ANALISIS SENTIMEN PENGGUNA APLIKASI CAPCUT DARI PERSPEKTIF INTERAKSI MANUSIA DAN KOMPUTER MENGGUNAKAN SUPPORT VECTOR MACHINE

Fadillah Harya Jatmiko, Ilham Fannani
Informatika, Universitas Tiga Serangkai
Jl. K.H Samanhudi No.84-86, Purwosari, Kec. Laweyan, Kota Surakarta, Jawa Tengah, Indonesia
fadillahmiko12@gmail.com

ABSTRAK

Perkembangan aplikasi digital mendorong meningkatnya jumlah ulasan pengguna sebagai sumber informasi terkait pengalaman penggunaan, termasuk pada aplikasi *CapCut*. Namun, data ulasan yang berbentuk teks tidak terstruktur menyulitkan proses analisis secara manual. Permasalahan ini menyebabkan sulitnya memahami persepsi pengguna secara cepat dan akurat. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen pengguna aplikasi *CapCut* dari perspektif *Human–Computer Interaction* menggunakan algoritma *Support Vector Machine (SVM)*. Metode yang digunakan meliputi pengumpulan data ulasan dari *Google Play Store* melalui teknik *scraping*, kemudian dilakukan tahap *text preprocessing* yang terdiri dari *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Selanjutnya, data direpresentasikan menggunakan pembobotan *TF-IDF* dan diklasifikasikan ke dalam sentimen positif, netral, dan negatif menggunakan *SVM*. Evaluasi model dilakukan menggunakan *confusion matrix* dengan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa model *SVM* mampu mengklasifikasikan sentimen dengan akurasi terbaik sebesar 71,55%. Selain itu, mayoritas ulasan pengguna didominasi oleh sentimen negatif, yang mengindikasikan adanya permasalahan pada kualitas pengalaman pengguna aplikasi *CapCut*.

Kata kunci : *Analisis Sentimen, Support Vector Machine, CapCut, TF-IDF*

1. PENDAHULUAN

Perkembangan teknologi informasi telah membawa perubahan signifikan dalam pola interaksi masyarakat dengan berbagai layanan digital, termasuk aplikasi berbasis media sosial dan hiburan. Salah satu aplikasi yang banyak digunakan di Indonesia adalah *CapCut*, sebuah aplikasi *editing video* yang populer karena kemudahan fitur dan integrasinya dengan platform seperti *TikTok* dan *Instagram*. Seiring bertambahnya jumlah pengguna, semakin banyak ulasan dan opini yang diunggah melalui platform seperti *Google Play Store*. Ulasan tersebut menjadi sumber informasi penting untuk memahami tingkat kepuasan dan persepsi pengguna terhadap kinerja aplikasi, namun data yang bersifat tidak terstruktur menyebabkan proses analisis secara manual menjadi sulit dilakukan[1].

Analisis sentimen digunakan sebagai pendekatan untuk mengekstraksi dan mengklasifikasikan opini pengguna ke dalam kategori positif, negatif, atau netral. Pendekatan ini membantu peneliti maupun pengembang aplikasi dalam memahami pola emosi dan tanggapan pengguna berdasarkan data tekstual dari ulasan pengguna. Melalui teknik ini, opini publik dapat diubah menjadi data terstruktur yang dapat digunakan untuk pengambilan keputusan dan perbaikan fitur aplikasi. Selain itu, analisis sentimen juga telah diterapkan secara luas pada berbagai aplikasi digital untuk mengidentifikasi persepsi dan pengalaman pengguna secara otomatis[2].

Dalam konteks perkembangan teknologi data, analisis sentimen menjadi bagian penting dari data mining yang mampu mengekstraksi informasi tersembunyi dari kumpulan data besar, sehingga mendukung pengambilan keputusan berbasis data[3].

Penelitian terdahulu menunjukkan bahwa algoritma *Support Vector Machine (SVM)* merupakan metode yang efektif dalam melakukan klasifikasi sentimen karena mampu menemukan *hyperplane* optimal untuk memisahkan data berdasarkan polaritas sentimen. Penelitian analisis sentimen pada aplikasi *Halodoc* menunjukkan bahwa algoritma *SVM* mampu mencapai akurasi hingga 88,32%[4].

Sementara itu, penelitian lain pada aplikasi *TikTokShop* juga mencatat bahwa *SVM* memiliki performa lebih baik dibandingkan *Naïve Bayes* dengan akurasi 68,86%[5]. Penelitian lain yang dilakukan pada aplikasi *OYO* mencatat akurasi sebesar 80,36% ketika menggunakan kernel *Radial Basis Function*[6].

Studi lain pada aplikasi *Wattpad*, *Noveltoon*, dan *Joylada* membuktikan bahwa *SVM* dapat mengelompokkan ulasan pengguna dengan akurasi berkisar antara 63% hingga 70%[7].

Penelitian terbaru mengenai *Quantum-Enhanced Support Vector Machine* (*QE-SVM*) menunjukkan peningkatan performa klasifikasi hingga mencapai 93,33%[8], yang menegaskan potensi perkembangan algoritma ini dalam konteks analisis sentimen.

Berdasarkan berbagai penelitian tersebut, dapat disimpulkan bahwa algoritma *SVM* konsisten memberikan performa yang baik dalam analisis sentimen pada berbagai jenis aplikasi digital. Namun, penelitian analisis sentimen yang secara khusus membahas aplikasi *CapCut* masih terbatas, padahal aplikasi ini memiliki karakteristik unik sebagai aplikasi editing video dengan fokus pada kreativitas dan interaksi sosial pengguna. Permasalahan utama

dalam penelitian ini adalah bagaimana mengolah data ulasan pengguna yang tidak terstruktur menjadi informasi yang bermakna serta bagaimana mengklasifikasikan sentimen pengguna secara akurat untuk mengetahui kecenderungan persepsi terhadap aplikasi *CapCut*. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen pengguna aplikasi *CapCut* dengan mengelompokkan ulasan ke dalam kategori positif, netral, dan negatif menggunakan algoritma *SVM*, serta mengevaluasi kinerja model yang dihasilkan.

Rencana penyelesaian masalah dalam penelitian ini dilakukan melalui beberapa tahapan, yaitu pengumpulan data ulasan menggunakan teknik *scraping* dari *Google Play Store*, pelabelan data berdasarkan rating pengguna, serta proses *text preprocessing* yang meliputi *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Selanjutnya, data direpresentasikan menggunakan pembobotan *TF-IDF* dan diklasifikasikan menggunakan algoritma *SVM*. Kinerja model kemudian dievaluasi menggunakan *confusion matrix* dengan metrik akurasi, presisi, *recall*, dan *F1-score* untuk mengetahui tingkat keberhasilan klasifikasi yang dihasilkan.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Tabel 1. Penelitian Terdahulu

Penelitian	Persamaan	Perbedaan
Sentiment Analysis of the Halodoc Application Using the Support Vector Machine (Svm) Algorithm [4]	Sama-sama menggunakan *SVM* pada ulasan aplikasi	Objek penelitian aplikasi kesehatan, bukan *CapCut*
Sentiment Analysis of OYO App Reviews Using the Support Vector Machine Algorithm [6]	Metode sama: analisis sentimen berbasis ulasan pengguna	Domain aplikasi perhotelan, bukan aplikasi editing video
Sentiment Analysis using Support Vector Machine and Random Forest [9]	Menguatkan bahwa *SVM* unggul untuk klasifikasi teks	Menganalisis ulasan nyata CapCut dan menghasilkan rekomendasi yang dapat digunakan pengembang.
Quantum-Enhanced Support Vector Machine for Sentiment Classification [8]	Relevan sebagai pengembangan teori *SVM*	Menggunakan metode kuantum; tidak digunakan dalam penelitian ini
Sentiment Analysis of Free Online Novel Applications Using the Support Vector Machine Method [7]	Sama-sama menganalisis aplikasi hiburan	Preprocessing lebih lengkap dan berpotensi menghasilkan akurasi lebih tinggi.

Berdasarkan Tabel 1, penelitian terdahulu menunjukkan bahwa algoritma Support Vector Machine (SVM) telah banyak diterapkan dalam analisis sentimen pada berbagai domain seperti aplikasi kesehatan, perhotelan, dan hiburan, namun masih memiliki keterbatasan dari segi objek penelitian, jenis data, dan pendekatan metode yang digunakan. Beberapa penelitian belum secara spesifik membahas aplikasi berbasis kreativitas seperti CapCut, sementara penelitian lain lebih berfokus pada pengembangan metode tanpa memanfaatkan data ulasan nyata yang relevan, serta terdapat pula pendekatan lanjutan seperti metode berbasis kuantum yang kurang aplikatif dalam implementasi umum. Oleh karena itu, penelitian ini menawarkan keunggulan dengan menggunakan data ulasan pengguna aplikasi CapCut yang lebih relevan, menerapkan tahapan preprocessing yang lebih sistematis untuk meningkatkan kualitas data, serta memanfaatkan algoritma SVM yang sesuai dengan karakteristik data teks, sehingga diharapkan mampu menghasilkan klasifikasi sentimen yang lebih akurat sekaligus memberikan insight yang berguna bagi pengembang dalam meningkatkan kualitas aplikasi.

2.2. Analisis Sentimen

Analisis sentimen merupakan proses untuk mengidentifikasi dan mengklasifikasikan opini atau emosi pengguna ke dalam kategori tertentu seperti positif, negatif, atau netral[9].

Teknik ini digunakan untuk memahami persepsi publik terhadap suatu aplikasi atau layanan digital dengan mengolah data ulasan yang tersedia secara daring. Selain itu, analisis sentimen juga dapat menghasilkan informasi terstruktur yang membantu pengembang aplikasi dalam melakukan evaluasi dan peningkatan kualitas layanan. Melalui hasil analisis ini, pengembang dapat mengetahui aspek-aspek yang menjadi kelebihan maupun kekurangan suatu aplikasi berdasarkan sudut pandang pengguna, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan untuk pengembangan fitur, perbaikan sistem, serta peningkatan pengalaman pengguna secara keseluruhan.

2.3. Text Preprocessing

Proses *text preprocessing* yang terdiri dari *case folding*, *tokenizing*, *stopword removal*, dan *stemming* diperlukan untuk menyiapkan data teks agar siap diolah oleh algoritma klasifikasi[10].

Proses *case folding* digunakan untuk menyeragamkan seluruh teks menjadi huruf kecil.

Selanjutnya, *tokenizing* berfungsi untuk memecah teks menjadi unit kata, yang kemudian diikuti oleh *stopword removal* guna menghapus kata-kata umum yang tidak memiliki makna sentimen yang signifikan. Proses *stemming* dilakukan untuk mengembalikan kata ke bentuk dasarnya sehingga mengurangi variasi kata yang memiliki makna serupa.Tahapan ini membantu meningkatkan akurasi analisis sentimen dengan menghilangkan elemen teks yang tidak relevan dan menyeimbangkan struktur data.

2.4. Pembobotan Kata dengan TF-IDF

TF-IDF (*Term Frequency–Inverse Document Frequency*) merupakan metode pembobotan yang digunakan untuk menentukan tingkat kepentingan suatu kata dalam sebuah dokumen. Teknik ini mengubah teks menjadi representasi numerik sehingga dapat diproses oleh model pembelajaran mesin dengan lebih efektif. Pembobotan TF-IDF banyak digunakan dalam penelitian analisis sentimen karena mampu menyoroti kata-kata yang memiliki nilai informatif dalam ulasan pengguna[11].

2.5. Algoritma Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan salah satu metode dalam machine learning yang banyak digunakan untuk proses klasifikasi karena kemampuannya dalam menghasilkan tingkat akurasi yang tinggi. Metode ini bekerja dengan membangun sebuah model klasifikasi berbasis teori pembelajaran statistik yang mampu melakukan generalisasi data dengan baik[12].

Secara umum, *Support Vector Machine* (*SVM*) adalah algoritma klasifikasi yang bekerja dengan menentukan *hyperplane* terbaik untuk memisahkan data pada setiap kelas sentimen. Algoritma ini efektif untuk data teks berdimensi tinggi karena mampu menghasilkan pemisahan kelas yang akurat[13].

2.6. Ulasan Pengguna sebagai Sumber Data

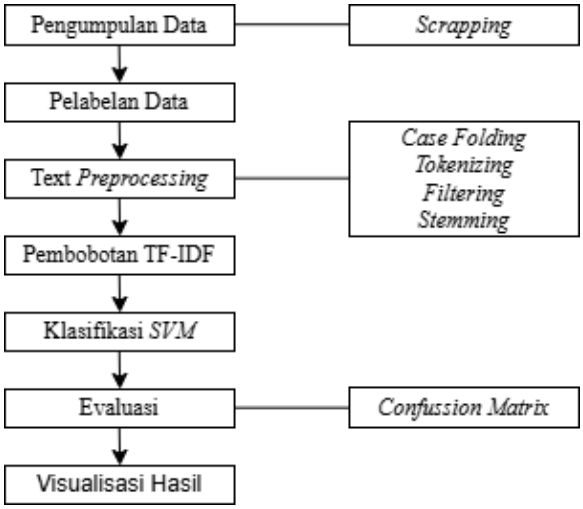
Ulasan pengguna pada platform digital seperti *Google Play Store* merupakan sumber data yang penting karena mencerminkan pengalaman dan persepsi pengguna terhadap suatu aplikasi. Selain itu, ulasan pengguna biasanya mencakup berbagai aspek, seperti performa aplikasi, tampilan antarmuka, kemudahan penggunaan, hingga kualitas fitur yang disediakan. Informasi ini dapat dimanfaatkan untuk mengetahui tingkat kepuasan pengguna secara lebih mendalam. Data ulasan yang bersifat autentik dan berasal langsung dari pengguna memungkinkan analisis sentimen untuk menghadirkan gambaran yang akurat terkait kualitas layanan aplikasi[14].

2.7. Confusion Matrix

Confusion matrix merupakan alat evaluasi dalam machine learning yang digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan label aktual dan hasil prediksi dalam bentuk matriks dua dimensi, di mana baris menunjukkan label sebenarnya dan kolom menunjukkan label prediksi sehingga distribusi kesalahan dan keberhasilan klasifikasi dapat terlihat dengan jelas. Confusion matrix umumnya terdiri dari empat komponen utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN)[15].

True Positive menunjukkan jumlah data yang diprediksi benar sebagai kelas positif, sedangkan *True Negative* adalah jumlah data yang diprediksi benar sebagai kelas negatif. Sementara itu, *False Positive* merupakan kesalahan ketika data negatif diprediksi sebagai positif, dan *False Negative* adalah kesalahan ketika data positif diprediksi sebagai negatif.

3. METODE PENELITIAN



Gambar 1. Alur Proses Penelitian

Berdasarkan gambar tersebut, penelitian ini dilakukan melalui beberapa tahapan utama, dimulai dari pengumpulan data menggunakan teknik *scraping* dan dilanjutkan dengan pelabelan data. Data teks kemudian diproses melalui tahap *preprocessing* untuk meningkatkan kualitas data sebelum dilakukan pembobotan menggunakan metode TF-IDF. Selanjutnya, data diklasifikasikan dengan algoritma Support Vector Machine (SVM) dan dievaluasi menggunakan *confusion matrix*. Hasil akhir dari penelitian ini disajikan dalam bentuk visualisasi untuk memudahkan pemahaman terhadap kinerja model.

Pengumpulan data dilakukan dengan teknik *scraping* menggunakan pustaka *Google Play Scraper* berbasis *Python* yang dijalankan pada platform *Google Colab*. Pada tahap awal, diperoleh sebanyak 10.000 ulasan pengguna, kemudian dilakukan penyaringan data berdasarkan rentang waktu tahun 2023 hingga 2026. Hasil dari proses penyaringan tersebut menghasilkan 6.395 ulasan yang memenuhi kriteria penelitian, selanjutnya diolah dan diurutkan berdasarkan tingkat relevansi untuk mendukung proses analisis lebih lanjut.

Setelah pengumpulan data yang diperoleh dari proses *scraping* berhasil, data tersebut kemudian dikelompokkan ke dalam kategori sentimen

berdasarkan nilai penilaian pengguna. Penilaian tersebut menggunakan skala 1 sampai 5 yang tersedia pada aplikasi. Dalam penelitian ini, ulasan dengan nilai 1 dan 2 dikelompokkan sebagai sentimen negatif dan diberi kode -1, nilai 3 dikategorikan sebagai sentimen netral dengan kode 0, sedangkan ulasan dengan nilai 4 dan 5 dimasukkan ke dalam kategori sentimen positif dengan kode 1.

Tahapan selanjutnya adalah *text preprocessing*, yang bertujuan untuk mengubah data teks yang bersifat tidak terstruktur menjadi data yang lebih siap untuk dianalisis. *Text preprocessing* mencakup *case folding* untuk menyeragamkan teks ke dalam bentuk huruf kecil, *tokenizing* untuk memecah teks menjadi unit kata, serta *filtering* melalui *stopword removal* guna menghapus kata-kata yang tidak memiliki kontribusi signifikan terhadap analisis. Selain itu, dilakukan proses *stemming* untuk mengembalikan kata ke bentuk dasarnya.

Setelah *text preprocessing* selesai, data dibagi ke menjadi dua bagian, yaitu data pelatihan dan data pengujian. Tahap selanjutnya adalah membangun model klasifikasi dengan menerapkan teknik pembobotan kata *Term Frequency–Inverse Document Frequency* (*TF-IDF*) pada data pelatihan. Model *Support Vector Machine* (*SVM*) kemudian dilatih menggunakan data pelatihan yang telah direpresentasikan dalam bentuk vektor *TF-IDF*. Data pelatihan digunakan untuk membentuk model klasifikasi, sedangkan data pengujian digunakan untuk menilai performa model berdasarkan nilai akurasi, presisi, dan *recall* pada klasifikasi sentimen.

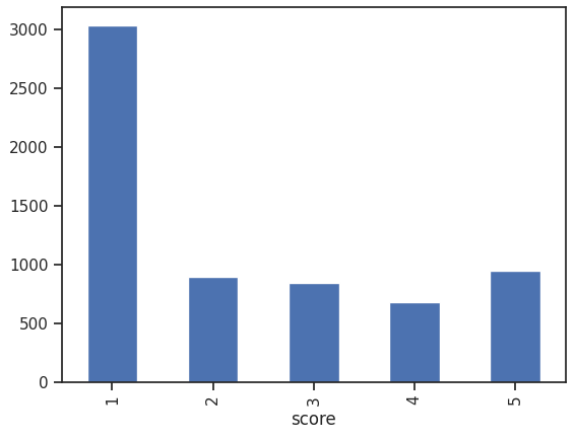
Selanjutnya, tingkat kinerja model dievaluasi dengan menghitung nilai akurasi berdasarkan hasil klasifikasi pada data pengujian. Proses evaluasi dilakukan menggunakan *confusion matrix* yang digunakan untuk memperoleh nilai akurasi, presisi, dan *recall*. Nilai presisi mencerminkan ketepatan model dalam mengklasifikasikan sentimen positif maupun negatif, sementara *recall* menunjukkan kemampuan model dalam menangkap seluruh ulasan yang relevan secara menyeluruh.

Tahap visualisasi dilakukan dengan memanfaatkan grafik serta *word cloud* untuk seluruh kumpulan ulasan. *Word cloud* digunakan untuk menampilkan visualisasi dari data teks dan berfungsi sebagai sarana pendukung dalam analisis teks. Pada visualisasi tersebut, ukuran setiap kata merepresentasikan frekuensi kemunculannya dalam data, sehingga kata-kata yang paling sering muncul akan terlihat lebih dominan dibandingkan kata lainnya.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Visualisasi pada gambar 2 menunjukkan data distribusi rating sebagai bentuk penilaian pengguna terhadap penggunaan aplikasi *CapCut*.Hasil pengumpulan data melalui proses *scraping* terhadap 10.000 ulasan yang kemudian difilter berdasarkan periode tahun 2023 hingga 2026 menghasilkan sebanyak 6.395 ulasan yang digunakan sebagai data penelitian. Proses penyaringan ini dilakukan untuk memastikan data yang digunakan relevan dengan konteks waktu penelitian dan memiliki kualitas yang lebih baik.



Gambar 2. Hasil Pengumpulan Data

4.2. Pelabelan Data

Data hasil *scraping* selanjutnya diberi label sentimen berdasarkan skor rating pengguna pada skala 1 hingga 5. Rating 1–2 diklasifikasikan sebagai sentimen negatif dengan kode -1, rating 3 sebagai sentimen netral dengan kode 0, dan rating 4–5 sebagai sentimen positif dengan kode 1.

		sentiment
Year	sentiment	
2023	-1	155
	0	76
	1	198
2024	-1	702
	0	152
	1	195
2025	-1	2201
	0	454
	1	841
2026	-1	880
	0	159
	1	382

Gambar 3. Pelabelan Data Sentimen

Berdasarkan gambar 3, pada tahun 2023 sentimen didominasi oleh ulasan positif (198), diikuti negatif (155) dan netral (76), yang menunjukkan persepsi pengguna masih cenderung baik. Namun pada tahun 2024 terjadi perubahan dengan dominasi sentimen negatif (702), sementara sentimen positif (195) dan netral (152) berada di bawahnya. Tren ini semakin meningkat pada tahun 2025, di mana sentimen negatif mencapai puncaknya (2201), jauh

lebih tinggi dibandingkan sentimen positif (841) dan netral (454). Pada tahun 2026, jumlah sentimen negatif menurun menjadi 880, tetapi masih menjadi yang paling dominan dibandingkan sentimen positif (382) dan netral (159). Secara keseluruhan, dapat disimpulkan bahwa terjadi pergeseran sentimen dari dominasi positif pada tahun 2023 menjadi dominasi negatif pada tahun 2024 hingga 2026, dengan puncaknya pada tahun 2025. Hal ini menunjukkan adanya perubahan persepsi pengguna yang perlu dianalisis lebih lanjut untuk mengetahui faktor penyebabnya.

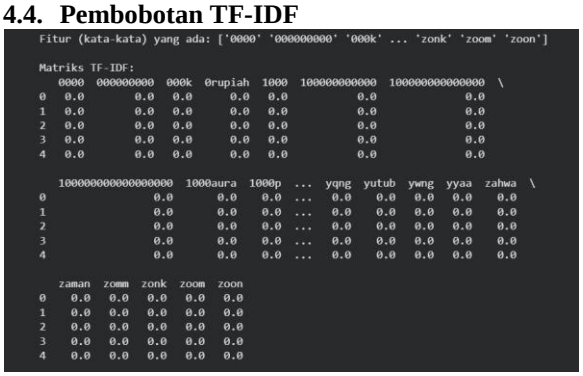
4.3. Text Preprocessing

Tabel 2. Hasil Text Preprocesing

Tahapan	Hasil
Case Folding	saya kasih bintang 3 dulu ya semenjak saya download jadi banyak bug nya sering ngeleg trus macet banyak iklan lgi trus g bisa buat nulis teks lagi ya menurut saya udah bagus cuman ada sedikit kekurangan nya tolong yang bikin aplikasi ini diperbaiki yah
Tokenizing	['kasih', 'bintang', '3', 'ya', 'semenjak', 'download', 'bug', 'nya', 'ngeleg', 'trus', 'macet', 'iklan', 'lgi', 'trus', 'g', 'nulis', 'teks', 'ya', 'udah', 'bagus', 'cuman', 'kekurangan', 'nya', 'tolong', 'bikin', 'aplikasi', 'diperbaiki', 'yah']
Filtering	['kasih', 'bintang', '3', 'ya', 'semenjak', 'download', 'bug', 'nya', 'ngeleg', 'trus', 'macet', 'iklan', 'lgi', 'trus', 'g', 'nulis', 'teks', 'ya', 'udah', 'bagus', 'cuman', 'kekurangan', 'nya', 'tolong', 'bikin', 'aplikasi', 'diperbaiki', 'yah']
Stemming	['kasih', 'bintang', '3', 'ya', 'semenjak', 'download', 'bug', 'nya', 'ngeleg', 'trus', 'macet', 'iklan', 'lgi', 'trus', 'g', 'nulis', 'teks', 'ya', 'udah', 'bagus', 'cuman', 'kurang', 'nya', 'tolong', 'bikin', 'aplikasi', 'baik', 'yah']

Berdasarkan Tabel 2, proses *text preprocessing* dilakukan melalui beberapa tahapan penting untuk membersihkan dan menyiapkan data teks agar dapat diolah dengan lebih optimal oleh model klasifikasi. Pada tahap *case folding*, seluruh teks diubah menjadi huruf kecil untuk menyeragamkan penulisan dan menghindari perbedaan makna akibat penggunaan huruf kapital. Selanjutnya, tahap *tokenizing* memecah kalimat menjadi unit-unit kata sehingga memudahkan proses analisis pada setiap kata secara individu. Setelah itu, dilakukan *filtering* atau *stopword removal* untuk menyaring kata-kata yang tidak memiliki makna penting atau tidak berkontribusi terhadap sentimen, meskipun dalam kasus ini sebagian besar kata tetap dipertahankan karena dianggap relevan dengan konteks ulasan. Tahap terakhir adalah *stemming*, yaitu mengubah kata ke bentuk dasarnya, seperti "kekurangan" menjadi "kurang" dan "diperbaiki" menjadi "baik", sehingga dapat mengurangi variasi kata dengan makna serupa. Secara keseluruhan, tahapan preprocessing ini berperan penting dalam meningkatkan kualitas data dengan menghilangkan noise, menyederhanakan struktur teks, serta membantu model dalam mengenali pola kata secara lebih efektif pada proses analisis sentimen.

4.4. Pembobotan TF-IDF



Gambar 4. Hasil Pembobotan TF-IDF

Gambar 4 menunjukkan data yang telah melalui tahap text preprocessing, kemudian dilakukan proses seleksi untuk menghilangkan data yang memiliki nilai kosong atau tidak relevan sehingga kualitas data menjadi lebih baik. Setelah itu, data direpresentasikan menggunakan teknik pembobotan kata Term Frequency–Inverse Document Frequency (TF-IDF), yang bertujuan untuk mengubah data teks menjadi bentuk numerik agar dapat diproses oleh algoritma machine learning.

4.5. Klasifikasi SVM

Tabel 3. Hasil Akurasi Perbandingan Data

Data Latih	Data Uji	Akurasi
90% (5755)	10% (640)	70,47%
80% (5116)	20% (1279)	71,54%
70% (4476)	30% (1919)	71,55%

Berdasarkan Tabel 3, hasil pengujian menunjukkan bahwa variasi pembagian data latih dan data uji memberikan pengaruh terhadap nilai akurasi model yang dihasilkan. Pengujian dilakukan dalam tiga skenario pembagian data untuk mengetahui kombinasi terbaik dalam membangun model klasifikasi menggunakan algoritma *Support Vector Machine*. Pada pembagian 90% data latih dan 10% data uji, diperoleh akurasi sebesar 70,47%. Ketika proporsi data latih dikurangi menjadi 80% dan data uji ditingkatkan menjadi 20%, akurasi mengalami peningkatan menjadi 71,54%. Hal ini menunjukkan bahwa jumlah data latih yang besar tidak selalu menjamin peningkatan performa model, terutama jika data uji yang digunakan terlalu sedikit sehingga kurang merepresentasikan variasi data secara menyeluruh. Selanjutnya, pada pembagian 70% data latih dan 30% data uji, akurasi kembali meningkat meskipun tidak signifikan, yaitu sebesar 71,55%. Secara keseluruhan, hasil dari ketiga skenario tersebut menunjukkan bahwa model *SVM* memiliki performa yang cukup konsisten dengan rentang akurasi sekitar

70% hingga 71%. Selain itu, dapat disimpulkan bahwa pembagian data latih dan data uji yang proporsional berpengaruh terhadap kualitas evaluasi model, sehingga pemilihan skenario pembagian data menjadi salah satu faktor penting dalam proses klasifikasi sentimen. Lalu hasil 3 skenario pengujian di atas diklasifikasikan dengan *SVM*.

Tabel 4. Hasil Klasifikasi Skenario Pertama

Label Sentimen	Positif	Netral	Negatif	Rata - rata
Presisi	0,72	0,00	0,70	0,62
Recall	0,51	0,00	0,95	0,70
F1 Score	0,59	0,00	0,81	0,65

Tabel 4 menunjukkan klasifikasi hasil pengujian skenario pertama yang mempunyai akurasi 70,47%.

Tabel 5. Hasil Klasifikasi Skenario Kedua

Label Sentimen	Positif	Netral	Negatif	Rata - rata
Presisi	0,70	0,00	0,72	0,69
Recall	0,56	0,00	0,93	0,72
F1 Score	0,62	0,00	0,81	0,66

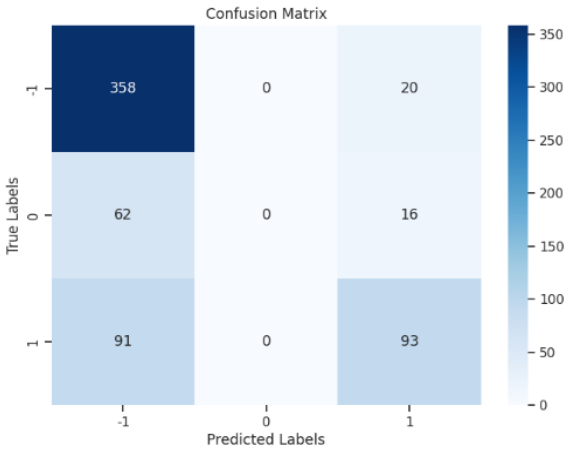
Tabel 5 menunjukkan klasifikasi hasil pengujian skenario kedua yang mempunyai akurasi 71,54%.

Tabel 6. Hasil Klasifikasi Skenario Ketiga

Label Sentimen	Positif	Netral	Negatif	Rata - rata
Presisi	0,67	0,00	0,73	0,62
Recall	0,56	0,00	0,93	0,72
F1 Score	0,61	0,00	0,82	0,66

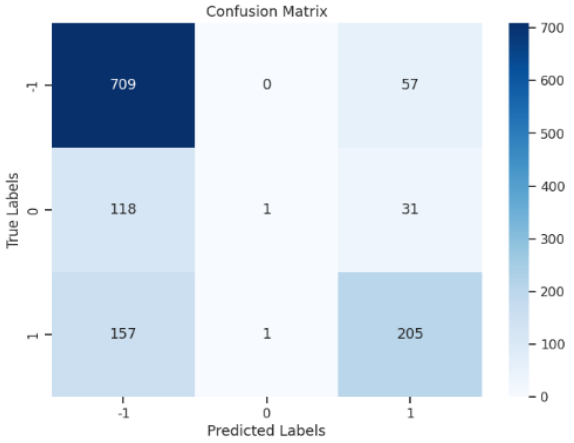
Tabel 6 menunjukkan klasifikasi hasil pengujian skenario ketiga yang mempunyai akurasi 71,55%.

4.6. Evaluasi



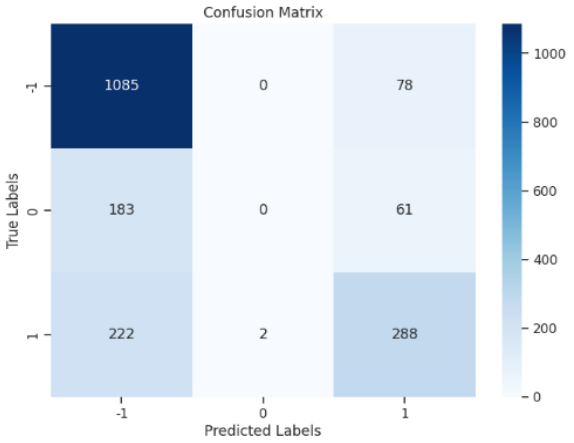
Gambar 5. Confusion Matrix Skenario 1

Gambar 5 menampilkan hasil evaluasi skenario pertama menggunakan *Confusion Matrix*. Pada skenario ini model mendapatkan tingkat nilai akurasi 70,47%. Nilai akurasi tersebut dihitung berdasarkan jumlah data yang berhasil diklasifikasikan dengan benar dari keseluruhan data uji, yaitu sebanyak 451 dari 640 data, yang terdiri dari 358 *true* negatif dan 93 *true* positif. Berdasarkan hasil evaluasi, maka dapat disimpulkan bahwa data uji yang gagal diklasifikasikan berjumlah 189.



Gambar 6. Confusion Matrix Skenario 2

Gambar 6 menampilkan hasil evaluasi skenario kedua menggunakan *Confusion Matrix*. Pada skenario ini model mendapatkan tingkat akurasi 71,54%. Nilai akurasi tersebut dihitung berdasarkan jumlah data yang berhasil diklasifikasikan dengan benar dari keseluruhan data uji, yaitu sebanyak 914 dari 1279 data, yang terdiri dari 709 *true* negatif dan 205 *true* positif. Berdasarkan hasil evaluasi, maka dapat disimpulkan bahwa data uji yang gagal diklasifikasikan berjumlah 365.



Gambar 7. Confusion Matrix Skenario 3

Gambar 7 menampilkan hasil evaluasi skenario ketiga menggunakan *Confusion Matrix*. Pada skenario ini model mendapatkan tingkat akurasi 71,55%. Nilai akurasi tersebut dihitung berdasarkan jumlah data yang berhasil diklasifikasikan dengan benar dari keseluruhan data uji, yaitu sebanyak 1373 dari 1919 data, yang terdiri dari 1085 *true* negatif dan 288 *true* positif. Berdasarkan hasil evaluasi, maka dapat disimpulkan bahwa data uji yang gagal diklasifikasikan berjumlah 546.

4.7. Visualisasi Hasil

Visualisasi hasil dalam penelitian ini menggunakan Wordcloud untuk menampilkan representasi visual dari data teks sebagai sarana pendukung dalam analisis sentimen. Pada Wordcloud, ukuran setiap kata mencerminkan frekuensi kemunculannya dalam kumpulan data, sehingga kata-kata yang paling sering muncul akan terlihat lebih besar dan dominan dibandingkan kata lainnya. Selain itu, visualisasi ini mampu menyederhanakan data teks yang kompleks menjadi bentuk yang lebih mudah dipahami secara cepat dan intuitif, sehingga peneliti dapat dengan mudah mengidentifikasi pola kata, topik yang sering dibahas, serta kecenderungan opini pengguna. Dalam penelitian ini, Wordcloud dikelompokkan menjadi tiga kategori berdasarkan kelas sentimen, yaitu Wordcloud sentimen positif, negatif, dan netral. Melalui pengelompokan tersebut, perbedaan karakteristik kata pada masing-masing sentimen dapat diamati dengan lebih jelas, seperti kata-kata yang mencerminkan kepuasan pengguna pada sentimen positif, keluhan atau masalah pada sentimen negatif, serta kata-kata yang bersifat informatif atau netral. Dengan demikian, Wordcloud tidak hanya berfungsi sebagai visualisasi data, tetapi juga sebagai alat eksplorasi yang membantu dalam menginterpretasikan hasil analisis sentimen dan memberikan gambaran yang lebih komprehensif mengenai persepsi pengguna terhadap aplikasi.

Gambar 8. Wordcloud Positif

Pada gambar 8 ditampilkan visualisasi untuk wordcloud sentimen positif.

Gambar 9. Wordcloud Negatif

Pada gambar 9 ditampilkan visualisasi untuk wordcloud sentimen negatif.

Gambar 10. Wordcloud Netral

Pada gambar 10 ditampilkan visualisasi untuk wordcloud sentimen netral.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma *Support Vector Machine (SVM)* mampu digunakan untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi *CapCut* dengan kinerja yang cukup baik. Hal ini ditunjukkan oleh nilai akurasi terbaik sebesar 71,55% pada skenario pembagian data latih 70% dan data uji 30%, yang mengindikasikan bahwa model memiliki kemampuan yang cukup stabil dalam mengklasifikasikan data teks. Selain itu, tahapan text preprocessing dan pembobotan *TF-IDF* terbukti berperan penting dalam meningkatkan kualitas data, sehingga model dapat mengenali pola kata dengan lebih efektif dan menghasilkan klasifikasi yang lebih optimal, khususnya dalam mendeteksi sentimen negatif yang memiliki nilai recall relatif tinggi. Hasil analisis juga menunjukkan bahwa mayoritas ulasan pengguna didominasi oleh sentimen negatif, terutama pada periode tahun 2024 hingga 2026 dengan puncaknya pada tahun 2025, yang mengindikasikan adanya penurunan kualitas pengalaman pengguna terhadap aplikasi *CapCut* serta perlunya perhatian lebih dari pengembang dalam mengevaluasi fitur dan performa aplikasi.

Berdasarkan temuan tersebut, disarankan untuk penelitian selanjutnya agar melakukan optimasi parameter pada model *SVM*, seperti penentuan kernel, nilai C, dan *gamma*, guna meningkatkan performa klasifikasi. Selain itu, penggunaan algoritma pembanding seperti *Naïve Bayes*, *Random Forest*, atau metode berbasis *deep learning* seperti *LSTM* dan *BERT* dapat dipertimbangkan untuk memperoleh hasil yang lebih komprehensif. Pengembangan analisis sentimen berbasis aspek *(aspect-based sentiment analysis)* juga direkomendasikan agar dapat mengidentifikasi secara lebih spesifik bagian atau fitur aplikasi yang menjadi sumber kepuasan maupun keluhan pengguna. Di samping itu, penambahan jumlah dan variasi data, serta pemanfaatan teknik

preprocessing yang lebih lanjut, diharapkan mampu meningkatkan akurasi model dan memberikan insight yang lebih mendalam bagi pengembang dalam upaya meningkatkan kualitas dan pengalaman pengguna aplikasi.

DAFTAR PUSTAKA

[1] F. Halim, H. Ramadhan, D. O. Sitepu, S. Megawan, and H. Gohzali, "Analisis Sentimen dari Review Aplikasi Triv Menggunakan Algoritma Support Vector Machine," *J-COSINE (Journal of Computer Science and Informatics Engineering)*, vol. 04, no. 3, pp. 188–200, 2025.

[2] C. Anilkumar, S. V E., S. Kanchana, and S. B. Kumar, "Sentimental Analysis on Product Reviews Using Support Vector Machine and Nave Bayes," *Applied and Computational Engineering*, vol. 2, no. 1, pp. 1067–1073, 2023, doi: 10.54254/2755-2721/2/20220586.

[3] A. M. Fajria, A. Faqih, and G. Dwilestari, "The Impact of Principal Component Analysis Dimensionality Reduction on Sentiment Classification Performance Using Support Vector Machine," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 4, no. 2, 2025.

[4] F. A. Rachmadi Putri and S. Siswanti, "Sentiment Analysis of the Halodoc Application Using the Support Vector Machine (Svm) Algorithm," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 11, no. 2, pp. 353–360, 2025, doi: 10.33330/jurteksi.v11i2.3748.

[5] A. M. Rizki, B. Bustami, and S. F. Anshari, "Comparison of Support Vector Machine and NaÃ¯ve Bayes Algorithms in Sentiment Analysis of Tiktokshop Application User Reviews," *Journal of Renewable Energy, Electrical, and Computer Engineering*, vol. 5, no. 1, pp. 18–29, 2025, doi: 10.29103/jreece.v5i1.21342.

[6] Z. Zaenal and I. R. I. Astutik, "Sentiment Analysis of OYO App Reviews Using the Support Vector Machine Algorithm," *Procedia of Engineering and Life Science*, vol. 3, no. 12, 2023, doi: 10.21070/pels.v3i0.1338.

[7] Yulidayanti, Safwandi, and Fajriana, "Sentiment Analysis of Free Online Novel Applications Using the Support Vector Machine Method," *International Journal of Engineering, Science and Information Technology*, vol. 5, no. 1, pp. 340–349, 2025, doi: 10.52088/ijesty.v5i1.732.

[8] F. Z. Ruskanda, M. R. Abiwardani, R. Mulyawan, I. Syafalni, and H. T. Larasati, "Quantum-Enhanced Support Vector Machine for Sentiment Classification," *IEEE Access*, vol. 11, no. 2, pp. 87520–87532, 2023, doi: 10.1109/ACCESS.2023.3304990.

[9] T. Ahmed Khan, R. Sadiq, Z. Shahid, M. M. Alam, and M. Mohd Su'ud, "Sentiment Analysis using Support Vector Machine and Random Forest," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, pp. 67–75, 2024, doi: 10.33093/jiwe.2024.3.1.5.

[10] M. A. Palomino and F. Aider, "Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis," *Applied Sciences (Switzerland)*, vol. 12, no. 17, 2022, doi: 10.3390/app12178765.

[11] D. F. Surianto and D. F. Surianto, "Enhancing K-Means Clustering for Journal Articles using TF-IDF and LDA Feature Extraction," *Brilliance: Research of Artificial Intelligence*, vol. 4, no. 2, pp. 964–972, 2025, doi: 10.47709/brilliance.v4i2.5547.

[12] N. H. Ovirianti, M. Zarlis, and H. Mawengkang, "Support Vector Machine Using A Classification Algorithm," *Sinkron : Jurnal dan Penelitian Teknik Informatika*, vol. 7, no. 3, pp. 2103–2107, 2022, doi: 10.33395/sinkron.v7i3.11597.

[13] A. Febriani, K. Umam, and M. I. Mustofa, "Implementation of Support Vector Machine for Classifying User Reviews on the Sentuh Tanahku Application," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 4, pp. 1551–1558, 2025.

[14] I. Nawawi, K. F. Ilmawan, M. R. Maarif, and M. Syafrudin, "Exploring Tourist Experience through Online Reviews Using Aspect-Based Sentiment Analysis with Zero-Shot Learning for Hospitality Service Enhancement," *Information (Switzerland)*, vol. 15, no. 8, 2024, doi: 10.3390/info15080499.

[15] Jude Chukwura Obi, "A comparative study of several classification metrics and their performances on data," *World Journal of Advanced Engineering Technology and Sciences*, vol. 8, no. 1, pp. 308–314, 2023, doi: 10.30574/wjaets.2023.8.1.0054