



# Analisis Sentimen Aplikasi Gojek Pada Ulasan Pengguna di Google Play Store Menggunakan Metode Support Vector Machine

Yoga Adi Nugroho\*, Sudarno

Pascasarjana, Program Studi Teknik Informatika S-2, Universitas Pamulang, Tangerang Selatan

Jl. Raya Puspitek No.46, Buaran, Serpong, Kota Tangerang Selatan, Banten, Indonesia

Email: <sup>1,\*</sup>noegraha9@gmail.com, <sup>2</sup>dosen00119@unpam.ac.id

Email Penulis Korespondensi: noegraha9@gmail.com

Submitted: 07/07/2025; Accepted: 22/07/2025; Published: 31/07/2025

**Abstrak**—Gojek merupakan aplikasi yang mendapatkan perhatian besar di Indonesia dimana platform ini menawarkan berbagai layanan, termasuk transportasi, pengantaran makanan, dan layanan on-demand lainnya. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi Gojek yang diunduh dari Play Store tahun 2024 menggunakan metode Support Vector Machine (SVM). Data dikumpulkan melalui web scrapping sebanyak 8.000 ulasan. Tahapan preprocessing yang dilakukan meliputi cleaning, case folding, tokenizing, stopword removal, stemming, pelabelan menggunakan lexicon-based, selanjutnya ke tahap TF-IDF dan pengujian dengan confusion matrix menggunakan python. Hasil pelabelan menunjukkan mayoritas ulasan bersentimen negatif sebesar 44,25%, disusul sentimen positif sebesar 36,62%, dan netral sebesar 19,12%. Skema pengujian 80% untuk training dan 20% untuk testing. Hasil analisis menunjukkan akurasi 93,69%, recall 93,66%, presisi 93,65%, dan F1-Score 93,67%. Hasil ini menunjukkan bahwa model SVM mampu mengklasifikasikan sentimen dengan tingkat akurasi yang tinggi. Temuan ini dapat menjadi masukan bagi pengembang Gojek dalam meningkatkan kualitas layanan aplikasi.

**Kata Kunci:** Gojek; Analisis Sentimen; Play Store; Support Vector Machine; Machine Learning

**Abstract**—Gojek is an application that has garnered significant attention in Indonesia, offering a variety of services, including transportation, food delivery, and other on-demand services. This study aims to analyze user sentiment towards the Gojek app downloaded from the Play Store in 2024 using the Support Vector Machine (SVM) method. Data was collected through web scrapping, comprising 8,000 reviews. The preprocessing steps included cleaning, case folding, tokenizing, stopword removal, stemming, and lexicon-based labeling, followed by TF-IDF and testing using a confusion matrix with Python. The sentiment labeling results revealed that the majority of reviews were negative at 44.25%, followed by positive sentiment at 36.62%, and neutral sentiment at 19.12%. The testing scheme applied 80% for training and 20% for testing. The analysis results showed an accuracy of 93.69%, recall of 93.66%, precision of 93.65%, and F1-Score of 93.67%. These results indicate that the SVM model is capable of classifying sentiment with a high level of accuracy. These findings can provide valuable insights for Gojek developers to enhance the app's service quality.

**Keywords:** Gojek, Sentiment Analysis; Play Store; Support Vector Machine; Machine Learning

## 1. PENDAHULUAN

Perkembangan teknologi digital telah mengubah perilaku konsumen secara drastis. Aplikasi seperti Gojek memudahkan masyarakat dalam memenuhi kebutuhan harian melalui satu platform terpadu—transportasi, logistik, hingga pembayaran. Ulasan pengguna di Play Store, baik yang memuji maupun mengkritik, menjadi sumber data utama untuk evaluasi kualitas layanan dan kepuasan pengguna [1]. Penelitian Hanafi menunjukkan bahwa model LSTM dengan representasi vektor kata cukup efektif dalam mengklasifikasi sentimen ulasan Gojek [2]. Sejalan dengan itu, studi oleh Rahman mengaplikasikan Naïve Bayes pada 5.000 ulasan dan mencapai akurasi hingga 82% serta F1 score 78%. Namun, analisis survei komparatif pada pertengahan 2025 mengungkap bahwa hanya 49% ulasan Gojek yang positif, sedangkan 51% bersifat negatif, mencerminkan apresiasi terhadap driver, namun ada kekurangan dari sisi performa aplikasi dan metode pembayaran [3]. Temuan dari Pahmi Amri juga mengonfirmasi adanya dominasi nuansa negatif terkait error aplikasi dan perilaku driver berdasarkan 5.000 ulasan [4]. Di sisi lain, tren terbaru dalam sistem rekomendasi berbasis sentimen menunjukkan potensi peningkatan akurasi dan interpretabilitas model analisis ulasan jika menggunakan embeddings sentimen dan pendekatan NLP modern [5]. Ini menunjukkan bahwa strategi pemasaran maupun operasional Gojek dapat diperkuat melalui insights dari analisis sentimen mendalam.

Support Vector Machine (SVM) telah lama diakui sebagai algoritma efektif dalam klasifikasi teks, terutama berkat kemampuannya mengoptimalkan margin pemisahan melalui prinsip structural risk minimization [6]. Keunggulan ini membuat SVM populer dalam analisis sentimen, di mana teks—seperti ulasan pengguna—dikelompokkan ke dalam kelas sentimen positif, negatif, atau netral. Abdullah & Abdulazeez dalam kajian komprehensif mereka menyimpulkan bahwa SVM menghasilkan akurasi tinggi dalam berbagai domain, termasuk sentiment analysis, berkat kinerjanya yang andal dibanding teknik lain [7]. Kernelling inovatif, seperti Fisher-Kernel SVM yang menggabungkan probabilistic latent semantic features, terbukti meningkatkan efektivitas klasifikasi teks lewat pengayaan representasi latennya. Evaluasi terbaru (1,6 tahun lalu) dari Atlantis-Press menyoroti bahwa SVM menawarkan keseimbangan antara margin maksimal dan generalisasi, menjadikannya pilihan kuat dalam klasifikasi emosi dan opini teks. Bahwa feature-engineering—seperti penggunaan TF-IDF dan seleksi fitur yang cermat—sangat mempengaruhi performa SVM, seperti dibahas dalam penelitian Avinash &

Sivasankar [8]. Dalam konteks penelitian ini, SVM digunakan untuk menganalisis sentimen pengguna Gojek dari ulasan Play Store tahun 2024. Pendekatan ini mengadopsi teori margin maksimal dengan kernel yang sesuai, dan disertai proses preprocessing plus feature-engineering (TF-IDF + seleksi fitur). Dengan dataset yang besar dan evaluasi komprehensif—termasuk akurasi, precision, recall, F1, penelitian ini diharapkan menghasilkan insight baru dan performa lebih kuat dibanding studi sebelumnya.

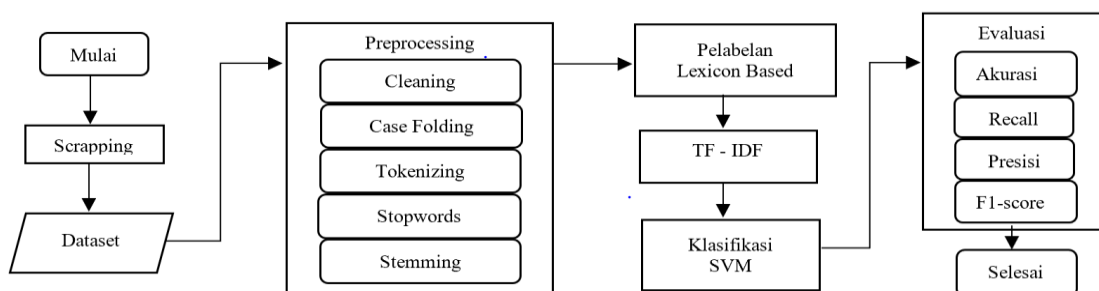
Penelitian ini memiliki keunggulan signifikan dibanding studi terdahulu. Walaupun Setiawan telah menggunakan SVM untuk analisis sentiment marketplace dan menunjukkan akurasi tinggi, fokus utama mereka adalah media sosial [9]. Sementara itu, Pang & Lee menawarkan kerangka teoretis opini mining yang mendasari pengklasifikasian teks opini [10], dan Liu memberikan fondasi metodologis yang penting [5]. Pendekatan hybrid seperti NBSVM oleh Lestari & Setiawan membuktikan kombinasi SVM dan FastText efektif untuk bahasa Indonesia [11], sedangkan Hamka & Tukiran memverifikasi efisiensi kernel SVM bervariasi pada data e-commerce dari Twitter [9]. Meskipun begitu, studi-studi sebelumnya belum secara khusus menangani ulasan Gojek tahun 2024 dari Play Store, yang menjadi kebaruan utama penelitian ini. Fokus pada dataset besar dan terbaru—ditambah evaluasi komprehensif seperti analisis metrik kinerja (akurasi, precision, recall, F1-score) dan mungkin analisis aspek (aspect-based sentiment)—diperkirakan akan memberikan insight yang lebih tajam tentang persepsi pengguna Gojek terkini.

Penelitian ini bertujuan untuk menganalisis sentimen pengguna aplikasi Gojek dengan menggunakan metode Support Vector Machine (SVM) sebagai algoritma klasifikasi teks. Fokus utama penelitian ini adalah mengelompokkan ulasan pengguna di Play Store ke dalam kategori sentimen positif, negatif, dan netral secara akurat, serta mengevaluasi performa model klasifikasi yang dibangun melalui pengukuran akurasi, precision, recall, dan F1-score. Hasil dari penelitian ini diharapkan dapat memberikan rekomendasi strategis yang berguna bagi pengembang aplikasi dalam meningkatkan kualitas layanan, khususnya dalam merespons kebutuhan dan ekspektasi pengguna berdasarkan hasil analisis data terbaru. Beberapa penelitian sebelumnya telah membahas penerapan SVM dalam analisis sentimen pada berbagai platform. Setiawan melakukan analisis sentimen marketplace berbasis media sosial dengan SVM, namun fokusnya belum spesifik pada aplikasi layanan on-demand seperti Gojek [12]. Pang dan Lee membahas secara luas tentang klasifikasi opini dan metodologi sentiment analysis, tetapi tidak menerapkannya secara spesifik pada platform dan layanan berbasis aplikasi di Indonesia. Liu juga mengembangkan landasan metodologi analisis sentimen secara umum, tanpa fokus pada konteks aplikasi mobile lokal. Selain itu, S. D. Lestari dan E. B. Setiawan menggunakan kombinasi NBSVM dan FastText untuk data Twitter pengguna Telkomsel, namun tidak meneliti ulasan aplikasi di Play Store yang cenderung lebih panjang dan detail [13]. Hamka dan Tukiran juga membahas analisis sentimen marketplace dari media sosial dengan berbagai kernel SVM [9], tetapi belum mengeksplorasi data ulasan aplikasi Gojek secara spesifik. Gap penelitian ini yaitu belum adanya penelitian sebelumnya yang secara khusus menganalisis sentimen ulasan pengguna aplikasi Gojek di Play Store pada tahun 2024. Sebagian besar penelitian terdahulu menggunakan data dari media sosial seperti Twitter, yang secara struktur dan konteks berbeda dengan data ulasan di Play Store. Penelitian sebelumnya umumnya berfokus pada analisis sentimen umum, belum menyoroti secara spesifik kebutuhan perbaikan layanan Gojek berdasarkan ulasan aplikasi terbaru. Dengan demikian, penelitian ini menawarkan kebaruan (novelty) melalui fokus pada dataset Play Store Gojek tahun 2024, penerapan SVM dengan evaluasi yang komprehensif, serta penyusunan rekomendasi yang lebih relevan dan aplikatif bagi pengembang Gojek dalam meningkatkan kualitas layanan ke depan.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen berbasis machine learning. Penelitian ini dimulai dengan scrapping data untuk mendapatkan dataset, kemudian data dibersihkan pada tahap preprocessing. Tahap selanjutnya pelabelan menggunakan teknik lexicon, kemudian pembobotan kata menggunakan TF-IDF [14]. Tahap berikutnya klasifikasi menggunakan SVM dan tahap terakhir evaluasi model. Alur penelitian ini dapat dilihat pada Gambar 1 berikut:



**Gambar 1.** Diagram Alur Penelitian



Tahapan-tahapan dalam penelitian ini dijelaskan sebagai berikut:

- a. **Scrapping**  
Data dikumpulkan melalui teknik web scrapping pada ulasan aplikasi Gojek yang tersedia di Play Store. Pengambilan data difokuskan pada ulasan pengguna pada tahun 2024 agar data yang diperoleh relevan dan terkini. Total data yang berhasil dikumpulkan sebanyak 8.000 ulasan
- b. **Preprocessing Data**  
Tahapan preprocessing bertujuan untuk membersihkan dan menyiapkan data agar dapat diolah oleh algoritma machine learning. Langkah-langkah preprocessing yang dilakukan antara lain: Cleaning (menghapus karakter yang tidak diperlukan seperti angka, tanda baca, dan simbol), Case Folding (mengubah seluruh teks menjadi huruf kecil agar seragam), Tokenizing (memisahkan kalimat menjadi satuan kata), Stopword Removal (menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis, seperti “yang”, “dan”, “adalah”), Stemming (mengembalikan setiap kata ke bentuk dasarnya menggunakan algoritma stemming Bahasa Indonesia), Labeling (proses pelabelan sentimen dilakukan menggunakan lexicon-based dengan bantuan kamus kata positif, negatif dan netral).
- c. **Pembobotan Data**  
Data yang telah melalui tahap preprocessing selanjutnya diolah menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF digunakan untuk memberikan bobot pada setiap kata berdasarkan frekuensi kemunculan kata dalam suatu dokumen dibandingkan dengan keseluruhan korpus data. Kata yang sering muncul dalam dokumen tetapi jarang muncul di keseluruhan data akan memiliki bobot yang lebih tinggi.
- d. **Pembagian Data**  
Dataset yang telah dibobot kemudian dibagi menjadi dua bagian yaitu 80% data training (digunakan untuk membangun model SVM) dan 20% data testing (digunakan untuk menguji performa model yang telah dibangun). Pembagian data dilakukan secara acak untuk menghindari bias dalam proses pelatihan dan pengujian model
- e. **Pemodelan menggunakan SVM**  
Support Vector Machine (SVM) merupakan algoritma supervised learning yang efektif dalam memecahkan permasalahan klasifikasi teks. SVM bekerja dengan mencari hyperplane terbaik yang memisahkan data berdasarkan kelasnya. Model SVM dibangun menggunakan data training yang telah diproses sebelumnya.
- f. **Evaluasi Model**  
Model dievaluasi menggunakan confusion matrix untuk mengukur kinerja model dalam mengklasifikasikan sentimen. Parameter evaluasi yang digunakan adalah Akurasi (mengukur proporsi prediksi yang tepat dari keseluruhan data), Presisi (mengukur akurasi prediksi positif dari semua prediksi positif yang dibuat), Recall (mengukur kemampuan model dalam menemukan data positif yang sebenarnya), F1-Score (mengukur keseimbangan antara presisi dan recall). Evaluasi ini bertujuan untuk memastikan model yang dibangun memiliki performa yang baik dan dapat diandalkan untuk menganalisis sentimen ulasan pengguna.

## 2.2 Perangkat Penelitian

Penelitian ini menggunakan beberapa perangkat dan aplikasi sebagai berikut: Bahasa pemrograman Python 3.10, Library scikit-learn untuk membangun model SVM dan preprocessing teks, Google Colab sebagai platform pengolahan data, Google Play Store sebagai sumber data ulasan aplikasi Gojek. Seluruh tahapan dilakukan secara sistematis untuk memastikan kualitas data dan hasil penelitian yang valid.

## 3. HASIL DAN PEMBAHASAN

Berdasarkan pendahuluan dan metode penelitian yang telah dipaparkan di atas. Berikut hasil penelitian di jelaskan berdasarkan tahapan penelitian.

### 3.1 Hasil Preprocessing dan Labeling Data

- a. **Pengumpulan Dataset**  
Proses pengumpulan dataset menggunakan web scrapping dengan bahasa pemrograman python dan ulasan yang diambil berfokus pada tahun 2024. Dataset yang didapat sebanyak 8000 ulasan. Hasil pengumpulan data dalam format csv. Hanya kolom review text yang diambil karena penelitian ini menggunakan ulasan pengguna untuk pengoahan text. Data tersebut kemudian melalui serangkaian tahapan preprocessing untuk memastikan kualitas dan kesiapannya dalam analisis lebih lanjut. Berikut hasil web scrapping ulasan pengguna aplikasi gojek di google playstore:

**Tabel 1.** Dataset Analisis Sentimen Gojek 2024

Review Text
Sering dapat drivernya jauh. Jadi lama nunggu
Sangat membantu sekali.terima kasih Recommended



---

**Review Text**

---

Kenapa sering lemot alias masih lelet  
Gojek makin sulit dengan jadwal pengiriman baru  
Tambahin lagi promo potongan goride dan gocar nya  
Pokonya Mantap Betul.... aplikasi bagus

---

Dari hasil web scrapping yang kemudian digunakan sebagai dataset untuk penelitian akan tetapi dari data tersebut terdapat penulisan yang tidak sesuai, terdapat tanda baca dan simbol maka diperlukan proses selanjutnya agar dapat diolah dalam machine learning.

b. Cleaning

Pada tahap cleaning, karakter-karakter yang tidak relevan seperti tanda baca, simbol, angka berhasil dihapus untuk menghasilkan data yang lebih bersih.

**Tabel 2.** Cleaning Dataset Analisis Sentimen Gojek 2024

---

**Cleaning**

---

Sering dapat drivernya jauh Jadi lama nunggu  
Sangat membantu sekaliterima kasih Recommended  
Kenapa sering lemot alias masih lelet  
Gojek makin sulit dengan jadwal pengiriman baru  
Tambahin lagi promo potongan goride dan gocar nya  
Pokonya Mantap Betul aplikasi bagus

---

Pada tahap cleaning semua tanda baca dan simbol sudah dibersihkan, sehingga dataset tidak mengandung tanda baca dan simbol.

c. Case Folding

Selanjutnya, case folding dilakukan dengan mengubah seluruh huruf menjadi huruf kecil guna menjaga konsistensi penulisan.

**Tabel 3.** Case Folding Dataset Analisis Sentimen Gojek 2024

---

**Case Folding**

---

sering dapat drivernya jauh jadi lama nunggu  
sangat membantu sekaliterima kasih recommended  
kenapa sering lemot alias masih lelet  
gojek makin sulit dengan jadwal pengiriman baru  
tambahin lagi promo potongan goride dan gocar nya  
pokonya mantap betul aplikasi bagus

---

Pada tahap case folding semua huruf besar diubah menjadi huruf kecil, sehingga dataset tidak memuat huruf besar yang dapat mengganggu konsistensi penulisan.

d. Tokenizing

Proses tokenizing kemudian memecah kalimat menjadi kata-kata tunggal, sehingga mempermudah dalam tahap pembobotan dan analisis berikutnya. Tahapan preprocessing ini menjadi langkah penting untuk mendapatkan data yang terstruktur dan siap digunakan dalam pemodelan analisis sentimen.

**Tabel 4.** Tokenizing Dataset Analisis Sentimen Gojek 2024

---

**Tokenize**

---

['sering', 'dapat', 'drivernya', 'jauh', 'jadi', 'lama', 'nunggu']  
['sangat', 'membantu', 'sekaliterima', 'kasih', 'recommended']  
['kenapa', 'sering', 'lambat', 'alias', 'masih', 'lelet']  
['gojek', 'makin', 'sulit', 'dengan', 'jadwal', 'pengiriman', 'baru']  
['tambahin', 'lagi', 'promo', 'potongan', 'goride', 'dan', 'gocar']  
['pokonya', 'bagus', 'betul', 'aplikasi', 'bagus']

---

Pada tahap Tokenizing semua kalimat sudah terpecah menjadi kata-kata tunggal, sehingga akan mempermudah dalam pembobotan.

e. Stopword Removal

Selanjutnya, proses stopwords removal dilakukan untuk mengurangi kata-kata umum yang tidak memiliki makna signifikan dalam analisis sentimen, seperti 'yang', 'dari', 'dan', serta 'adalah'. Dengan menghapus kata-kata tersebut, fokus analisis menjadi lebih terarah pada kata-kata yang membawa makna penting.

**Tabel 5.** Stopword Removal Dataset Analisis Sentimen Gojek 2024

---

**Stopword Removal**

---

['sering', 'dapat', 'drivernya', 'jauh', 'jadi', 'lama', 'nunggu']

---

**Stopword Removal**

['sangat', 'membantu', 'sekaliterima', 'kasih', 'recommended']  
 ['kenapa', 'sering', 'lambat', 'alias', 'masih', 'lelet']  
 ['gojek', 'makin', 'sulit', 'jadwal', 'pengiriman', 'baru']  
 ['tambahin', 'lagi', 'promo', 'potongan', 'goride', 'gocar']  
 ['pokonya', 'bagus', 'betul', 'aplikasi', 'bagus']

Pada tahap Stopword Removal kata-kata umum yang tidak memiliki makna signifikan dalam analisis sentimen dihapus atau dihilangkan. Sehingga tidak ada lagi kata-kata umum yang tidak memiliki makna signifikan dalam analisis sentimen.

f. Stemming

Setelah itu, proses stemming diterapkan untuk mengembalikan setiap kata ke bentuk dasarnya. Sebagai contoh, kata-kata seperti 'berjalan', 'berjalanlah', dan 'berjalanannya' disederhanakan menjadi 'jalan'. Tahapan ini bertujuan untuk mengurangi variasi kata sehingga mempermudah dalam proses analisis dan pembobotan kata selanjutnya.

**Tabel 6.** Stemming Dataset Analisis Sentimen Gojek 2024

Stemming
sering dapat pengemudi jauh lama nunggu
membantu sekali terima kasih gojek mantap
sering lambat alias masih lelet
gojek makin menyulitkan jadwal pengiriman baru
banyak promo potongan goride gocar
pokoknya bagus betul aplikasi bagus

Pada tahap Stemming setiap kata telah dikembalikan kebentuk dasarnya, sehingga variasi kata yang tidak perlu telah dihilangkan.

g. Pelabelan

Setelah melalui tahapan preprocessing, data ulasan kemudian dilanjutkan ke proses pelabelan sentimen. Proses ini dilakukan dengan menggunakan pendekatan lexicon-based, yaitu metode yang memanfaatkan kamus kata positif dan negatif dalam Bahasa Indonesia untuk menentukan jenis sentimen pada masing-masing ulasan. Setiap kata dalam ulasan dianalisis berdasarkan daftar kata yang sudah diklasifikasikan sebelumnya, sehingga memungkinkan sistem untuk mengidentifikasi apakah ulasan tersebut mengandung sentimen positif, negatif, atau netral.

**Tabel 7.** Pelabelan Dataset Analisis Sentimen Gojek 2024

Stemming	Pos Score	Neg Score	Sentimen
sering dapat pengemudi jauh lama nunggu	3	4	negatif
membantu sekali terima kasih gojek mantap	3	1	positif
sering lambat alias masih lelet	0	2	negatif
gojek makin menyulitkan jadwal pengiriman baru	1	2	negatif
banyak promo potongan goride gocar	3	1	positif
pokoknya bagus betul aplikasi bagus	1	3	positif

Dari hasil pelabelan terhadap 8.000 ulasan yang dikumpulkan, diketahui bahwa sebanyak 44,25% ulasan termasuk dalam kategori negatif, 36,62% ulasan tergolong positif, dan sisanya sebesar 19,12% merupakan ulasan netral. Proporsi distribusi sentimen ini memberikan gambaran bahwa selama tahun 2024, sebagian besar pengguna cenderung memberikan kritik, keluhan, atau pengalaman kurang menyenangkan terhadap layanan aplikasi Gojek. Temuan ini mengindikasikan adanya perhatian serius dari pengguna terhadap aspek-aspek layanan yang mungkin masih perlu diperbaiki oleh pihak Gojek. Dengan demikian, hasil pelabelan ini apat menjadi dasar penting bagi pengembang aplikasi dalam meningkatkan kualitas layanan dan kepuasan pelanggan di masa mendatang.

**3.1.1 Hasil Pembobotan TF-IDF**

Setelah proses pelabelan selesai dilakukan, tahap berikutnya adalah pembobotan data dengan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini berfungsi untuk menghitung tingkat kepentingan suatu kata dalam sebuah ulasan dibandingkan dengan keseluruhan kumpulan data. TF-IDF memberikan bobot yang lebih besar pada kata-kata yang sering muncul dalam satu dokumen (ulasan) namun jarang muncul di ulasan lainnya, sehingga kata tersebut dianggap memiliki makna yang lebih spesifik dan penting dalam konteks ulasan tersebut. Melalui proses pembobotan ini, sistem dapat mengidentifikasi kata-kata kunci yang mewakili sentimen pengguna secara lebih akurat. Sebagai contoh, kata 'driver' dan 'delay' ditemukan sering muncul dalam ulasan dengan sentimen negatif, yang mengindikasikan adanya masalah terkait layanan pengemudi atau keterlambatan. Sebaliknya, kata 'cepat' dan 'mudah' lebih sering muncul dalam ulasan positif, yang menunjukkan kepuasan pengguna terhadap kecepatan dan kemudahan layanan aplikasi Gojek. Dengan pembobotan TF-IDF,

proses analisis data menjadi lebih terarah karena setiap kata yang muncul dinilai berdasarkan relevansi dan frekuensinya dalam keseluruhan data, bukan hanya sekadar jumlah kemunculannya.

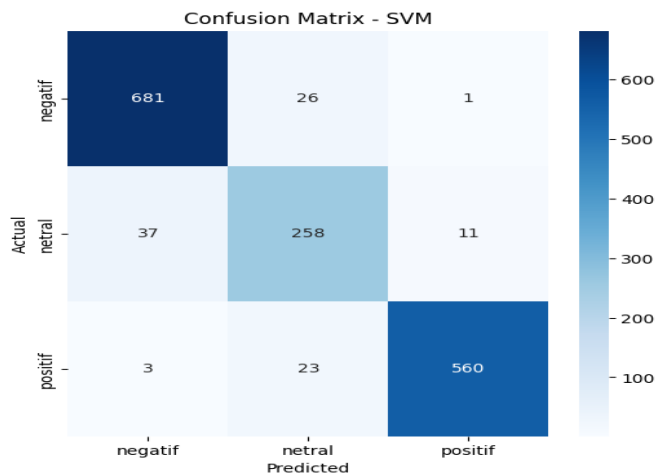
Proses pembobotan menggunakan TF-IDF menghasilkan sebuah matriks yang memuat nilai bobot untuk setiap kata pada masing-masing ulasan. Matriks ini berfungsi sebagai representasi numerik dari teks yang telah diolah, di mana setiap baris mewakili satu ulasan dan setiap kolom mewakili satu kata yang telah dihitung bobotnya. Matriks TF-IDF ini menjadi komponen penting yang digunakan sebagai input dalam tahap klasifikasi data. Dalam penelitian ini, algoritma Support Vector Machine (SVM) digunakan untuk melakukan klasifikasi sentimen berdasarkan pola kata yang telah dibobotkan. TF-IDF membantu algoritma SVM dalam mengenali dan membedakan pola distribusi kata yang khas pada masing-masing kelas sentimen, baik positif, negatif, maupun netral. Dengan adanya bobot dari TF-IDF, SVM dapat menentukan garis pemisah (hyperplane) yang optimal untuk memisahkan data ke dalam kelas sentimen yang sesuai. Dengan demikian, pembobotan TF-IDF tidak hanya berfungsi untuk mengukur pentingnya kata, tetapi juga memegang peran krusial dalam meningkatkan akurasi dan ketepatan hasil klasifikasi. Berikut contoh perhitungan TF-IDF:

**Tabel 8.** Hasil TF-IDF

Kata	TF-IDF D1	TF-IDF D2	TF-IDF D3
sangat	0.0352	0	0
membantu	0.0352	0	0
sekali	0.0352	0	0
dalam	0	0	0
perjalanan	0.0352	0	0
cepat	0	0	0
mudah	0	0	0
murah	0	0.044	0
bayarnya	0	0.044	0
lebih	0	0	0.0294
mencari	0	0	0.0294
pengemudi	0	0	0.0294

### 3.2 Hasil Pemodelan dan Evaluasi Model

Setelah proses pembobotan selesai, data kemudian dibagi menjadi dua bagian, yaitu 80% sebagai data training dan 20% sebagai data testing. Pembagian ini bertujuan untuk melatih dan menguji model secara terpisah, sehingga dapat mengevaluasi performa model dengan lebih objektif. Pada tahap ini, model Support Vector Machine (SVM) dilatih menggunakan data training yang sebelumnya telah melalui proses pembobotan dengan metode TF-IDF. Dalam proses pelatihan, digunakan kernel linear yang dianggap paling sesuai dan efisien untuk tugas klasifikasi teks. Kernel linear mampu memisahkan data secara efektif dalam ruang vektor berdimensi tinggi yang dihasilkan dari TF-IDF, tanpa memerlukan transformasi data yang kompleks. Pemilihan kernel ini juga didasarkan pada karakteristik data teks yang umumnya bersifat linier, sehingga algoritma dapat bekerja dengan lebih cepat dan akurat. Dengan pendekatan ini, model SVM diharapkan mampu mempelajari pola-pola kata yang menjadi ciri khas masing-masing sentimen dan membangun garis pemisah yang optimal untuk mengklasifikasikan ulasan dengan tepat.



**Gambar 2.** Confusion Matrix

Hasil evaluasi model yang dilakukan menggunakan confusion matrix menunjukkan performa yang sangat baik dalam klasifikasi sentimen. Berdasarkan hasil perhitungan, model menghasilkan akurasi sebesar 93,69%,



presisi sebesar 93,65%, recall sebesar 93,66%, dan F1-Score sebesar 93,67%. Nilai-nilai ini mencerminkan keseimbangan yang optimal antara kemampuan model dalam mengidentifikasi sentimen secara tepat (presisi), mendeteksi semua data yang relevan (recall), serta menjaga konsistensi antara kedua ukuran tersebut (F1-Score). Dengan akurasi yang tinggi, model terbukti mampu mengklasifikasikan ulasan pengguna ke dalam kategori sentimen positif, negatif, dan netral dengan tingkat kesalahan yang sangat rendah. Hasil ini menunjukkan bahwa algoritma Support Vector Machine (SVM) yang dilatih dengan pembobotan TF-IDF dan kernel linear telah bekerja secara efektif dalam mengenali pola-pola kata yang menjadi karakteristik masing-masing sentimen. Kinerja yang baik ini menjadi indikator bahwa model yang dikembangkan dapat diandalkan untuk menganalisis sentimen pengguna terhadap aplikasi Gojek dengan tingkat akurasi yang memadai. Gambar 2 berikut adalah matriks confusion yang dihasilkan:

Berdasarkan hasil confusion matrix pada model SVM, diperoleh hasil sebagai berikut: data aktual negatif sebagian besar berhasil diklasifikasikan dengan benar sebanyak 681 data, dengan kesalahan klasifikasi sebanyak 26 data yang diprediksi sebagai netral dan 1 data yang diprediksi sebagai positif. Data aktual netral sebagian besar juga berhasil diklasifikasikan dengan benar sebanyak 258 data, namun terdapat 37 data yang salah diklasifikasikan sebagai negatif dan 11 data yang salah diklasifikasikan sebagai positif. Data aktual positif sebagian besar diklasifikasikan dengan benar sebanyak 560 data, meskipun terdapat 3 data yang salah diklasifikasikan sebagai negatif dan 23 data yang salah diklasifikasikan sebagai netral.

### 3.3 Pembahasan

Hasil analisis sentimen menunjukkan bahwa sebagian besar pengguna memberikan ulasan dengan sentimen negatif terhadap aplikasi Gojek. Temuan ini sejalan dengan pernyataan M. R. Patel bahwa analisis sentimen dapat memberikan wawasan yang mendalam mengenai persepsi pelanggan terhadap suatu produk atau layanan melalui pola kata yang sering muncul dalam ulasan [15]. Dalam penelitian ini, beberapa kata kunci yang paling sering muncul dalam ulasan negatif antara lain 'driver lambat', 'layanan buruk', 'aplikasi error', dan 'saldo tertunda'. Kata-kata tersebut mencerminkan keluhan yang berkaitan dengan kecepatan pelayanan, kualitas layanan, gangguan teknis pada aplikasi, dan keterlambatan pemrosesan saldo yang menjadi perhatian pengguna. Sebaliknya, ulasan dengan sentimen positif didominasi oleh kata kunci seperti 'pelayanan cepat', 'driver ramah', 'fitur lengkap', dan 'proses mudah'. Pola ini menunjukkan bahwa pengguna menghargai kecepatan layanan, keramahan pengemudi, kelengkapan fitur aplikasi, serta kemudahan dalam penggunaan layanan. Hal ini sejalan dengan pendapat C. A. Haryani, A. E. Widjaja, H. Hery, and F. V. Ferdinand yang menyatakan bahwa sentimen positif biasanya mencerminkan pengalaman yang memuaskan dan cenderung berkaitan dengan aspek kecepatan, kemudahan, serta kenyamanan [16]. Melalui temuan ini, pihak Gojek dapat memanfaatkan hasil analisis untuk memperbaiki aspek layanan yang sering dikritik serta mempertahankan keunggulan yang sudah diapresiasi oleh pengguna.

Dari hasil analisis yang telah dilakukan, dapat diinterpretasikan bahwa keluhan utama yang disampaikan oleh pengguna aplikasi Gojek berkaitan dengan performa driver dan stabilitas aplikasi. Permasalahan seperti keterlambatan driver, layanan yang kurang memuaskan, serta gangguan teknis pada aplikasi menjadi fokus utama keluhan yang muncul dalam ulasan negatif. Temuan ini menunjukkan bahwa aspek operasional dan teknis masih menjadi tantangan yang perlu mendapatkan perhatian serius dari pihak pengembang. Aspek ini sangat penting bagi pengembang aplikasi untuk melakukan evaluasi dan perbaikan, khususnya dalam meningkatkan kinerja layanan pengemudi, memperbaiki bug aplikasi, serta mengoptimalkan sistem agar lebih stabil dan responsif. Dengan adanya perbaikan pada area tersebut, diharapkan tingkat kepuasan pengguna dapat meningkat, sehingga kepercayaan dan loyalitas pelanggan terhadap aplikasi Gojek dapat terjaga dengan baik. Hasil ini juga memperkuat pendapat yang disampaikan oleh O. E. Putri, V. H. Pranatawijaya, and N. Kristianti bahwa analisis sentimen dapat menjadi alat yang efektif untuk mengidentifikasi titik-titik kritis yang perlu ditingkatkan dalam layanan berbasis teknologi [17].

Model Support Vector Machine (SVM) yang dibangun dalam penelitian ini menunjukkan kemampuan klasifikasi yang sangat baik, meskipun data yang digunakan bersifat teks yang tidak terstruktur. SVM mampu mengolah dan memetakan data teks ke dalam ruang vektor berdimensi tinggi secara efektif, sehingga tetap dapat membedakan pola-pola kata yang menjadi ciri khas masing-masing sentimen dengan akurasi yang tinggi [18]. Kemampuan SVM dalam menangani data berdimensi tinggi, seperti yang dijelaskan oleh T. Joachims, merupakan salah satu keunggulan utamanya dalam klasifikasi teks [19]. Dalam konteks penelitian ini, keunggulan tersebut terbukti dengan performa model yang mampu mengklasifikasikan ulasan pengguna ke dalam kategori sentimen positif, negatif, dan netral dengan tingkat kesalahan yang minimal. Hasil ini menunjukkan bahwa SVM adalah algoritma yang andal dan dapat diandalkan dalam menganalisis data teks yang kompleks dan tidak terstruktur.

Beberapa penelitian sebelumnya mendukung efektivitas algoritma Support Vector Machine (SVM) dalam analisis sentimen. Salah satunya adalah penelitian yang dilakukan oleh Setiawan [3], yang menunjukkan bahwa SVM mampu mencapai akurasi sebesar 87% dalam menganalisis sentimen ulasan pada platform marketplace. Selain itu, penelitian terdahulu oleh Pang dan Lee [4] yang menerapkan metode Naïve Bayes dalam analisis sentimen film berhasil memperoleh akurasi sebesar 82%. Perbandingan hasil penelitian ini dengan studi sebelumnya menunjukkan adanya peningkatan akurasi yang signifikan. Hal ini dapat dicapai berkat penerapan tahapan preprocessing yang lebih komprehensif, mulai dari pembersihan data hingga proses stemming yang dilakukan secara menyeluruh. Selain itu, pembobotan menggunakan metode TF-IDF yang optimal dan pemilihan



kernel linear yang tepat pada algoritma SVM turut berkontribusi dalam meningkatkan performa model [20]. Keunggulan lain dari penelitian ini terletak pada jumlah dataset yang besar, yaitu 8.000 ulasan pengguna aplikasi Gojek, serta periode data yang lebih mutakhir, yaitu tahun 2024. Data yang lebih baru memberikan relevansi yang lebih tinggi terhadap kondisi layanan terkini, sehingga hasil analisis yang diperoleh menjadi lebih akurat dan dapat dijadikan dasar pengambilan keputusan yang tepat bagi pengembang aplikasi [21].

Hasil penelitian ini memberikan beberapa implikasi penting yang dapat menjadi perhatian bagi pengembang aplikasi Gojek. Temuan bahwa mayoritas ulasan pengguna mengandung sentimen negatif menjadi sinyal kuat bahwa perbaikan perlu diprioritaskan, khususnya pada aspek layanan driver dan stabilitas sistem aplikasi. Pengembang dapat memanfaatkan hasil ini sebagai dasar dalam menyusun strategi peningkatan kualitas layanan yang lebih tepat sasaran dan sesuai dengan kebutuhan pengguna. Selain itu, penelitian ini juga menunjukkan bahwa penerapan analisis sentimen berbasis machine learning merupakan pendekatan yang efektif untuk memahami kebutuhan, harapan, dan tingkat kepuasan pengguna secara real-time. Dengan memanfaatkan model Support Vector Machine (SVM) yang didukung oleh tahapan preprocessing yang komprehensif dan pembobotan TF-IDF yang optimal, analisis sentimen dapat dilakukan dengan tingkat akurasi yang tinggi dan hasil yang dapat diandalkan. Keberhasilan model dalam penelitian ini dapat menjadi acuan yang berguna bagi penelitian selanjutnya, khususnya yang berfokus pada analisis ulasan aplikasi mobile. Mengingat data ulasan bersifat dinamis dan terus berkembang, penggunaan metode serupa sangat relevan untuk diterapkan secara berkelanjutan guna memantau persepsi dan kepuasan pengguna di masa mendatang. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi teoritis dalam bidang analisis sentimen, tetapi juga menawarkan manfaat praktis yang dapat langsung diterapkan dalam pengembangan dan evaluasi layanan aplikasi.

Berdasarkan hasil perhitungan confusion matrix tersebut didapat akurasi sebesar 93,69%, recall 93,66%, presisi 93,65% dan F1-score 93,67%.

## 4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi Gojek yang diunduh dari Play Store pada tahun 2024 dengan menggunakan metode Support Vector Machine (SVM). Permasalahan utama yang diangkat dalam penelitian ini adalah bagaimana mengidentifikasi persepsi pengguna secara akurat berdasarkan ulasan yang diberikan di Play Store serta bagaimana mengukur performa model klasifikasi yang mampu menangani data teks tidak terstruktur dalam jumlah besar. Berdasarkan hasil penelitian, ditemukan bahwa mayoritas ulasan pengguna memiliki sentimen negatif, yang menunjukkan adanya ketidakpuasan yang perlu menjadi perhatian pengembang aplikasi. Model SVM yang diterapkan dalam penelitian ini terbukti mampu mengklasifikasikan sentimen dengan akurasi tinggi sebesar 93,69%, yang didukung oleh hasil evaluasi presisi, recall, dan F1-Score yang seimbang. Hal ini menunjukkan bahwa metode yang digunakan efektif dalam menyelesaikan permasalahan penelitian dan memberikan solusi yang tepat dalam menganalisis persepsi pengguna. Namun, penelitian ini memiliki keterbatasan pada metode pelabelan yang masih berbasis lexicon dan belum sepenuhnya mampu menangkap konteks kalimat, seperti ironi atau sarkasme yang sering muncul dalam ulasan pengguna. Selain itu, penelitian hanya memanfaatkan ulasan dari satu platform, yaitu Google Play Store, sehingga cakupan data masih terbatas. Untuk penelitian selanjutnya, disarankan untuk menggunakan metode labeling berbasis supervised learning dengan pelabelan manual serta memperluas sumber data dari berbagai platform seperti media sosial dan App Store untuk memperoleh hasil yang lebih komprehensif dan representatif.

## REFERENCES

- [1] Y. Amirkhalili and H. Y. Wong, "Banking on Feedback: Text Analysis of Mobile Banking iOS and Google App Reviews," arxiv, 2025, [Online]. Available: <http://arxiv.org/abs/2503.11861>
- [2] M. Hanafi, S. Adi, and A. Setiawan, "A Model of Sentiment Analysis on Gojek Application Review using Word Vector Representation and Long Short-Term Memory (LSTM)," 2025 Int. Conf. Comput. Sci. Eng. Technol. Innov., pp. 944–949, 2025, doi: 10.1109/ICoCSETI63724.2025.11019757.
- [3] Z. Rahman, P. Sakinah, Y. Hendra, B. Satria, F. Maulana, and A. Q. Ayun, "Sentiment Analysis of Gojek App Reviews on Google Play Store with Natural Language Processing Using Naive Bayes Algorithm," *Jatilima J. Multimed. Dan Teknol. Inf.*, vol. 06, no. 03, pp. 60–69, 2024, [Online]. Available: <https://journal.cattleyadf.org/index.php/jatilima/index>
- [4] P. Amri, D. M. Suri, and Syuhada, "The analysis of ride hailing user characteristics from app reviews," *J. Siasat Bisnis*, vol. 28, no. 2, pp. 241–262, 2024, doi: 10.20885/jsb.vol28.iss2.art7.
- [5] C. Anilkumar, S. V E., S. Kanchana, and S. B. Kumar, "Sentimental Analysis on Product Reviews Using Support Vector Machine and Nave Bayes," *Appl. Comput. Eng.*, vol. 2, no. 1, pp. 66–72, 2023, doi: 10.54254/2755-2721/2/20220586.
- [6] M. Ahmad, S. Aftab, M. S. Bashir, and N. Hameed, "Sentiment analysis using SVM: A systematic literature review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 2, pp. 182–188, 2018, doi: 10.14569/IJACSA.2018.090226.
- [7] D. M. Abdullah and A. M. Abdulazeez, "Machine Learning Applications based on SVM Classification: A Review," *Qubahan Acad. J.*, vol. 1, no. 2, pp. 81–90, 2021, doi: 10.48161/qaj.v1n2a50.
- [8] A. Madasu and S. Elango, "Efficient feature selection techniques for sentiment analysis," *Multimed. Tools Appl.*, vol. 79, no. 9–10, pp. 6313–6335, 2020, doi: 10.1007/s11042-019-08409-z.
- [9] M. Hamka and Tukiran, "Analisis Sentimen Pengguna E-Commerce dan Marketplace Menggunakan Support Vector Machine," *J. Rekayasa Sist. Inf. dan Teknol.*, vol. 1, no. 4, pp. 273–282, 2024, doi: 10.59407/jrsit.v1i4.555.



- [10] D. Bandorski et al., “Contraindications for video capsule endoscopy,” *World J. Gastroenterol.*, vol. 22, no. 45, pp. 9898–9908, 2016, doi: 10.3748/wjg.v22.i45.9898.
- [11] A. I. Ramadhan and E. B. Setiawan, “Aspect-based Sentiment Analysis on Social Media Using Convolutional Neural Network (CNN) Method,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, pp. 1828–1836, 2023, doi: 10.47065/bits.v4i4.3103.
- [12] Moh. Heri Setiawan, I Gede Aris Gunadi, and Gede Indrawan, “Klasifikasi Pelayanan Kesehatan Berdasarkan Data Sentimen Pelayanan Kesehatan menggunakan Multiclass Support Vector Machine,” *J. Sist. dan Inform.*, vol. 17, no. 1, pp. 47–54, 2023, doi: 10.30864/jsi.v17i1.512.
- [13] S. D. Lestari and E. B. Setiawan, “Sentiment Analysis Based on Aspects Using FastText Feature Expansion and NBSVM Classification Method,” *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 469–477, 2022, doi: 10.47065/josyc.v3i4.2202.
- [14] Tukino and Fifi, “Penerapan Support Vector Machine Untuk Analisis Sentimen Pada Layanan Ojek Online,” *JDDAT*, vol. 3, no. 2, pp. 104–113, 2024.
- [15] M. R. Patel, “Analytics and Research Project: Analyzing Retail Sentiment with Current Methodology and Emerging Technology,” Iowa State University, 2024.
- [16] C. A. Haryani, A. E. Widjaja, H. Hery, and F. V. Ferdinand, “Sentiment Analysis of User Satisfaction Towards Sales Promotion of Gojek Application Service Using Support Vector Machine (SVM),” *Ultim. InfoSys J. Ilmu Sist. Inf.*, vol. 14, no. 2, pp. 66–70, 2023, doi: 10.31937/si.v14i2.3398.
- [17] O. E. Putri, V. H. Pranatawijaya, and N. Kristianti, “Analisis Sentimen Berbasis Aspek Dan Deteksi Emosi Pada Coffee Shop Palangka Raya Menggunakan Deep Learning,” *J. Inf. Technol. Comput. Sci.*, vol. 4, no. 3, pp. 249–262, 2024, doi: 10.47111/jointecom.v4i3.19185.
- [18] K. Suresh Kumar et al., “Sentiment Analysis of Short Texts Using SVMs and VSMs-Based Multiclass Semantic Classification,” *Appl. Artif. Intell.*, vol. 38, no. 1, 2024, doi: 10.1080/08839514.2024.2321555.
- [19] T. Joachims, “UNIVERSIT AT DORTMUND Fachbereich Informatik Lehrstuhl VIII K unstliche Intelligenz Making Large-Scale SVM Learning Practical LS { 8 Report 24,” no. October 1999, 2018, doi: 10.17877/DE290R-5097.
- [20] A. A. J. Karim, K. H. M. Asad, and A. Azam, “Strengthening Fake News Detection: Leveraging SVM and Sophisticated Text Vectorization Techniques. Defying BERT?,” *arxiv*, August, 2024, [Online]. Available: <http://arxiv.org/abs/2411.12703>
- [21] Y. Guo, C. Hu, and Y. Yang, “Predict the Future from the Past? On the Temporal Data Distribution Shift in Financial Sentiment Classifications,” *EMNLP 2023 - 2023 Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 1029–1038, 2023, doi: 10.18653/v1/2023.emnlp-main.65.