Jurnal Teknologi Informasi dan Multimedia

e-ISSN: 2684-9151
p-ISSN: 2715-2529

https://journal.sekawan-org.id/index.php/jtim

Klasifikasi Sentimen Ulasan Pengguna Aplikasi Alfagift Menggunakan Metode *Support Vector Machine* Berbasis *TF-IDF*

Nazwa Asyifa[1*], Susanto[1]

[1] Program Studi Teknik Informatika, Universitas Semarang, Indonesia
[*] Corresponding author: nazwaasyifa0905@gmail.com

Sitasi: N. Asyifa, and S. Susanto, "Klasifikasi Sentimen Ulasan Pengguna Aplikasi Alfagift Menggunakan Metode *Support Vector Machine* Berbasis *TF-IDF*," *Jurnal Teknologi Informasi dan Multimedia*, vol. 8, no. 4, pp. 777–789, 2026.

Diterima : 19-05-2026
Direvisi : 29-06-2026
Disetujui : 09-07-2026
Diterbitkan : 02-09-2026

Abstrak. Perkembangan teknologi digital telah mendorong meningkatnya penggunaan aplikasi mobile di sektor ritel, salah satunya aplikasi Alfagift yang dikembangkan oleh PT Sumber Alfaria Trijaya Tbk. Ulasan pengguna di Google Play Store mengandung informasi berharga terkait kepuasan dan keluhan, namun volumenya yang sangat besar menyulitkan analisis manual sehingga diperlukan pendekatan otomatis berbasis kecerdasan buatan. Penelitian ini bertujuan mengklasifikasikan sentimen ulasan pengguna aplikasi Alfagift ke dalam tiga kelas yaitu Positif, Negatif, dan Netral, menggunakan metode Support Vector Machine (SVM) dengan representasi fitur TF-IDF. Sebanyak 4.937 data ulasan dikumpulkan melalui web scraping dan diproses melalui tahapan preprocessing meliputi case folding, normalisasi slang, tokenisasi, penghapusan stopword, dan stemming menggunakan PySastrawi. Ekstraksi fitur menggunakan TF-IDF konfigurasi unigram-bigram dengan pembatasan 10.000 fitur dan pembagian data 80:20. Model dievaluasi menggunakan confusion matrix dan 10-fold Stratified Cross Validation. Hasil menunjukkan akurasi data uji 87,15% dan rata-rata cross validation 87,28% dengan standar deviasi 1,0368%. Kelas Positif dan Negatif berhasil diklasifikasikan dengan baik (F1-score 0,9172 dan 0,8676), namun kelas Netral gagal diklasifikasikan akibat ketidakseimbangan distribusi data yang ekstrem, di mana kelas tersebut hanya mewakili 4,7% dari keseluruhan dataset.

Kata kunci: Analisis Sentimen; Support Vector Machine; TF-IDF; Google Play Store; Klasifikasi Teks

Abstract. The rapid development of digital technology has driven the increasing use of mobile applications in the retail sector, one of which is the Alfagift application developed by PT Sumber Alfaria Trijaya Tbk. User reviews on Google Play Store contain valuable information regarding user satisfaction and complaints, yet their large volume makes manual analysis inefficient, necessitating an automated approach. This study aims to classify the sentiment of Alfagift user reviews into three classes: Positive, Negative, and Neutral, using the Support Vector Machine (SVM) method with Term Frequency-Inverse Document Frequency (TF-IDF) feature representation. A total of 4,937 reviews were collected through web scraping and processed through preprocessing stages including case folding, slang normalization, tokenization, stopword removal, and stemming. Feature extraction used TF-IDF with unigram-bigram configuration and a maximum of 10,000 features, with an 80:20 data split ratio. The model was evaluated using a confusion matrix and 10-fold Stratified Cross Validation. The results show a test accuracy of 87.15% and a cross-validation average of 87.28% (standard deviation 1.0368%). The Positive and Negative classes were classified well (F1-score 0.9172 and 0.8676), while the Neutral class was poorly classified due to extreme class imbalance, representing only 4.7% of the entire dataset.

Keywords: Sentiment Analysis; Support Vector Machine; TF-IDF; Google Play Store; Text Classification.

1. Pendahuluan

Pertumbuhan teknologi digital yang cepat telah mendorong timbulnya bermacam aplikasi mobile di berbagai bidang, termasuk sektor ritel. Alfagift adalah aplikasi loy-al-itas yang dibuat oleh PT Sumber Alfaria Trijaya Tbk (Alfagift), yang memungkinkan pengguna untuk mengumpulkan poin, mendapatkan berbagai promosi eksklusif, melakukan pembayaran secara digital, serta memantau transaksi secara daring. Dengan populasi pengguna yang sangat besar, mencapai lebih dari satu juta pengguna aktif[1], Alfagift menjadi salah satu aplikasi belanja paling terkenal di Indonesia.

Tanggapan pengguna di Google Play Store adalah salah satu sumber informasi yang sangat berharga bagi pengembang aplikasi. Ulasan itu menggambarkan pengalaman nyata pengguna, masalah teknis, harapan akan fitur baru, dan kepuasan secara kese-luruhan. Permasalahan utama yang dihadapi adalah volume ulasan yang sangat besar dan terus bertambah setiap hari membuat proses identifikasi sentimen secara manual menjadi tidak praktis, memakan waktu lama, dan rawan bias subjektif dari penilai, sehingga keluhan kritis terkait bug atau fitur bermasalah berpotensi terlambat ditangani oleh pengembang. Untuk itu, diperlukan pendekatan otomatis berbasis kecerdasan buatan untuk menganalisis sentimen dari ulasan-ulasannya[2].

Analisis sentimen adalah bidang dalam *Natural Language Processing* (NLP) yang bertujuan untuk mengenali dan mengambil opini subjektif dari teks[3]. Dalam konteks penilaian aplikasi, analisis sentimen dapat membantu pengembang memahami elemen mana yang dihargai atau dikritik oleh pengguna secara otomatis dan terukur[4]. Beragam teknik *Machine Learning* telah diterapkan untuk analisis sentimen, termasuk Naïve Bayes, *Decision Tree, Random Forest, dan Support Vector Machine* (SVM).

Support Vector Machine (SVM) adalah algoritma *supervised learning* yang telah terbukti efektif dalam berbagai tugas klasifikasi teks[5]. Di antara berbagai metode tersebut, SVM dipilih berdasarkan bukti empiris dari penelitian terdahulu yang secara spesifik membandingkan performanya dengan algoritma lain pada domain analisis sentimen ulasan aplikasi di Google Play Store. Penelitian perbandingan pada ulasan aplikasi CapCut menunjukkan SVM mencapai akurasi 90,00%, lebih tinggi dibandingkan Naive Bayes yang hanya mencapai 84,50%[6], sementara penelitian lain melaporkan SVM mengalami peningkatan akurasi tertinggi (dari 75% menjadi 91%) dibandingkan Naive Bayes dan Random Forest pada dataset sentimen yang sama[7]. SVM bekerja sebagai *hyperplane* terbaik yang memisahkan kategori data di ruang fitur berdimensi tinggi. Kelebihan SVM dalam mengelola data berdimensi tinggi menjadikannya pilihan yang ideal untuk klasifikasi teks, terutama saat dipadukan dengan representasi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) yang dapat menangkap bobot pentingnya kata dalam kumpulan dokumen[8].

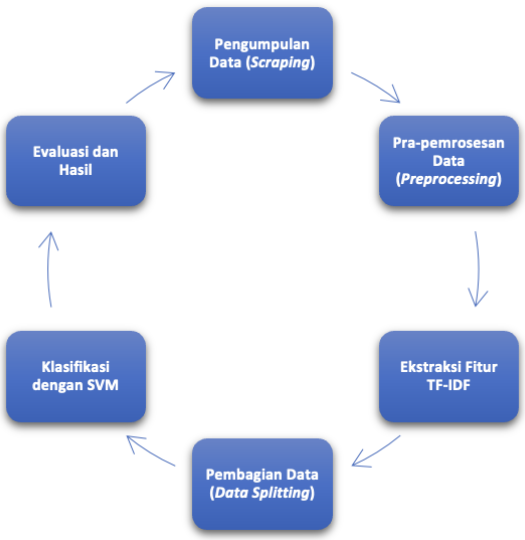
TF-IDF merupakan teknik penilaian kata yang memperhatikan seberapa sering kata muncul dalam dokumen (TF) dan menilai sejauh mana kata itu khas dalam keseluruhan korpus (IDF). Meskipun model berbasis transformer seperti BERT dan IndoBERT mampu menangkap konteks semantik kalimat secara lebih kaya, penelitian perbandingan pada klasifikasi sentimen melaporkan bahwa IndoBERT membutuhkan sumber daya komputasi yang signifikan lebih besar dibandingkan TF-IDF, sementara selisih akurasinya relatif tidak terlalu besar (IndoBERT 90,3% berbanding TF-IDF 82,0% pada kombinasi dengan algoritma yang sama)[9]. Mengingat keterbatasan sumber daya komputasi dan ukuran dataset penelitian ini yang tergolong menengah, TF-IDF dipilih karena lebih efisien secara komputasi, tidak memerlukan GPU atau proses finetuning model besar, serta tetap mampu menangkap bobot kepentingan kata dalam korpus dokumen secara memadai untuk dikombinasikan dengan SVM.

Penelitian analisis sentimen ulasan aplikasi *e-commerce* dan ritel di Indone-

sia menggunakan SVM dan TF-IDF telah banyak dilakukan, di antaranya pada aplikasi Shopee[10]. Namun, sebagian besar penelitian tersebut berfokus pada platform *e-commerce* umum dan belum ada yang secara spesifik mengkaji aplikasi loyalitas ritel seperti Alfagift. Selain itu, beberapa penelitian terdahulu yang menghadapi ketidakseimbangan distribusi kelas menerapkan teknik penanganan *imbalanced* data seperti SMOTE[11] atau *Random Oversampling* untuk meningkatkan performa pada kelas minoritas. Berbeda dengan penelitian-penelitian tersebut, penelitian ini berfokus secara khusus pada ulasan aplikasi Alfagift dengan mengevaluasi performa SVM-TF-IDF pada kondisi distribusi kelas yang sangat tidak seimbang tanpa teknik oversampling tambahan, sehingga dapat memberikan gambaran objektif mengenai keterbatasan pendekatan `class_weight` saja dalam menangani kelas minoritas pada konteks ulasan aplikasi ritel berbahasa Indonesia. Inilah kesenjangan (*research gap*) yang menjadi fokus penelitian ini. Studi ini bertujuan untuk menggunakan kombinasi metode SVM yang didasarkan pada TF-IDF untuk mengklasifikasikan sentimen ulasan pengguna aplikasi Alfagift ke dalam kategori positif, negatif, dan netral. Dataset diperoleh melalui proses pengambilan data dari *Google Play Store*, selanjutnya melalui tahap preprocessing yang komprehensif sebelum dilakukan pelatihan dan evaluasi model.

2. Bahan dan Metode

Pendekatan eksperimental dengan metode supervised learning digunakan dalam penelitian ini untuk mengklasifikasikan sentimen ulasan pengguna aplikasi Alfagift ke dalam tiga kelas: Positif, Negatif, dan Netral. Proses penelitian disusun secara sistematis melalui berbagai tahap terintegrasi untuk memastikan model klasifikasi memiliki tingkat akurasi yang optimal. Secara visual, seluruh tahapan dan proses sistematis dalam penelitian ini ditunjukkan pada Gambar 1 berikut:



Gambar 1. Alur Penelitian

Untuk mengubah ulasan tidak terstruktur menjadi informasi sentimen yang terukur, penelitian ini dirancang secara sistematis dalam enam tahapan utama, seperti yang ditunjukkan pada Gambar 1. Proses dimulai dengan Pengumpulan Data (*Data Acquisition*) melalui teknik *scraping* ulasan dari *Google Play Store* untuk mendapatkan dataset primer yang representatif. Selanjutnya, dilaksanakan Pra-pemrosesan Data (*Text Preprocessing*) yang mencakup tahapan case folding, tokenisasi, penyaringan, normalisasi kata, sampai stemming untuk membersihkan data dari gangguan (*noise*) bahasa. Langkah selanjutnya adalah Ekstraksi Fitur TF-IDF untuk mengonversi teks menjadi bentuk vektor numerik berdasarkan

bobot signifikansi kata, dilanjutkan dengan Pembagian Data menjadi data pelatihan dan data pengujian secara proporsional. Inti dari penelitian ini terletak pada tahap Klasifikasi dengan memanfaatkan algoritma *Support Vector Machine* (SVM) untuk secara akurat memetakan kategori sentimen, yang kemudian diakhiri dengan tahap Evaluasi Model untuk memvalidasi kinerja sistem menggunakan metrik ilmiah[12]

2.1. Pengumpulan Data (Scraping)

Tahap pertama adalah mengumpulkan data ulasan aplikasi Alfagift dari Google Play Store. Proses ini memanfaatkan pustaka `google-play-scraper` dengan parameter `app_id` yaitu com.alfamart.alfagift. Data yang dikumpulkan difokuskan pada ulasan dalam bahasa Indonesia. Atribut yang diambil mencakup konten ulasan, penilaian bintang (1-5), dan identitas ulasan. Proses *scraping* dilakukan dalam rentang waktu 9 Febuari 2026 hingga 10 Mei 2026, menghasilkan total 5.000 ulasan. Ulasan tersebut kemudian diidentifikasi sentimennya secara awal berdasarkan *rating*: bintang 4-5 dikategorikan sebagai positif, bintang 3 sebagai netral, dan bintang 1-2 sebagai negatif.

2.2. Pra-pemrosesan Data (Preprocessing)

Tahap pra-pemrosesan data adalah langkah penting dalam alur klasifikasi teks yang bertujuan untuk mengubah teks ulasan yang belum diproses menjadi representasi bersih yang siap untuk diolah oleh algoritma pembelajaran mesin. Langkah ini meliputi berbagai proses mulai dari membersihkan teks dari unsur yang tidak relevan, penyeragaman huruf, pemecahan kata, normalisasi kata, penghapusan kata umum, hingga stemming untuk merubah kata berimbuhan menjadi bentuk dasar. Dalam penelitian ini, proses persiapan data dilakukan secara berurutan menggunakan bahasa pemrograman Python dengan pustaka NLTK dan PySastrawi.

Tabel 1. Tahapan Pipeline Preprocessing Teks Ulasan Alfagift

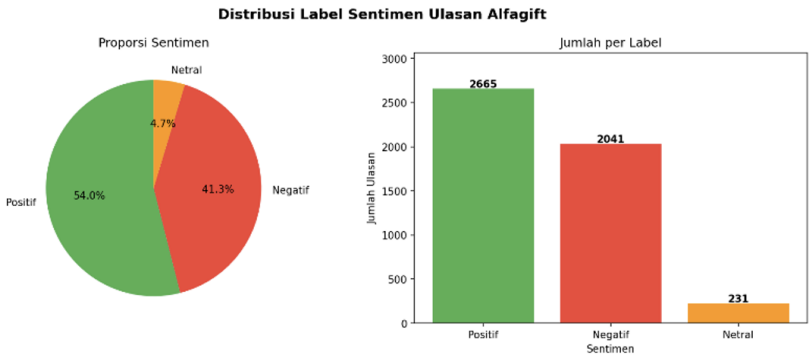
No	Tahapan	Fungsi	Contoh Input	Contoh Output
1	Case Folding	Menyeragamkan huruf menjadi huruf kecil	“Aplikasi BAGUS banget!!”	“aplikasi bagus banget!!”
2	Hapus Karakter Khusus	Menghapus simbol, angka, URL, dan emoji	“aplikasi bagus banget”	“aplikasi bagus banget”
3	Normalisasi Slang	Menyamakan kata tidak baku ke kata baku	“gak bisa checkout”	“tidak bisa checkout”
4	Tokenisasi	Memecah teks menjadi token kata	“tidak bisa checkout”	[’tidak’, ’bisa’, ’checkout’]
5	Hapus Stopword	Menghapus kata umum yang tidak bermakna	[’tidak’, ’bisa’, ’checkout’]	[’checkout’]
6	Stemming	Mengubah kata berimbuhan ke kata dasar	[’pengiriman’, ’checkout’]	[’kirim’, ’checkout’]

Sebagai mana terlihat pada Tabel 1, *case folding* merupakan langkah yang mengubah semua karakter teks menjadi huruf kecil untuk menyamakan representasi kata, sehingga kata ’BAGUS’ dan ’bagus’ diperlakukan sebagai entitas yang identik oleh model[13]. Dalam penelitian ini, case folding diterapkan dengan metode `.lower()` di Python yang digunakan pada semua teks ulasan sebelum langkah selanjutnya. Langkah selanjutnya menghilangkan elemen yang tidak berkaitan dengan analisis sentimen melalui ekspresi reguler (*regex*), termasuk URL, penyebutan, hashtag, emoji, angka, dan karakter non-alfabet. Ulasan dari pengguna di platform digital sering ditulis dengan cara yang informal, menggunakan singkatan dan bahasa gaul. Normalisasi dilakukan dengan membandingkan setiap kata dengan kamus slang yang telah disusun, lalu menggantinya dengan padanan kata baku yang tepat untuk meningkatkan akurasi model[14].

Tokenisasi merupakan langkah memecah teks menjadi unit kata yang lebih kecil yang disebut token. Langkah ini dilakukan supaya setiap kata bisa dianalisis secara terpisah oleh algoritma klasifikasi[15]. Dalam penelitian ini, tokenisasi dila-

kukan dengan menggunakan fungsi `word_tokenize` dari pustaka NLTK, sehingga kalimat 'tidak bisa checkout' akan terpisah menjadi token ['tidak', 'bisa', 'checkout']. Selanjutnya, hapus stopword adalah langkah menghapus kata-kata umum yang sering muncul tetapi tidak memberikan kontribusi signifikan terhadap makna teks, seperti 'yang', 'dan', 'di', serta 'ke'. Lalu yang terkahir, stemming adalah prosedur mengubah kata berprefiks atau berinfiks menjadi bentuk dasar supaya berbagai variasi kata yang memiliki arti serupa dapat ditampilkan secara konsisten, contohnya 'pengiriman' dan 'dikirim' jadi 'kirim'[16].

Setelah menyelesaikan semua tahap preprocessing, dari total 5.000 ulasan yang dikumpulkan, didapatkan 4.937 data valid yang siap digunakan dalam proses klasifikasi. Sebanyak 63 ulasan dihapus karena menghasilkan konten kosong setelah pemrosesan awal, misalnya ulasan yang hanya terdiri dari emoji atau symbol. Perlu dicatat bahwa rasio ketidakseimbangan antar kelas pada dataset ini tergolong ekstrem, yaitu sekitar 11,5:1 antara kelas Positif dan kelas Netral. Penelitian ini secara sengaja membatasi penanganan ketidakseimbangan kelas hanya pada parameter `class_weight='balanced'` tanpa menerapkan teknik resampling seperti SMOTE[11], dengan tujuan untuk mengevaluasi secara objektif sejauh mana pendekatan pembobotan kelas saja dapat menangani kondisi ekstrem semacam ini, sebagaimana akan dibahas lebih lanjut pada bagian Hasil dan Pembahasan.



Gambar 2. Distribusi Label Sentimen Ulasan Alfagift Setelah Preprocessing

Pada Gambar 2, distribusi data menunjukkan bahwa kelas Positif mendominasi dengan 2.665 ulasan (54,0%), kelas Negatif sebanyak 2.041 ulasan (41,3%), dan kelas Netral sebanyak 231 ulasan (4,7%).

2.3. Ekstraksi Fitur TF-IDF

Setelah tahap *preprocessing*, teks diubah menjadi vektor angka dengan menerapkan metode Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF menilai seberapa penting suatu kata dalam dokumen dibandingkan dengan keseluruhan korpus kata yang muncul sering dalam satu dokumen tetapi jarang di dokumen lain akan memiliki bobot yang tinggi[17].

Term Frequency mengukur seberapa sering sebuah kata `t` muncul dalam dokumen `d`. Semakin sering kata muncul dalam satu ulasan, semakin tinggi nilai TF-nya. Pada penelitian ini, diterapkan *sublinear TF scaling* dengan transformasi logaritmik agar kata yang muncul 100 kali tidak dianggap 100× lebih penting dari kata yang muncul 1 kali, sesuai dengan implementasi pada kode `processing.py` dan `klasifikasi_svm.py` (parameter `sublinear_tf=True`). Rumusnya adalah:

$$TF(t,d) = 1 + \log(f_{t,d}) \text{ jika } f_{t,d} > 0$$

di mana $f_{t,d}$ adalah sefrekuensi kemunculan kata *t* dalam dokumen *d*, selanjutnya *Inverse Document Frequency* mengukur seberapa unik atau langka suatu kata di seluruh korpus dokumen. Kata yang muncul di hampir semua dokumen (misalnya kata "belanja" yang ada di mayoritas ulasan) mendapat bobot IDF

rendah, sementara kata yang hanya muncul di sebagian kecil dokumen mendapat bobot IDF tinggi. Rumusnya adalah:

$$IDF(t, D) = \log\left(\frac{N}{df_t}\right) + 1$$

di mana N adalah jumlah total dokumen dalam korpus dan df_t adalah jumlah dokumen yang mengandung kata *t*. Nilai TF-IDF akhir untuk sebuah kata diperoleh dengan mengalihkan kedua komponen:

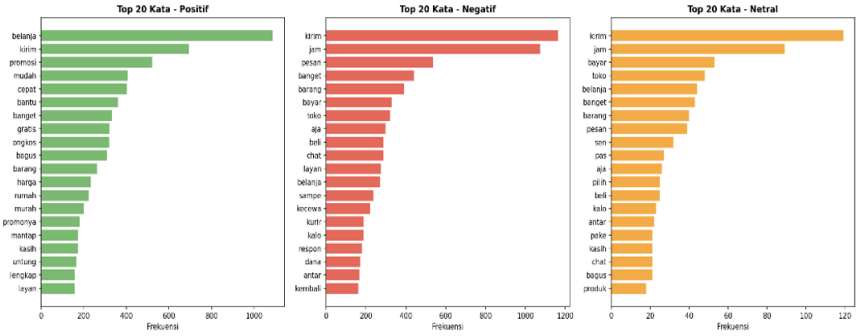
$$TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D)$$

Nilai TF-IDF yang tinggi menunjukkan bahwa kata tersebut muncul dengan frekuensi tinggi dalam satu dokumen tertentu, tetapi jarang dijumpai di dokumentasi lain ini berarti kata tersebut memiliki sifat informatif dan diskriminatif untuk dokumen tersebut[18].

Tabel 2. Konfigurasi Parameter TF-IDF

Parameter	Nilai	Keterangan
`max_features`	10.000	Membatasi 10.000 fitur terpenting
`ngram_range`	(1,2)	Unigram dan bigram
`min_df`	2	Kata minimal muncul di 2 dokumen
`max_df`	0,95	Abaikan kata di > 95% dokumen
`sublinear_tf`	`true`	Log-scaling pada nilai TF

Pada Tabel 2, bigram pada parameter `ngram_range=(1,2)` memungkinkan model menangkap konteks frasa seperti “tidak bisa” atau “sangat membantu” yang tidak dapat ditangkap jika hanya menggunakan unigram[19]. Hasil akhir ekstraksi menghasilkan matriks sparse berukuran 4.937×10.000 yang menjadi input langsung ke model SVM.



Gambar 3. Top 20 Kata per Kelas Sentimen Berdasarkan Frekuensi setelah TF-IDF

Berdasarkan Gambar 3, kata ”belanja”, ”kirim”, dan ”promosi” mendominasi kelas Positif; kata ”kirim”, ”jam”, dan ”error” mendominasi kelas Negatif; sedangkan pola kata di kelas Netral lebih beragam dan tumpang tindih dengan kedua kelas yang lain.

2.4. Pembagian Data (Data Splitting)

Setelah fitur TF-IDF diekstraksi dan menghasilkan matriks angka, dataset dibagi menjadi dua subset yang terpisah, yaitu data latih (*training set*) dan data uji (*testing set*). Pembagian data adalah langkah penting untuk menilai kemampuan generalisasi model pada data yang belum dikenali sebelumnya, sehingga hasil evaluasi yang diperoleh dapat merepresentasikan kinerja model secara objektif[20].

Menggunakan rasio pembagian 80:20, di mana 80% dari data dipakai untuk melatih model dan 20% sisanya untuk pengujian. Dari keseluruhan 4.937

data yang valid, didapatkan 3.949 data latih dan 988 data uji. Pengalokasian ini dilakukan dengan memanfaatkan fungsi `train_test_split` dari pustaka `scikit-learn` dengan parameter `stratify=y`, yang menjaga agar proporsi setiap kelas sentimen (Positif, Negatif, Netral) tetap seimbang antara data pelatihan dan data pengujian.

Tabel 3. Distribusi Data Latih dan Data Uji

Kelas Sentimen	Total	Data Latih (80%)	Data Uji (20%)
Positif	2.665	2.132	533
Negatif	2.041	1.632	409
Netral	231	185	46

Penerapan parameter `stratify` sangat penting dalam mengingat adanya distribusi kelas yang tidak seimbang, di mana kelas Netral hanya mencakup 4,7% dari total data. Tanpa adanya stratifikasi, pembagian acak dapat berisiko menghasilkan subset uji yang tidak mencerminkan proporsi kelas minoritas, sehingga dapat mempengaruhi hasil evaluasi[21].

2.5. Klasifikasi dengan SVM

Support Vector Machine (SVM) merupakan algoritma pembelajaran terawasi yang berfungsi untuk mencari hyperplane terbaik yang memisahkan kelas-kelas data dengan margin maksimum dalam ruang fitur berdimensi tinggi. Bentuk umum hyperplane pemisah SVM dapat dituliskan sebagai:

$$W \cdot X + b = 0$$

di mana W merupakan vektor bobot, X adalah vektor fitur dari data masukan, dan b adalah nilai bias. Studi ini memanfaatkan kernel *Radial Basis Function* (RBF) karena pada beberapa penelitian terdahulu yang mengkaji ulasan aplikasi di *Google Play Store*, kernel RBF memberikan akurasi yang lebih tinggi disbanding kernel linear (86, 1% berbanding 83,1%)[22]. Meskipun performa kernel dapat bervariasi tergantung karakteristik dataset, pemilihan RBF pada penelitian ini didasarkan pada bukti empiris tersebut serta kemampuannya menangani hubungan non-linear pada data teks berdimensi tinggi. Parameter `class_weight='balanced'` diterapkan untuk mengatasi ketidakseimbangan distribusi kelas, terutama kelas Netral yang hanya berkontribusi 4, 7% dari total data, agar model tidak memihak pada kelas mayoritas. Seluruh langkah-langkah diterapkan dengan menggunakan `sklearn.pipeline.Pipeline` yang menggabungkan TF-IDF dan SVM dalam satu alur terintegrasi untuk mencegah kebocoran data saat evaluasi [23].

2.6. Evaluasi dan Hasil

Evaluasi model dilakukan menggunakan dua pendekatan. Pertama, evaluasi pada data uji menggunakan *confusion matrix* yang membandingkan hasil prediksi model dengan label aktual untuk menghasilkan empat komponen yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dari komponen tersebut dihitung metrik akurasi, presisi, recall, dan F1-score dengan rumus berikut:

$$\begin{aligned} \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \\ \text{Presisi} &= \frac{TP}{TP + FP} \\ \text{Recall} &= \frac{TP}{TP + FN} \\ \text{F1-Score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \end{aligned}$$

Kedua, dilaksanakan *10-fold Stratified Cross Validation* di mana data dibagi menjadi 10 segmen secara bergantian sebagai data tes, sementara 9 segmen lainnya dipakai sebagai data pelatihan." Pendekatan stratifikasi diterapkan agar proporsi setiap kelas di setiap *fold* tetap mencerminkan distribusi asli dataset, sehingga evaluasi menjadi lebih adil terutama untuk kelas minoritas. Output akhir berupa rata-rata akurasi serta deviasi standar dari 10 lipatan yang mencerminkan kestabilan dan kemampuan generalisasi model.

3. Hasil

Bagian ini menyajikan hasil dari setiap langkah penelitian mulai dari pengumpulan data, pemrosesan awal, pelatihan model, hingga evaluasi akhir dengan menggunakan data nyata yang didapat dari eksperimen. Semua hasil yang ditampilkan adalah output langsung dari penerapan kode yang dijalankan pada dataset ulasan aplikasi Alfagift yang diambil dari *Google Play Store*.

3.1. Hasil Pelatihan Model

Model SVM-TF-IDF dilatih dengan 3.949 data latihan (80%) dari keseluruhan 4.937 data yang valid. Proses pelatihan menghasilkan bobot kelas yang dihitung secara otomatis berdasarkan distribusi data, yaitu Negatif: 0,8066, Netral: 7,1153, dan Positif: 0,6174. Bobot kelas Netral yang lebih tinggi menunjukkan usaha model untuk memberi perhatian lebih pada kelas minoritas selama pelatihan.

3.2. Hasil Evaluasi pada Data Uji

Akurasi sebesar 87,15% dihasilkan dari evaluasi model yang melibatkan 988 data uji. Tabel 4 menunjukkan laporan klasifikasi lengkap untuk setiap kelas sentimen.

Tabel 4. Distribusi Data Latih dan Data Uji

Kelas	Presisi	Recall	F1-Score	Support
Negatif	0,7959	0,9535	0,8676	409
Netral	0,0000	0,0000	0,0000	46
Positif	0,9534	0,8837	0,9172	533
Macro avg	0,5831	0,6124	0,5950	988
Weighted avg	0,8438	0,8715	0,8540	988

Tabel 4 menunjukkan bahwa kelas Positif memiliki kinerja terbaik dengan F1-score 0,9172 dan presisi tertinggi sebesar 0,9534. Kelas Negatif memiliki recall tertinggi sebesar 0,9535, yang menunjukkan bahwa model mendeteksi ulasan negatif dengan sangat baik, meskipun presisinya lebih rendah (0,7959). Akibat jumlah sampel yang sangat kecil (46 data uji, atau hanya 4,7% dari total), Kelas Netral menghasilkan nilai 0,0000 pada seluruh metrik. Meskipun telah diterapkan `class_weight= 'balanced'`, model tidak dapat mempelajari pola sentimen netral dengan baik.

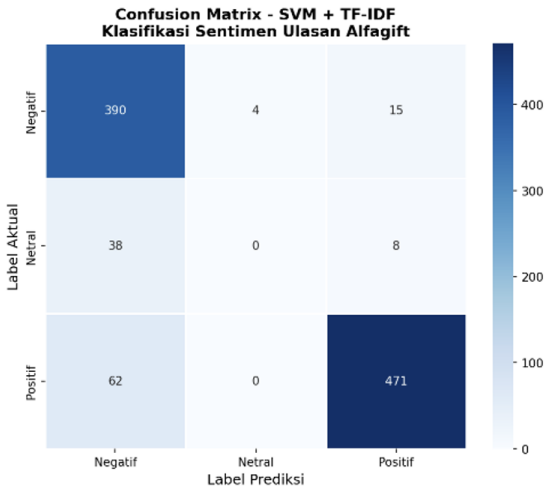
3.3. Hasil Confusion Matrix

Untuk menemukan pola kesalahan klasifikasi antar kelas, analisis lebih lanjut dilakukan dengan menggunakan *confusion matrix*, yang ditunjukkan pada Gambar 4.

Menurut Gambar 4, dari 409 ulasan Negatif, model mengklasifikasikan 390 dengan benar dan 19 salah diprediksi sebagai Positif. Dari 533 ulasan Positif, 471 diklasifikasikan dengan benar dan 62 diprediksi sebagai Negatif. Dari 46 ulasan Netral, seluruhnya salah diklasifikasikan 38 diprediksi sebagai Negatif dan 8 sebagai Positif. Pola kesalahan ini menunjukkan bahwa model paling sulit membedakan batas keputusan antara kelas Negatif dan Netral.

3.4. Hasil Cross Validation

Evaluasi menggunakan *10-fold Stratified Cross Validation* menghasilkan akurasi per fold sebagai berikut: 87,65%, 85,83%, 86,64%, 88,46%, 87,04%, 88,87%,



Gambar 4. Confusion Matrix Model SVM-TF-IDF

85,43%, 87,22%, 88,03%, dan 87,63%. Tabel 5 merangkum hasil keseluruhan *cross validation*.

Tabel 5. Distribusi Data Latih dan Data Uji

Metrik	Nilai
Akurasi	87,15%
Rata-rata CV (10-Fold)	87,28%
Standar Deviasi CV	1,0368%
Akurasi Tertinggi (Fold 6)	88,87%
Akurasi Terendah (Fold 7)	85,43%

Rata-rata akurasi validasi silang sebesar 87,28% dengan standar deviasi 1,0368% mengindikasikan bahwa model menunjukkan performa yang stabil di seluruh lipatan. Perbedaan kecil antara akurasi data uji (87,15%) dan rata-rata *cross validation* (87,28%) menandakan bahwa model tidak *overfitting* dan memiliki kemampuan generalisasi yang baik terhadap data baru. Dibanding dengan penelitian sejenis pada domain ulasan *e-commerce* Indonesia dengan struktur tiga kelas yang serupa, akurasi 87,15% pada penelitian ini tergolong kompetitif sedikit lebih tinggi disbanding studi pada ulasan Shopee yang mencapai akurasi terbaik 85,11% menggunakan Logistic Regression[24]. Namun, perbandingan akurasi absolut semata berpotensi menyesatkan apabila tidak memperhitungkan performa perkelas, sebagaimana ditunjukkan pada penelitian ulasan marketplace Indonesia lain yang melaporkan SVM mencapai akurasi tertinggi 94,54% namun dengan recall kelas minoritas yang sangat rendah (5,71%) akibat bias terhadap kelas mayoritas[25]. Fenomena serupa turut teramati pada penelitian ini, di mana akurasi keseluruhan yang tinggi (87,15%) tidak mencerminkan kegagalan total model dalam mengklasifikasikan kelas Netral.

4. Pembahasan

Penelitian ini menghasilkan klasifikasi sentimen ulasan aplikasi Alfagift dengan memanfaatkan kombinasi SVM dan TF-IDF, mencapai akurasi data uji sebesar 87,15% dan rata-rata 10-fold cross validation sebesar 87,28% (dengan standar deviasi 1,0368%). Temuan ini sejalan dengan beberapa penelitian terdahulu yang menerapkan kombinasi SVM dan TF-IDF pada domain ulasan aplikasi berbahasa Indonesia. Penelitian pada ulasan aplikasi CapCut melaporkan SVM mencapai akurasi 90,00%[6], penelitian pada ulasan terkait pertamina mencatat SVM mencapai akurasi 91% setelah optimasi[7], dan penelitian pada ulasan KAI Access menggunakan kernel RBF memperoleh akurasi 86,1%[22]. Akurasi 87,15%

yang diperoleh pada penelitian ini berada pada rentang yang sebanding dengan ketiga penelitian tersebut, mengonfirmasi bahwa kombinasi SVM dan TF-IDF secara konsisten memberikan performa kompetitif dalam klasifikasi sentiment teks berbahasa Indonesia. Perbedaan yang sangat kecil antara akurasi data uji dan rata-rata cross validation (0,13%) menunjukkan bahwa model tidak mengalami overfitting dan memiliki kemampuan generalisasi yang baik terhadap data yang baru.

Kelas Positif mencatat kinerja tertinggi dengan F1-score 0,9172 dan presisi maksimum 0,9534, sementara kelas Negatif mencapai recall tertinggi 0,9535, yang menunjukkan bahwa model sangat peka dalam mengidentifikasi ulasan negatif. Ini berkaitan dengan ciri-ciri kata-kata utama dalam setiap kelas-kelas Positif didominasi oleh kata "belanja", "promosi", dan "mudah", sementara kelas Negatif didominasi oleh kata "kirim", "jam", dan "error" yang memiliki nilai TF-IDF tinggi dan bersifat memisahkan. Pola ini menunjukkan bahwa konfigurasi TF-IDF dengan `ngram_range=(1,2)` efektif dalam menangkap frasa-frasa kunci yang menandai sentimen pada setiap kelas.

Tantangan utama dalam penelitian ini adalah klasifikasi kelas Netral yang menghasilkan F1-score sebesar 0,0000 pada semua metrik. Dari 46 data uji kategori Netral, 38 diprediksi sebagai Negatif dan 8 sebagai Positif. Situasi ini disebabkan oleh ketidakseimbangan distribusi data yang parah, di mana kategori Netral hanya terdiri dari 4,7% dari keseluruhan 4.937 data. Walaupun parameter `class_weight='balanced'` telah diterapkan untuk memberikan bobot lebih pada kelas minoritas, jumlah sampel yang sangat terbatas (231 data) tidak mencukupi bagi model untuk memahami pola linguistik khas sentimen netral dengan memadai. Ketidakseimbangan kelas serupa adalah tantangan umum dalam analisis sentimen yang didasarkan pada ulasan pengguna di platform digital.

Kelas Netral sulit dibedakan oleh model karena secara linguistik ulasan netral tidak memiliki tanda yang jelas. Ulasan netral cenderung memiliki makna yang kabur, sering kali mencakup ungkapan yang tidak tampak mendukung baik sisi positif maupun negatif secara jelas, seperti 'aplikasi biasa' atau 'cukup baik tetapi masih bisa ditingkatkan'. Tidak seperti kelas Positif yang memiliki istilah seperti 'baik', 'cepat', dan 'senang', serta kelas Negatif yang ditandai dengan kata-kata seperti 'kesalahan', 'tidak berhasil', dan 'kecewa' kelas Netral tidak menunjukkan pola kata yang jelas dan teratur. Keadaan ini menjadikan representasi TF-IDF untuk kelas Netral kurang mampu membedakan dengan jelas, sehingga model tidak dapat memisahkan kelas Netral dari kelas yang lain.

Masalah pada kelas Netral semakin diperburuk oleh tumpang tindih kata yang signifikan antara ketiga kelas sentimen. Merujuk pada Gambar 3, terdapat kata-kata yang muncul secara dominan di lebih dari satu kategori dengan konteks yang berbeda. Contohnya, kata 'kirim' terdapat di kelas Positif dalam konteks 'pengiriman segera' dan di kelas Negatif dalam konteks 'pengiriman terlambat atau cacat'. Demikian juga istilah 'belanja' dan 'promo' yang muncul di kategori Positif maupun Netral. Tumpang tindih ini mengakibatkan model mengalami kesulitan dalam mencari hyperplane pemisah yang jelas, terutama antara kelas Netral dengan kelas Positif dan Negatif. Ini terlihat dari *confusion matrix* pada Gambar 4 di mana semua 46 ulasan Netral pada data uji terklasifikasi salah 38 diprediksi sebagai Negatif dan 8 diprediksi sebagai Positif.

Keterbatasan lain yang perlu dicermati adalah cara pelabelan data yang diotomatisasi berdasarkan penilaian bintang. Pendekatan ini menganggap bahwa penilaian 1–2 selalu menunjukkan sentimen negatif, penilaian 3 selalu netral, dan penilaian 4–5 selalu positif. Akan tetapi, dalam praktiknya, penilaian bintang tidak selalu mencerminkan konten dari ulasan yang ditulis oleh pengguna. Seorang pengguna bisa memberi rating 3 tetapi menuliskan ulasan yang cenderung positif seperti 'aplikasi ini cukup baik dan mudah dipakai', atau sebaliknya memberikan rating 4 namun mengeluhkan fitur tertentu yang tidak berfungsi. Fenomena ini dikenal sebagai ketidaksesuaian antara penilaian dan ulasan yang sering terjadi

pada platform ulasan digital. Inkonsistensi ini dapat menghasilkan label yang tidak tepat, terutama pada kelas Netral yang hanya dipisahkan oleh satu nilai rating (rating 3), sehingga dapat mengurangi kualitas data latih dan mempengaruhi performa model secara keseluruhan.

Selain itu, validitas label pada penelitian ini perlu diperhatikan karena pelabelan dilakukan secara otomatis berdasarkan rating bintang tanpa verifikasi manual. Fenomena ketidaksesuaian antara rating dan ulasan yang sering terjadi di platform ulasan digital dapat mengakibatkan label tidak selalu mencerminkan sentimen dari teks yang sebenarnya contohnya, pengguna yang memberi rating 3 tetapi menulis ulasan yang cenderung positif, atau sebaliknya. Sebaiknya, validasi label dilakukan oleh manusia yang menganotasi atau dengan pendekatan lexicon-based untuk memastikan kesesuaian antara label dan konten teks ulasan.

Temuan dari penelitian ini mengindikasikan bahwa metode SVM yang menggunakan TF-IDF cocok untuk menganalisis sentimen pada ulasan aplikasi ritel mobile berbahasa Indonesia, terutama pada kelas Positif dan Negatif yang dominan dalam dataset. Di masa mendatang, riset lanjutan dapat mempertimbangkan penggunaan teknik augmentasi data atau oversampling untuk menangani masalah ketidakseimbangan kelas Netral, serta mengeksplorasi metode representasi fitur yang lebih kontekstual seperti word embedding agar dapat meningkatkan kemampuan model dalam mengenali nuansa sentimen yang tidak jelas.

Studi ini memiliki sejumlah keterbatasan yang harus diakui sebagai bahan pertimbangan dan kemajuan di masa depan. Pertama, pelabelan data dilakukan secara otomatis tanpa pemeriksaan manual oleh anotator, sehingga kualitas label sangat tergantung pada konsistensi perilaku pengguna dalam memberikan penilaian. Kedua, ketidakseimbangan distribusi kelas yang sangat ekstrem terutama kelas Netral yang hanya mencakup 4,7% dari keseluruhan data dengan rasio 11,5:1 dibandingkan dengan kelas Positif tidak diatasi dengan menggunakan teknik oversampling seperti SMOTE, yang bisa meningkatkan performa model pada kelas minoritas secara signifikan. Ketiga, kamus normalisasi slang yang diterapkan bersifat statis dengan 60 entri tetap, sehingga tidak dapat mengakomodasi kata slang baru yang terus bermunculan seiring dengan dinamika bahasa pengguna digital. Keempat, model SVM yang menggunakan kernel RBF berfungsi sebagai black-box, sehingga tidak mampu menjelaskan secara langsung fitur-fitur mana yang paling berperan dalam keputusan klasifikasi. Kelima, data yang dikumpulkan hanya berasal dari satu platform, yaitu Google Play Store, sehingga hasil penelitian mungkin tidak dapat diterapkan pada platform lain seperti App Store atau media social.

5. Kesimpulan

Penelitian ini sukses menciptakan model klasifikasi sentimen untuk ulasan aplikasi Alfagift dengan memanfaatkan SVM yang memakai kernel RBF dan TF-IDF. Dari 5.000 ulasan yang diperoleh, sebanyak 4.937 data yang valid diproses melalui tahapan pra-pemrosesan, menghasilkan akurasi data uji 87,15% dan rata-rata 10-fold Cross Validation 87,28% dengan deviasi standar 1,0368%, mencerminkan model yang stabil dan tidak mengalami overfitting. Kelas Positif dan Negatif telah berhasil dikategorikan dengan baik (F1-score 0,9172 dan 0,8676), namun kelas Netral tidak dapat diklasifikasikan (F1-score 0,0000) karena adanya ketidakseimbangan distribusi data yang sangat ekstrim, di mana kelas Netral hanya mencakup 4,7% dari keseluruhan dataset. Untuk penelitian selanjutnya, disarankan menggunakan teknik oversampling seperti SMOTE dan menjelajahi metode representasi fitur yang lebih kontekstual seperti IndoBERT guna meningkatkan kemampuan model dalam mendeteksi nuansa sentimen yang tidak jelas.

Pernyataan Penggunaan *Artificial Intelligence*. Penulis menggunakan **Claude** (**Anthropic**) untuk membantu penyusunan, penyuntingan bahasa, dan revisi naskah berdasarkan masukan reviewer. Seluruh keluaran yang dihasilkan oleh AI

telah diperiksa dan diverifikasi oleh penulis. Penulis bertanggung jawab sepenuhnya atas keakuratan, orisinalitas, dan integritas isi artikel.

Pernyataan Kontribusi Penulis. **NA** (Nazwa Asyifa) berkontribusi dalam konseptualisasi, metodologi, pengumpulan data, analisis penelitian, dan penyusunan naskah. **S** (Susanto) berkontribusi dalam peninjauan, penyuntingan, dan pembimbingan naskah. Seluruh penulis telah membaca, meninjau, dan menyetujui versi akhir artikel serta bertanggung jawab atas seluruh aspek penelitian dan publikasi.

Pernyataan Konflik Kepentingan. Penulis menyatakan bahwa tidak terdapat konflik kepentingan yang berkaitan dengan penelitian, kepengarangan, dan publikasi artikel ini.

Pernyataan Pendanaan. Penelitian ini tidak menerima pendanaan khusus dari lembaga pemerintah, komersial, maupun organisasi nirlaba.

Daftar Pustaka

[1] W. Ningrat and S. Syahriani, "Analisis tingkat kepuasan pengguna aplikasi alfagift menggunakan metode end user computing satisfaction dan net promoter score," *Jurnal Komputer Antartika*, vol. 3, no. 3, pp. 87–98, 2025. https://doi.org/10.70052/jka.v3i3.1036.

[2] A. Lukito, F. Nugraha, and N. Latifah, "Analisis sentimen ulasan di playstore terhadap aplikasi marketplace menggunakan naive bayes dan logistic regression," *Jurnal UNITEK*, vol. 18, no. 2, pp. 322–334, 2025. https://ejurnal.sttdumai.ac.id/index.php/unitek/article/view/1625.

[3] R. Afif, A. Supriyanto, R. Fitri Damaryanti, and W. Prasetya Adi, "Analisis sentimen aplikasi adiraku di google play store menggunakan metode support vectore machine," *Jurnal FASILKOM (teknologi inFormASi dan ILmu KOMputer)*, vol. 15, no. 1, pp. 163–171, 2025. https://doi.org/10.37859/jf.v15i1.8510.

[4] E. Indrayuni, A. Nurhadi, and D. A. Kristiyanti, "Implementasi algoritma naive bayes, support vector machine, dan k-nearest neighbors untuk analisa sentimen aplikasi halodoc," *Faktor Exacta*, vol. 14, no. 2, p. 64, 2021. http://dx.doi.org/10.30998/faktorexacta.v14i2.9697.

[5] A. B. Muzayyanah, R. E. Pawening, and Z. Arifin, "Analisis sentimen pada ulasan aplikasi ehadrah di google playstore menggunakan support vector machine," *IDEALIS: InDonEsiA journaL Information System*, vol. 7, no. 2, pp. 258–266, 2024. https://doi.org/10.36080/idealis.v7i2.3250.

[6] I. Irawan, Wardianto, M. H. Wathan, and M. B. Prayogi, "Studi perbandingan: Algoritma random forest, naive bayes dan support vector machine dalam analisis sentimen pada aplikasi capcut di google play store," *Jurnal Pengembangan Sistem Informasi Dan Informatika*, vol. 5, no. 4, pp. 189–201, 2024. https://doi.org/10.47747/jpsii.v5i4.1959.

[7] H. Vazli and R. R. Suryono, "Komparasi model svm, naive bayes, dan random forest dalam analisis sentimen terhadap percakapan publik tentang pertamina," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 11, no. 1, pp. 92–105, 2026. https://doi.org/10.29100/jipi.v11i1.7699.

[8] U. Brawijaya, M. R. Ghiffari, A. Pinandito, and R. S. Sianturi, "Studi perbandingan metode ekstraksi fitur teks tf-idf, word2vec, fasttext dan pengaruh klasifikasi multi-aspek terhadap akurasi svm," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 1, 2026. https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/15969.

[9] T. Nabarian, S. Nurhalizah, and S. El Farisi, "Analisis perbandingan kinerja indobert dan tf-idf dalam mengklasifikasikan sentimen edom menggunakan algoritma k-nearest neighbor," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 6, no. 2, pp. 640–650, 2026. https://www.journal.irpi.or.id/index.php/malcom/article/view/2614.

[10] T. P. R. Sanjaya, A. Fauzi, and A. F. N. Masruriyah, "Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naive bayes dan support vector machine," *INFOTECH: Jurnal Informatika & Teknologi*, vol. 4, no. 1, pp. 16–26, 2023. https://doi.org/10.37373/infotech.v4i1.422.

[11] N. D. Ramadhanti and R. N. Handayani, "Analisis sentimen ulasan pengguna aplikasi shopee di goggle play store menggunakan support vector machine dengan teknik smote," *Jurnal Informatika Polinema*, vol. 12, no. 2, pp. 313–322, 2026. https://doi.org/10.33795/jip.v12i2.9073.

[12] N. P. Husain, S. Sukirman, and S. Sajilah, "Analisis sentimen ulasan pengguna tiktok pada google play store berbasis tf-idf dan support vector machine," *Journal of System and Computer Engineering*, vol. 5, no. 1, pp. 313–322, 2024. https://doi.org/10.61628/jsce.v5i1.1105.

[13] M. M. Rachmatullah, B. Irawan, A. Faqih, D. Pratama, and D. A. Kurnia, "Analisis komparatif multinomial naive bayes dan logistic regression untuk klasifikasi sentimen ulasan pengguna aplikasi tix id," *JSI (Jurnal Sistem Informasi) Universitas Suryadarma*, vol. 13, no. 1, pp. 62–70, 2026. https://doi.org/10.35968/jsi.v13i1.1723.

[14] I. F. Rahman, A. N. Hasanah, and N. Heryana, "Analisis sentimen ulasan pengguna aplikasi samsat digital nasional (signal) dengan menggunakan metode naive bayes classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, 2024. https://doi.org/10.23960/jitet.v12i2.4073.

[15] A. A. Qureshi, M. Ahmad, S. Ullah, M. N. Yasir, F. Rustam, and I. Ashraf, "Performance evaluation of machine learning models on large dataset of android applications reviews," *Multimedia Tools and Applications*, vol. 82, no. 24, pp. 37197–37219, 2023. https://doi.org/10.1007/s11042-023-14713-6.

[16] A. Setiawan, F. Firman, N. Hasan, and R. Artikel, "Analisis sentimen tanggapan pengguna aplikasi bale by btn menggunakan metode support vector machine (svm)," *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 4, no. 4, pp. 315–326, 2025. https://doi.org/10.55123/storage.v4i4.6469.

[17] R. Wati, S. Ernawati, and H. Rachmi, "Pembobotan tf-idf menggunakan naive bayes pada sentimen masyarakat mengenai isu kenaikan bipih," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 13, no. 1, pp. 84–93, 2023. https://doi.org/10.34010/jamika.v13i1.9424.

[18] V. W. D. Thomas and F. Rumaisa, "Analisis sentimen ulasan hotel bahasa indonesia menggunakan support vector machine dan tf-idf," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 3, p. 1767, 2022. https://doi.org/10.30865/mib.v6i3.4218.

[19] R. Rahmadani, A. Rahim, and R. Rudiman, "Analisis sentimen ulasan ojol the game di google play store menggunakan algoritma naive bayes dan model ekstraksi fitur tf-idf untuk meningkatkan kualitas game," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, 2024. https://doi.org/10.23960/jitet.v12i3.4988.

[20] Y. A. Prasetyo, E. Utami, and A. Yaqin, "Pengaruh komposisi split data terhadap performa akurasi analisis sentimen algoritma naive bayes dan svm," *Journal of Electrical Engineering and Computer (JEECOM)*, vol. 6, no. 2, 2024. https://doi.org/10.33650/jeecom.v6i2.9188.

[21] A. Mudya Yolanda and R. Tri Mulya, "Implementasi metode support vector machine untuk analisis sentimen pada ulasan aplikasi sayurbox di google play store," *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, vol. 6, no. 2, pp. 76–83, 2024. https://jurnalvariansi.unm.ac.id/index.php/variansi/article/view/258.

[22] E. R. M. Sholihah, I. G. S. M. Diyasa, and E. Y. Puspaningrum, "Perbandingan kinerja kernel linear dan rbf support vector machine untuk analisis sentimen ulasan pengguna kai access pada google play store," *JATI*, vol. 8, no. 1, 2024. https://doi.org/10.36040/jati.v8i1.8800.

[23] S. Widodo, H. Brawijaya, and S. Samudi, "Stratified k-fold cross validation optimization on machine learning for prediction," *Sinkron*, vol. 7, no. 4, pp. 2407–2414, 2022. https://doi.org/10.33395/sinkron.v7i4.11792.

[24] S. F. Kadir and A. Fairuzabadi, "Analisis sentimen ulasan shopee di google play dengan tf-idf dan logistic regression," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 2, pp. 7940–7945, 2025. https://doi.org/10.31004/riggs.v4i2.2850.

[25] L. Ardin Gulo and A. Wibowo, "Analisis sentimen ulasan produk marketplace indonesia menggunakan naive bayes dan svm dengan label berdasarkan rating," *Jurnal Algoritma*, vol. 23, no. 1, 2026. https://doi.org/10.33364/algoritma/v.23-1.3206.