

Klasifikasi Sentimen Ulasan Pengguna Aplikasi Dompot Digital Menggunakan Metode *Support Vector Machine* (SVM)

Tiara Bela Harahap^{1*}, Naina Nazwa Hasibuan²

^{1,2} Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Indonesia

*Penulis Korespondensi: tiarabelahrp@gmail.com

Abstract. This study was motivated by the large volume of user reviews for the DANA application on the Google Play Store, which contain diverse opinions regarding user experiences. While these reviews offer insights into user satisfaction and perceptions, systematic analysis is required to identify the underlying sentiments. This research aims to classify DANA user reviews into positive, negative, and neutral sentiment categories using the Support Vector Machine (SVM) algorithm. The dataset was sourced from the "DANA Sentiment Analysis from Indonesian Play Store" dataset, comprising 50,000 reviews, from which a subset of 1,500 reviews was selected using stratified sampling. The research process involved text preprocessing, TF-IDF feature weighting, data splitting, and sentiment classification using SVM with a linear kernel. Model performance was evaluated using a confusion matrix, accuracy, precision, recall, and F1-score. The results indicate that the SVM model achieved an accuracy of 82.00%, demonstrating the method's effectiveness in classifying the sentiment of DANA user reviews. These findings can serve as supporting information for identifying user responses and evaluating the user experience with the DANA application.

Keywords: DANA Application; Google Play Store; Sentiment Analysis; Support Vector Machine; TF-IDF.

Abstrak. Penelitian ini dilatar belakangi oleh banyaknya ulasan pengguna aplikasi DANA pada Google Play Store yang memuat berbagai pendapat mengenai pengalaman pengguna terhadap aplikasi. Ulasan tersebut dapat memberikan informasi mengenai tanggapan pengguna, tetapi diperlukan analisis secara sistematis untuk mengetahui sentimen yang terkandung di dalamnya. Penelitian ini bertujuan untuk mengklasifikasikan ulasan pengguna aplikasi DANA ke dalam sentimen positif, negatif, dan netral menggunakan algoritma Support Vector Machine (SVM). Dataset yang digunakan berasal dari "DANA Sentiment Analysis from Playstore Indonesia" yang terdiri dari 50.000 ulasan, dengan 1.500 ulasan dipilih sebagai sampel menggunakan teknik stratified sampling. Tahapan penelitian meliputi reprocessing teks, pembobotan fitur menggunakan TF-IDF, pembagian data, dan klasifikasi menggunakan SVM. Untuk mengevaluasi model, matrix confusion, accuracy, precision, recall, dan F1-score digunakan. Hasil penelitian menunjukkan bahwa model SVM memiliki kemampuan untuk mengklasifikasikan sentimen ulasan pengguna DANA dengan akurasi 82,00%. Hasil klasifikasi tersebut dapat dimanfaatkan sebagai informasi untuk mengetahui kecenderungan tanggapan pengguna serta mendukung evaluasi terhadap pengalaman pengguna aplikasi DANA.

Kata kunci: Analisis Sentimen; Aplikasi DANA; Google Play Store; Support Vector Machine; TF-IDF.

1. LATAR BELAKANG

Berbagai aspek kehidupan sosial, termasuk transaksi keuangan, telah diubah secara signifikan oleh kemajuan teknologi digital. Masyarakat dapat melakukan berbagai transaksi dengan lebih mudah dengan menggunakan dompet digital. DANA adalah salah satu layanan dompet digital yang paling populer di Indonesia.

Ulasan yang diberikan pengguna menunjukkan beragam pengalaman dalam menggunakan aplikasi DANA. Sebagian pengguna menyampaikan kepuasan terhadap manfaat aplikasi, sedangkan pengguna lainnya menyampaikan keluhan mengenai transaksi, akses akun, layanan pelanggan, maupun kendala teknis. Untuk menemukan tren sentimen pengguna, diperlukan metode yang memungkinkan analisis menyeluruh dari data teks ini.

Salah satu pendekatan yang dapat digunakan untuk mengidentifikasi adalah analisis sentimen. Identifikasi yang digunakan pada ulasan aplikasi DANA menggunakan metode Support Vector Machine (SVM). Untuk klasifikasi sentimen ulasan pengguna DANA, Prapasta dan Zainuddin (2026) menggunakan SVM dengan tahapan preprocessing dan pembobotan TF-IDF. (Nawulansih et al. 2025) juga menggunakan SVM pada ulasan aplikasi DANA dengan metode SMOTE untuk mengatasi ketidakseimbangan data. Selain itu, analisis sentimen terhadap ulasan pengguna aplikasi DANA menggunakan SVM yang dikombinasikan dengan Word2Vec (Acuan Supian, 2024). Penelitian telah menunjukkan bahwa SVM dapat digunakan dalam proses klasifikasi sentimen pada ulasan aplikasi DANA.

Penelitian tentang analisis sentimen aplikasi DANA dengan metode SVM telah banyak dilakukan dengan berbagai pendekatan dan karakteristik data. Berdasarkan pertimbangan tersebut, penelitian ini bertujuan untuk mengklasifikasikan ulasan ke dalam kategori "positif", "negatif", dan "netral" dengan menggunakan metode Support Vector Machine. Hasil penelitian ini diharapkan dapat memberikan wawasan tentang sentimen pengguna, yang dapat digunakan sebagai dasar untuk evaluasi.

2. KAJIAN TEORITIS

Analisis Sentimen

Analisis sentimen adalah metode pemrosesan data teks yang digunakan untuk mengidentifikasi tren dalam pendapat atau reaksi seseorang terhadap objek tertentu. Ini memungkinkan informasi dalam teks dikelompokkan menurut kategori sentimen tertentu, seperti positif, negatif, atau netral (Sari 2025).

Pada data ulasan aplikasi, analisis sentimen dapat dimanfaatkan untuk mengetahui pengalaman dan tanggapan pengguna berdasarkan komentar yang mereka berikan. Ulasan dengan sentimen positif umumnya menunjukkan kepuasan atau penilaian yang baik terhadap layanan, sedangkan sentimen negatif dapat menunjukkan ketidakpuasan maupun keluhan pengguna. Sementara itu, sentimen netral merupakan ulasan yang tidak menunjukkan kecenderungan positif ataupun negatif secara dominan (Dometian and Jatmiko 2025). Proses pengelompokan dapat dilakukan menggunakan metode *machine learning*, salah satunya adalah *Support Vector Machine (SVM)* (Willyarnandi et al. 2025).

Text Mining

Text mining adalah proses untuk mendapatkan informasi, pola, atau pengetahuan dari kumpulan data teks. Data teks pada umumnya memiliki struktur yang tidak teratur sehingga tidak dapat langsung digunakan dalam proses analisis (Alva, Zuliarso, and Stikubank 2025).

Oleh karena itu, diperlukan beberapa langkah pengolahan untuk membuat data teks lebih terorganisir dan dapat diproses dengan komputasi (Hadiani 2023).

Preprocessing Teks

Preprocessing teks merupakan tahapan awal dalam pengolahan data yang bertujuan untuk membersihkan dan mempersiapkan teks sebelum digunakan dalam proses analisis. Beberapa tahapan yang dapat digunakan dalam preprocessing meliputi *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming* (Wanny and Iqbal 2026). Seluruh huruf dilipat menjadi huruf kecil dengan case folding. Karakter yang tidak diperlukan seperti tanda baca, angka, simbol, dan URL dihilangkan dengan menggunakan pembersihan. Proses memecah teks menjadi bagian yang disebut kata atau token disebut tokenisasi. Dalam proses klasifikasi, kata-kata umum yang dianggap tidak memberikan informasi penting dihilangkan dengan *stopword removal*. *Sstemming* adalah langkah terakhir, di mana kata yang memiliki imbuhan diubah menjadi bentuk kata dasarnya (Fransiska and Gufroni 2020).

Term Frequency–Inverse Document Frequency (TF-IDF)

Term Frequency–Inverse Document Frequency (TF-IDF) adalah metode yang digunakan untuk menentukan bobot suatu kata berdasarkan frekuensinya di dalam sebuah dokumen dan di seluruh kumpulan dokumen. Metode ini mengubah data teks menjadi nilai numerik agar dapat diproses oleh algoritma pembelajaran mesin (Rifa and Mukti 2025).

TF-IDF terdiri dari dua komponen utama: *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)*. *TF* digunakan untuk menentukan frekuensi kemunculan suatu kata dalam sebuah dokumen, sedangkan *IDF* berfungsi menentukan relevansi kata tersebut berdasarkan jumlah dokumen yang memuat kata itu (Aen, Tyas, and Rakhmawati 2025).

Secara umum, perhitungan *TF-IDF* dapat dituliskan dalam persamaan berikut:

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

Keterangan:

1. $TF(t,d)$ = jumlah kemunculan kata t pada dokumen d .
2. $IDF(t)$ = nilai kepentingan kata t berdasarkan keseluruhan dokumen.
3. $TF-IDF(t,d)$ = bobot kata t pada dokumen d .

Dalam penelitian ini, *TF-IDF* digunakan untuk memberikan bobot pada setiap kata setelah tahap pra-pemrosesan. Nilai bobot yang dihasilkan kemudian digunakan sebagai fitur masukan untuk model *Support Vector Machine*.

Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang dapat digunakan untuk menyelesaikan masalah klasifikasi. Algoritma ini menentukan *hyperplane*

optimal untuk memisahkan data yang termasuk dalam kelas-kelas yang berbeda (Setyani et al. 2026). Dalam menentukan batas tersebut, *SVM* mempertimbangkan data yang memiliki posisi paling dekat dengan *hyperplane*. Data tersebut dikenal sebagai *support vector* (Rahmasari et al. 2025).

Dalam penelitian ini, algoritma *SVM* digunakan untuk mengklasifikasikan ulasan ke dalam tiga kategori sentimen: positif, negatif, dan netral.

Confusion Matrix

Salah satu cara untuk mengevaluasi kinerja model klasifikasi adalah dengan menggunakan confusion matrix, yang membandingkan kelas aktual dengan kelas hasil prediksi model. Confusion matrix memungkinkan kita untuk mengetahui berapa banyak titik data yang diklasifikasikan dengan benar dalam setiap kategori dan berapa banyak yang salah diklasifikasikan. Hasil dari confusion matrix tersebut kemudian dapat digunakan untuk menghitung jumlah titik data yang salah diklasifikasikan. (Aginta et al. 2025).

Akurasi digunakan untuk menentukan persentase titik data yang diprediksi dengan benar. Presisi menunjukkan seberapa akurat model dalam memprediksi kelas tertentu. Recall menunjukkan sejauh mana model mampu mengidentifikasi titik data yang sebenarnya termasuk dalam kelas tertentu. Di sisi lain, F1-score merupakan ukuran yang menggabungkan presisi dan recall, sehingga memberikan gambaran mengenai keseimbangan di antara keduanya. (Budianto et al. 2024). Pada penelitian ini, *confusion matrix* digunakan sebagai dasar untuk mengevaluasi hasil klasifikasi sentimen ulasan menggunakan *SVM*.

3. METODE PENELITIAN

Jenis Penelitian

Karena pemrosesan data dan penghitungan nilai numerik yang diperlukan, pendekatan kuantitatif dipilih untuk penelitian ini untuk mengklasifikasikan sentimen dalam ulasan pengguna aplikasi DANA.

Pelaksanaan penelitian dilakukan melalui beberapa tahapan yang saling berurutan. Tahapan tersebut dimulai dari penyiapan dataset, pengambilan sampel data, *preprocessing*, pembobotan *TF-IDF*, pembagian data, proses klasifikasi menggunakan *SVM*, serta evaluasi model menggunakan *confusion matrix*.

Sumber dan Dataset

Penelitian ini menggunakan data ulasan pengguna DANA dari platform Kaggle. Dataset tersebut terdiri atas 1500 ulasan dan telah dilengkapi dengan label sentimen. Beberapa atribut yang terdapat dalam dataset meliputi *UserName*, *score*, *at*, *content*, dan *sentimen*. Pada

penelitian ini, atribut *content* digunakan sebagai sumber data teks yang akan dianalisis, sedangkan atribut *sentimen* digunakan sebagai target atau label klasifikasi.

Tahapan Penelitian

Penelitian dilakukan melalui beberapa langkah yang dilakukan secara sistematis untuk mempersiapkan data. Setiap langkah dilakukan untuk menghasilkan model klasifikasi yang dapat digunakan untuk memprediksi perasaan pengguna saat melakukan ulasan.

Preprocessing

Sebelum proses pembobotan dan klasifikasi dimulai, tahap preprocessing dilakukan untuk mempersiapkan teks. Proses ini bertujuan menghilangkan bagian-bagian teks yang tidak diperlukan serta menyeragamkan bentuk penulisan pada data ulasan.

Proses *preprocessing* yang diterapkan terdiri atas beberapa proses berikut:

1. *Case Folding*

Setiap karakter huruf dalam teks diubah menjadi huruf kecil.

2. *Cleaning*

Tahap ini digunakan untuk menghilangkan karakter yang dianggap tidak penting selama proses analisis.

3. *Tokenizing*

Pada tahap ini, teks yang telah dibersihkan dibagi menjadi token.

4. *Stopword Removal*

menghapus kata-kata umum yang dianggap tidak mengandung informasi penting.

5. *Stemming*

proses mengembalikan kata yang memiliki imbuhan ke bentuk aslinya.

Data teks yang lebih bersih dan terstruktur dapat diperoleh untuk digunakan pada tahap berikutnya setelah seluruh tahapan tersebut diselesaikan.

Pembobotan *TF-IDF*

Setelah proses preprocessing, Term Frequency–Inverse Document Frequency (TF-IDF) digunakan untuk mengkonversi data menjadi format numerik. Selanjutnya, ini digunakan sebagai fitur masukan dalam proses pembuatan model Support Vector Machine.

Pembagian Data

Setelah proses pembobotan selesai, dataset dibagi menjadi dua bagian: data uji untuk mengevaluasi kemampuan model untuk melakukan prediksi pada data, dan data latih untuk melatih model.

Klasifikasi Menggunakan SVM

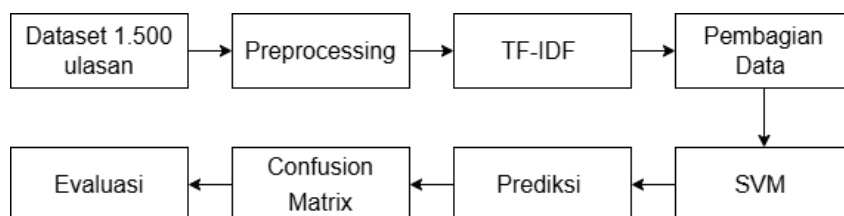
Klasifikasi dilakukan menggunakan algoritma Support Vector Machine. Data hasil pembobotan TF-IDF digunakan sebagai fitur masukan, dan kategori sentimen digunakan sebagai target yang harus dipelajari oleh model. Untuk mengetahui seberapa baik kemampuan model untuk melakukan klasifikasi, hasil prediksi dibandingkan dengan label sebenarnya.

Evaluasi Model

Dalam proses evaluasi, matrix confusion digunakan; ini berarti membandingkan hasil prediksi model dengan hasil kelas sebenarnya pada data uji. Precision, accuracy, recall, dan F1-score adalah beberapa metrik yang digunakan untuk mengukur kinerja model. Persentase keseluruhan prediksi yang benar dihitung dengan accuracy. Sementara recall mengukur kemampuan model untuk mengenali data yang sebenarnya termasuk dalam suatu kelas, precision mengukur ketepatan model dalam memprediksi kelas, dan skor F1 mengukur keseimbangan antara recall dan precision.

Teknik Analisis Data

Pengolahan dan analisis data dalam penelitian ini dilakukan menggunakan bahasa pemrograman *Python* melalui *Google Colaboratory (Google Colab)*. Seluruh proses dilakukan secara terkomputerisasi. Secara keseluruhan, tahapan penelitian dapat dirangkum dalam alur berikut:



Gambar 1. Alur pengolahan data.

Hasil dari setiap tahapan tersebut kemudian digunakan sebagai dasar untuk melakukan analisis.

4. HASIL DAN PEMBAHASAN

Dataset

Dataset yang digunakan dalam penelitian ini merupakan kumpulan ulasan pengguna aplikasi DANA yang diperoleh dari *Kaggle*. Dataset terdiri dari 1.500 ulasan dan memiliki lima atribut, yaitu *UserName*, *score*, *at*, *content*, dan *sentimen*. Dataset tersebut memiliki tiga kategori sentimen, yaitu *positive*, *negative*, dan *neutral*.

Tabel 1. Data Mentah Ulasan Pengguna.

No.	Nama Pengguna	Score	Tanggal	Ulasan	Sentimen
1	dadang s	1	31 Desember 2023, 23:23:42	Tambah buruk sekarang, loading sangat lambat, sering logout sendiri, segera diperbaiki.	NEGATIVE
2	Bayu Setiawan	3	12 Februari 2024, 05:13:38	Tolong pihak dana untuk verifikasi pake scan muka untuk bisa diakses lagi	NEUTRAL
3	Fathur Zahraaa	5	4 Februari 2024, 12:26:28	bagus banget aku suka sekali terimakasih mang dana	POSITIVE
4	Elyada Pasuang	5	20 Januari 2024, 03:33:01	puas banget dengan dana	POSITIVE
5	Egi_ "Exp"	5	28 Januari 2024, 16:13:16	Ok	POSITIVE

Tabel 2. Distribusi Dataset Penelitian.

Sentimen	Jumlah Ulasan	Persentase
Positive	797	53,13%
Negative	512	34,13%
Neutral	191	12,73%
Total	1.500	

Prediksi Sentimen

Distribusi sentimen digunakan untuk mengetahui komposisi masing-masing kelas yang terdapat pada data penelitian. Berdasarkan 1.500 ulasan yang telah dipilih, kelas *positive* memiliki jumlah paling banyak, yaitu 797 ulasan atau 53,13% dari keseluruhan data. Selanjutnya, kelas *negative* terdiri dari 512 ulasan atau 34,13%, sedangkan kelas *neutral* memiliki jumlah paling sedikit, yaitu 191 ulasan atau 12,73%.

Hasil Preprocessing

Data ulasan memerlukan preprocessing untuk mengubah teks mentah menjadi data yang lebih terstruktur. Dalam penelitian ini, prosedur preprocessing yaitu:

Case Folding

Proses mengubah seluruh huruf pada teks menjadi huruf kecil. Hasil tahap *case folding* ditunjukkan pada Tabel 3.

Tabel 3. Hasil Case Folding.

No.	Sebelum Case Folding	Sesudah Case Folding
1	Tambah buruk sekarang, loading sangat lambat, sering logout sendiri, segera diperbaiki.	tambah buruk sekarang, loading sangat lambat, sering logout sendiri, segera diperbaiki.
2	Tolong pihak dana untuk verifikasi pake scan muka untuk bisa diakses lagi	tolong pihak dana untuk verifikasi pake scan muka untuk bisa diakses lagi

3	bagus banget aku suka sekali terimakasih mang dana	bagus banget aku suka sekali terimakasih mang dana
4	puas banget dengan dana	puas banget dengan dana
5	Ok	ok

Berdasarkan Tabel 3, perubahan yang terjadi pada tahap ini terutama terlihat pada penggunaan huruf kapital.

Cleaning

Tahap ini bertujuan untuk menghilangkan tanda baca, simbol, dan karakter khusus. Hasil *cleaning* ditampilkan pada Tabel 4.

Tabel 4. Hasil *Cleaning*.

No.	Sebelum <i>Cleaning</i>	Sesudah <i>Cleaning</i>
1	tambah buruk sekarang, loading sangat lambat, sering logout sendiri, segera diperbaiki.	tambah buruk sekarang loading sangat lambat sering logout sendiri segera diperbaiki
2	tolong pihak dana untuk verivikasi pake scan muka untuk bisa diakses lagi	tolong pihak dana untuk verivikasi pake scan muka untuk bisa diakses lagi
3	bagus banget aku suka sekali terimakasih mang dana	bagus banget aku suka sekali terimakasih mang dana
4	puas banget dengan dana	puas banget dengan dana
5	ok	ok

Berdasarkan Tabel 4., tanda koma dan titik pada ulasan pertama dihilangkan sehingga teks hanya menyisakan kata-kata yang digunakan dalam proses analisis.

Tokenizing

Tahap selanjutnya adalah *tokenizing*, yaitu memecah teks menjadi token. Hasil proses *tokenizing* ditunjukkan pada Tabel 5.

Tabel 4. Hasil *Tokenizing*.

No.	Sebelum <i>Tokenizing</i>	Hasil <i>Tokenizing</i>
1	tambah buruk sekarang loading sangat lambat sering logout sendiri segera diperbaiki	[tambah, buruk, sekarang, loading, sangat, lambat, sering, logout, sendiri, segera, diperbaiki]
2	tolong pihak dana untuk verivikasi pake scan muka untuk bisa diakses lagi	[tolong, pihak, dana, untuk, verivikasi, pake, scan, muka, untuk, bisa, diakses, lagi]
3	bagus banget aku suka sekali terimakasih mang dana	[bagus, banget, aku, suka, sekali, terimakasih, mang, dana]
4	puas banget dengan dana	[puas, banget, dengan, dana]
5	ok	[ok]

Berdasarkan Tabel 5, setiap ulasan telah dipecah menjadi beberapa token.

Stopword Removal

Proses ini dilakukan dengan menghilangkan kata penghubung, kata depan, atau kata umum lainnya yang termasuk dalam daftar *stopword* bahasa Indonesia. Hasil proses tersebut ditampilkan pada Tabel 6.

Tabel 5. Hasil *Stopword Removal*

No.	Sebelum <i>Stopword Removal</i>	Sesudah <i>Stopword Removal</i>
1	tambah, buruk, sekarang, loading, sangat, lambat, sering, logout, sendiri, segera, diperbaiki	tambah, buruk, sekarang, loading, lambat, logout, segera, diperbaiki
2	tolong, pihak, dana, untuk, verifikasi, pake, scan, muka, untuk, bisa, diakses, lagi	tolong, pihak, dana, verifikasi, pake, scan, muka, diakses
3	bagus, banget, aku, suka, sekali, terimakasih, mang, dana	bagus, banget, suka, terimakasih, mang, dana
4	puas, banget, dengan, dana	puas, banget, dana
5	ok	ok

Berdasarkan Tabel 6, beberapa kata umum seperti “untuk”, “bisa”, “lagi”, “dengan”, dan “sekali” dihilangkan sesuai dengan daftar *stopword* yang digunakan.

Stemming

Proses mengubah kata yang memiliki imbuhan menjadi bentuk kata dasarnya. Hasil *stemming* ditampilkan pada Tabel 7.

Tabel 6. Hasil *Stemming*

No.	Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
1	tambah, buruk, sekarang, loading, lambat, logout, segera, diperbaiki	tambah, buruk, sekarang, loading, lambat, logout, segera, baik
2	tolong, pihak, dana, verifikasi, pake, scan, muka, diakses	tolong, pihak, dana, verifikasi, pake, scan, muka, akses
3	bagus, banget, suka, terimakasih, mang, dana	bagus, banget, suka, terimakasih, mang, dana
4	puas, banget, dana	puas, banget, dana
5	ok	ok

Berdasarkan Tabel 7, seperti kata “diperbaiki” menjadi “baik”. Setelah seluruh tahapan *preprocessing* dilakukan, data dapat digunakan untuk tahap pembobotan kata menggunakan *TF-IDF*.

Hasil Pembobotan *TF-IDF*

Data ulasan dapat diubah menjadi data numerik dengan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Setiap kata diberi nilai berdasarkan seberapa penting kata tersebut dalam dokumen secara keseluruhan.

Pembobotan *Term Frequency (TF)*

Tahap pertama adalah menghitung *Term Frequency (TF)*. Nilai *TF* menunjukkan seberapa sering suatu kata muncul dalam sebuah dokumen.

Tabel 7. Perhitungan *Term Frequency*.

Kata	Dokumen 1	Dokumen 2	Dokumen 3	Dokumen 4	Dokumen 5
dana	0	1	1	1	0
buruk	1	0	0	0	0
loading	1	0	0	0	0
lambat	1	0	0	0	0
bagus	0	0	1	0	0

puas	0	0	0	1	0
ok	0	0	0	0	1

Berdasarkan Tabel 8, nilai *TF* menunjukkan jumlah kemunculan suatu kata pada dokumen. Sebagai contoh, kata “dana” muncul satu kali pada dokumen 2, 3, dan 4, sehingga memperoleh nilai *TF* sebesar 1.

Perhitungan *Inverse Document Frequency (IDF)*

Tahap berikutnya adalah menghitung *Inverse Document Frequency (IDF)* setelah memperoleh nilai *TF*. *IDF* dihitung berdasarkan jumlah dokumen yang mengandung kata tertentu.

Tabel 8. Perhitungan *IDF*.

Kata	Jumlah Dokumen (<i>N</i>)	<i>Document Frequency (df)</i>	<i>IDF</i>
dana	5	3	0,222
buruk	5	1	0,699
loading	5	1	0,699
lambat	5	1	0,699
bagus	5	1	0,699
puas	5	1	0,699
ok	5	1	0,699

Berdasarkan Tabel 9, kata “dana” memiliki nilai *IDF* lebih rendah karena kata tersebut muncul pada tiga dokumen. Sebaliknya, kata seperti “buruk” hanya muncul pada satu dokumen sehingga memiliki nilai *IDF* lebih tinggi.

Perhitungan *TF-IDF*

Tahap terakhir adalah menggabungkan nilai *TF* dan *IDF* untuk memperoleh bobot *TF-IDF*. Sebagai contoh, kata “buruk” memiliki *TF* sebesar 1 pada dokumen 1 dan nilai *IDF* sebesar 0,699. Dengan demikian, bobot *TF-IDF* kata tersebut adalah: $TF-IDF = 1 \times 0,699 = 0,699$

Hasil pembobotan beberapa kata dari lima dokumen ditunjukkan pada Tabel 10.

Tabel 9. Hasil Pembobotan *TF-IDF*.

Kata	Dokumen 1	Dokumen 2	Dokumen 3	Dokumen 4	Dokumen 5
dana	0,000	0,222	0,222	0,222	0,000
buruk	0,699	0,000	0,000	0,000	0,000
loading	0,699	0,000	0,000	0,000	0,000
lambat	0,699	0,000	0,000	0,000	0,000
bagus	0,000	0,000	0,699	0,000	0,000
puas	0,000	0,000	0,000	0,699	0,000
ok	0,000	0,000	0,000	0,000	0,699

Berdasarkan hasil pembobotan pada Tabel 10, setiap kata yang terdapat dalam dokumen memperoleh nilai numerik yang dapat digunakan dalam proses klasifikasi.

Hasil Klasifikasi Menggunakan *Support Vector Machine*

Setelah data ulasan diubah menjadi bentuk numerik, tahap selanjutnya adalah melakukan klasifikasi *Support Vector Machine (SVM)* ke dalam tiga sentimen, yaitu *positive*, *negative*, dan *neutral*.

Pembagian Data

Sebelum proses klasifikasi dilakukan, dataset sebanyak 1.500 ulasan data dibagi dengan perbandingan 80% : 20%.

Tabel 10. Pembagian Data.

Jenis Data	Jumlah Data	Persentase
Data Latih	1.200	80%
Data Uji	300	20%
Total	1.500	100%

Berdasarkan Tabel 11, sebagian besar data digunakan untuk proses pembelajaran model, yaitu sebanyak 1.200 ulasan dan 300 ulasan digunakan sebagai data uji untuk mengukur kemampuan model.

Untuk memastikan proporsi setiap kelas sentimen pada kedua kelompok data, pembagian sampel stratifikasi digunakan. Hasil pembagian data ditunjukkan pada Gambar 4.1.

```

*** Jumlah data training : 1200
    Jumlah data testing  : 300

Distribusi sentimen data training:
sentimen
POSITIVE    638
NEGATIVE    409
NEUTRAL     153
Name: count, dtype: int64

Distribusi sentimen data testing:
sentimen
POSITIVE    159
NEGATIVE    103
NEUTRAL      38
Name: count, dtype: int64

```

Gambar 2. Output Pembagian data.

Berdasarkan Gambar 2, sebanyak 1.200 data *training* dan 300 data *testing*.

Proses Klasifikasi *SVM*

Selanjutnya, model SVM dilatih menggunakan data pelatihan, yang direpresentasikan sebagai bobot TF-IDF. SVM mencari hyperplane yang mampu memisahkan data yang termasuk dalam kelas sentimen yang berbeda. Karena tiga kelas sentimen digunakan dalam penelitian ini, hasil pelatihan SVM dibandingkan dengan label sentimen aktual untuk menentukan jumlah titik data yang diproses.

Tabel 11. Hasil Prediksi SVM.

No.	Ulasan	Sentimen Aktual	Hasil Prediksi
1	Tambah buruk sekarang, loading sangat lambat, sering logout sendiri, segera diperbaiki.	NEGATIVE	NEGATIVE
2	Tolong pihak dana untuk verifikasi pake scan muka untuk bisa diakses lagi	NEUTRAL	NEUTRAL
3	bagus banget aku suka sekali terimakasih mang dana	POSITIVE	POSITIVE
4	puas banget dengan dana	POSITIVE	POSITIVE
5	Ok	POSITIVE	POSITIVE

Berdasarkan Tabel 12, hasil prediksi menunjukkan bahwa model dapat memberikan klasifikasi yang sesuai dengan sentimen pada data tersebut.

Confusion Matrix

Kemampuan model Support Vector Machine (SVM) ditentukan dengan menggunakan matrix confusion. Label sentimen aktual dan label hasil prediksi model dibandingkan oleh matriks ini. Dengan demikian, kami dapat mengetahui berapa banyak data yang diklasifikasikan dengan benar dan berapa banyak yang mengalami kesalahan klasifikasi pada masing-masing kelas.

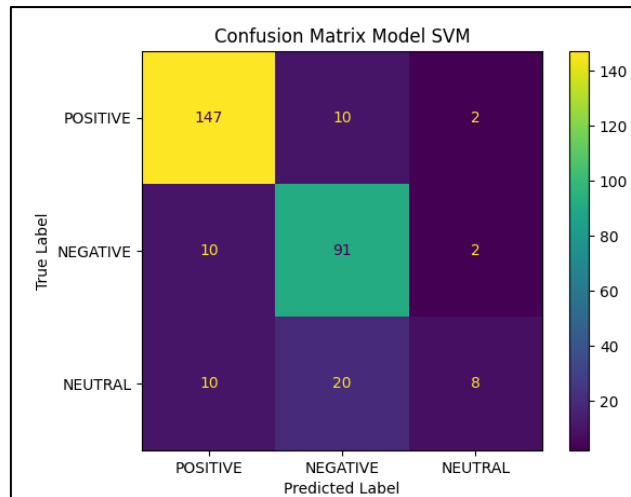
Berdasarkan hasil pengujian terhadap 300 data, diperoleh *confusion matrix* seperti pada Tabel 13.

Tabel 12. Confusion Matrix.

Aktual \ Prediksi	Positive	Negative	Neutral
Positive	147	10	2
Negative	10	91	2
Neutral	10	20	8

Berdasarkan Tabel 13, sebanyak 147 data *positive* berhasil diprediksi sebagai *positive*, 91 data *negative* berhasil diprediksi sebagai *negative*, dan 8 data *neutral* berhasil diprediksi sebagai *neutral*. Nilai-nilai tersebut berada pada diagonal utama *confusion matrix* dan menunjukkan jumlah prediksi yang benar dari masing-masing kelas.

Untuk memperjelas hasil tersebut, *confusion matrix* juga divisualisasikan dalam bentuk grafik pada Gambar 3.



Gambar 3. Visualisasi *confusion matrix*.

Berdasarkan visualisasi pada Gambar 3, nilai prediksi yang benar terlihat pada diagonal utama matriks.

Hasil Evaluasi

Proses evaluasi dilakukan untuk menentukan tingkat kinerja model SVM dalam membagi ulasan ke dalam tiga kelas sentimen. Metrik yang digunakan adalah ketepatan, precision, recall, dan F1-score. Ketepatan mengukur ketepatan model secara keseluruhan, sedangkan precision, recall, dan F1-score mengukur kemampuan model pada masing-masing kelas sentimen. Gambar 4 menunjukkan hasil laporan klasifikasi model SVM.

*** Accuracy : 82.00%				
Classification Report:				
	precision	recall	f1-score	support
POSITIVE	0.8802	0.9245	0.9018	159
NEGATIVE	0.7521	0.8835	0.8125	103
NEUTRAL	0.6667	0.2105	0.3200	38
accuracy			0.8200	300
macro avg	0.7663	0.6728	0.6781	300
weighted avg	0.8092	0.8200	0.7975	300

Gambar 4. Hasil *classification report*.

Berdasarkan Gambar 4, model *SVM* memperoleh nilai akurasi sebesar 82,00% dari 300 data *testing*. Artinya, sebanyak 246 data berhasil diklasifikasikan, sedangkan 54 data lainnya mengalami kesalahan klasifikasi. Untuk melihat performa model pada kelas secara lebih rinci, hasil disajikan pada Tabel 14.

Tabel 13. Hasil Evaluasi Model *SVM*.

Kelas Sentimen	Precision	Recall	F1-Score	Support
Positive	88,02%	92,45%	90,18%	159
Negative	75,21%	88,35%	81,25%	103
Neutral	66,67%	21,05%	32,00%	38
Accuracy			82,00%	300

Berdasarkan Tabel 14, secara keseluruhan model mampu memberikan hasil klasifikasi yang cukup baik dengan akurasi sebesar 82,00%.

Pembahasan

Hasil penelitian menunjukkan bahwa penerapan *Support Vector Machine (SVM)* dengan pembobotan *TF-IDF* mampu melakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi DANA dengan akurasi sebesar 82,00%. Nilai tersebut diperoleh dari 300 data *testing*, dengan 246 data berhasil diklasifikasikan secara tepat dan 54 data mengalami kesalahan klasifikasi. Hasil ini menunjukkan bahwa model dapat mengenali pola sentiment.

Berdasarkan hasil evaluasi, model memiliki kinerja terbaik dalam mengenali sentimen positif. Kelas tersebut memperoleh *precision* 88,02%, *recall* 92,45%, dan *F1-score* 90,18%. Sementara itu, kelas negatif memperoleh *precision* 75,21%, *recall* 88,35%, dan *F1-score* 81,25%, yang menunjukkan bahwa model juga cukup mampu mengidentifikasi sentimen negatif.

Berbeda dengan dua kelas tersebut, kelas *neutral* menunjukkan hasil yang lebih rendah. Kelas *neutral* memiliki *precision* 66,67%, *recall* 21,05% dan *F1-score* 32,00%. Berdasarkan *confusion matrix*, dari 38 data *neutral* pada data *testing*, hanya 8 data yang berhasil diprediksi sebagai *neutral*. Sebanyak 20 data *neutral* justru diprediksi sebagai *negative* dan 10 data diprediksi sebagai *positive*. Kondisi ini menunjukkan bahwa karakteristik ulasan netral lebih sulit dibedakan oleh model dibandingkan dengan ulasan positif dan negatif.

Secara keseluruhan, model *SVM* menghasilkan akurasi sebesar 82,00% dan menunjukkan kemampuan yang cukup baik dalam mengklasifikasikan sentimen *positive* dan *negative*.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, *Support Vector Machine (SVM)* dengan pembobotan *TF-IDF* mampu mengelompokkan ulasan pengguna aplikasi DANA ke dalam sentimen positif, negatif, dan netral dengan tingkat akurasi sebesar 82,00%. Hasil tersebut menunjukkan bahwa *SVM* dapat digunakan untuk mengidentifikasi kecenderungan sentimen pada ulasan pengguna. Untuk penelitian berikutnya, pengembangan dapat dilakukan dengan menambah jumlah data, meningkatkan kualitas tahap preprocessing, serta melakukan optimasi parameter *SVM*. Selain itu, pembobotan *TF-IDF* sebaiknya diterapkan setelah pembagian data latih dan data uji agar potensi data leakage dapat diminimalkan dan hasil evaluasi model menjadi lebih optimal.

DAFTAR REFERENSI

- Acuan Supian, Bagus Tri Revaldo, Nanda Marhadi, Rahmaddeni, dan Lusiana Efrizoni, and Program. 2024. "Penerapan SVM Dan Word2Vec Untuk Analisis Sentimen Ulasan Pengguna Aplikasi DANA." 23(September):321–32.
- Aen, Khurotul, Fitri Ayuning Tyas, and Hidayatur Rakhmawati. 2025. "OPTIMASI ALGORITMA SUPPORT VECTOR MACHINE (SVM) MENGGUNAKAN TEKNIK SMOTE UNTUK ANALISIS SENTIMEN PUBLIK TERHADAP DAMPAK PENGGUNAAN ARTIFICIAL INTELLIGENCE (AI)." 9(5):8287–95.
- Aginta, Yudhi, Pranata Ginting, Esterina Ndraha, Mohammad Irfan Fahmi, and Universitas Prima Indonesia. 2025. "Analisis Sentimen Ulasan Aplikasi Media Sosial Whatsapp Dan Telegram Berdasarkan Ulasan Play Store Menggunakan Algoritma Support Vector Machine." 8:172–84. doi: 10.37600/tekinkom.v8i1.1979.
- Alva, Putra, Eri Zuliarso, and Universitas Stikubank. 2025. "ANALISIS SENTIMEN BERDASARKAN ULASAN GOOGLE PLAY STORE PADA APLIKASI DANA MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) SENTIMENT." 8.
- Budianto, Akhmad Ghiffary, Arief Trisno, Eko Suryo, and Gunawan Rudi. 2024. "Perbandingan Performa Algoritma Support Vector Machine (SVM) Dan Logistic Regression Untuk Analisis Sentimen Pengguna Aplikasi Retail Di Android." 10(November):1–10. doi: 10.34128/jsi.v10i2.911.
- Dometian, Charles, and Singgih Jatmiko. 2025. "Analisis Sentimen Ulasan Google Play Store : Studi Komparatif Algoritma SVM , Naïve Bayes , Dan Logistic Regression." 15(3):610–20.
- Fransiska, Susanti, and Acep Irham Gufroni. 2020. "Sentiment Analysis Provider by . U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method." 7(2):203–12.
- Hadianti, Adinda Nazalia. 2023. "Hybrid Machine Learning for Sentiment Analysis of Dana Application Reviews." 7(2):51–58. doi: 10.21070/jicte.v7i2.1651.
- Nawulansih, Dewi Fajar, Nirma Ceisa Santi, Ita Aristia, Teknik Informatika, Universitas Nahdlatul, Ulama Sunan, Giri Bojonegoro, Sistem Informasi, Universitas Nahdlatul, Ulama Sunan, and Giri Bojonegoro. 2025. "Analisis Sentimen Ulasan Aplikasi DANA Di Google Play Store : Penerapan Support Vector Machine Dan Synthetic Minority Over-Sampling Technique Sentiment Analysis of DANA Application Reviews on Google Play Store : Implementation of Support Vector Machine and Synthetic Minority Over-Sampling Technique." 5(9):2660–71.
- Prapasta, Dentar, and Mohammad Zainuddin. 2026. "Klasifikasi Sentimen Terhadap Ulasan Pengguna Aplikasi DANA Menggunakan Metode Support Vector Machine." 5(1):89–99.
- Rahmasari, Fanny, Nining Rahaningsih, Raditya Danar Dana, and Cep Lukman Rohmat. 2025. "OPTIMASI ANALISIS SENTIMEN APLIKASI GLINTS MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)." 13(1).
- Rifa, Zanuar, and Bayu Priya Mukti. 2025. "Weakly Supervised Sentiment Analysis of Indonesian Rural Tourism Reviews : A TF-IDF Baseline for Melung Tourism Village." 12(1):48–60.

- Sari, Ariska. 2025. "Analisis Sentimen Ulasan Pengguna Aplikasi Kredivo Menggunakan Algoritma Support Vector Machine (SVM) Dengan Metode TF – IDF." 2(4):590–98.
- Setyani, Tria, Kevinda Sari, Helma Nopijani Heidy, and Ryan Randy Suryono. 2026. "Sentiment Analysis of Jamsostek Mobile Application Reviews on Google Play Store Using Support Vector Machine and Naive Bayes Algorithms Analisis Sentimen Pengguna Aplikasi Jamsostek Mobile Berdasarkan Ulasan Google Play Store Menggunakan Algoritma Support Vector Machine Dan Naive Bayes." 6(January):373–84.
- Wanny, Puspita, and Muhammad Iqbal. 2026. "Analisis Sentimen Ulasan Pengguna Aplikasi DANA Pada Google Play Store Menggunakan TF-IDF Dan Naïve Bayes Puspita." 7(1):1–12.
- Willyarnandi, Muchammad Chadavi, Lenny Margaretta Huizen, Jurusan Teknik Informatika, and Universitas Semarang. 2025. "Analisis Sentimen Terhadap Ulasan Pengguna Aplikasi Dana Di Google Playstore Menggunakan Algoritma Support Vector Machines." 8:349–57. doi: 10.37600/tekinkom.v8i1.2253.