

ANALISIS SENTIMEN KEPUASAN PELANGGAN TRANSPORTASI ONLINE PADA GRAB MENGGUNAKAN SUPPORT VECTOR MACHINE

Denny Sitanggang¹, Hotma Pangaribuan²

¹Mahasiswa Program Studi Teknik Informatika, Universitas Putera Batam

²Dosen Program Studi Teknik Informatika, Universitas Putera Batam

email: pb210210109@upbatam.ac.id

ABSTRACT

Online transportation services such as Grab have become an essential part of urban mobility in Indonesia, generating a wide range of user reviews that reflect levels of satisfaction. This study aims to analyze the sentiment of these reviews using the Support Vector Machine (SVM) algorithm. Data were collected from the Google Play Store and processed through several stages, including text preprocessing, automatic labeling based on rating scores (≤ 3 as negative, ≥ 4 as positive), and feature representation using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The dataset was split into training data (80%) and testing data (20%), and the SVM model was trained using a linear kernel. Evaluation results showed an accuracy of 82%, precision of 84%, recall of 78%, F1-score of 79%, and an AUC of 0.9015. Further analysis of negative reviews revealed that the aspects of “drivers,” “application,” and “payment” were the main sources of complaints. These findings demonstrate the effectiveness of SVM in sentiment classification and its potential as a data-driven service evaluation tool. The study also recommends manual labeling or semantic-based approaches to address inconsistencies between review scores and content.

Keywords: Sentiment Analysis, Grab, Support Vector Machine, TF-IDF, User Review

PENDAHULUAN

Pemanfaatan internet di Indonesia terus menunjukkan tren peningkatan seiring dengan meluasnya jangkauan jaringan dan adopsi perangkat pintar. Kemudahan akses ini mendorong digitalisasi di berbagai sektor, termasuk transportasi. Layanan ojek daring seperti Grab telah menjadi solusi mobilitas harian yang praktis dan efisien bagi masyarakat urban, sekaligus menciptakan peluang kerja baru dan mengurangi ketergantungan pada kendaraan pribadi (Wahyudi & Kusumawardana, 2021).

Sebagai salah satu platform transportasi daring terbesar di Indonesia, Grab menawarkan pengalaman pengguna yang fleksibel melalui fitur-fitur digital berbasis aplikasi. Namun demikian, peningkatan jumlah pengguna tidak selalu diiringi dengan peningkatan kualitas layanan secara merata. Banyaknya ulasan pengguna yang tersebar di Google Play Store menunjukkan adanya variasi persepsi terhadap layanan Grab, namun data tersebut belum dimanfaatkan secara

optimal sebagai dasar evaluasi layanan (Wahyudi & Kusumawardana, 2021).

Beberapa studi sebelumnya telah menerapkan pendekatan analisis sentimen untuk memahami opini pelanggan. Misalnya, penelitian oleh (Laksono & Suryono, 2025) membandingkan algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* dalam mengklasifikasikan sentimen pengguna Tinder, dan menunjukkan bahwa SVM mencapai akurasi lebih tinggi (85% dibandingkan 84%) setelah penerapan teknik *SMOTE* untuk menangani ketidakseimbangan data. Hasil tersebut mengindikasikan bahwa SVM memiliki stabilitas yang lebih baik dalam mengenali sentimen positif dan negatif dibanding algoritma lainnya.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma SVM dalam klasifikasi sentimen ulasan pengguna terhadap layanan Grab. Ulasan diambil dari Google Play Store, kemudian diproses melalui tahapan pra-pemrosesan, pelabelan otomatis berdasarkan skor bintang, dan representasi fitur dengan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Penelitian ini diharapkan dapat memberikan kontribusi dalam pengukuran objektif kepuasan pelanggan serta menjadi masukan strategis bagi peningkatan kualitas layanan transportasi daring.

KAJIAN TEORI

2.1 Machine Learning

Machine Learning (ML) adalah salah satu disiplin dalam kecerdasan buatan yang memungkinkan komputer untuk belajar dari data dan menyesuaikan kinerjanya secara dinamis tanpa instruksi pemrograman yang spesifik. (Diana et al., 2023). Dalam praktiknya, ML terbagi

menjadi beberapa pendekatan, di antaranya *Supervised Learning* yang umum digunakan untuk klasifikasi sentimen berbasis data berlabel seperti pada ulasan Grab, *Unsupervised Learning* untuk segmentasi pengguna melalui pola tersembunyi, serta *Reinforcement Learning* yang relevan dalam optimasi rute dinamis. Selain itu, pendekatan *Semi-Supervised* dan *Self-Supervised Learning* juga mulai banyak diterapkan, terutama ketika data berlabel terbatas atau dalam pengembangan sistem berbasis pemahaman konteks seperti chatbot layanan pelanggan (Sitohang, 2021). Penelitian terdahulu juga menunjukkan bahwa algoritma seperti *Support Vector Machine* (SVM) dalam skema *supervised learning* mampu menghasilkan akurasi yang kompetitif dalam tugas klasifikasi sentimen dibandingkan dengan metode lain seperti *Naïve Bayes* (Laksono & Suryono, 2025).

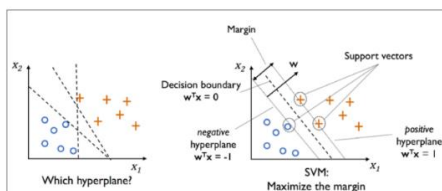
2.2 Sentiment Analysis

Sentiment analysis merupakan studi komputasional yang berfokus pada penggalan sikap, opini, emosi, dan penilaian dari sekumpulan teks menggunakan pendekatan *Natural Language Processing* (NLP), statistik, maupun machine learning (Manullang & Prianto, 2023). Teknik ini bertujuan untuk mengklasifikasikan dokumen ke dalam kategori sentimen seperti positif, negatif, atau netral. Meskipun telah berkembang pesat karena manfaatnya dalam pengambilan keputusan strategis, analisis sentimen tetap menghadapi tantangan seperti ambiguitas makna, tidak adanya intonasi dalam teks, serta dinamika bahasa yang terus berkembang. Dalam konteks penelitian ini, analisis sentimen diterapkan pada

ulasan layanan ojek online Grab dengan memanfaatkan algoritma *Support Vector Machine* (Alfandi Safira & Hasan, 2023).

2.3 Support Vector Machine

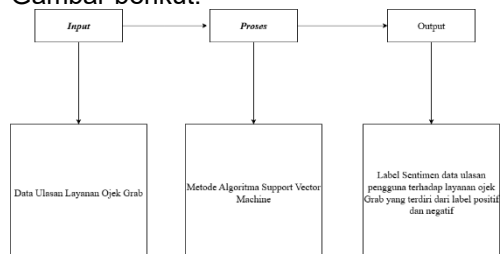
Support Vector Machine (SVM) adalah algoritma pembelajaran terawasi (*supervised learning*) yang digunakan untuk klasifikasi dan regresi, pertama kali dikenalkan oleh Boser, Guyon, dan Vapnik pada tahun 1992 dalam ajang Workshop on Computational Learning Theory. Algoritma ini bekerja dengan mencari hyperplane terbaik yang dapat memisahkan data ke dalam kelas-kelas yang berbeda secara optimal (Wahyudi & Kusumawardana, 2021). SVM dikenal efektif untuk tugas klasifikasi karena kemampuannya dalam mencari hyperplane optimal yang memisahkan dua kelas data dengan margin maksimum. Metode ini bekerja pada ruang berdimensi tinggi, bahkan mampu menangani data non-linier melalui fungsi kernel yang memetakan data ke ruang fitur baru (Junellson & Sitohang, 2024). SVM banyak digunakan dalam berbagai bidang seperti klasifikasi citra, deteksi spam, dan analisis biomedis karena kestabilannya dalam generalisasi pola data (Styawati et al., 2021). Gambar berikut menggambarkan konsep hyperplane terbaik yang memisahkan dua kelas dengan margin maksimum, inti dari prinsip kerja SVM.



Gambar 1. Konsep Algoritma *Support Vector Machine*
(Sumber: Data Penelitian, 2025)

2.4 Kerangka Pemikiran

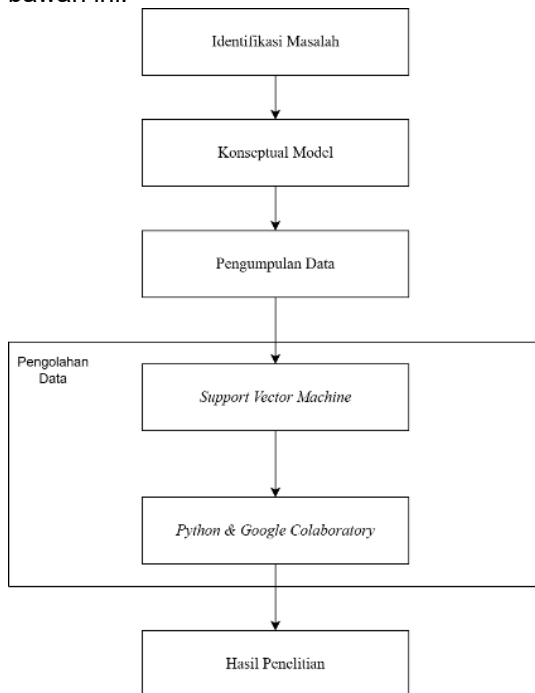
Kerangka pemikiran dalam penelitian ini berfokus pada analisis sentimen untuk mengevaluasi kepuasan pelanggan terhadap layanan transportasi online Grab, berdasarkan ulasan pengguna yang diperoleh dari platform seperti Google Play Store. Ulasan tersebut mencerminkan opini publik terkait aspek layanan seperti kenyamanan, kecepatan, dan interaksi dengan pengemudi. Penelitian ini memanfaatkan algoritma *Support Vector Machine* (SVM) untuk mengklasifikasikan sentimen ke dalam dua kategori utama: positif dan negatif. Proses dimulai dengan tahapan *preprocessing* teks terdiri dari *case folding*, tokenisasi, *stopword removal*, dan *stemming* diikuti dengan pembobotan fitur menggunakan TF-IDF. Data yang telah diproses digunakan untuk melatih model klasifikasi sehingga dapat mengenali pola sentimen dalam data ulasan. Tujuan akhir dari proses ini adalah untuk menghasilkan klasifikasi otomatis terhadap opini pelanggan yang dapat memberikan insight objektif bagi peningkatan layanan. Kerangka ini menjadi dasar metodologis pemanfaatan *machine learning* dalam memahami persepsi pengguna, sebagaimana divisualisasikan pada Gambar berikut.



Gambar 2. Kerangka Pemikiran
(Sumber: Data Penelitian, 2025)

METODE PENELITIAN

Pada penelitian ini peneliti menerapkan desain penelitian yang terdiri dari beberapa tahapan seperti gambar di bawah ini.



Gambar 3. Desain Penelitian
(Sumber: Data Penelitian, 2025)

Berdasarkan gambar di atas, penelitian ini dimulai dengan mengidentifikasi permasalahan dari ulasan pengguna aplikasi Grab di Google Play Store, yang mencerminkan sentimen terhadap aspek layanan seperti kualitas pengemudi, tarif, dan keandalan aplikasi. Peneliti merancang model konseptual dengan tahapan mulai dari pengumpulan data melalui web scraping, preprocessing teks, hingga klasifikasi sentimen menggunakan algoritma Support Vector Machine (SVM). Proses dilakukan dengan Python di platform Google Colab,

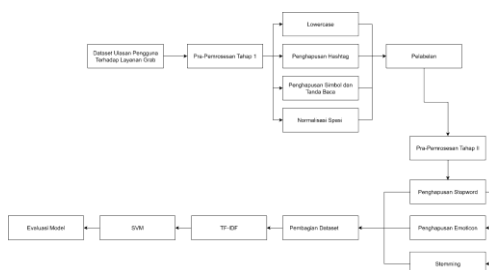
memanfaatkan pustaka seperti Scikit-learn dan Pandas. SVM dipilih karena kemampuannya menangani data teks berdimensi tinggi, untuk mengklasifikasikan sentimen menjadi positif dan negatif. Hasilnya divisualisasikan dalam aplikasi interaktif berbasis Streamlit, yang memungkinkan eksplorasi hasil klasifikasi dan analisis kata kunci secara real-time, menjadikan penelitian ini tidak hanya akademis, tetapi juga aplikatif bagi peningkatan layanan Grab secara data-driven.

3.1 Metode Pengumpulan Data

Metode pengumpulan data dalam penelitian ini dilakukan secara sistematis untuk memperoleh informasi yang relevan dengan topik analisis sentimen terhadap layanan aplikasi Grab. Dua teknik utama digunakan, yaitu studi pustaka dan web scraping. Studi pustaka dilakukan dengan menelaah literatur ilmiah seperti jurnal, buku, dan artikel terkait analisis sentimen, algoritma Support Vector Machine (SVM), dan teknik data mining untuk memperkuat landasan teori. Sementara itu, data utama diperoleh melalui teknik web scraping dengan menggunakan skrip Python untuk mengekstraksi komentar pengguna aplikasi Grab di Google Play Store, mencakup ulasan teks, rating, dan tanggal unggahan. Teknik ini memungkinkan pengumpulan data real-time yang autentik dari pengguna aktif, dan digunakan sebagai dataset utama dalam analisis sentimen, dengan tetap memperhatikan etika dan kebijakan platform sumber data.

3.2 Metode Perancangan

Metode perancangan yang digunakan dalam penelitian ini dirancang seperti pada gambar diagram alur di bawah ini.



Gambar 4. Kerangka Perancangan
(Sumber: Data Penelitian, 2025)

Berdasarkan gambar di atas, kerangka perancangan yang digunakan mencakup tahapan pra-pemrosesan, pelabelan, pembobotan, pelatihan model, dan evaluasi. Proses dimulai dengan pengumpulan 1000 ulasan pengguna Grab yang kemudian dibersihkan melalui pra-pemrosesan tahap I menggunakan Python. Langkah ini mencakup normalisasi teks, penghapusan simbol, hashtag, dan spasi berlebih. Setelah data bersih, dilakukan pelabelan otomatis berdasarkan skor bintang, dengan ulasan skor ≤ 3 dikategorikan negatif dan skor ≥ 4 sebagai positif, menggunakan fungsi `label_sentiment`.

Selanjutnya, pada pra-pemrosesan tahap II, dilakukan tokenisasi, penghapusan stopword, dan stemming dengan pustaka Sastrawi untuk meningkatkan kualitas representasi data. Data yang telah dibersihkan dibagi menjadi data latih dan data uji menggunakan rasio 80:20 melalui `train_test_split` dari Scikit-learn. Tahapan berikutnya adalah pembobotan teks menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency), yang menekankan kata-kata penting berdasarkan kemunculannya dalam dokumen dan keseluruhan korpus.

Model klasifikasi dibangun dengan algoritma SVM ber-kernel linear karena

data TF-IDF bersifat linier. Model dilatih dengan data latih dan diuji dengan data uji untuk menghasilkan prediksi sentimen. Evaluasi kinerja dilakukan menggunakan confusion matrix dan metrik seperti akurasi, precision, recall, F1-score, serta AUC. Akhirnya, model disimpan dalam format .pkl agar dapat digunakan ulang. Evaluasi ini bertujuan menilai efektivitas model dalam membedakan sentimen positif dan negatif secara objektif terhadap ulasan layanan Grab.

HASIL DAN PEMBAHASAN

4.1 Hasil Pengumpulan Data

Data diperoleh melalui teknik *web scraping* menggunakan pustaka `google-play-scraper` di Google Colaboratory. Sumber data adalah ulasan pengguna aplikasi Grab di Google Play Store, difokuskan pada ulasan berbahasa Indonesia tanpa penyaringan skor, sehingga mencakup opini dari bintang 1 hingga 5. Sebanyak 1.000 ulasan paling relevan berhasil dikumpulkan, lalu dikonversi ke format `DataFrame` menggunakan `pandas`, dengan kolom utama berupa `content` dan `score`. Dataset tersebut kemudian disimpan sebagai file CSV untuk digunakan pada tahap pra-pemrosesan dan analisis sentimen.

4.2 Pra-Pemrosesan Tahap 1

Pra-pemrosesan awal dilakukan terhadap 1.000 ulasan pengguna Grab untuk membersihkan teks dari karakter tidak relevan seperti simbol, tanda baca, dan hashtag. Proses ini menggunakan Python dengan pustaka `pandas` dan `re`, melalui empat tahap utama: konversi huruf ke lowercase, penghapusan hashtag, penghapusan simbol dan tanda baca, serta normalisasi spasi. Hasil pembersihan disimpan dalam format CSV

dan Excel untuk keperluan analisis lanjutan. Contoh perubahan teks sebelum dan sesudah pembersihan ditampilkan pada gambar berikut, menunjukkan

peningkatan kualitas data untuk analisis sentimen.

Tabel 1. Sesudah dan Sebelum Pemrosesan Data

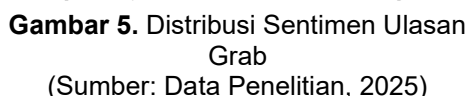
Data Sebelum Pra-Proses Tahap 1	Data Sesudah Pra-Proses Tahap 1
<p>banyak driver tolol gak sih, di aplikasinya aturan lalu lintas harus diperketat, yang ngelanggar aturan lalu lintas kalo bisa akunya langsung di suspend ya, udah sering banget gw hampir kecelakaan gara gara driver yang tiba tiba pelan di jalur cepat/kanan untung aja gw masih bisa belok kiri dengan cepat, dan paling banyak itu dari driver mau itu Maxim,gojek,grab banyak banget yang main naik-naik trotoar aja, dan lawan arah</p>	<p>banyak driver tolol gak sih di aplikasinya aturan lalu lintas harus diperketat yang ngelanggar aturan lalu lintas kalo bisa akunya langsung di suspend ya udah sering banget gw hampir kecelakaan gara gara driver yang tiba tiba pelan di jalur cepat/kanan untung aja gw masih bisa belok kiri dengan cepat dan paling banyak itu dari driver mau itu maximgojekgrab banyak banget yang main naiknaik trotoar aja dan lawan arah</p>
<p>aplikasi ini sekarang tidak bersahabat lagi dengan tunanetra. aplikasi ini tidak kompatibel dengan pembaca layar TalkBack sangat berat digunakan dan tidak akses terutama daftar makanan tidak dibacakan Jadi kami tidak bisa melakukan pemesanan harapannya pengembang bisa memperbaiki dan membuat aplikasi ini aksesibel bagi tunanetra yang menggunakan pembaca layar TalkBack</p>	<p>aplikasi ini sekarang tidak bersahabat lagi dengan tunanetra aplikasi ini tidak kompatibel dengan pembaca layar talkback sangat berat digunakan dan tidak akses terutama daftar makanan tidak dibacakan jadi kami tidak bisa melakukan pemesanan harapannya pengembang bisa memperbaiki dan membuat aplikasi ini aksesibel bagi tunanetra yang menggunakan pembaca layar talkback</p>

(Sumber: Data Penelitian, 2025)

4.3 Tahap Pelabelan Otomatis

Setelah tahap pra-pemrosesan, data ulasan diklasifikasikan secara otomatis menjadi sentimen positif atau negatif berdasarkan skor penilaian pengguna. Skor ≤ 3 dikategorikan sebagai negatif, sedangkan skor ≥ 4 sebagai positif. Proses ini diimplementasikan menggunakan Python dengan fungsi `label_sentiment(score)` dan hasilnya disimpan dalam kolom baru bernama *Label Aktual*.

Dari 1.000 ulasan, diperoleh 633 ulasan negatif dan 367 positif. Distribusi ini menunjukkan dataset tidak seimbang, yang dapat memengaruhi performa klasifikasi. Dataset hasil pelabelan disimpan dengan nama `Dataset_Ulasan_Grab_Cleaned_Labeled.csv` untuk digunakan pada tahap vektorisasi dan pelatihan model.



Model klasifikasi sentimen dibangun menggunakan algoritma Support Vector Machine (SVM) dengan kernel linear.

Pemilihan SVM didasarkan pada kemampuannya dalam memisahkan dua kelas secara optimal pada data berdimensi tinggi seperti teks. Model diinisialisasi menggunakan SVC() dari pustaka scikit-learn, lalu dilatih pada data latih yang telah diubah menjadi representasi TF-IDF. Setelah pelatihan, model digunakan untuk memprediksi data uji. Evaluasi dilakukan dengan confusion matrix, classification report (precision, recall, f1-score), dan

akurasi guna mengetahui performa klasifikasi terhadap data baru.

Setelah diuji dengan data penelitian yang bersifat tidak seimbang (*imbalanced*), dengan 633 ulasan berlabel negatif dan 367 berlabel positif dari total 1.000 data. Setelah dibagi 80:20, data latih terdiri dari 509 label negatif dan 291 label positif, sementara data uji terdiri dari 124 label negatif dan 76 label positif.

Tabel 2. Evaluasi Pengujian dengan *Confusion Matrix*

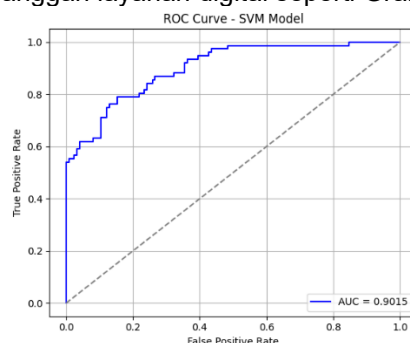
	Prediksi Positif	Prediksi Negatif
Aktual Positif	47	29
Aktual Negatif	7	117

(Sumber: Data Penelitian, 2025)

Berdasarkan tabel di atas yaitu *confusion matrix* yang digunakan untuk mengukur kinerja model berdasarkan hasil prediksi terhadap label aktual. Hasil evaluasi menunjukkan model mengklasifikasikan 47 data positif dan 117 data negatif dengan benar. Namun, terdapat kesalahan klasifikasi yaitu 29 false negative dan 7 false positive. Model cukup akurat dalam mengenali sentimen negatif, namun masih kurang dalam mengidentifikasi sentimen positif, terutama ketika teks ulasan ambigu, mengandung ironi, atau memiliki campuran sentimen. Selain itu, penggunaan skor bintang sebagai dasar pelabelan menyebabkan ketidaksesuaian antara label dan isi ulasan pada beberapa kasus, seperti pada ulasan bernada keluhan namun diberi skor.

Evaluasi model menggunakan akurasi, precision, recall, f1-score, dan AUC (Area Under Curve). Hasil menunjukkan bahwa model mencapai akurasi 82%, precision

84%, recall 78%, dan f1-score 79%. AUC sebesar 0,9015 menunjukkan kemampuan yang sangat baik dalam membedakan ulasan positif dan negatif, sebagaimana ditunjukkan oleh kurva ROC yang jauh di atas baseline. Secara keseluruhan, model SVM menunjukkan performa yang andal untuk diterapkan dalam klasifikasi sentimen otomatis terhadap ulasan pelanggan layanan digital seperti Grab.



Gambar 6. AUC Model SVM Terhadap Data Uji



(Sumber: Data Penelitian, 2025)

R. (2021). Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(3), 150–155. <https://doi.org/10.30591/jpit.v6i3.2870>

Tri Putra, K., Amin Hariyadi, M., & Crysdiyan, C. (2023). Perbandingan Feature Extraction Tf-Idf Dan Bow Untuk Analisis Sentimen Berbasis Svm. *Jurnal Cahaya MAndalika*, 1449.

Wahyudi, R., & Kusumawardana, G. (2021). Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. *Jurnal Informatika*, 8(2), 200–207.

<https://doi.org/10.31294/ji.v8i2.9681>

	Penulis pertama, Denny Sitanggang, merupakan mahasiswa Prodi Teknik Informatika Universitas Putera Batam
	Penulis kedua, Bapak Hotma Pangaribuan S.Kom., M.Si, merupakan Dosen Prodi Teknik Informatika Universitas Putera Batam.