

Analisis Sentimen Ulasan Aplikasi Gojek Menggunakan Support Vector Machine

DOI: <http://dx.doi.org/10.35889/progresif.v22i3.3969>

Creative Commons License 4.0 (CC BY –NC)



Muhammad Fariz¹, Ade Davy Wiranata^{2*}

Teknik Informatika, Universitas Muhammadiyah Prof. DR. HAMKA, Jakarta, Indonesia

*e-mail Corresponding Author: adedavy@uhamka.ac.id

Abstract

Evaluating Gojek's service quality can be achieved by extracting valuable insights from the Google Play Store review corpus. However, the surging volume of data necessitates computational automation to substitute inefficient manual analysis. This study develops a sentiment analysis model using the Support Vector Machine (SVM) algorithm applied to 2,000 scraped reviews. The corpus undergoes comprehensive text preprocessing, rating-based polarity labeling, and lexical feature extraction via Term Frequency-Inverse Document Frequency (TF-IDF). Model validation demonstrates an impressive predictive performance, achieving an accuracy of 86.21%. The distribution of public opinion reveals a dominance of positive sentiment at 57.47%, outperforming the negative cluster at 42.53%. Empirically, the methodological fusion of TF-IDF and SVM is highly reliable as an analytical instrument to dissect consumer perceptions for future service improvements.

Keywords: Sentiment Classification; Support Vector Machine; TF-IDF Weighting; Service Evaluation; Gojek.

Abstrak

Evaluasi kualitas layanan Gojek dapat diekstraksi dari korpus ulasan Google Play Store. Namun, lonjakan volume data menuntut otomatisasi komputasional guna mensubstitusi analisis manual yang inefisien. Penelitian ini mengonstruksi model analisis sentimen menggunakan algoritma *Support Vector Machine* (SVM) terhadap 2.000 entri ulasan hasil *web scraping*. Korpus diproses melalui tahapan prapemrosesan teks komprehensif, pelabelan polaritas berbasis *rating*, serta ekstraksi fitur leksikal mendayagunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Validasi arsitektur klasifikasi membuktikan performa prediktif yang impresif, dengan raihan akurasi menyentuh 86,21%. Peta distribusi opini publik mendemonstrasikan dominasi sentimen afirmatif (positif) sebesar 57,47%, mengungguli klaster negatif di angka 42,53%. Secara empiris, fusi metodologis antara pembobotan TF-IDF dan SVM terbukti sangat andal sebagai instrumen analitik komputasional guna membedah persepsi konsumen demi mendukung evaluasi perbaikan layanan Gojek di masa depan.

Kata Kunci: Klasifikasi Sentimen; Support Vector Machine; Pembobotan TF-IDF; Evaluasi Layanan; Gojek.

1. Pendahuluan

Perkembangan teknologi digital telah mendorong transformasi yang signifikan pada berbagai sektor kehidupan, termasuk industri transportasi berbasis aplikasi. Salah satu platform yang berkembang pesat di Indonesia adalah Gojek, yang tidak lagi hanya menyediakan layanan transportasi daring, tetapi juga menghadirkan berbagai layanan digital seperti pesan antar makanan, logistik, pembayaran elektronik, hingga berbagai layanan pendukung lainnya. Menurut [1] dan [2], perkembangan ekosistem digital tersebut memberikan kontribusi yang besar terhadap peningkatan aktivitas ekonomi digital nasional sekaligus mengubah pola interaksi masyarakat dalam memenuhi kebutuhan sehari-hari. Seiring meningkatnya jumlah pengguna, volume ulasan yang diberikan melalui Google Play Store juga terus bertambah sehingga menjadi sumber

informasi yang penting dalam menggambarkan pengalaman pengguna terhadap kualitas layanan aplikasi.

Ulasan pengguna pada Google Play Store tidak hanya merepresentasikan tingkat kepuasan, tetapi juga mengandung berbagai kritik, keluhan, maupun apresiasi terhadap layanan yang diberikan. Pengguna umumnya menyampaikan pengalaman mengenai kegagalan sistem (*error*), kendala transaksi pembayaran, keterlambatan layanan, permasalahan mitra pengemudi, hingga kualitas fitur aplikasi. Informasi tersebut memiliki nilai yang sangat penting bagi pengembang sebagai dasar evaluasi dan pengambilan keputusan untuk meningkatkan kualitas layanan. Namun demikian, jumlah ulasan yang sangat besar serta terus bertambah setiap waktu menyebabkan proses analisis secara manual menjadi tidak efektif. Selain itu, karakteristik data yang berbentuk teks tidak terstruktur menjadikan proses identifikasi opini semakin kompleks sehingga memerlukan pendekatan komputasional yang mampu melakukan ekstraksi informasi secara otomatis melalui teknik *text mining* dan *Natural Language Processing* (NLP) [3].

Salah satu penerapan *text mining* yang banyak digunakan untuk mengolah opini pengguna adalah analisis sentimen. Metode ini bertujuan mengelompokkan suatu opini ke dalam kategori sentimen positif maupun negatif berdasarkan isi teks yang dituliskan pengguna [4]. Agar proses klasifikasi menghasilkan performa yang baik, data teks terlebih dahulu harus melalui tahapan *preprocessing* yang meliputi *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming* sehingga diperoleh data yang lebih bersih dan konsisten [5]. Selanjutnya, representasi teks diubah menjadi bentuk numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Menurut [6], TF-IDF mampu memberikan bobot pada setiap kata berdasarkan tingkat kepentingannya di dalam dokumen sehingga fitur yang dihasilkan menjadi lebih representatif dan dapat meningkatkan kualitas proses klasifikasi.

Berbagai penelitian sebelumnya telah menunjukkan bahwa algoritma Support Vector Machine (SVM) merupakan salah satu metode yang memiliki performa tinggi dalam analisis sentimen. Hal tersebut disebabkan SVM memiliki kemampuan yang baik dalam menangani data berdimensi tinggi seperti data teks [7]. Iqrom dkk. [8] menerapkan kombinasi TF-IDF dan SVM untuk menganalisis sentimen ulasan Gojek, Grab, dan Maxim menggunakan data Google Play Store, serta memperoleh tingkat akurasi sebesar 94% pada objek Gojek. Penelitian lain oleh Andreansyah [9] juga memanfaatkan algoritma SVM untuk mengklasifikasikan sentimen ulasan Gojek dengan capaian akurasi sebesar 67,6%. Sementara itu, Dasuki dkk. [10] membuktikan bahwa SVM mampu menghasilkan akurasi sebesar 88,20% pada analisis sentimen aplikasi pinjaman daring, sehingga semakin memperkuat bahwa algoritma tersebut layak digunakan untuk klasifikasi opini berbasis teks.

Penelitian mengenai analisis sentimen terus mengalami perkembangan melalui penerapan berbagai teknik optimasi. Bima Aditya dkk. [11] mengombinasikan TF-IDF, SMOTE, SVM, dan Random Forest sehingga memperoleh akurasi sebesar 96% pada klasifikasi sentimen ulasan Gojek. Selain itu, Nugraha dan Gustian [12] membandingkan performa beberapa platform transportasi daring menggunakan SVM, sedangkan penelitian mengenai layanan GoPay menunjukkan bahwa SVM memiliki performa yang lebih baik dibandingkan Decision Tree maupun Random Forest [12]. Meskipun demikian, sebagian besar penelitian terdahulu lebih berfokus pada perbandingan algoritma, penerapan teknik penyeimbangan data, maupun optimasi model klasifikasi. Kajian yang menyajikan implementasi analisis sentimen secara lengkap mulai dari proses *web scraping*, *preprocessing*, pelabelan sentimen berbasis *rating*, ekstraksi fitur TF-IDF, klasifikasi menggunakan SVM, hingga evaluasi model menggunakan data ulasan Gojek yang diperoleh langsung dari Google Play Store masih relatif terbatas. Kondisi tersebut menunjukkan adanya *research gap* yang menjadi dasar dilakukannya penelitian ini. Oleh karena itu, *state of the art* penelitian ini terletak pada penyajian alur analisis sentimen yang terintegrasi dan komprehensif menggunakan data aktual hasil *scraping* Google Play Store dengan implementasi penuh pada lingkungan Google Colaboratory.

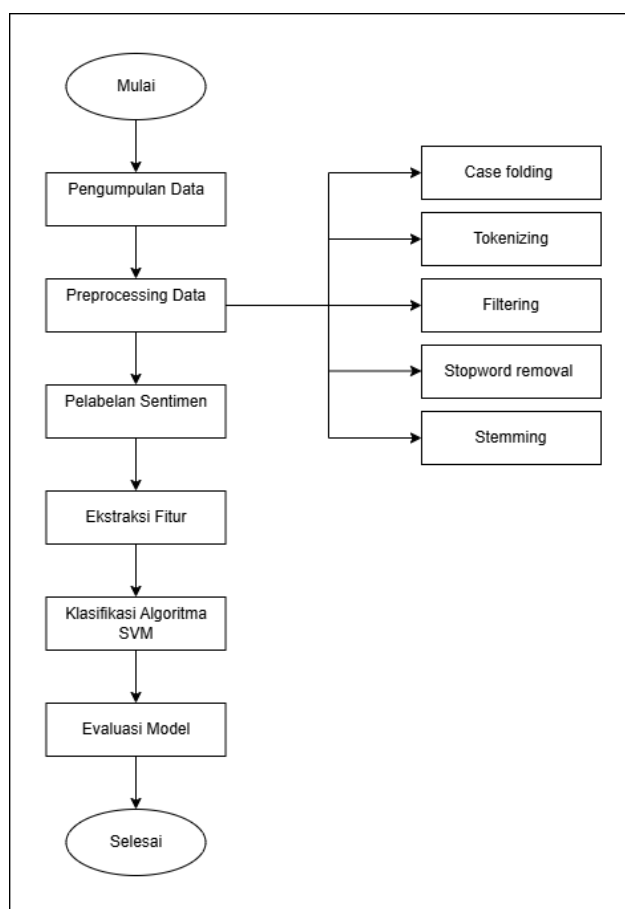
Berdasarkan kesenjangan penelitian tersebut, penelitian ini mengusulkan kombinasi Term Frequency–Inverse Document Frequency (TF-IDF) dan Support Vector Machine (SVM) sebagai pendekatan untuk mengklasifikasikan sentimen ulasan pengguna aplikasi Gojek. Metode TF-IDF dipilih karena mampu menghasilkan representasi numerik yang menggambarkan tingkat kepentingan setiap kata dalam dokumen [6]. Selanjutnya, algoritma SVM digunakan karena memiliki kemampuan membangun *hyperplane* optimal untuk memisahkan data pada ruang berdimensi tinggi sehingga banyak digunakan dalam penelitian analisis sentimen dengan tingkat akurasi yang tinggi [7]. Efektivitas kombinasi kedua metode tersebut juga telah dibuktikan oleh

beberapa penelitian sebelumnya, seperti yang dilakukan oleh Iqrom dkk. [8], Dasuki dkk. [10], dan Bima Aditya dkk. [11], sehingga pendekatan ini dinilai memiliki dasar teoritis dan empiris yang kuat untuk menyelesaikan permasalahan klasifikasi sentimen pada ulasan Gojek.

Berdasarkan uraian tersebut, penelitian ini bertujuan menganalisis sentimen pengguna aplikasi Gojek melalui tahapan yang meliputi pengumpulan data menggunakan *web scraping*, *preprocessing* teks, pelabelan sentimen berdasarkan *rating*, ekstraksi fitur menggunakan TF-IDF, serta klasifikasi menggunakan algoritma *Support Vector Machine* (SVM). Kinerja model selanjutnya dievaluasi menggunakan *confusion matrix* dengan parameter *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian diharapkan mampu memberikan gambaran yang lebih komprehensif mengenai persepsi pengguna terhadap kualitas layanan Gojek, sekaligus menjadi masukan yang dapat dimanfaatkan oleh pengembang dalam meningkatkan kualitas layanan berbasis data ulasan pengguna.

2. Metodologi

Riset ini bersandar pada paradigma kuantitatif dengan mendayagunakan algoritma *Support Vector Machine* (SVM) sebagai instrumen utama untuk mengategorikan polaritas sentimen pada komentar konsumen Gojek. Rangkaian metodologi dalam studi ini diatur secara sekuensial, yang diinisiasi melalui tahap pengumpulan korpus teks dari platform Google Play Store. Pascakoleksi, data mentah tersebut diolah lebih lanjut melewati fase prapemrosesan tekstual, anotasi kelas opini, serta transformasi bobot fitur bahasa berbasis *Term Frequency–Inverse Document Frequency* (TF-IDF). Babak selanjutnya mencakup eksekusi pemilihan kelas oleh model SVM, yang berujung pada pengukuran keandalan arsitektur komputasi, dan dipungkas dengan representasi grafis persebaran kosakata dominan memanfaatkan *WordCloud*. Secara komprehensif, kerangka kerja operasional riset ini diilustrasikan melalui Gambar 1.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Data

Data Korpus data primer yang dieksploitasi dalam kajian ini dihimpun melalui mekanisme ekstraksi otomatis (*web scraping*). Eksekusi penarikan data tersebut dioperasikan di atas persekitaran komputasi awan Google Colaboratory dengan mendayagunakan modul *google-play-scraper*. Material utama bersumber dari jejak pendapat konsumen platform Gojek yang bertebaran di etalase Google Play Store, di mana proses filtrasinya dikonfigurasi secara spesifik untuk menjangkau lalu lintas teks berbahasa Indonesia (*lang='id'*) pada teritorial nasional (*country='id'*). Penarikan ini membuahkan total 2.000 entri opini yang mencakup tiga atribut esensial, yakni substansi narasi ulasan, kuantifikasi skor kepuasan (*rating*), serta jejak waktu pengunggahan (tanggal publikasi). Sebagai langkah finalisasi pada fase ini, keseluruhan himpunan data mentah tersebut diarsipkan ke dalam ekstensi *Comma Separated Values* (CSV) guna memfasilitasi tahapan prapemrosesan selanjutnya.

2.2 Tahap Preprocessing

Fase prapemrosesan diimplementasikan sebagai langkah esensial untuk memurnikan korpus tekstual sebelum memasuki eksekusi pemilahan kelas. Objektif utama dari prosedur tata olah ini adalah untuk mengeliminasi aneka elemen derau maupun atribut informasi yang nirfaedah, yang pada gilirannya akan melahirkan format representasi data yang jauh lebih presisi dan siap untuk dikomputasi. Pemaparan komprehensif mengenai rangkaian tahapan prapemrosesan yang diadopsi dalam riset ini direkapitulasi secara rinci pada Tabel 1.

Tabel 1. Tahapan *Preprocessing*

Tahapan	Deskripsi
Cleaning	Menghapus URL, angka, simbol, tanda baca, emoji, dan karakter khusus lainnya.
Case Folding	Mengubah seluruh huruf menjadi huruf kecil (<i>lowercase</i>).
Tokenizing	Memecah kalimat menjadi kumpulan token atau kata.
Stopword Removal	Menghapus kata-kata umum yang tidak memiliki makna penting.
Stemming	Mengubah kata berimbuhan menjadi bentuk kata dasar menggunakan algoritma stemming bahasa Indonesia.

Hasil dari tahapan *preprocessing* digunakan sebagai masukan pada proses pelabelan sentimen dan ekstraksi fitur.

2.3 Pelabelan Sentimen

Penentuan anotasi sentimen pada riset ini diorkestrasikan secara komputasional dengan menjadikan parameter skor kepuasan (*rating*) dari konsumen di Google Play Store sebagai rujukan absolut. Entri ulasan dengan raihan skor 1 dan 2 secara definitif dialokasikan ke dalam kelas polaritas negatif, berbanding terbalik dengan skor 4 beserta 5 yang diposisikan pada kutub sentimen positif. Lebih lanjut, data dengan skor 3 dieliminasi dari skema pemodelan. Kebijakan eksklusi ini diterapkan secara sengaja lantaran nilai tersebut mengindikasikan kecenderungan sikap netral, yang diyakini berisiko tinggi memantik bias interpretasi maupun ambiguitas matematis tatkala algoritma menjalankan tugas pengklasifikasian.

Tabel 2. Aturan Pelabelan Sentimen

Rating	Label Sentimen
1–2	Negatif
3	Tidak digunakan
4–5	Positif

2.4 Ekstraksi Fitur Menggunakan TF-IDF

Korpus teks yang telah tuntas melewati fase prapemrosesan selanjutnya ditransformasikan ke dalam wujud kuantitatif mendayagunakan skema *Term Frequency–Inverse Document Frequency* (TF-IDF). Secara konseptual, formulasi pembobotan ini bekerja dengan mengalkulasi rasio signifikansi suatu leksikon (kata) bersandar pada intensitas kemunculannya di dalam satu entri spesifik, yang kemudian dikomparasikan terhadap total populasi dokumen di dalam korpus keseluruhan. Hasil kalkulasi matematis inilah yang nantinya diekstraksi sebagai variabel independen (*fitur masukan*) bagi arsitektur pengklasifikasi. Melalui mekanisme tersebut, substansi linguistik dari masing-masing opini konsumen sepenuhnya terkonversi ke dalam susunan vektor angka yang siap dikomputasi oleh model algoritma.

Secara matematis, bobot *Term Frequency–Inverse Document Frequency* (TF-IDF) diperoleh melalui kombinasi nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Nilai TF menunjukkan frekuensi kemunculan suatu kata pada sebuah dokumen, sedangkan IDF menunjukkan tingkat kelangkaan kata tersebut terhadap seluruh dokumen pada korpus. Persamaan TF-IDF dirumuskan sebagai berikut.

$$\text{TF-IDF}_{t,d} = \text{TF}_{t,d} \times \text{IDF}_t \quad (1)$$

dengan

$$\text{IDF}_t = \log \left(\frac{N}{df_t} \right)$$

dengan keterangan:

- $\text{TF}_{t,d}$ = frekuensi kemunculan kata t pada dokumen d
- IDF_t = bobot *inverse document frequency* kata t
- N = jumlah seluruh dokumen
- df_t = jumlah dokumen yang mengandung kata t

Melalui proses tersebut setiap dokumen direpresentasikan dalam bentuk vektor numerik yang selanjutnya digunakan sebagai masukan (*feature vector*) pada proses klasifikasi menggunakan algoritma *Support Vector Machine*.

2.5 Klasifikasi Menggunakan Support Vector Machine

Fase pengelompokan opini dieksekusi secara terpusat dengan mengandalkan arsitektur *Support Vector Machine* (SVM). Operasionalisasi komputasi ini direalisasikan sepenuhnya di dalam lingkungan komputasi awan Google Colaboratory, dengan mengadopsi fungsionalitas dari modul pustaka *Scikit-learn*. Secara mekanis, algoritma SVM ditugaskan untuk mengonstruksi model matematis yang mampu membentuk batas pemisah (*hyperplane*) secara presisi. Garis pembatas inilah yang berperan mendikotomi ulasan konsumen ke dalam ranah positif atau negatif, dengan bersandar mutlak pada matriks bobot TF-IDF yang telah diproduksi sebelumnya.

Pada fase pelatihan model, keseluruhan korpus data yang telah melalui tahapan prapemrosesan, pelabelan, dan ekstraksi fitur pembobotan TF-IDF dibagi menjadi dua himpunan bagian, yaitu data latih (*training set*) dan data uji (*testing set*). Pembagian komposisi data dilakukan dengan rasio 85:15, di mana 85% data dimanfaatkan sebagai data latih dan 15% sisa data dialokasikan sebagai data uji. Dari total populasi data hasil pelabelan (setelah pengeluaran ulasan ber-rating 3), sebanyak 1.642 sampel ulasan difungsikan dalam fase pelatihan untuk mengarahkan algoritma SVM mempelajari pola, membentuk vektor bobot, serta mengoptimalkan margin pemisah antarkelas. Selanjutnya, sebanyak 290 sampel ulasan dialokasikan khusus sebagai data uji (*testing support set*) guna mengevaluasi keandalan dan generalisasi model dalam mengklasifikasikan polaritas sentimen pada data baru yang belum pernah dipelajari sebelumnya.

Secara matematis, *Support Vector Machine* (SVM) bekerja dengan membentuk *hyperplane* optimal yang mampu memisahkan dua kelas data secara maksimal. Persamaan *hyperplane* dirumuskan sebagai berikut.

$$w^T x + b = 0 \quad (2)$$

dengan:

- w = vektor bobot (*weight vector*)
- x = vektor fitur
- b = bias

Tujuan utama algoritma SVM adalah memperoleh *hyperplane* optimal dengan memaksimalkan margin antarkelas melalui fungsi optimasi berikut.

$$\min \frac{1}{2} \|w\|^2$$

dengan kendala:

$$y_i(w^T x_i + b) \geq 1, \quad i = 1, 2, \dots, n \quad (3)$$

dengan:

- x_i = data latih ke- i
- y_i = label kelas (+1 atau -1)
- n = jumlah data latih

Setelah proses pelatihan selesai dilakukan, kelas suatu data baru ditentukan menggunakan fungsi keputusan (*decision function*) berikut.

$$f(x) = \text{sign}(w^T x + b) \quad (4)$$

Apabila nilai $f(x)$ bernilai positif maka data diklasifikasikan ke dalam kelas positif, sedangkan apabila bernilai negatif maka data diklasifikasikan ke dalam kelas negatif. Pada penelitian ini, representasi fitur hasil pembobotan TF-IDF digunakan sebagai masukan bagi algoritma SVM untuk melakukan klasifikasi sentimen ulasan pengguna aplikasi Gojek.

2.6 Evaluasi Model

Pengukuran keandalan arsitektur model diuji secara komprehensif melalui representasi matriks konfusi (*confusion matrix*), dengan memproyeksikan empat parameter evaluasi utama: akurasi (*accuracy*), presisi (*precision*), perolehan kembali (*recall*), serta skor F1 (*F1-score*). Parameter akurasi merefleksikan rasio kebenaran prediksi model secara menyeluruh terhadap total entri data uji. Di sisi lain, presisi menitikberatkan pada tingkat ketepatan model dalam memproyeksikan label positif dari keseluruhan estimasi yang diprediksi sebagai kelas positif. Sementara itu, nilai *recall* mengukur efisiensi algoritma dalam menjaring kembali seluruh sampel aktual yang sejatinya berada pada kutub positif. Guna menyelaraskan kedua parameter tersebut, skor F1 dihitung untuk memotret nilai rata-rata harmonis antara presisi dan *recall*. Akumulasi hasil kalkulasi pada fase evaluasi ini menjadi parameter objektif dalam menilai kepiawaian fungsional algoritma SVM ketika mengklasifikasikan polaritas ulasan konsumen Gojek.

Nilai evaluasi model dihitung menggunakan parameter *accuracy*, *precision*, *recall*, dan *F1-score* yang dirumuskan sebagai berikut.

Accuracy

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Precision

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

Recall

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

F1-Score

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

dengan

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

Nilai keempat metrik tersebut digunakan untuk mengukur performa model Support Vector Machine dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Gojek.

2.7 Visualisasi WordCloud

Sebagai fase konklusi dari seluruh rangkaian metodologi, dilakukan representasi visual atas luaran analisis dengan mengonstruksi awan kata (*WordCloud*). Teknik penyajian grafis ini ditujukan untuk memproyeksikan leksikon-leksikon yang memiliki intensitas kemunculan tertinggi

pada tiap-tiap polaritas opini. Secara mekanis, skala dimensi fisik dari setiap kata di dalam diagram *WordCloud* tersebut berkorelasi linear dengan tingkat frekuensi kemunculannya di dalam korpus data. Konseptualisasi visual ini sengaja diterapkan untuk memudahkan penelaah dalam mengidentifikasi secara cepat intisari topik, keluhan utama, ataupun bentuk apresiasi dominan yang melingkupi kubu sentimen positif maupun negatif.

3. Hasil dan Pembahasan

3.1 Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini diperoleh melalui proses *web scraping* terhadap ulasan pengguna aplikasi Gojek pada Google Play Store menggunakan pustaka Google Play Scraper yang dijalankan pada lingkungan Google Colaboratory. Proses pengambilan data dikonfigurasi menggunakan parameter bahasa Indonesia (*lang='id'*) dan wilayah Indonesia (*country='id'*), sehingga data yang diperoleh merepresentasikan ulasan pengguna aplikasi Gojek di Indonesia.

Sebanyak 2.000 data ulasan berhasil dikumpulkan dan disimpan dalam format Comma Separated Values (CSV). Dataset terdiri atas tiga atribut utama, yaitu ulasan, rating, dan tanggal. Atribut ulasan berisi komentar pengguna terhadap layanan Gojek, atribut *rating* menunjukkan penilaian pengguna dengan rentang nilai 1 hingga 5, sedangkan atribut tanggal menunjukkan waktu publikasi ulasan.

Data hasil *scraping* masih berupa data mentah (*raw data*), sehingga belum dapat langsung digunakan dalam proses analisis sentimen. Oleh karena itu, dataset selanjutnya melalui tahapan *preprocessing* untuk membersihkan dan menyiapkan data sebelum dilakukan pelabelan sentimen, ekstraksi fitur menggunakan Term Frequency–Inverse Document Frequency (TF-IDF), serta klasifikasi menggunakan algoritma Support Vector Machine (SVM).

Distribusi kelas sentimen pada dataset diperoleh melalui proses pelabelan berdasarkan nilai *rating*. Hasil pelabelan menunjukkan bahwa 57,47% data termasuk ke dalam sentimen positif dan 42,53% termasuk ke dalam sentimen negatif.



ulasan	rating	tanggal
GK pernah bisa dapat playletter y	2	2026-06-09 16:22:56
Aplikasi ini sangat membantu sekali untuk berpergian mudah diakses , ditingkatkan terus untuk pelayanan lebih baik lagi terimakasih	5	2026-06-09 16:12:19
mantap	5	2026-06-09 15:51:18
membantu	5	2026-06-09 15:17:30
ko skarang makin naik ya tarifnya.. mohon sesuaikan sperti semula	3	2026-06-09 14:25:28
yg jelas dan yg penting selalu menyenangkan karena tepat waktu /cepat tidak be Baartele tele	5	2026-06-09 14:20:51
aplikasi yang sangat membantu dan memudahkan masyarakat dalam bidang transportasi,keungan dan membantu membuka lapangan tenaga kerja 🙏	5	2026-06-09 14:10:58
sekarang kalo cari gojar tambah susah dan lama banget, giliran udah dapet drivernya yg banyak alasan	2	2026-06-09 13:54:44
aplikasi TAIK, pakai gojek express barang hilang karna ulah driver, pihak gojek tidak ada tanoquna jawab, respon lambat, tidak mau tau, beda banget sama	1	2026-06-09 13:51:20

Gambar 1. Dataset Hasil Scraping Ulasan Pengguna Aplikasi Gojek dari Google Play Store

Gambar 1 menunjukkan sebagian data hasil *web scraping* yang berhasil dikumpulkan. Dataset terdiri atas tiga atribut utama, yaitu ulasan, rating, dan tanggal. Atribut *ulasan* berisi komentar pengguna terhadap layanan Gojek, *rating* menunjukkan penilaian pengguna pada skala 1–5, sedangkan atribut *tanggal* menunjukkan waktu publikasi ulasan. Dataset inilah yang selanjutnya digunakan sebagai data awal sebelum memasuki tahap *preprocessing*.

3.2 Hasil Preprocessing

Tahap *preprocessing* dilakukan untuk mempersiapkan data teks sebelum memasuki proses ekstraksi fitur dan klasifikasi sentimen. Data hasil *web scraping* masih mengandung berbagai elemen yang tidak diperlukan, seperti tanda baca, angka, simbol, emoji, serta variasi penulisan kata yang dapat memengaruhi proses analisis. Oleh karena itu, diterapkan tahapan *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming* untuk menghasilkan data yang lebih bersih dan terstruktur.

Penelitian ini berhasil mengumpulkan 2.000 ulasan pengguna aplikasi Gojek dari Google Play Store yang terdiri atas atribut teks ulasan, *rating*, dan tanggal publikasi. Dataset tersebut kemudian diproses melalui tahapan *preprocessing* sebelum digunakan pada proses pelabelan

sentimen, ekstraksi fitur TF-IDF, dan klasifikasi menggunakan Support Vector Machine (SVM). Hasil transformasi pada setiap tahapan ditunjukkan pada Tabel 1.

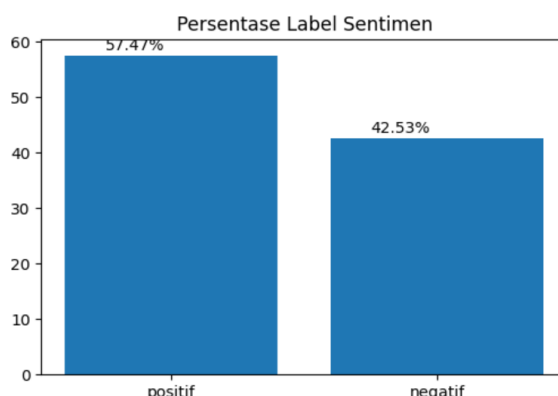
Tabel 1. Hasil Transformasi Data Ulasan pada Tahapan Preprocessing

Tahapan	Hasil Transformasi
Cleaning	GK pernah bisa dapat playletter y
Case Folding	gk pernah bisa dapat playletter y
Tokenizing	[gk, pernah, bisa, dapat, playletter, y]
Stopword Removal	[gk, playletter, y]
Stemming	gk playletter y

Berdasarkan Tabel 1, setiap tahapan *preprocessing* menghasilkan perubahan pada struktur data teks. Setelah seluruh proses selesai dilakukan, diperoleh korpus yang lebih bersih dan terstandarisasi sehingga siap digunakan pada tahap ekstraksi fitur menggunakan metode TF-IDF.

3.3 Pelabelan Sentimen

Setelah tahap *preprocessing* selesai dilakukan, data ulasan diberi label sentimen berdasarkan nilai *rating* pada Google Play Store. Ulasan dengan *rating* 4 dan 5 dikategorikan sebagai sentimen positif, sedangkan *rating* 1 dan 2 dikategorikan sebagai sentimen negatif. Sementara itu, ulasan dengan *rating* 3 tidak digunakan karena dianggap merepresentasikan sentimen netral sehingga tidak diikutsertakan dalam proses klasifikasi.



Gambar 2. Hasil Pelabelan Sentimen Pengguna Aplikasi Gojek

Berdasarkan Gambar 2, distribusi kelas sentimen menunjukkan bahwa **57,47%** data termasuk ke dalam sentimen positif, sedangkan **42,53%** termasuk ke dalam sentimen negatif. Hasil pelabelan ini selanjutnya digunakan sebagai dasar dalam proses ekstraksi fitur menggunakan metode TF-IDF dan klasifikasi sentimen menggunakan algoritma Support Vector Machine (SVM).

3.4 Transformasi Fitur Menggunakan TF-IDF

Setelah proses pelabelan sentimen selesai dilakukan, data teks ditransformasikan ke dalam bentuk numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Tahap ini bertujuan untuk mengubah setiap dokumen menjadi representasi vektor berdasarkan bobot setiap kata sehingga dapat digunakan sebagai masukan bagi algoritma Support Vector Machine (SVM).

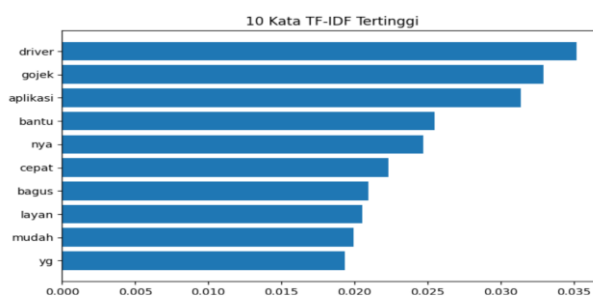
Hasil pembobotan TF-IDF menghasilkan sejumlah kata dengan nilai bobot tertinggi yang disajikan pada Tabel 2.

Tabel 2. Daftar Kata dengan Bobot TF-IDF Tertinggi

No.	Kata	Bobot TF-IDF
1	driver	0,035198
2	gojek	0,032958
3	aplikasi	0,031370
4	bantu	0,025469
5	nya	0,024733
6	cepat	0,022318
7	bagus	0,020954
8	layan	0,020534
9	mudah	0,019965
10	yg	0,019330

Sumber: Hasil pengolahan data menggunakan metode TF-IDF (2026).

Berdasarkan Tabel 2, kata "driver" memiliki bobot TF-IDF tertinggi sebesar 0,035198, diikuti oleh kata "gojek" dan "aplikasi". Selain itu, kata "bantu", "cepat", "bagus", "layan", dan "mudah" juga termasuk dalam sepuluh kata dengan bobot tertinggi pada dataset.

**Gambar 3.** Visualisasi Sepuluh Kata dengan Bobot TF-IDF Tertinggi

Gambar 3 memperlihatkan visualisasi bobot TF-IDF dari sepuluh kata yang memiliki nilai tertinggi. Hasil pembobotan tersebut selanjutnya digunakan sebagai representasi fitur (*feature representation*) pada proses klasifikasi sentimen menggunakan algoritma Support Vector Machine (SVM).

3.5 Hasil Klasifikasi Support Vector Machine (SVM)

Setelah ekstraksi fitur TF-IDF selesai, dilakukan pemodelan klasifikasi menggunakan Support Vector Machine (SVM). Proses ini diawali dengan fase pelatihan menggunakan 1.642 sampel data latih (85% data) di Google Colaboratory. Algoritma SVM mengalkulasikan matriks bobot (w) dan bias (b) guna mengonstruksi batas pemisah (*hyperplane*) optimal yang memisahkan kelas positif dan negatif secara maksimal. Selanjutnya, dilakukan fase pengujian menggunakan 290 sampel data uji (15% data) yang belum pernah dipelajari model untuk menguji keandalan prediksinya.

Sebelum mengevaluasi performa secara agregat, dilakukan analisis terhadap sampel ulasan yang berhasil maupun gagal dikenali oleh model SVM, sebagaimana disajikan pada Tabel 3.

Tabel 3. Sampel Hasil Klasifikasi SVM pada Data Uji

Teks Ulasan Pengguna	Label Awal	Prediksi SVM	Status Klasifikasi
"Aplikasi sangat membantu untuk bepergian, driver ramah dan cepat."	Positif	Positif	Berhasil (True Positive)
"Sangat kecewa, driver sering cancel orderan dan aplikasi sering error."	Negatif	Negatif	Berhasil (True Negative)
"Aplikasi bagus tapi tarifnya mahal dan kadang driver susah didapat."	Negatif	Positif	Salah (False Positive)
"Drivernya sih ramah banget, cuma sayang sistem aplikasinya sering lelet."	Positif	Negatif	Salah (False Negative)

Tabel 3 menunjukkan bahwa model berhasil mengidentifikasi opini yang tegas (seperti ramah, kecewa, cancel), namun cenderung mengalami kesalahan prediksi (*misclassification*) pada kalimat ber-sentimen campuran (*mixed sentiment*) atau kalimat ambigu.

Kinerja model dievaluasi menggunakan *Confusion Matrix* pada Gambar 5 serta metrik evaluasi pada Gambar 4.

Akurasi: 86.21 %

	precision	recall	f1-score	support
negatif	0.829630	0.868217	0.848485	129.000000
positif	0.890323	0.857143	0.873418	161.000000
accuracy	0.862069	0.862069	0.862069	0.862069
macro avg	0.859976	0.862680	0.860951	290.000000
weighted avg	0.863325	0.862069	0.862327	290.000000

Gambar 4. Hasil Klasifikasi Menggunakan Support Vector Machine

Berdasarkan hasil klasifikasi pada Gambar 4, model SVM memperoleh akurasi sebesar 86,21%. Pada kelas sentimen negatif diperoleh nilai precision sebesar 82,96%, recall sebesar 86,82%, dan F1-score sebesar 84,85%. Sementara itu, pada kelas sentimen positif diperoleh nilai precision sebesar 89,03%, recall sebesar 85,71%, dan F1-score sebesar 87,34%. Nilai macro average F1-score sebesar 86,10% dan weighted average F1-score sebesar 86,23% menunjukkan hasil evaluasi model secara keseluruhan.

Untuk melihat distribusi hasil prediksi pada masing-masing kelas sentimen, dilakukan evaluasi menggunakan Confusion Matrix yang ditunjukkan pada Gambar 5.

Confusion Matrix

Aktual	Negatif	112	17
	Positif	23	138
	Prediksi	Negatif	Positif

Gambar 5. Confusion Matri Model Support Vector Machine

Berdasarkan Gambar 5, model berhasil mengklasifikasikan 112 data sentimen negatif sebagai True Negative dan 138 data sentimen positif sebagai True Positive. Selain itu, terdapat 17 data yang termasuk False Positive dan 23 data yang termasuk False Negative.

Tabel 4. Ringkasan Hasil Evaluasi Model Support Vector Machine

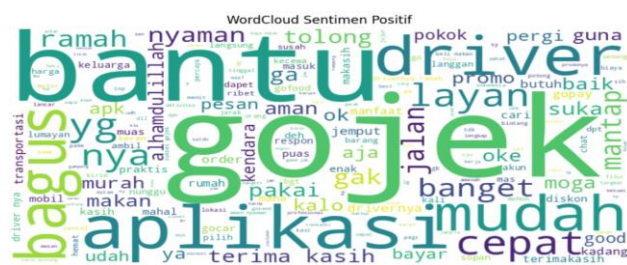
Metrik	Sentimen Positif	Sentimen Negatif
Precision	89,03%	82,96%
Recall	85,71%	86,82%
F1-Score	87,34%	84,85%

Berdasarkan hasil evaluasi tersebut, model Support Vector Machine mampu mengklasifikasikan sentimen ulasan pengguna aplikasi Gojek dengan tingkat akurasi sebesar 86,21%.

3.6 Analisis WordCloud

Sebagai pelengkap hasil klasifikasi sentimen, penelitian ini juga menyajikan visualisasi WordCloud untuk menggambarkan kata-kata yang paling sering muncul pada ulasan pengguna aplikasi Gojek. Visualisasi dibuat menggunakan data yang telah melalui tahapan *preprocessing*, sehingga kata-kata yang ditampilkan merupakan kata yang telah dibersihkan dan dinormalisasi. Ukuran setiap kata pada WordCloud menunjukkan frekuensi kemunculannya di dalam dataset.

Untuk mempermudah analisis, WordCloud dibagi menjadi dua kelompok, yaitu sentimen positif dan sentimen negatif. Pemisahan ini bertujuan untuk memperlihatkan karakteristik kata yang dominan pada masing-masing kelas sentimen.



Gambar 6. WordCloud Sentimen Positif

Berdasarkan Gambar 6, kata-kata yang paling dominan pada sentimen positif antara lain "gojek", "bantu", "aplikasi", "mudah", "layan", "cepat", dan "bagus". Kata-kata tersebut merupakan kosakata yang paling sering muncul pada ulasan pengguna yang diklasifikasikan sebagai sentimen positif.



Gambar 7. WordCloud Sentimen Negatif

Berdasarkan Gambar 7, kata-kata yang dominan pada sentimen negatif meliputi "driver", "gojek", "aplikasi", "gak", "bayar", "cancel", "susah", "pesan", dan "sistem". Kata-kata tersebut merupakan kosakata yang paling sering muncul pada ulasan pengguna yang diklasifikasikan sebagai sentimen negatif.

Hasil visualisasi WordCloud memberikan gambaran mengenai kata-kata yang dominan pada masing-masing kelas sentimen dan digunakan sebagai pendukung hasil analisis sentimen yang diperoleh dari proses klasifikasi menggunakan algoritma *Support Vector Machine* (SVM).

3.6 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma Support Vector Machine (SVM) yang dikombinasikan dengan pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) mampu mengklasifikasikan sentimen ulasan pengguna aplikasi Gojek dengan tingkat akurasi sebesar 86,21%. Capaian tersebut menunjukkan bahwa pendekatan yang digunakan mampu mengidentifikasi pola sentimen positif dan negatif pada ulasan pengguna secara efektif. Selain menghasilkan performa klasifikasi yang baik, penelitian ini juga berhasil memetakan karakteristik opini pengguna melalui distribusi sentimen dan visualisasi WordCloud, sehingga tujuan penelitian untuk menganalisis sentimen pengguna aplikasi Gojek dapat tercapai.

Temuan penelitian ini sejalan dengan konsep text mining dan Natural Language Processing (NLP) yang menyatakan bahwa data teks tidak terstruktur dapat diubah menjadi informasi yang bernilai melalui tahapan *preprocessing*, ekstraksi fitur, dan klasifikasi. Penerapan metode TF-IDF mampu merepresentasikan kata-kata penting dalam dokumen ke dalam bentuk numerik, sedangkan algoritma SVM bekerja dengan membentuk *hyperplane* optimal untuk memisahkan kelas sentimen positif dan negatif. Kombinasi kedua metode tersebut

memungkinkan model memperoleh representasi fitur yang lebih baik sehingga mampu menghasilkan performa klasifikasi yang tinggi.

Hasil penelitian ini juga mendukung beberapa penelitian terdahulu. Iqrom dkk. [8] melaporkan bahwa algoritma SVM mampu memperoleh akurasi sebesar 94% pada analisis sentimen ulasan aplikasi transportasi daring. Penelitian Dasuki dkk. [10] juga menunjukkan bahwa SVM menghasilkan akurasi sebesar 88,20% pada analisis sentimen aplikasi pinjaman daring. Selain itu, Bima Aditya dkk. [11] memperoleh akurasi hingga 96% melalui kombinasi TF-IDF, SMOTE, dan SVM. Meskipun akurasi pada penelitian ini sedikit lebih rendah dibandingkan kedua penelitian tersebut, hasil yang diperoleh masih menunjukkan bahwa SVM merupakan algoritma yang memiliki performa baik untuk klasifikasi sentimen ulasan aplikasi berbasis teks.

Visualisasi WordCloud memperlihatkan bahwa kata-kata seperti "driver", "gojek", dan "aplikasi" merupakan kata yang paling sering muncul pada ulasan pengguna. Pada kelompok sentimen positif, kata-kata seperti "bantu", "mudah", "cepat", dan "bagus" lebih dominan, sedangkan pada sentimen negatif lebih banyak muncul kata "cancel", "bayar", "susah", dan "sistem". Temuan ini menunjukkan bahwa pengguna tidak hanya memberikan penilaian terhadap aplikasi, tetapi juga terhadap kualitas layanan, performa mitra pengemudi, dan pengalaman penggunaan aplikasi secara keseluruhan.

Penelitian ini memberikan kontribusi dalam penerapan metode TF-IDF dan Support Vector Machine untuk analisis sentimen ulasan aplikasi Gojek menggunakan data yang diperoleh langsung melalui proses *web scraping* dari Google Play Store. Hasil penelitian diharapkan dapat menjadi masukan bagi pengembang aplikasi dalam memahami persepsi pengguna terhadap layanan yang diberikan, sekaligus menjadi referensi bagi penelitian selanjutnya yang mengembangkan analisis sentimen pada aplikasi layanan digital.

Penelitian ini masih memiliki beberapa keterbatasan. Pelabelan sentimen dilakukan berdasarkan nilai *rating* sehingga belum sepenuhnya merepresentasikan isi teks ulasan. Selain itu, penelitian hanya menggunakan algoritma SVM tanpa melakukan perbandingan dengan algoritma klasifikasi lainnya maupun teknik optimasi model. Oleh karena itu, penelitian selanjutnya dapat mempertimbangkan penggunaan metode pelabelan yang lebih representatif, menambahkan algoritma pembanding seperti Random Forest atau Naïve Bayes, serta menerapkan teknik optimasi atau *hyperparameter tuning* untuk meningkatkan performa model.

4. Simpulan

Berdasarkan seluruh rangkaian eksperimen komputasional yang telah diselenggarakan, dapat disimpulkan bahwa integrasi metodologis antara skema *Term Frequency–Inverse Document Frequency* (TF-IDF) sebagai teknik ekstraksi fitur dan arsitektur *Support Vector Machine* (SVM) sebagai pengklasifikasi biner sukses diaplikasikan untuk membedah polaritas sentimen ulasan pengguna aplikasi Gojek pada Google Play Store dengan raihan akurasi akumulatif mencapai 86,21% serta sebaran metrik evaluasi yang stabil pada kedua kelas. Peta distribusi opini publik secara distributif mencatat dominasi sentimen afirmatif (positif) sebesar 57,47% yang mengungguli sentimen negatif sebesar 42,53%, di mana temuan kualitatif melalui proyeksi *WordCloud* mempertegas bahwa kepuasan konsumen berakar pada aspek simplifikasi antarmuka dan reliabilitas layanan, sedangkan keluhan terfokus pada anomali teknis peranti lunak, hambatan transaksi, serta performa interpersonal mitra pengemudi. Guna mengakselerasi penyempurnaan studi ini di masa mendatang, agenda penelitian selanjutnya sangat disarankan untuk mengekspansi skala volume korpus data ulasan, menginisiasi teknik optimasi hiperparameter kernel model, serta menyelenggarakan studi komparatif lintas algoritma tingkat lanjut demi mendongkrak batas atas ketajaman serta reliabilitas sistem analisis sentimen secara menyeluruh.

Daftar Referensi

- [1] H. Aldabbas, A. Bajahzar, M. Alruily, A. A. Qureshi, R. M. A. Latif, and M. Farhan, "Google Play Content Scraping and Knowledge Engineering Using Natural Language Processing Techniques with the Analysis of User Reviews," *Journal of Intelligent Systems*, vol. 30, no. 1, pp. 192–208, 2020. doi: 10.1515/jisys-2019-0197.
- [2] H. Andreansyah, "Klasifikasi Sentimen Positif dan Negatif pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine (SVM)," *Jurnal Riset Sistem Informasi dan Teknik Informatika (JURASIK)*, vol. 9, no. 1, pp. 329–336, 2024.

- [3] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 1, pp. 147–156, 2021. doi: 10.25126/jtiik.202183944
- [4] A. B. Aditya, S. Samsudin, W. P. Rizki, M. Mahendra, and A. Setiawan, "Analisis Sentimen Ulasan Aplikasi Gojek Menggunakan Support Vector Machine dan Random Forest," *Jurnal Informatika Terpadu*, vol. 11, no. 2, pp. 134–143, 2025.
- [5] L. E. Dewi and D. Sundiman, "Development of Sharing Economy Business Model for Company Sustainability: A Case Study on Gojek," *Binus Business Review*, vol. 14, no. 1, pp. 29–38, 2022. doi: 10.21512/bbr.v14i1.8598.
- [6] M. D. Dhiyaulhaq and P. H. Gunawan, "Sentiment Analysis of the Jakarta–Bandung Fast Train Project Using the SVM Method," *Jurnal Media Informatika Budidarma*, vol. 7, no. 4, 2023. doi: 10.30865/mib.v7i4.6855.
- [7] T. Elizabeth, "Analisis Sentimen Ulasan Aplikasi Primaku Menggunakan Metode Support Vector Machine," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 4, 2022.
- [8] L. Gunawan, "Exploring Impacts of Gojek Application as a Sharing Economy Tool on Indonesian Economics and Society: Systematic Literature Review Approach," *Journal of Ecohumanism*, vol. 3, no. 4, pp. 1390–1405, 2024. doi: 10.62754/joe.v3i4.3670.
- [9] M. H. Mahendra, D. T. Murdiansyah, and K. M. Lhaksana, "Analisis Sentimen Tweet COVID-19 Menggunakan Metode K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan CountVectorizer," *Dike: Jurnal Ilmu Multidisiplin*, 2023.
- [10] N. Hendrastuty, A. R. Isnain, and A. Y. Rahmadhani, "Analisis Sentimen Masyarakat terhadap Program Kartu Prakerja pada Twitter dengan Metode Support Vector Machine," vol. 6, no. 3, 2021.
- [11] I. K. S. Irma, M. A. Yasin, and I. A. Salihi, "Analisis Sentimen terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Journal of Information Science*, vol. 36, no. 6, pp. 823–848, 2023. doi: 10.1177/0165551510388123.
- [12] M. Iqrom, M. Afdal, R. Novita, M. Rahmawita, and T. K. Ahsyar, "Sentiment Analysis of Gojek, Grab, and Maxim Applications Using Support Vector Machine Algorithm," vol. 10, no. 1, 2023.
- [13] U. Jazuli and A. D. Wiranata, "Analisis Sentimen Pengguna terhadap Aplikasi MyPertamina Berdasarkan Ulasan Google Play Store Menggunakan Algoritma Support Vector Machine," SVM, n.d.
- [14] M. Khoirul, U. Hayati, and O. Nurdiawan, "Analisis Sentimen Aplikasi BRIMO pada Ulasan Pengguna di Google Play Menggunakan Algoritma Naïve Bayes," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 1, 2023.
- [15] K. A. Dasuki, S. S. Hilabi, F. N. Apriani, and Tukino, "Analisis Sentimen Pinjaman Online Akulaku dan Kredivo dengan Metode Support Vector Machine (SVM)," *Journal of Mandalika Literature*, vol. 4, no. 4, pp. 323–332, 2023. doi: 10.36312/jml.v4i4.2045.
- [16] F. A. Kusumo, D. R. S. Saputro, and P. Widyaningsih, "Sentiment Analysis of Reviews on Apps on Google Play Store Using Support Vector Machine and N-Gram Feature Selection," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 19, no. 2, pp. 1037–1046, 2025. doi: 10.30598/barekengvol19iss2pp1037-1046.
- [17] D. Nugraha and D. Gustian, "SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika) Universitas Nusa Putra," *Prosiding Seminar Nasional SISMATIK*, 2023.
- [18] D. S. Putri and T. Ridwan, "Analisis Sentimen Ulasan Aplikasi Pospay dengan Algoritma Support Vector Machine," n.d.
- [19] H. Tuhuteru, "Analisis Sentimen Masyarakat terhadap Pembatasan Sosial Berskala Besar Menggunakan Algoritma Support Vector Machine," n.d.

- [20] J. A. Wibowo, V. C. Mawardi, and T. Sutrisno, "Visualisasi Word Cloud Hasil Analisis Sentimen Berbasis Fitur Layanan Aplikasi Gojek dengan Support Vector Machine," *Jurnal Serina Sains, Teknik dan Kedokteran*, vol. 2, no. 1, pp. 61–70, 2024. doi: 10.24912/jsstk.v2i1.32058.
- [21] M. Y. Siregar, A. D. Wiranata, and R. A. Saputra, "Analisis Sentimen pada Ulasan Pengguna Aplikasi Streaming Vidio Menggunakan Metode Naïve Bayes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 5, pp. 2419–2429, 2024.
- [22] A. D. Prastiko and A. D. Wiranata, "Analisis Sentimen Publik terhadap Fenomena Judi Online di Media Sosial dengan SVM," *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, vol. 5, 2025.
- [23] M. Iqbal, A. D. Wiranata, R. Suwito, and R. F. Ananda, "Perbandingan Algoritma Naïve Bayes, KNN, dan Decision Tree terhadap Ulasan Aplikasi Threads dan Twitter," *Media Online*, vol. 4, no. 3, pp. 1799–1807, 2023.
- [24] I. Indriani and A. D. Wiranata, "Comparison of Accuracy Levels of SVM, Decision Tree and Random Forest Algorithms in Sentiment Analysis of User Responses of the GoPay Application," *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 3, pp. 777–787, 2024.
- [25] A. D. Wiranata, S. Soleman, I. Irwansyah, I. K. Sudaryana, and R. Rizal, "Klasifikasi Data Mining untuk Menentukan Kualitas Udara di Provinsi DKI Jakarta Menggunakan Algoritma K-Nearest Neighbors (K-NN)," *Infotech: Journal of Technology Information*, vol. 9, no. 1, pp. 95–100, 2023.