

Klasifikasi Sentimen Ulasan Transportasi Umum Kota Surabaya pada Media Sosial X (*Twitter*)

1st Novan Ardani Agustian
Sistem Informasi

Universitas Telkom Surabaya
Surabaya, Indonesia

novanardani@student.telkomuniversity.
ac.id

2nd Anisa Dzulkarnain, S.Kom.,
M.Kom.

Sistem Informasi
Universitas Telkom Surabaya

Surabaya, Indonesia
anisazhulkarnain@
telkomuniversity.ac.id

3rd Dr. Berlian Rahmy Lidiawaty,
S.ST., M.MT.

Sistem Informasi
Universitas Telkom Surabaya

Surabaya, Indonesia
berlianerel@telkomuniversity.ac.id

Abstrak — Transportasi umum merupakan aspek vital dalam infrastruktur perkotaan yang mendukung mobilitas dan pertumbuhan ekonomi. Di Kota Surabaya, meskipun pemerintah telah berupaya meningkatkan layanan melalui Suroboyo Bus, Trans Semanggi, dan Wira Wiri, kemacetan dan ketidakpuasan publik masih menjadi isu utama. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap layanan transportasi umum di Surabaya dengan memanfaatkan ulasan pada media sosial X (sebelumnya Twitter). Metode yang digunakan adalah *machine learning* dengan algoritma *Support Vector Machine* (SVM) untuk mengklasifikasikan ulasan ke dalam sentimen positif, negatif, dan netral. Tahapan penelitian mencakup pengumpulan data (*web scraping*), *pre-processing* data teks, pembobotan fitur menggunakan TF-IDF, penanganan ketidakseimbangan kelas dengan teknik SMOTE, serta optimasi model menggunakan *GridSearchCV*. Hasil penelitian menunjukkan bahwa model SVM dengan kernel RBF mampu menghasilkan akurasi sebesar 82.62%. Analisis lebih lanjut mengungkapkan bahwa sentimen negatif mendominasi (43.76%) dengan fokus pada masalah kemacetan dan waktu tunggu, sementara sentimen positif (33.44%) menyoroti kenyamanan armada, dan sentimen netral (22.80%) berkaitan dengan informasi rute. Temuan ini memberikan wawasan berbasis data bagi pemangku kepentingan untuk perbaikan strategis layanan transportasi publik.

Kata kunci — analisis sentimen, *support vector machine*, transportasi umum, surabaya, *twitter*, *text mining*

I. PENDAHULUAN

Transportasi umum memegang peranan krusial dalam kehidupan perkotaan modern, berfungsi sebagai tulang punggung mobilitas yang memungkinkan konektivitas antarwilayah serta mendukung dinamika sosial dan ekonomi masyarakat [1]. Tujuan utama dari sistem transportasi publik adalah menyediakan aksesibilitas yang efisien, terjangkau, dan merata bagi seluruh lapisan penduduk untuk berpindah dari satu lokasi ke lokasi lain [2]. Di kota metropolitan sebesar Surabaya, yang merupakan pusat ekonomi terbesar kedua di Indonesia, keberadaan sistem transportasi yang handal menjadi kebutuhan mutlak untuk mencegah stagnasi mobilitas akibat kemacetan lalu lintas.

Namun, Surabaya saat ini menghadapi tantangan serius terkait manajemen lalu lintas. Berdasarkan data *TomTom Traffic Index* tahun 2024, Surabaya tercatat sebagai salah satu kota dengan tingkat kemacetan yang mengkhawatirkan, menempati peringkat ke-44 kota termacet di dunia [3]. Kondisi ini memaksa pemerintah kota dan masyarakat untuk mengevaluasi kembali efektivitas moda transportasi massal yang ada sebagai solusi utama pengurai kemacetan. Pemerintah Kota Surabaya telah meluncurkan berbagai inisiatif, termasuk pengoperasian Suroboyo Bus, Trans Semanggi Suroboyo, dan angkutan *feeder* Wira Wiri Suroboyo untuk menjangkau wilayah yang lebih luas.

Meskipun berbagai upaya telah dilakukan, realitas di lapangan menunjukkan adanya kesenjangan yang signifikan antara harapan kebijakan dan pengalaman pengguna. Studi yang dilakukan oleh Widyastuti dan Santoso (2023) mengungkapkan bahwa tingkat kepuasan pengguna terhadap integrasi antarmoda di Surabaya masih rendah, yakni di bawah 60%, terutama pada aspek *first mile* dan *last mile* yang belum terlayani secara optimal [4]. Lebih lanjut, laporan dari media lokal menyoroti bahwa transportasi umum di Surabaya dinilai masih tertinggal dibandingkan kota besar lainnya karena minimnya integrasi fisik maupun sistem pembayaran, serta aksesibilitas yang sulit bagi warga di wilayah pinggiran [5]. Hal ini menyebabkan masyarakat enggan beralih dari kendaraan pribadi ke transportasi umum.

Kompleksitas permasalahan ini tercermin jelas dalam diskursus publik di media sosial. Platform X (sebelumnya Twitter) menjadi saluran utama bagi masyarakat, yang dideskripsikan sebagai demografi yang kritis, untuk menyampaikan keluhan, kritik, dan saran secara *real-time* [6]. Berbeda dengan survei konvensional yang memakan waktu lama, data ulasan di media sosial menawarkan kecepatan dan kejujuran ekspresi pengguna saat menghadapi kendala operasional seperti keterlambatan jadwal (*headway*), rute yang membingungkan, hingga perilaku staf di lapangan [7]. Oleh karena itu, analisis terhadap data ulasan ini menjadi sangat urgen untuk menangkap respons publik secara akurat.

Penelitian ini mengusulkan penggunaan Analisis Sentimen untuk mengekstraksi informasi berharga dari ribuan ulasan tersebut. Analisis sentimen merupakan teknik pemrosesan bahasa alami yang bertujuan untuk

mengidentifikasi dan mengklasifikasikan opini dalam teks ke dalam kategori sentimen positif, negatif, atau netral [8]. Dalam konteks ini, algoritma *Support Vector Machine* (SVM) dipilih sebagai metode klasifikasi utama. SVM dikenal memiliki performa yang unggul dalam menangani data teks berdimensi tinggi dan tidak terstruktur, serta efektif dalam mengatasi masalah *overfitting* dibandingkan metode lain seperti *Naive Bayes* [9], [10]. Penelitian terdahulu menunjukkan bahwa SVM mampu menghasilkan akurasi hingga 84.08% dalam analisis sentimen transportasi *online*, jauh lebih tinggi dibandingkan metode probabilistik lainnya [11].

Tujuan utama penelitian ini adalah untuk: (1) Mengklasifikasikan ulasan pengguna transportasi umum di Surabaya pada media sosial X menggunakan metode SVM; dan (2) Mengevaluasi kinerja model SVM menggunakan *Confusion Matrix* untuk mengukur tingkat akurasi, presisi, *recall*, dan *F1-score*. Hasil dari penelitian ini diharapkan dapat memberikan peta permasalahan yang komprehensif bagi operator transportasi dan pemerintah kota dalam merumuskan kebijakan perbaikan layanan yang tepat sasaran.

II. KAJIAN TEORI

A. Penelitian Terdahulu

Sejumlah penelitian telah dilakukan untuk mengeksplorasi penerapan SVM dalam analisis sentimen. Ladayya et al. (2022) menganalisis sentimen program JakLingko menggunakan SVM dan menemukan bahwa topik utama sentimen positif berkaitan dengan tarif gratis, sedangkan keluhan berfokus pada penggunaan kartu [12]. Alhaq et al. (2021) juga menerapkan SVM untuk analisis ulasan *e-commerce* dengan hasil akurasi mencapai 93%, membuktikan kehandalan algoritma ini pada data teks pendek [13]. Dalam perbandingan algoritma, Dwianto & Sadikin (2021) membuktikan bahwa SVM mengungguli *Naive Bayes* dengan selisih akurasi yang signifikan dalam klasifikasi ulasan transportasi *online* [11]. Berdasarkan tinjauan literatur tersebut, penelitian ini memposisikan diri untuk mengisi celah analisis spesifik pada ekosistem transportasi umum di Surabaya yang memiliki karakteristik unik dengan menggunakan optimasi *hyperparameter*.

B. Text Mining dan Analisis Sentimen

Text mining adalah proses ekstraksi pola dan informasi bermakna dari data teks tidak terstruktur [14]. Salah satu aplikasi utamanya adalah analisis sentimen, yang bertujuan untuk memahami orientasi opini publik. Proses ini melibatkan teknik *Natural Language Processing* (NLP) untuk membedah struktur bahasa. Tantangan utama dalam analisis sentimen pada media sosial X adalah penggunaan bahasa tidak baku, singkatan, slang, dan *noise* karakter yang memerlukan tahap *pre-processing* yang ketat [15].

C. Support Vector Machine (SVM)

SVM adalah algoritma *supervised learning* yang bekerja berdasarkan prinsip pencarian *hyperplane* pemisah terbaik antara dua kelas data [16]. Konsep kunci SVM adalah maksimasi margin, yaitu jarak tegak lurus antara *hyperplane* dengan data terdekat dari masing-masing kelas yang disebut *support vectors*. Untuk data yang tidak dapat dipisahkan secara linear (*non-linear separable*), SVM memanfaatkan teknik *kernel trick*. Fungsi kernel memetakan data dari ruang input berdimensi rendah ke ruang fitur berdimensi tinggi di mana pemisahan linear dimungkinkan [17]. Kernel yang umum digunakan meliputi *Linear*, *Polynomial*, dan *Radial Basis Function* (RBF). Persamaan matematis untuk fungsi keputusan SVM dapat dituliskan sebagai: $f(x) = \text{sign}(\sum \alpha_i y_i K(x_i, x) + b)$ di mana $K(x_i, x)$ adalah fungsi kernel, α_i adalah bobot lagrange, dan b adalah bias.

D. Support Vector Machine (SVM)

SVM adalah algoritma *supervised learning* yang bekerja berdasarkan prinsip pencarian *hyperplane* pemisah terbaik antara dua kelas data [16]. Konsep kunci SVM adalah maksimasi margin, yaitu jarak tegak lurus antara *hyperplane* dengan data terdekat dari masing-masing kelas yang disebut *support vectors*. Untuk data yang tidak dapat dipisahkan secara linear (*non-linear separable*), SVM memanfaatkan teknik *kernel trick*. Fungsi kernel memetakan data dari ruang input berdimensi rendah ke ruang fitur berdimensi tinggi di mana pemisahan linear dimungkinkan [17]. Kernel yang umum digunakan meliputi *Linear*, *Polynomial*, dan *Radial Basis Function* (RBF). Persamaan matematis untuk fungsi keputusan SVM dapat dituliskan sebagai: $f(x) = \text{sign}(\sum \alpha_i y_i K(x_i, x) + b)$ di mana $K(x_i, x)$ adalah fungsi kernel, α_i adalah bobot lagrange, dan b adalah bias.

E. Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF adalah metode pembobotan fitur yang digunakan untuk merepresentasikan pentingnya sebuah kata dalam dokumen relatif terhadap kumpulan dokumen (korpus). Nilai TF menghitung frekuensi kemunculan kata, sedangkan IDF mengukur seberapa jarang kata tersebut muncul di seluruh dokumen. Kata yang sering muncul di satu dokumen tetapi jarang di dokumen lain akan memiliki nilai TF-IDF yang tinggi, yang mengindikasikan bahwa kata tersebut merupakan fitur pembeda yang kuat [18].

F. SMOTE (Synthetic Minority Over-sampling Technique)

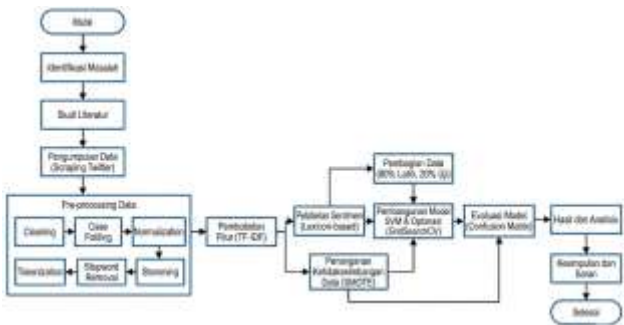
Ketidakseimbangan kelas (*imbalance class*) adalah masalah umum dalam dataset sentimen di mana satu kategori (misalnya negatif) jauh lebih banyak daripada kategori lain. Hal ini dapat menyebabkan model bias ke kelas mayoritas. SMOTE mengatasi hal ini dengan cara membangkitkan data sintesis baru untuk kelas minoritas berdasarkan tetangga terdekat (*k-nearest neighbors*) dari data yang ada, bukan sekadar menduplikasi data [19].

III. METODE PENELITIAN

A. Alur Penelitian

Penelitian ini mengikuti metodologi sistematis yang terdiri dari tahapan: identifikasi masalah, studi literatur, pengumpulan data, *pre-processing*, pembobotan fitur,

pelabelan sentimen, penyeimbangan data (SMOTE), pembangunan model SVM, dan evaluasi.



GAMBAR 1
(FLOWCHART TAHAPAN PENELITIAN)

B. Pengumpulan Data

Data diperoleh melalui teknik *web scraping* pada platform X menggunakan pustaka *Tweepy* dengan bahasa pemrograman Python. Pengambilan data dibatasi pada periode Januari hingga Juli 2024, yang merupakan masa transisi dan evaluasi layanan *feeder* Wira Wiri. Kata kunci (*keywords*) yang digunakan telah divalidasi melalui *pilot search* dan pertimbangan ahli, meliputi: "Angkutan umum Surabaya", "Transportasi umum Surabaya", "Suroboyo Bus", "Trans Semanggi", dan "Feeder Surabaya" [20]. Total data mentah yang berhasil dikumpulkan sebanyak 1.932 ulasan.

C. Pre-processing Data

Penelitian Tahap ini krusial untuk membersihkan *noise* agar kualitas data terjamin. Langkah-langkahnya meliputi:

1. **Cleaning:** Menghapus elemen yang tidak relevan seperti URL, *username* (@user), *hashtag* (#), angka, tanda baca, emoji, dan karakter non-alfanumerik menggunakan *Regular Expressions* (Regex).
2. **Case Folding:** Mengonversi seluruh huruf menjadi huruf kecil (*lowercase*) untuk menyeragamkan data.
3. **Normalization:** Mengubah kata tidak baku, singkatan, atau bahasa gaul (*slang*) menjadi bentuk baku bahasa Indonesia menggunakan kamus leksikon yang telah disusun (contoh: "yg" menjadi "yang", "gak" menjadi "tidak").
4. **Tokenization:** Memecah kalimat menjadi potongan kata (*token*) individual.
5. **Stopword Removal:** Menghapus kata-kata umum yang sering muncul namun tidak memiliki makna sentimen yang signifikan (seperti "dan", "yang", "di", "ke") menggunakan pustaka NLTK Bahasa Indonesia.
6. **Stemming:** Mengubah kata berimbuhan menjadi kata dasar menggunakan algoritma Sastrawi (contoh: "menunggu" menjadi "tunggu").

D. Pembobotan TF-IDF dan Pelabelan

Setelah bersih, data teks diubah menjadi vektor numerik menggunakan TF-IDF. Pelabelan sentimen dilakukan secara otomatis berbasis leksikon (*lexicon-based*) dengan kamus kata positif dan negatif. Skor sentimen dihitung dengan pengurangan jumlah kata positif dengan kata negatif. Jika skor > 0 maka "Positif", skor < 0 maka "Negatif", dan skor = 0 maka "Netral".

E. Pemodelan

Dataset dibagi menjadi data latih (80%) dan data uji (20%). Sebelum pelatihan, data latih yang tidak seimbang ditangani menggunakan metode SMOTE untuk menyamakan proporsi kelas. Model SVM dibangun dan dioptimasi menggunakan *GridSearchCV* dengan skema *K-Fold Cross Validation* (k=5) untuk mencari parameter terbaik (Kernel, C, dan Gamma).

IV. HASIL DAN PEMBAHASAN

A. Hasil Pre-processing Data

Proses *pre-processing* berhasil mereduksi *noise* secara signifikan. Dari 1.932 data awal, dilakukan penghapusan duplikat sehingga tersisa 1.861 data unik. Tabel berikut menunjukkan contoh transformasi data pada beberapa tahapan.

TABEL 1
(HASIL CLEANING)

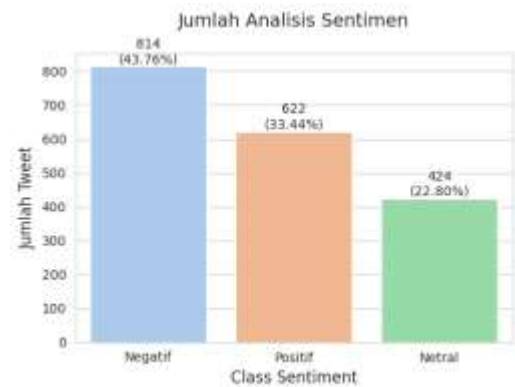
No.	Username	Full Text (Asli)	Hasil Cleaning
1.	stravenues	Kesimpulan dari 2 jam sambat soal Surabaya (dan lain-lainnya): Komitmen transum masih rendah Prioritas pembangunan ga terarah 3. Bapuknya transum mempengaruhi ke sektor lain misal olahraga dan kesehatan 4. Kuliah kedokteran Next spaces enaknya bahas apa lagi ya?	Kesimpulan dari jam sambat soal Surabaya dan lainlainnya 1. Komitmen transum masih rendah 2. Prioritas pembangunan ga terarah 3. Bapuknya transum mempengaruhi ke sektor lain misal olahraga dan kesehatan 4. Kuliah kedokteran Next spaces enaknya bahas apa lagi ya
2.	archiefferry	@tfsurabaya Ini yang aku sering notice min. Setiap aku naik transum di Surabaya khususnya armada wirawiri itu mayoritas bapak/ibu yang udah sepuh-sepuh. Dan mereka naik mereka 'kenal' sama helpernya yang indikasinya mereka udah menjadikan moda ini se	Ini yang aku sering notice min Setiap aku naik transum di Surabaya khususnya armada wirawiri itu mayoritas bapak/ibu yang udah sepuh-sepuh Dan mereka naik mereka kenal sama helpernya yang indikasinya mereka udah menjadikan moda ini se
3.	robotdiesel	@thecancerg0at bowleeeh transum surabaya walau jadwalnya gak pasti tapi ttp ngelewatin	bowleeeh transum surabaya walau jadwalnya gak pasti tapi ttp ngelewatin

		semua mall jadinya enak bgt	semua mall jadinya enak bgt
4.	pravasta_rpp	@tfsurabaya Bebal kata yg tepat utk mereka para pemangku jabatan yg berwenang utk pembangunan transportasi massal yg cepat tepat dan terjangkau semua lapisan masyarakat sehingga transum di Surabaya Raya msh apa adanya.	Bebal kata yg tepat utk mereka para pemangku jabatan yg berwenang utk pembangunan transportasi massal yg cepat tepat dan terjangkau semua lapisan masyarakat sehingga transum di Surabaya Raya msh apa adanya
5.	by_inkriya	COBA SINI LO LIAT HALTE DI SURABAYA KEK APAAN. PETUGAS MANA YANG BISA DITANYA ORANG GOIB SEMUA SAMPE WUJUD HALTENYA CUMA PLANG DOANG. GOOGLING JUGA APAAN YANG BISA DICARI ORANG PEMKOTNYA AJA GAK NIAT NGURUS TRANSUM. EMOSI NI GUE ARIE MENDING LO MUNDUR MENKOMINFO AMPAS	COBA SINI LO LIAT HALTE DI SURABAYA KEK APAAN. PETUGAS MANA YANG BISA DITANYA ORANG GOIB SEMUA SAMPE WUJUD HALTENYA CUMA PLANG DOANG. GOOGLING JUGA APAAN YANG BISA DICARI ORANG PEMKOTNYA AJA GAK NIAT NGURUS TRANSUM. EMOSI NI GUE ARIE MENDING LO MUNDUR MENKOMINFO AMPAS

B. Analisis Distribusi Sentimen

Hasil pelabelan sentimen menunjukkan bahwa opini publik terhadap transportasi umum di Surabaya cenderung terpolarisasi dengan dominasi sentimen negatif. Sebanyak 814 ulasan (43.76%) diklasifikasikan sebagai Negatif, 622 ulasan (33.44%) sebagai Positif, dan 424 ulasan (22.80%) sebagai Netral.

Dominasi sentimen negatif mengindikasikan bahwa masih banyak aspek layanan yang belum memenuhi harapan masyarakat, khususnya terkait keandalan operasional. Namun, persentase sentimen positif yang mencapai sepertiga dari total data menunjukkan adanya apresiasi terhadap peningkatan fasilitas yang telah dilakukan.



GAMBAR 2
(JUMLAH ANALISIS SENTIMEN)

C. Visualisasi Word Cloud

Untuk memahami konteks di balik sentimen, dilakukan analisis *Word Cloud*.

1. Sentimen Negatif: Kata-kata yang paling dominan muncul adalah "macet", "butuh", "lambat", "susah", "lama", dan "jarang". Kemunculan kata "macet" dan "lambat" berkorelasi kuat dengan keluhan mengenai waktu tempuh dan *headway* bus yang tidak pasti. Kata "susah" sering dikaitkan dengan akses pembayaran atau lokasi halte yang tidak ramah pejalan kaki.



GAMBAR 3
(SENTIMEN NEGATIF)

2. Sentimen Positif: Didominasi oleh kata "bagus", "nyaman", "integrasi", "bersih", "dingin" (AC), dan "murah". Ini menunjukkan bahwa dari sisi fisik armada (kenyamanan, kebersihan, pendingin udara) dan tarif, layanan transportasi Surabaya sudah dinilai baik oleh masyarakat. Kata "integrasi" muncul positif, kemungkinan merujuk pada kemudahan transfer di titik-titik tertentu yang sudah terhubung.



GAMBAR 4
(SENTIMEN POSITIF)

3. **Sentimen Netral:** Kata kunci meliputi "rute", "akses", "info", dan "halte". Ulasan ini umumnya berisi pertanyaan tentang rute, informasi operasional, atau saran objektif tanpa muatan emosi.



GAMBAR 5
(SENTIMEN NETRAL)

D. Optimasi Hyperparameter SVM

Penerapan *GridSearchCV* pada data latih menghasilkan kombinasi parameter terbaik untuk model SVM sebagai berikut:

- Kernel: RBF (*Radial Basis Function*)
- C (Regularization parameter): 10
- Gamma: 1 Penggunaan kernel RBF menunjukkan bahwa batas keputusan antar kelas sentimen pada data ini bersifat non-linear, dan parameter C yang cukup besar (10) mengindikasikan model berusaha meminimalisir kesalahan klasifikasi pada data latih seoptimal mungkin.

E. Evaluasi Kinerja Model

Pengujian model dilakukan pada 20% data uji (489 data) yang tidak dilihat oleh model selama proses pelatihan. Evaluasi menggunakan *Confusion Matrix* menghasilkan metrik performa sebagai berikut:

TABEL 3
(HASIL CLASSIFICATION REPORT MODEL SVM)

Kelas Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.76	0.91	0.83	183
Netral	0.90	0.73	0.81	165
Positif	0.86	0.84	0.85	141
Accuracy			0.83	489
Macro Avg	0.84	0.82	0.83	489
Weighted Avg	0.84	0.83	0.83	489

Berdasarkan Tabel I, model SVM mencapai akurasi keseluruhan sebesar 82.62%. Nilai ini dikategorikan sebagai kinerja yang "Sangat Baik" (*Good to Excellent*) menurut skala interpretasi Landis & Koch.

- Analisis Kelas Negatif: Model memiliki *Recall* yang sangat tinggi (0.91), artinya model sangat sensitif dan mampu mendeteksi 91% dari total ulasan negatif yang ada. Ini sangat penting bagi penyedia layanan agar tidak ada keluhan pelanggan yang terlewat.
- Analisis Kelas Netral: Memiliki *Precision* tertinggi (0.90), yang berarti jika model memprediksi netral, kemungkinannya 90% benar. Namun, *Recall*-nya

terendah (0.73), menunjukkan model agak konservatif dalam melabeli data netral dan sering salah mengklasifikasikannya ke kelas lain.

- Analisis Kelas Positif: Menunjukkan keseimbangan terbaik antara *Precision* (0.86) dan *Recall* (0.84) dengan *F1-Score* tertinggi (0.85).

F. Validasi Silang (K-Fold Cross Validation)

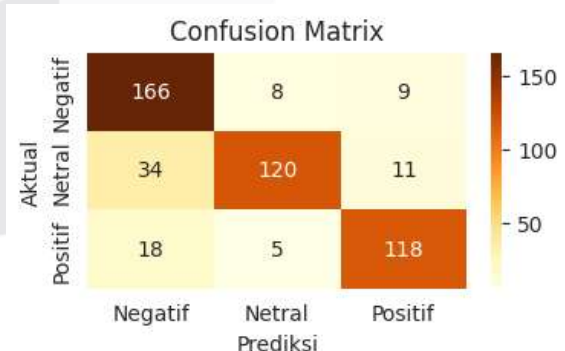
Untuk memastikan model yang dibangun stabil dan tidak *overfitting* (hanya bagus pada data tertentu), dilakukan pengujian *5-Fold Cross Validation* pada data latih yang telah di-*oversampling* (SMOTE). Hasilnya menunjukkan rata-rata akurasi sebesar **96.60%** dengan standar deviasi rendah (0.038). Tingginya akurasi pada validasi silang dibandingkan data uji (82.62%) adalah wajar karena validasi silang dilakukan pada lingkungan data latih yang diseimbangkan secara sintesis, namun hal ini membuktikan bahwa model memiliki kemampuan generalisasi yang baik dan robust terhadap variasi data latih. Hasil *K-Fold Cross Validation*

TABEL 4
(HASIL K-FOLD CROSS VALIDATION)

Iterasi (Fold)	Akurasi
Fold-1	91.23%
Fold-2	92.76%
Fold-3	99.86%
Fold-4	99.16%
Fold-5	100.00%

G. Confusion Matrix

Confusion Matrix memperlihatkan detail kesalahan prediksi. Dari 183 data aktual negatif, model berhasil menebak 166 dengan benar. Kesalahan terbesar terjadi pada kelas Netral, di mana 34 data netral diprediksi sebagai negatif. Hal ini wajar terjadi dalam analisis sentimen teks pendek karena ulasan netral yang berisi saran sering kali mengandung kata-kata yang beririsan dengan keluhan (misal: "tolong rute diperbaiki" bisa dianggap negatif oleh mesin, padahal netral/saran).



GAMBAR 6
(CONFUTION MATRIX)

V. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem analisis sentimen untuk evaluasi transportasi umum di Surabaya menggunakan algoritma *Support Vector Machine*. Berdasarkan hasil eksperimen dan analisis yang telah dilakukan, dapat ditarik beberapa kesimpulan utama:

1. Performa Model: Algoritma SVM dengan kernel RBF terbukti efektif untuk klasifikasi sentimen ulasan transportasi berbahasa Indonesia dengan akurasi mencapai 82.62%. Penerapan teknik SMOTE dan optimasi *GridSearchCV* berkontribusi signifikan dalam menghasilkan model yang seimbang (*balanced*) dengan nilai *F1-Score* di atas 0.80 untuk semua kelas.
2. Persepsi Publik: Masyarakat Surabaya masih cenderung kritis terhadap layanan transportasi umum, dibuktikan dengan dominasi sentimen negatif sebesar 43.76%. Isu utama yang menjadi sorotan adalah aspek operasional (kemacetan, waktu tunggu lama, jadwal tidak pasti). Namun, apresiasi positif (33.44%) terhadap kualitas fisik armada menjadi modal bagus bagi pemerintah kota.
3. Implikasi Praktis: Pemerintah Kota Surabaya dan operator (Trans Semanggi/Suroboyo Bus) disarankan untuk memprioritaskan perbaikan pada manajemen waktu (*headway*) dan integrasi jalur untuk mengurangi sentimen negatif. Sistem deteksi dini berbasis sentimen ini dapat diadopsi sebagai alat evaluasi kinerja yang lebih cepat dibandingkan survei manual.

REFERENSI

- [1] J. A. Ariesandi, R. Resita, and Z. Salsabila, "Kebijakan Transportasi Umum (Angkot) Untuk Menanggulangi Kemacetan Jalan," *Jurnal Kebijakan Publik*, vol. 11, no. 2, 2020.
- [2] A. R. Riawan and E. Ahyudanari, "Analisis Aksesibilitas dalam Penggunaan Transportasi Umum, di Kota Bekasi dengan Metode Competition Measure," *Jurnal Aplikasi Teknik Sipil*, vol. 18, no. 2, 2020.
- [3] TomTom, "TomTom Traffic Index 2024," *TomTom*, 2024. [4] H. Widyastuti and E. Santoso, "Analisis Integrasi Moda Transportasi Bus Suroboyo dan Kereta Api Komuter Menggunakan Metode Importance-Performance Analysis," *Jurnal Aplikasi Teknik Sipil*, vol. 21, no. 2, pp. 115-124, 2023.
- [5] Jurno.id, "Transportasi Umum di Surabaya Biasa Saja," *Jurno.id*, 2023.
- [6] T. Hidayat and R. K. Wardani, "Pemanfaatan Media Sosial Twitter sebagai Media Evaluasi Pelayanan Publik Menggunakan Analisis Sentimen," *Jurnal Sistem Informasi dan Bisnis Cerdas*, vol. 15, no. 2, pp. 88-96, 2022.
- [7] A. Z. Rasyida, I. D. Wijaya, and Y. Yunhasnawa, "Analisis Sentimen Kualitas Layanan Online Marketplace di Indonesia Menggunakan Metode Support Vector Machine," in *Seminar Informatika Aplikatif Polinema (SIAP)*, 2020.
- [8] M. Fitra Rizki, K. Auliasari, and R. Primaswara Prasetya, "Analisis Sentiment Cyberbullying Pada Sosial Media Twitter Menggunakan Metode Support Vector Machine," *Jurnal Mahasiswa Teknik Informatika*, vol. 5, no. 2, 2021.
- [9] T. M. P. Aulia, N. Arifin, and R. Mayasari, "Perbandingan Kernel Support Vector Machine (SVM) Dalam Penerapan Analisis Sentimen Vaksinasi Covid-19," *SINTECH JOURNAL*, vol. 4, no. 2, 2021.
- [10] F. Junian Wahidna and P. Nerisafitra, "Analisis Sentimen Pengguna Sistem Pay Later Menggunakan Support Vector Machine Metode Pembobotan Lexicon," *Journal of Informatics and Computer Science*, vol. 04, 2023.
- [11] E. Dwianto and M. Sadikin, "Analisis Sentimen Transportasi Online pada Twitter Menggunakan Metode Klasifikasi Naïve Bayes dan Support Vector Machine," *Jurnal Format*, vol. 10, no. 1, 2021.
- [12] F. Ladayya, D. Siregar, W. E. Pranoto, and H. D. Muchtar, "Analisis Sentimen pada Program Transportasi Publik JakLingko dengan Metode Support Vector Machine," *Jurnal Statistika Dan Aplikasinya*, vol. 6, no. 2, 2022.
- [13] Z. Alhaq, A. Mustopa, and J. D. Santoso, "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter," *Journal of Information System Management (JOISM)*, vol. 3, no. 1, 2021. [14] N. Andriani and A. Wibowo, "Implementasi Text Mining Klasifikasi Topik Tugas Akhir Mahasiswa Teknik Informatika Menggunakan Pembobotan TF-IDF dan Metode Cosine Similarity Berbasis Web," in *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, 2021.
- [15] A. K. Fauziyyah, "Analisis Sentimen Pandemi Covid19 Pada Streaming Twitter Dengan Text Mining Python," *Jurnal Ilmiah SINUS*, vol. 18, no. 2, p. 31, 2020.
- [16] D. Agustina, E. Putri, F. Fauzi, S. N. Alawiyah, and R. Wasono, "Penerapan Metode Support Vector Machine (SVM) Untuk Klasifikasi Data Ekspresi Gen Microarray," *Prosiding Seminar Edusainstech*, vol. 4, 2020.
- [17] S. S. M. Putri, M. Arhami, and H. Hendrawaty, "Penerapan Metode SVM pada Klasifikasi Kualitas Air," *JAISE: Journal of Artificial Intelligence and Software Engineering*, vol. 3, no. 2, 2023.
- [18] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, pp. 18-26, 2021.
- [19] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, 2021.
- [20] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, vol. 8, no. 1, pp. 147-156, 2021.