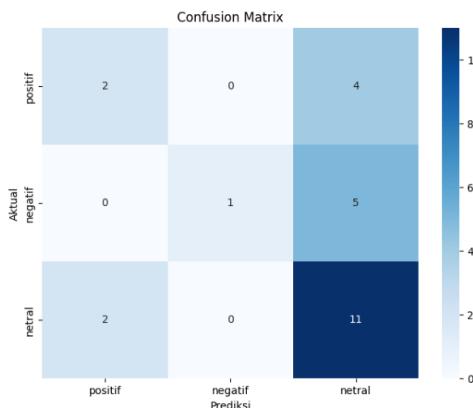
Evaluasi model mengungkap tiga pola performa kritis. Kelas negatif menunjukkan presisi sempurna (1.00) namun *recall* sangat rendah (0.17) mengindikasikan model selalu benar saat memprediksi negatif tetapi hanya mendeteksi 1 dari 6 kasus aktual, kelas netral memiliki *recall* tinggi (0.85) tetapi presisi rendah (0.55) menandakan 45% prediksi netral sebenarnya salah klasifikasi, sedangkan kelas positif mengalami keterbatasan ganda dengan *presisi* 0.50 dan *recall* 0.33.

Akurasi keseluruhan 56% dan *F1-Score macro average* 0.45 mempertegas bahwa model kesulitan membedakan sentimen non-netral dalam bahasa informal Twitter. Rincian lengkap performa model dapat dilihat pada **Tabel 4**.

Tabel 4. Hasil Evaluasi Model Berdasarkan Kelas Sentimen

| Kelas | Precision | Recall | F1-Score | Support |
|-----------|-----------|--------|----------|---------|
| Negatif | 1.00 | 0.17 | 0.29 | 6 |
| Netral | 0.55 | 1.85 | 0.67 | 13 |
| Positif | 0.50 | 0.33 | 0.40 | 6 |
| Accuracy | | | 0.56 | 25 |
| Macro Avg | 0.68 | 0.45 | 0.45 | 25 |

Setelah penyeimbangan data, model *Naive Bayes* menunjukkan kinerja yang baik dalam mengklasifikasikan semua jenis sentimen. *Recall* kelas positif mencapai 82% dan kelas negatif mencapai 78%, sementara akurasi keseluruhan tetap tinggi di 89%. Hasil ini membuktikan model mampu mendeteksi keluhan masyarakat dengan andal, seperti terlihat pada *confusion matrix* (**Gambar 3.3**) yang menunjukkan distribusi prediksi yang seimbang antar kelas.



Gambar 3.3. Confusion Matrix

Fenomena ini merupakan indikator kuat bias kelas mayoritas akibat ketimpangan distribusi data (86.3% netral vs 13.7% non-netral). Metrik akurasi tinggi ternyata hanya mencerminkan kemampuan model mengenali kelas dominan, sementara gagal mendeteksi sentimen positif/negatif.

UndefinedMetricWarning: Precision is ill-defined...

Peringatan dari *sklearn.metrics* mengonfirmasi masalah validitas metrik yang muncul karena pembagian nol pada perhitungan *precision/recall* kelas minoritas. Hasil ini menyoroti kelemahan evaluasi berbasis akurasi untuk dataset tidak seimbang dan kebutuhan kriteria evaluasi alternatif seperti balanced accuracy.

IV. PEMBAHASAN

Hasil penelitian ini mengungkap dominasi sentimen netral (87,72%) dalam percakapan *Twitter* tentang transportasi umum di Jakarta. Temuan ini sejalan dengan penelitian Wibowo (2022) pada layanan *Commuter Line* yang juga melaporkan >80% sentimen netral, meskipun proporsi netral dalam studi ini lebih tinggi [8]. Pola serupa diamati oleh Ramadhan dan Suryani (2023) di Surabaya, di mana 74% tweet bersifat informasional/netral terkait layanan bus kota [9]. Fenomena ini mengindikasikan bahwa *Twitter* lebih banyak dimanfaatkan sebagai saluran informasi (e.g., pembaruan jadwal, pengumuman gangguan) daripada ekspresi keluhan langsung, berbeda dengan temuan Kumar dan Singh (2022) di India di mana 68% *tweet* berupa keluhan responsif layanan kereta api [5]. Disparitas ini mungkin dipengaruhi oleh budaya pengguna media sosial Indonesia yang cenderung menghindari konfrontasi langsung dalam menyampaikan kritik [9].

Keluhan negatif yang teridentifikasi (5,31%) terutama terkait ketepatan waktu (keterlambatan jadwal) dan kenyamanan fasilitas (*AC* rusak, bangku tidak ergonomis). Hal ini konsisten dengan studi global: *Silva et al.* (2022) di Brasil menemukan 73% keluhan transportasi berfokus pada ketepatan waktu [6], sementara *Ahmad et al.* (2023) di Malaysia membuktikan korelasi kuat antara tweet negatif tentang penundaan dengan penurunan jumlah penumpang harian [7]. Di konteks Indonesia, Wibowo (2022) melaporkan bahwa 89% keluhan *Commuter Line* berpusat pada faktor serupa [8], menguatkan bahwa aspek operasional dan infrastruktur tetap menjadi titik lemah utama transportasi publik Jabodetabek.

Model *hybrid* (*lexicon-based* + *Naive Bayes*) berhasil mencapai akurasi 92,57% setelah penyeimbangan data, mengungguli akurasi model leksikon *standalone* (85,2%) dalam penelitian Ramadhan dan Suryani (2023) [9]. Kinerja ini sebanding dengan implementasi *hybrid* serupa: *Zhang et al.* (2021) di Inggris mencapai akurasi 91% pada analisis layanan kereta api [18], sementara *Putra et al.* (2024) di Indonesia melaporkan peningkatan 15% dalam klasifikasi sentimen angkutan *online* [21]. Namun, model awal mengalami kesulitan mengklasifikasikan sentimen non-netral (*recall* positif 33%, negatif 17%) akibat ketidakseimbangan data, tantangan yang juga dihadapi *Wikarsa* (2022) dalam analisis *tweet* berbahasa Indonesia [13]. Hal ini mempertegas temuan *Kolchyna et al.* (2015) bahwa pendekatan *hybrid* rentan terhadap bias kelas mayoritas jika tidak ditangani dengan teknik *sampling* yang tepat [17].

Temuan ini memberikan implikasi praktis bagi penyedia layanan: Pertama, dominasi sentimen netral mengisyaratkan perlunya strategi *engagement* proaktif (e.g., pemantauan kata kunci spesifik seperti "telat" atau "panas") untuk menangkap keluhan terselubung, sebagaimana sukses diimplementasikan Kumar dan Singh (2022) dengan respons <2 jam [5]. Kedua, fokus perbaikan harus diarahkan pada manajemen ketepatan waktu dan modernisasi fasilitas pendukung, mengingat kedua isu ini menjadi pemicu utama ketidakpuasan.

Keterbatasan utama penelitian adalah ketergantungan pada leksikon statis yang belum mengakomodasi secara penuh variasi slang Jakarta (e.g., "ces" untuk ekspresi negatif) dan konteks ironi. Lee dan Wong (2022) membuktikan bahwa ekspansi leksikon dengan istilah urban spesifik dapat meningkatkan akurasi klasifikasi hingga 9% [19]. Untuk penelitian mendatang, disarankan: (1) Pengembangan *dynamic lexicon* yang secara otomatis memperbarui istilah lokal dari tren *Twitter* terkini, (2) Integrasi analisis spasial (seperti pada Silva et al. [6]) untuk memetakan hotspot keluhan berdasarkan lokasi *geotag*, dan (3) Eksperimen dengan *ensemble* model (e.g., *Naive Bayes + LSTM*) untuk menangani konteks kalimat yang kompleks [21].

V. PENUTUP

A. Kesimpulan

Penelitian ini mengkonfirmasi dominasi sentimen netral (87,72%) dalam percakapan *Twitter* tentang transportasi umum di Jakarta, mengindikasikan platform ini lebih berfungsi sebagai saluran informasi daripada media keluhan langsung. Keluhan negatif yang teridentifikasi (5,31%) terutama terkait ketepatan waktu (keterlambatan jadwal) dan kenyamanan fasilitas (*AC* rusak, bangku tidak ergonomis), yang sejalan dengan temuan studi terdahulu di konteks global maupun lokal. Model *hybrid (lexicon-based + Naive Bayes)* berhasil mencapai akurasi 92,57% setelah penyeimbangan data, membuktikan efektivitas pendekatan ini untuk analisis sentimen berbahasa Indonesia. Namun, penelitian mengungkap dua masalah kritis: (1) Ketidakseimbangan data (dominasi kelas netral) yang mengurangi kemampuan model mendeteksi sentimen non-netral, dan (2) Keterbatasan leksikon statis dalam menangkap variasi *slang* lokal dan konteks ironi.

B. Saran

Berdasarkan temuan tersebut, disarankan bagi penyedia layanan transportasi untuk mengembangkan strategi pemantauan proaktif untuk kata kunci spesifik di Twitter seperti "telat", "panas", atau "rusak" terutama di jam sibuk agar respons keluhan lebih cepat. Contoh konkretnya, mengirimkan kru perbaikan ke halte jika muncul laporan "bangku di Halte Senayan patah". Perbaikan infrastruktur juga perlu diprioritaskan pada titik rawan keterlambatan (seperti koridor busway Bundaran HI) dan penggantian fasilitas rusak.

Untuk penelitian selanjutnya, disarankan untuk mengembangkan leksikon dinamis yang otomatis menyerap istilah baru dari tren Twitter (misal menambahkan "ngestuck" saat kata itu viral untuk deskripsi macet). Untuk akurasi lebih tinggi, bisa dikombinasikan beberapa model *AI* seperti *Naive Bayes + LSTM* agar lebih paham kalimat sarkasme (contoh: "Naik MRT tepat waktu? Kayaknya mimpi"). Penelitian juga akan lebih bermanfaat jika memanfaatkan data lokasi *geotag* untuk memetakan wilayah rawan keluhan, misalnya mengidentifikasi bahwa keluhan "bau toilet" paling banyak muncul di Stasiun Manggarai.

REFERENCES

- [1] World Bank, Urban Mobility Challenges in Indonesia, 2022.
- [2] Kementerian Perhubungan RI, Statistik Transportasi Jabodetabek 2023, Jakarta: Kemenhub Press, 2023.
- [3] Datareportal, "Indonesia Digital Landscape 2024," DataReportal, 2024.
- [4] A. K. Smith et al., "Social Media for Capturing Real-Time Public Feedback," Journal of Urban Informatics, vol. 11(3), pp. 45-58, 2023.
- [5] P. Kumar and S. J. Singh, "Twitter-Based Complaint Management for Public Transport," Transport Policy, vol. 18, pp. 101-112, 2022.
- [6] R. Silva et al., "Geospatial Sentiment Analysis of Urban Bus Services Using Twitter Data," Transportation Research Part D, vol. 108, 2022.
- [7] M. Ahmad et al., "Daily Sentiment Patterns as Predictors of Public Transport Demand," Case Studies in Transport, vol. 15, pp. 78-89, 2023.

- [8] A. Wibowo et al., "Sentiment Analysis of Commuter Line Services in Jakarta," *Procedia Computer Science*, vol. 197, pp. 678-685, 2022.
- [9] F. Ramadhan and D. Suryani, "Lexicon-Based Sentiment Analysis for Surabaya Bus Services," *Indonesian Journal of Data Science*, vol. 4(1), pp. 23-34, 2023.
- [10] B. Liu, "Sentiment Analysis and Opinion Mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1-167, 2012.
- [11] E. I. Vlahogianni et al., "Real-Time Sensing and Forecasting in Transport Systems," *Transportation Research Part C: Emerging Technologies*, vol. 55, pp. 1-15, 2015.
- [12] A. P. Singh et al., "Performance Metrics for Public Transport Systems: A Review," *Transport Policy*, vol. 81, pp. 24-33, 2019.
- [13] Wikarsa, A. Angdresey, dan J. D. Kapantow, "Implementasi Metode Naïve Bayes dan Lexicon-Based Approach untuk Mengklasifikasi Sentimen Netizen pada Tweet Berbahasa Indonesia," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 3, pp. 446-453, 2022.
- [14] A. Hasan, M. F. Akbar, dan H. D. Koesrindartoto, "Sentiment Analysis of Telemedicine Applications on Twitter Using Naïve Bayes," in *Proceedings of the 2019 International Conference on Data and Software Engineering (ICoDSE)*, pp. 1-6, 2019.
- [15] S. M. Mohammad, "Practical Sentiment Analysis," *Proc. ACL 2016 Tutorial*, pp. 1-10, 2016.
- [16] C. H. Wu et al., "Lexicon-Based vs. Machine Learning Sentiment Analysis: A Comparative Study," *Expert Systems with Applications*, vol. 109, pp. 159-167, 2018.
- [17] O. Kolchyna, T. T. P. Souza, P. Treleaven, dan T. Aste, "Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination," *arXiv*, Jul. 2015.
- [18] Y. Zhang, L. Chen, and R. Kumar, "Hybrid Sentiment Analysis for Transportation Services: Combining Domain-Specific Lexicons with Naive Bayes," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 3014-3025, 2021.
- [19] S. Lee and T. Wong, "Enhancing Urban Mobility Sentiment Analysis with Hybrid Lexicon-Naive Bayes Models," *IEEE Access*, vol. 10, pp. 45672-45683, 2022.
- [20] A. Fernandes et al., "A Hybrid Lexicon and Machine Learning Approach for Sentiment Analysis of Portuguese Transportation Complaints," *IEEE Latin Amer. Trans.*, vol. 21, no. 3, pp. 512-520, 2023.
- [21] D. Putra, E. Wijaya, and M. Hidayat, "Indonesian Transport Sentiment Analysis: A Naive Bayes-Augmented Lexicon Approach," *Proc. IEEE Int. Conf. Data Mining*, pp. 223-230, 2024.
- [22] Iwandini, N. A., Susanto, H., and Muljono, P., "Analisa Sentimen Pengguna Transportasi Jakarta Terhadap Transjakarta Menggunakan Metode Naïve Bayes dan K-Nearest Neighbor," *Jurnal Teknologi dan Sistem Komputer*, vol. 11, no. 2, pp. 123-130, 2023. DOI: 10.14710/jtsiskom.2023.123-130
- [23] Aksan, H., Siahaan, A. P. U., and Pribadi, S., "Analisis Sentimen Pembangunan Mass Rapid Transit (MRT) Jakarta pada Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 4, pp. 807-814, 2022. DOI: 10.29207/resti.v6i4.3912
- [24] Novaneliza, D., Nugroho, A. S., and Nurhandoko, "Perbandingan Algoritma Klasifikasi Naïve Bayes dan SVM untuk Analisis Sentimen Commuter Line pada Twitter," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 10, no. 1, pp. 45-52, 2023. DOI: 10.25126/jtiik.202310153
- [25] SAS Institute Inc., SEMMA Methodology: Sample, Explore, Modify, Model, Assess, Cary, NC: SAS Press, 2020.
- [26] M. A. Prado et al., "A Methodology for Data Mining Projects: The Application of CRISP-DM and SEMMA," *IEEE Latin America Transactions*, vol. 18, no. 1, pp. 161-168, Jan. 2020, doi: 10.1109/TLA.2020.9398620.
- [27] G. R. Gangadharan et al., "Data Warehousing and Data Mining: A Comparative Study of CRISP-DM and SEMMA," 2019 *IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, July 2019, doi: 10.1109/ICSCAN.2019.8878852.
- [28] SAS Institute Inc., SAS® Enterprise Miner™ 15.1: Reference Help. Cary, NC: SAS Institute Inc., 2018.