

## Analisis Sentimen Ulasan Aplikasi DANA di Google Play Store: Penerapan Support Vector Machine dan Synthetic Minority Over-sampling Technique

Dewi Fajar Nawulansih<sup>\*1</sup>, Nirma Ceisa Santi<sup>2</sup>, Ita Aristia Sa'ida<sup>3</sup>

<sup>1,3</sup>Teknik Informatika, Universitas Nahdlatul Ulama Sunan Giri Bojonegoro, Indonesia  
<sup>2</sup>Sistem Informasi, Universitas Nahdlatul Ulama Sunan Giri Bojonegoro, Indonesia  
Email: <sup>1</sup>dewifajar237@gmail.com, <sup>2</sup>nirmaceisa@unugiri.ac.id, <sup>3</sup>itaaristia@unugiri.ac.id

### Abstrak

Analisis sentimen ulasan pengguna aplikasi DANA di Google Play Store menggunakan algoritma Support Vector Machine (SVM). Ketidakseimbangan data ditangani dengan Synthetic Minority Over-sampling Technique (SMOTE), dan optimasi parameter dilakukan melalui GridSearchCV. Sebanyak 1.000 ulasan terbaru dianalisis setelah pre-processing dan transformasi TF-IDF. Model SVM dengan kernel linear menghasilkan akurasi tertinggi sebesar 90%, meningkat dari 84% sebelum penerapan SMOTE dan tuning. Uji paired t-test terhadap hasil 10-fold cross-validation menunjukkan peningkatan yang signifikan secara statistik ( $p\text{-value} < 0,05$ ). Recall kelas negatif meningkat dari 63% menjadi 82%, sementara recall positif mencapai 94%. Word cloud menunjukkan kata "dana" paling sering muncul pada ulasan positif dan "aplikasi" pada ulasan negatif. Kombinasi metode ini meningkatkan performa klasifikasi sentimen terhadap ulasan aplikasi DANA secara signifikan.

**Kata kunci:** Analisis Sentiment, Aplikasi Dana, GridSearchCV, SMOTE, Support Vector Machine.

### *Sentiment Analysis of DANA Application Reviews on Google Play Store: Implementation of Support Vector Machine and Synthetic Minority Over-sampling Technique*

#### Abstract

Sentiment analysis of user reviews for the DANA application on the Google Play Store was conducted using the Support Vector Machine (SVM) algorithm. Data imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE), and parameter optimization was performed through GridSearchCV. A total of 1,000 recent reviews were analyzed following preprocessing and TF-IDF transformation. The SVM model with a linear kernel achieved the highest accuracy of 90%, improving from 84% before applying SMOTE and hyperparameter tuning. A paired t-test on the results of 10-fold cross-validation indicated a statistically significant performance improvement ( $p\text{-value} < 0.05$ ). The recall for the negative class increased from 63% to 82%, while the recall for the positive class reached 94%. A word cloud visualization showed that the word "dana" most frequently appeared in positive reviews, while "aplikasi" was more common in negative reviews. These findings demonstrate that the combination of SMOTE and hyperparameter tuning significantly enhances sentiment classification performance on DANA app reviews.

**Keywords:** Sentiment Analysis, Dana Application, GridSearchCV, SMOTE, Support Vector Machine.

## 1. PENDAHULUAN

Di era digital, perkembangan teknologi semakin pesat dan mengubah cara masyarakat berinteraksi, bertransaksi, serta mengelola keuangan. Inovasi digital dalam bidang keuangan, yang dikenal sebagai *Financial Technology* (Fintech), telah menghadirkan berbagai kemudahan dalam sistem pembayaran dan transaksi keuangan. Salah satu inovasi yang berkembang pesat adalah dompet digital (*e-wallet*), yang memungkinkan pengguna untuk menyimpan informasi keuangan dan melakukan transaksi secara elektronik dengan lebih efisien dan praktis [1].

Saat ini, dompet digital telah menjadi salah satu metode pembayaran yang paling banyak digunakan di Indonesia. Dengan fitur keamanan yang terus ditingkatkan, integrasi dengan berbagai layanan, serta kemampuan transaksi secara real-time, penggunaan dompet digital semakin meningkat. Salah satu aplikasi dompet digital yang populer di Indonesia adalah DANA. Sebagai layanan keuangan berbasis digital, DANA merupakan dompet digital

yang populer dan banyak digunakan oleh masyarakat Indonesia [2]. DANA juga merupakan hasil ciptaan anak bangsa [3]. DANA dapat diunduh melalui Google Play Store atau App Store yang tersedia di *smartphone*. DANA menyediakan fitur unggulan yang memungkinkan pengguna untuk menyimpan kartu bank, baik debit maupun kredit, sehingga memudahkan dalam melakukan transaksi. Fitur ini tidak tersedia di aplikasi *fintech* pesaing seperti OVO dan Go-Pay [4]. Meskipun DANA memiliki banyak pengguna, keberhasilan aplikasi ini tidak hanya ditentukan oleh fitur-fiturnya, tetapi juga oleh *feedback* dan ulasan pengguna di platform seperti Google Play Store. Ulasan tersebut mencerminkan pengalaman pengguna dan menjadi indikator kualitas layanan. Oleh karena itu, analisis sentimen sangat penting untuk memahami opini pengguna terhadap aplikasi ini.

Analisis sentimen adalah teknik dalam pemrosesan bahasa alami (*Natural Language Processing/NLP*) yang bertujuan untuk mengelompokkan opini pengguna menjadi sentimen positif, negatif, atau netral. Dengan analisis ini, pengembang aplikasi dapat memahami kekuatan dan kelemahan aplikasinya, serta mengambil keputusan strategis untuk meningkatkan kualitas layanan [5]. Analisis sentimen merupakan penerapan teknik komputasi dan analisis teks untuk mengelola, memahami, dan mengkategorikan berbagai emosi yang terkandung di dalamnya [6]. Sejak awal tahun 2000-an, analisis sentimen telah menjadi area penelitian yang sangat aktif dalam pemrosesan bahasa alami. Analisis ini juga penting karena dapat mengidentifikasi pola dalam ulasan, memungkinkan pengembang membuat keputusan yang lebih tepat dalam meningkatkan kualitas aplikasi sesuai kebutuhan pengguna [1].

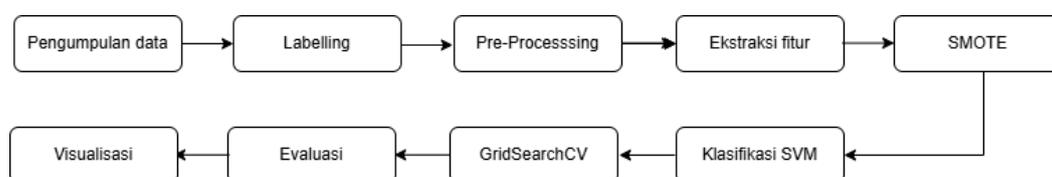
Berbagai penelitian sebelumnya telah menunjukkan efektivitas metode *Support Vector Machine (SVM)* dalam analisis sentimen. Misalnya, studi klasifikasi ulasan aplikasi Ruangguru dengan SVM mencapai akurasi hingga 90.2% dengan kernel Linear dan K-Fold Cross-Validation [7]. Penelitian lain juga berhasil mencapai akurasi 98% menggunakan SVM dan TF-IDF untuk analisis sentimen ulasan Shopee [8], serta menunjukkan efektivitas SVM dalam analisis sentimen pada ulasan Sayurbox [9] dan komentar berita Detik.com [10]. Khusus untuk aplikasi DANA, analisis sentimen pengguna dengan SVM telah dilakukan, menghasilkan akurasi 80% dengan mayoritas ulasan negatif [11]. Penelitian lain juga menunjukkan bahwa penggunaan seleksi fitur dengan SVM dapat meningkatkan akurasi hingga 89.41% untuk ulasan DANA [12]. Namun, penelitian tersebut menggunakan data yang relatif lama dan belum menggabungkan teknik *oversampling* untuk menangani ketidakseimbangan kelas serta tuning hyperparameter secara optimal. Penelitian ini menghadirkan pembaruan dengan menggunakan data ulasan terbaru, menggabungkan teknik *Synthetic Minority Oversampling Technique (SMOTE)* dengan GridSearchCV untuk optimasi parameter SVM, serta melakukan validasi statistik yang lebih komprehensif.

Dalam analisis sentimen, sering kali ditemukan masalah ketidakseimbangan jumlah ulasan positif dan negatif, yang dapat menyebabkan bias dan mengurangi akurasi model dalam mendeteksi sentimen minoritas. Untuk mengatasi ini, *Synthetic Minority Oversampling Technique (SMOTE)*, sebuah teknik *oversampling* yang menghasilkan data sintesis untuk kelas minoritas, dapat diterapkan untuk menyeimbangkan distribusi data. Penerapan SMOTE telah terbukti meningkatkan kinerja klasifikasi. Meskipun penelitian ini berfokus pada analisis sentimen, efektivitas SMOTE dalam menangani ketidakseimbangan data juga telah ditunjukkan dalam domain lain, peningkatan akurasi dengan SMOTE dan SVM dalam deteksi diabetes melitus tipe 2 [13]. Lebih lanjut, perbandingan efektivitas SVM dengan dan tanpa SMOTE dalam menganalisis opini terhadap Mixue menunjukkan bahwa SMOTE dapat meningkatkan akurasi dan presisi [14]. Secara lebih komprehensif, sebuah studi internasional membahas peningkatan performa SVM dengan SMOTE pada dataset tidak seimbang, di mana kombinasi SVM dan SMOTE mencapai akurasi tertinggi 93.25% [15].

Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi DANA berdasarkan ulasan terbaru dari Google Play Store dengan menggunakan metode *Support Vector Machine (SVM)* yang dioptimasi melalui penerapan *Synthetic Minority Oversampling Technique (SMOTE)* dan GridSearchCV.

## 2. METODE PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan utama: pengumpulan data, *labelling*, *pre-processing*, ekstraksi fitur, penanganan ketidakseimbangan data menggunakan SMOTE, pelatihan model SVM dengan GridSearchCV, evaluasi model, dan visualisasi hasil.



Gambar 1. Metode Penelitian

### 2.1. Pengumpulan Data

Data ulasan pengguna aplikasi DANA dikumpulkan dari Google Play Store pada tanggal 3 Maret 2025. Proses pengambilan data dilakukan menggunakan teknik *web scraping* dengan bantuan *library* google-play-scraper pada lingkungan pemrograman *Python*. Sebanyak 1.000 ulasan terbaru berhasil dikumpulkan, yang mencakup teks ulasan dan rating (bintang) dari pengguna.

Proses pengambilan data memastikan data yang diambil relevan dan terkini dengan menggunakan parameter spesifik dalam skrip *scraping*:

1. ID Aplikasi: Menggunakan ID aplikasi "id.dana" yang merupakan identifikasi unik untuk aplikasi DANA di Google Play Store.
2. Bahasa: Data yang diambil hanya yang berbahasa Indonesia (*lang='id'*), untuk menyesuaikan dengan populasi pengguna yang relevan dengan konteks penelitian ini.
3. Negara: Ulasan yang diambil hanya dari pengguna yang berada di Indonesia (*country='id'*), sehingga memastikan relevansi geografis data.
4. Jumlah Data: Sebanyak 1.000 ulasan terbaru (*count=1000*) berhasil dikumpulkan. Jumlah data ini dipilih untuk memastikan representasi yang memadai dan sejalan [16].

id	A	B	C
1	content	score	
2	sangat mudah untuk di gunakan	5	
3	APALAH GAJELAS MASA SAYA MAU TF KE QRIS GABISA BACAANNYA KENDALA TERUS MALAH UANG DI DANA SEMUA HADEH TOLONG SISTEMNYA DI PERBAIKIN LAH KOCAK	1	
4	Aplikasi yg sangat bermanfaat 9Y*	5	
5	langsung puse	5	
6	Good	4	
7	good peri good	5	
8	mantap	5	
9	sangat membantu banget	5	
10	baik	5	
11	Ok	5	
12	Terima kasih dana saya ampir 3 tahun menggunakan apk dana semua data * dana saldo kita terjamin terawasi nya TQ	5	
13	1 aplikasi banyak kegunaan nya dan cara pakainya juga simple dan yang paling memuaskan yaitu prosesnya cepat dan mudah thank's0Y~0Y"	5	
14	transaksi cepat dan aman	3	
15	lajutamuliah ini sangat membantu	5	
16	Min dah 2x TF pake Qris kok error min pertama 50 RB uang ga masuk juga ga balik ke dana sy trs sekarang TF lagi 100 error lagi tolong bantu proses min diana	4	
17	baik	5	
18	ok	5	
19	simpel dan aman	5	
20	keven	5	
21	Saldo dana saya tau sudah ada pembayaran penarikan dana padahal saya tidak melakukan pembayaran menggunakan QRIS, parah banget ini Apk dana sudah nyalong uang sai	1	
22	PELAYANAN BUSUK PENGADUAN TIDAK DI RESPON, SALDO TRANSFER BERHASIL TAPI SALDO TIDAK KUNJUNG MASUK!!	1	
23	udalah bentar lagi pada uninstal nih apk ga jelas, tiap bayar qris gangguan terus padahal pakai GOPAY lancar. semakin diupdate malah semakin kacau	5	
24	skrg aga lambat, dan pernah terjadi lg ckout pembayaran melalui dana malah byar dtemp. tolong dperbaiki ke amanannya biar nymaan	5	

Gambar 2. Cuplikan dataset ulasan aplikasi dana

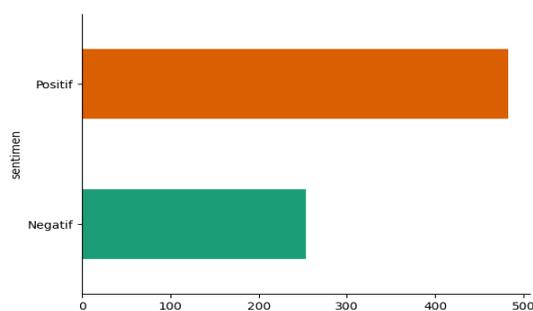
Gambar 2 menampilkan cuplikan dataset ulasan pengguna aplikasi DANA yang dikumpulkan dari Google Play Store. Dataset ini berisi dua atribut utama yaitu *content* dan *score*. Kolom *content* memuat isi ulasan dalam bentuk teks yang mencerminkan pendapat dan pengalaman pengguna terhadap aplikasi DANA. Sementara itu, kolom *score* berisi nilai *rating* dalam skala 1 hingga 5 yang diberikan oleh pengguna untuk menilai aplikasi tersebut.

### 2.2. Pelabelan (Labelling Data)

Pelabelan merupakan proses pelabelan dilakukan untuk menentukan kategori sentimen setiap ulasan. Pelabelan dilakukan secara manual berdasarkan aturan berikut:

- a. Sentimen positif : Ulasan dengan *rating* 4 dan 5 akan dikategorikan sebagai sentimen Positif.
- b. Sentimen negatif : Ulasan dengan *rating* 1 dan 2 akan dikategorikan sebagai sentimen Negatif.

namun, untuk fokus pada klasifikasi biner (positif/negatif), ulasan dengan *rating* 3 umumnya dapat dihilangkan atau tidak digunakan dalam dataset untuk pelatihan model dan dalam konteks penelitian ini yang berfokus pada dua kelas sentimen utama, ulasan dengan *rating* 3 akan dihapus dari dataset agar tidak mengganggu proses klasifikasi biner.



Gambar 3. Distribusi Sentimen

Berdasarkan Gambar 3 menunjukkan distribusi sentimen ulasan pengguna aplikasi DANA setelah proses pelabelan. Terlihat jelas adanya ketidakseimbangan (*imbalanced data*) dalam distribusi sentimen, dengan dominasi jumlah ulasan positif. Kondisi ini menjadi dasar perlunya penanganan ketidakseimbangan data dalam tahapan selanjutnya,

### 2.3. Pre-Processing Data

Tahap *pre-processing* data dilakukan untuk membersihkan dan mempersiapkan teks ulasan sebelum analisis. Tahapan ini meliputi:

1. Penghapusan Data Duplikat: Mengidentifikasi dan menghapus ulasan yang sama persis (duplikat) untuk menghindari bias dalam analisis.
2. *Case Folding*: Mengubah semua teks menjadi huruf kecil, menghapus karakter khusus yang digunakan oleh pengguna dalam ulasan Aplikasi DANA seperti, URL, tanda baca dan *emoticon*.
3. *Normalizing*: Mengonversi kata tidak baku atau bahasa slang ke dalam bentuk standar.
4. *Stopword Removal*: Menghapus kata-kata umum yang kurang memiliki makna signifikan (misalnya "yang", "dan", "di", dll.).
5. *Tokenizing*: Memisahkan teks menjadi unit kata individu.
6. *Stemming*: Mengubah kata menjadi bentuk dasar menggunakan algoritma *Porter Stemmer* atau *Sastrawi*.

### 2.4. Ekstraksi Fitur

Setelah *pre-processing*, teks ulasan diubah menjadi representasi numerik yang dapat dipahami oleh algoritma *machine learning*. Metode yang digunakan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF menghitung bobot kata berdasarkan frekuensi kemunculannya dalam sebuah dokumen dan invers frekuensinya dalam seluruh koleksi dokumen, sehingga kata-kata yang unik dan informatif mendapatkan bobot lebih tinggi.

$$TF - IDF(t, d) = tf(t, d) \times \log\left(\frac{N}{df(t)}\right) \quad (1)$$

Keterangan:

- $tf(t, d)$ : frekuensi kata  $t$  dalam dokumen  $d$ ,
- $df(t)$ : jumlah dokumen yang mengandung kata  $t$ ,
- $N$ : total jumlah dokumen.

### 2.5. Imbalance Data (SMOTE)

*Synthetic Minority Oversampling Technique* (SMOTE) adalah metode yang digunakan untuk menangani masalah ketidakseimbangan data (*imbalanced data*) dalam klasifikasi. Ketidakseimbangan ini terjadi ketika jumlah sampel pada salah satu kelas (biasanya kelas minoritas) jauh lebih sedikit dibandingkan dengan kelas lainnya, sehingga model cenderung bias terhadap kelas mayoritas dan mengabaikan informasi penting dari kelas minoritas [17].

SMOTE bekerja dengan cara menghasilkan data sintetis dari kelas minoritas, bukan hanya melakukan duplikasi. Algoritma ini mencari Tetangga terdekat dari setiap sampel minoritas, lalu membuat sampel baru di antara titik tersebut dalam ruang fitur. Teknik ini membantu menciptakan distribusi kelas yang lebih seimbang, meningkatkan kemampuan generalisasi model, serta mengurangi bias dalam proses pelatihan.

Dalam konteks penelitian ini, distribusi ulasan pengguna aplikasi DANA juga menunjukkan ketidakseimbangan—jumlah ulasan positif jauh lebih banyak dibandingkan ulasan negatif. Oleh karena itu, SMOTE digunakan untuk menyeimbangkan jumlah data antar kelas sebelum proses klasifikasi dilakukan.

### 2.6. Klasifikasi Support Vector Machine (SVM)

*Support Vector Machine* (SVM) adalah algoritma pembelajaran mesin yang efektif untuk tugas klasifikasi dan regresi. Prinsip dasar SVM adalah mencari *hyperplane* optimal yang mampu memisahkan kelas-kelas data dalam ruang berdimensi tinggi. *Hyperplane* ini dipilih sedemikian rupa sehingga jarak (*margin*) antara *hyperplane* dengan titik data terdekat dari setiap kelas (disebut *support vectors*) adalah maksimal. Pemilihan *hyperplane* dengan *margin* terbesar ini bertujuan untuk mencapai generalisasi yang baik pada data baru [18].

Salah satu komponen kunci dalam SVM adalah fungsi *kernel*. Fungsi *kernel* memungkinkan SVM untuk menangani data yang tidak dapat dipisahkan secara linear di ruang input dengan memetakan data tersebut ke ruang

fitur berdimensi lebih tinggi. Di ruang fitur yang baru ini, data mungkin menjadi dapat dipisahkan secara linear. Beberapa jenis *kernel* yang umum digunakan dalam SVM meliputi: *Kernel Linier*, *Radial Basis Function* (RBF), *Sigmoid* dan *Polynomial*. Pada tahap awal penelitian ini, dilakukan pelatihan model SVM dengan menguji berbagai kernel yang disebutkan di atas, yaitu *linear*, RBF, *Sigmoid* dan *polynomial*. Masing-masing kernel diuji menggunakan parameter default tanpa tuning lebih lanjut. Tujuan dari eksperimen awal ini adalah untuk memperoleh gambaran performa dasar dari masing-masing kernel dalam mengklasifikasikan sentimen ulasan, sebelum dilakukan optimasi parameter secara lebih mendalam.

## 2.7. Tuning Hyperparameter dengan GridSearchCV

Eksperimen awal dengan *kernel* SVM menggunakan parameter *default*, tahap selanjutnya adalah mengoptimalkan kinerja model melalui *tuning hyperparameter*. Metode yang digunakan untuk proses optimasi ini adalah *GridSearchCV*.

*Grid Search* adalah metode pencarian sistematis yang efektif untuk menemukan kombinasi *hyperparameter* terbaik dari suatu algoritma pembelajaran mesin [19]. Teknik ini bekerja dengan melakukan eksplorasi menyeluruh terhadap ruang parameter yang telah ditentukan. Setiap kombinasi parameter akan dievaluasi secara individual menggunakan proses pelatihan dan pengujian berulang melalui teknik validasi silang (*cross-validation*), sehingga model yang dihasilkan menjadi lebih andal dan mampu menghindari *overfitting*. Dalam konteks klasifikasi menggunakan algoritma *Support Vector Machine* (SVM), *Grid Search* sangat bermanfaat untuk mengoptimalkan performa model. Parameter penting yang umum ditentukan melalui *Grid Search* antara lain:

- C* (*cost*): Parameter regulasi yang menentukan seberapa besar penalti terhadap kesalahan klasifikasi. Nilai *C* yang besar dapat menghasilkan model yang mencoba mengklasifikasikan semua data dengan benar, tetapi berisiko *overfitting*.
- $\gamma$  (*gamma*): Parameter pada *kernel* non-linear (seperti RBF) yang mengatur seberapa jauh pengaruh dari satu data *training* terhadap data lainnya. Nilai *gamma* yang terlalu besar dapat menyebabkan model terlalu fokus pada data lokal dan kehilangan generalisasi.
- Kernel*: Fungsi *kernel* yang digunakan untuk mentransformasikan data ke dalam ruang berdimensi lebih tinggi agar lebih mudah dipisahkan. Jenis *kernel* umum mencakup *linear*, *polynomial*, RBF, dan *sigmoid*.

Penelitian sebelumnya telah menunjukkan efektivitas *Grid Search* dalam meningkatkan akurasi model. Peningkatan akurasi model klasifikasi penyakit jantung dari 87,65% menjadi 96,56% setelah kombinasi *Grid Search* dan seleksi fitur pada SVM [19]. Hasil serupa penerapan *Grid Search* pada model *Bagging-SVM* untuk klasifikasi penyakit diabetes, berhasil meningkatkan akurasi model hingga 92,1% dan meningkatkan presisi serta *recall* pada kelas minoritas [20].

Dalam penelitian ini, *GridSearchCV* mengevaluasi berbagai kombinasi *kernel*, nilai parameter *C*, dan parameter  $\gamma$  (untuk *kernel* non-linear) secara menyeluruh. Hasil *tuning* menunjukkan bahwa *kernel linear* dengan parameter optimal memberikan performa terbaik. Oleh karena itu, model akhir yang digunakan adalah SVM dengan *kernel linear* dan parameter yang telah dioptimalkan melalui *GridSearchCV*. Fungsi keputusan model SVM dengan kernel linear dapat dituliskan sebagai berikut:

$$f(x) = \text{sign}(w^T x + b) \quad (2)$$

di mana *w* adalah bobot model dan *b* adalah bias.

## 2.8. Evaluasi Model

Evaluasi kinerja model klasifikasi dilakukan menggunakan metrik-metrik berikut:

- Akurasi : Mengukur proporsi prediksi yang benar terhadap total prediksi.
  - Presisi : Mengukur akurasi prediksi positif, yaitu seberapa banyak data yang diklasifikasikan positif benar-benar positif.
  - Recall* : Mengukur kemampuan model dalam menemukan seluruh data positif yang benar.
  - F1-Score*: Rata-rata harmonis dari presisi dan *recall*, berguna saat terdapat ketidakseimbangan kelas data.
- Selain itu, digunakan *confusion matrix* untuk menganalisis kesalahan klasifikasi antara kelas positif dan negatif. Untuk mengukur pengaruh teknik augmentasi data dan *tuning hyperparameter*, dilakukan uji statistik paired t-test guna melihat signifikansi peningkatan akurasi setelah penerapan SMOTE (*Synthetic Minority Over-sampling Technique*) dan *GridSearchCV*.

### 2.9. Visualisasi Data (Word Cloud)

Wawasan tambahan mengenai kata-kata dominan dalam ulasan pengguna, dilakukan visualisasi dengan word cloud. Visualisasi ini membantu mengidentifikasi pola kata kunci yang sering muncul pada setiap kategori sentimen, baik positif maupun negatif. Dengan word cloud, peneliti dapat secara intuitif melihat representasi teks yang paling menonjol dan relevan dalam dataset.

### 3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari setiap tahapan penelitian, dimulai dari *pre-processing* data, penanganan *imbalance data* dengan SMOTE, hasil klasifikasi model SVM, evaluasi kinerja model, hingga visualisasi *word cloud*.

#### 3.1. Hasil Pre-processing

Tahap *pre-processing* dilakukan untuk membersihkan dan merapikan data ulasan yang telah dikumpulkan. Awalnya, terdapat 1.000 data ulasan yang dikumpulkan. Setelah tahap penghapusan duplikat, data berkurang menjadi 737 ulasan unik. Selanjutnya, proses *case folding* dilakukan, di mana semua teks diubah menjadi huruf kecil. Kemudian, proses *tokenizing* memisahkan kalimat menjadi kata-kata individual. Selanjutnya, *filtering* menghapus *stopwords* (kata-kata umum yang tidak relevan untuk analisis sentimen) dan karakter yang tidak dibutuhkan. Terakhir, *stemming* mengubah kata-kata menjadi bentuk dasarnya.

Tabel 1. Hasil *Pre-Processing*

Tahap <i>Pre-Processing</i>	Text
Text Asli	APALAH GAJELAS MASA SAYA MAU TF KE QRIS GABISA BACAANNYA KENDALA TERUS MALAH UANG DI DANA SEMUA HADEH TOLONG SISTEMNYA DI PERBAIKIN LAH KOCAK
<i>Case Foolding</i>	apalah gajelas masa saya mau tf ke qris gabisa bacaannya kendala terus malah uang di dana semua hadeh tolong sistemnya di perbaiki lah kocak
<i>Normalizing</i>	apalah masa saya mau transfer ke qris bacaannya kendala terus malah uang di dana semua hadeh tolong sistemnya di perbaiki lah kocak
<i>Stopword Removal</i>	apalah transfer qris bacaannya kendala uang dana hadeh tolong sistemnya perbaiki kocak
<i>Tokenizing</i>	['apalah', 'transfer', 'qris', 'bacaannya', 'kendala', 'uang', 'dana', 'hadeh', 'tolong', 'sistemnya', 'perbaiki', 'kocak']
<i>Stemming</i>	apa transfer qris baca kendala uang dana hadeh tolong sistem baik kocak

#### 3.2. Ekstraksi Fitur

Setelah tahap *pre-processing*, ulasan yang awalnya tidak terstruktur diubah menjadi format yang bersih dan siap dianalisis. Proses *stemming* dan *stopword removal* secara signifikan mengurangi jumlah kata unik, sementara TF-IDF mengubah teks menjadi vektor numerik.

Tabel 2. Hasil Transformasi TF-IDF

No	Term	TF-IDF
1	dana	53.142696
2	bagus	47.141800
3	bantu	40.160152
4	aplikasi	33.071627
5	mudah	28.377232
6	mantap	28.036142
	...	...
1049	dm	0.181835
1050	malam	0.181835
1051	rasai	0.181835

1052	nenanganinya	0.186463
1053	gws	0.190582
1054	padaahal	0.190582
1055	belah	0.190582
1056	biae	0.190789
1057	urung	0.190789

Berdasarkan tabel 2 Hasil transformasi TF-IDF menghasilkan sebuah matriks berukuran  $737 \times 1057$ , yang menunjukkan bahwa terdapat 737 dokumen (ulasan) yang dianalisis dan 1.057 kata unik yang diidentifikasi sebagai fitur. Setiap baris dalam matriks merepresentasikan satu dokumen, sedangkan setiap kolom menunjukkan satu kata unik dari seluruh korpus. Nilai-nilai dalam matriks tersebut merupakan skor TF-IDF, yaitu bobot numerik yang menunjukkan tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap seluruh koleksi dokumen. Semakin tinggi nilai TF-IDF suatu kata, semakin penting kata tersebut dalam konteks dokumen tertentu, dan semakin jarang kemunculannya di dokumen lain. Matriks ini menjadi representasi numerik dari teks yang kemudian digunakan sebagai masukan untuk proses pelatihan model klasifikasi sentimen.

### 3.3. Imbalance Data dengan SMOTE

Dataset awal menunjukkan ketidakseimbangan yang signifikan antara kelas sentimen. Untuk mempersiapkan data ini untuk pelatihan model, dataset dibagi menjadi data pelatihan dan data pengujian dengan rasio 80:20 sehingga memperoleh 589 data *training* dan 148 data *testing*. Tabel 8 menunjukkan perbandingan jumlah sampel per kelas sebelum dan sesudah SMOTE.

Tabel 3. Perbandingan Jumlah Sampel Sebelum dan Sesudah SMOTE

Sentimen	Sampel Awal	Sampel Setelah SMOTE
Positif	386	386
Negatif	205	386
Total	589	772

Berdasarkan tabel 3 diatas, penerapan SMOTE berhasil meningkatkan jumlah sampel pada kelas minoritas (negatif) dari 205 menjadi 386, sehingga jumlah sampel di kedua kelas (positif dan negatif) menjadi seimbang. Kondisi ini sangat penting untuk menghindari bias pada model klasifikasi, karena model yang dilatih dengan data yang seimbang cenderung memiliki performa yang lebih baik dalam mengenali kedua kelas secara adil.

### 3.4. Evaluasi Model SVM dan Optimasi Parameter

Model *Support Vector Machine* (SVM) diuji dengan berbagai jenis kernel, yaitu RBF, polynomial, dan linear, untuk mengevaluasi performa dasar masing-masing dalam mengklasifikasikan sentimen ulasan pengguna aplikasi DANA. Pada tahap awal, model dijalankan dengan parameter default, termasuk nilai  $C = 10$ . Hasil evaluasi menunjukkan bahwa kernel RBF mencatat akurasi sedikit lebih tinggi dibandingkan kernel linear, dengan selisih sekitar 1%, diikuti oleh kernel sigmoid, sementara kernel polynomial menunjukkan performa yang lebih rendah.

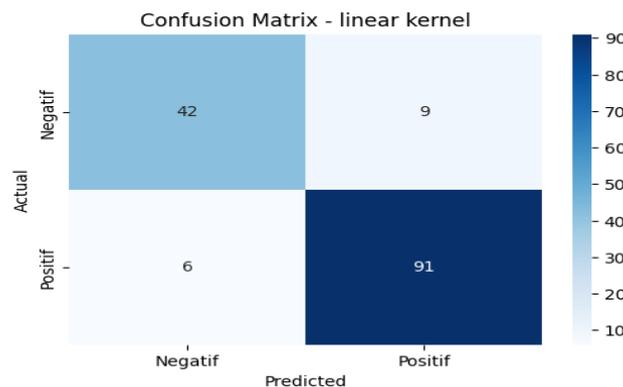
Selanjutnya, dilakukan proses tuning hyperparameter menggunakan GridSearchCV untuk mencari kombinasi parameter terbaik. Proses ini mencakup eksplorasi nilai-nilai parameter seperti C, gamma, dan pilihan kernel. Hasil tuning menunjukkan bahwa kernel linear dengan parameter  $C = 1$  memberikan performa paling optimal, baik dari segi akurasi keseluruhan maupun keseimbangan antar metrik evaluasi seperti precision, recall, dan F1-score. Berdasarkan hasil tersebut, model akhir yang digunakan dalam penelitian ini adalah SVM dengan kernel linear dan parameter  $C = 1$ , karena menunjukkan hasil terbaik secara konsisten setelah penyeimbangan data dengan SMOTE dan optimasi parameter.

Tabel 4. Perbandingan model sebelum dan sesudah SMOTE & GridSearchCV

Metrik	Sebelum	Sesudah
Presisi (Negatif)	0.86	0.88
Presisi (Positif)	0.83	0.91
Recall (Negatif)	0.63	0.82
Recall (Positif)	0.95	0.94
F1-Score (Negatif)	0.73	0.85

F1-Score (Positif)	0.88	0.92
Akurasi	0.84	0.90

Dari Tabel 4 terlihat bahwa penggunaan *kernel linear* menghasilkan akurasi keseluruhan tertinggi setelah penerapan SMOTE dan *GridSearchCV*, yaitu mencapai 90%. Peningkatan akurasi sebesar 6% dari nilai awal 84% menunjukkan efektivitas kombinasi metode ini dalam meningkatkan performa klasifikasi. Secara spesifik, peningkatan signifikan juga diamati pada metrik *recall* kelas negatif (dari 0.63 menjadi 0.82), yang mengindikasikan kemampuan model yang jauh lebih baik dalam mengidentifikasi ulasan negatif setelah penyeimbangan data menggunakan SMOTE. Hal ini penting karena ulasan negatif kerap menjadi minoritas namun sangat kritis dalam analisis layanan aplikasi.



Gambar 4. *Confusion Matrix*

Pada gambar 4 *Confusion Matrix* di atas, yang dihasilkan dari model SVM dengan *linear kernel* dari 148 data testing, secara jelas menunjukkan performa klasifikasi. Model berhasil mengidentifikasi 42 ulasan negatif dengan benar (*True Negatif*) dan 91 ulasan positif dengan benar (*True Positif*). Namun, terdapat 9 kasus di mana model salah memprediksi ulasan negatif sebagai positif (*False Positif*) dan 6 kasus di mana ulasan positif salah diprediksi sebagai negatif (*False Negatif*), menunjukkan kinerja yang solid namun dengan sedikit area untuk perbaikan pada kesalahan klasifikasi dari total data pengujian.

Tabel 5. Hasil Validasi *K-fold Validation*

Sebelum	0.81	0.83	0.75	0.86	0.79	0.86	0.85	0.79	0.76	0.82
Sesudah	0.82	0.89	0.85	0.87	0.83	0.87	0.86	0.80	0.80	0.86

Rata-rata akurasi dari validasi 10-fold sebelum optimasi tercatat sebesar 81,67%, sedangkan setelah penerapan SMOTE dan tuning parameter meningkat menjadi 85,06%, menegaskan peningkatan stabilitas dan generalisasi model.

### 3.5. Uji Signifikansi (*Paired T-test*)

Untuk menguji signifikansi peningkatan performa model tersebut, dilakukan uji *paired t-test* terhadap akurasi dari 10 *fold cross-validation*. Uji ini bertujuan untuk mengetahui apakah peningkatan akurasi yang diamati memiliki perbedaan yang signifikan secara statistik.

Tabel 6. Hasil Uji *Paired T-test*

<i>T-Statistic</i>	-4.051
<i>P-Value</i>	0.0029

Berdasarkan Tabel 6, hasil uji paired t-test terhadap akurasi dari proses 10-fold *cross-validation* sebelum dan sesudah penerapan teknik SMOTE serta *tuning hyperparameter* menunjukkan nilai *t-statistic* sebesar -4.051 dan p-value sebesar 0.0029 ( $p < 0.05$ ). Nilai p-value yang jauh lebih kecil dari tingkat signifikansi 0.05 mengindikasikan terdapat perbedaan performa yang signifikan secara statistik antara model sebelum dan sesudah optimasi. Dengan demikian, dapat disimpulkan bahwa kombinasi penerapan SMOTE dan *tuning hyperparameter* memberikan kontribusi nyata dalam meningkatkan akurasi klasifikasi sentimen terhadap ulasan pengguna aplikasi DANA.

### 3.6. Visualisasi Data

Visualisasi data dengan *word cloud* dilakukan untuk melihat kata-kata yang paling sering muncul dalam ulasan positif dan negatif setelah proses *pre-processing*.



Gambar 5. *Word Cloud* sebelum

Pada Gambar 5 terlihat bahwa kata-kata umum seperti “sangat”, “dan”, “saya”, “di”, “tidak”, dan “ada” muncul sangat sering. Kata-kata ini termasuk dalam kategori *stopwords*, yang umumnya tidak memberikan makna signifikan terhadap analisis sentimen, sehingga perlu dihapus dalam tahap *pre-processing*.



Gambar 6. *Word Cloud* sesudah

Pada gambar 6 Setelah proses *pre-processing* dan pemisahan berdasarkan label sentimen, word cloud memperlihatkan kata-kata yang dominan dalam setiap kategori:

- 1) Pada ulasan positif, kata-kata yang sering muncul mencerminkan kepuasan pengguna terhadap layanan dan kemudahan aplikasi. Kata-kata seperti “dana”, “bantu”, “bagus”, “mudah”, dan “aman” menandakan pengalaman yang baik dan persepsi positif terhadap fitur aplikasi.
- 2) Sementara itu, pada ulasan negatif, kata-kata yang muncul lebih banyak terkait masalah keuangan dan teknis, seperti “saldo”, “uang”, “hilang”, dan “akun”. Hal ini menunjukkan bahwa keluhan utama pengguna berfokus pada kehilangan dana, kesulitan akses akun, atau masalah transaksi.

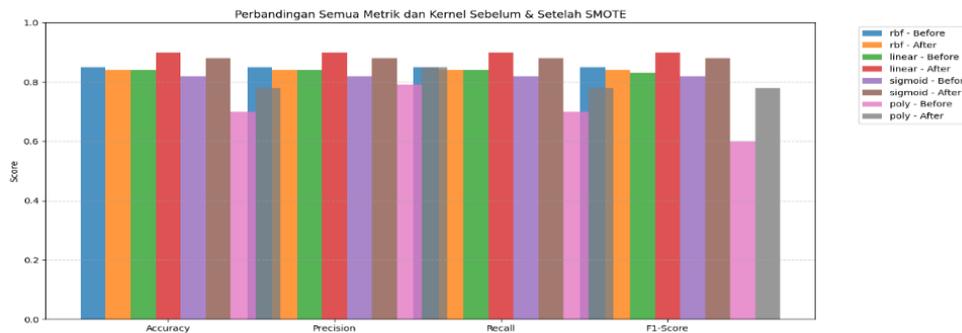
Selain itu, kata yang paling sering muncul pada:

- 1) Ulasan positif adalah “dana” dengan frekuensi kemunculan sebanyak 74 kali.
- 2) Ulasan negatif adalah “aplikasi” dengan frekuensi kemunculan sebanyak 104 kali.

Visualisasi ini membantu peneliti dalam memahami tema dominan dari masing-masing sentimen dan dapat menjadi dasar rekomendasi pengembangan aplikasi.

### 3.7. Diskusi

Hasil evaluasi model *Support Vector Machine* (SVM) dengan kernel linear menunjukkan peningkatan performa signifikan setelah penerapan *Synthetic Minority Over-sampling Technique* (SMOTE) dan *hyperparameter tuning* menggunakan GridSearchCV. Model yang dievaluasi dengan *cross-validation* (CV) memperoleh akurasi rata-rata 85%. Pada pengujian data terpisah, akurasi meningkat menjadi 90%, menandakan kemampuan model untuk generalisasi dengan baik tanpa risiko *overfitting* yang signifikan.



Gambar 7. Perbandingan metrik

Pada Gambar 7 ditunjukkan Peningkatan performa model SVM dengan kernel linear setelah penerapan SMOTE dan GridSearchCV menunjukkan efektivitas kombinasi teknik balancing dan optimasi parameter dalam mengatasi ketidakseimbangan data. Peningkatan *recall* pada kelas negatif dari 0.63 menjadi 0.82 menjadi indikator utama bahwa model kini lebih sensitif terhadap ulasan negatif yang sebelumnya sering terabaikan. Selain *recall*, peningkatan akurasi dari 84% menjadi 90% menegaskan bahwa proses tuning berkontribusi signifikan dalam meningkatkan generalisasi model terhadap data uji. Secara umum, kernel linear terbukti paling optimal dibandingkan kernel lainnya (*sigmoid*, *polynomial*, dan RBF), sebagaimana divisualisasikan pada Gambar 6. Kernel *sigmoid* dan *polynomial* menunjukkan performa cukup baik, sedangkan kernel RBF mengalami sedikit penurunan, kemungkinan karena bentuk distribusi data hasil SMOTE yang tidak sesuai dengan asumsi kernel RBF.

Temuan ini konsisten dengan penelitian sebelumnya. Studi [11] dan [13] melaporkan akurasi sekitar 80% hingga 82% lebih rendah dibandingkan hasil model ini. Penelitian oleh [14] dan [15] juga mendukung bahwa penerapan SMOTE dan GridSearchCV dapat meningkatkan performa klasifikasi, terutama untuk dataset tidak seimbang.

Dalam konteks implementasi nyata, peningkatan deteksi sentimen negatif menjadi sangat krusial. Ulasan negatif sering kali merepresentasikan pengalaman buruk pengguna yang perlu segera direspons oleh pengembang aplikasi untuk menjaga kualitas layanan. Oleh karena itu, keberhasilan model dalam mengidentifikasi sentimen negatif dengan akurat menjadi kontribusi penting dalam studi ini.

#### 4. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa penerapan *Synthetic Minority Over-sampling Technique* (SMOTE) dan *tuning hyperparameter* menggunakan GridSearchCV secara signifikan meningkatkan performa klasifikasi sentimen terhadap ulasan pengguna aplikasi DANA. Evaluasi performa dilakukan menggunakan metode *10-fold cross-validation*, sehingga hasil yang diperoleh lebih stabil dan dapat diandalkan. Model *Support Vector Machine* (SVM) dengan kernel linear menunjukkan performa terbaik, dengan akurasi tertinggi sebesar 90% pada data uji dan rata-rata akurasi sebesar 85,06% dari proses *cross-validation*, meningkat dari 81,67% sebelum optimasi. Peningkatan signifikan juga terjadi pada *recall* kelas negatif, menandakan efektivitas metode dalam menangani ketidakseimbangan data. Hasil uji paired t-test terhadap akurasi masing-masing fold menunjukkan nilai *p-value* < 0,05, yang mengindikasikan bahwa peningkatan performa model signifikan secara statistik. Selain itu, visualisasi melalui *word cloud* berhasil mengungkap kata-kata kunci utama yang mencerminkan sentimen pengguna, memberikan wawasan tambahan yang bermanfaat bagi pengembang aplikasi.

Berdasarkan penelitian tersebut, Dianjurkan bagi penelitian selanjutnya agar kombinasi SMOTE dan GridSearchCV dipertimbangkan dalam penelitian serupa yang menangani klasifikasi sentimen dengan data yang tidak seimbang. *Kernel linear* pada SVM dapat digunakan sebagai *baseline* karena memberikan hasil yang stabil dan mudah diinterpretasikan. Untuk penelitian selanjutnya, penggunaan dataset yang lebih besar dan beragam akan meningkatkan kemampuan generalisasi model. Selain itu, pengujian terhadap algoritma lain seperti Random Forest, XGBoost, atau metode *deep learning* seperti LSTM dapat menjadi alternatif untuk memperoleh hasil klasifikasi yang lebih optimal. Penelitian ini juga dapat menjadi dasar bagi pengembang aplikasi dalam mengevaluasi dan meningkatkan kualitas layanan berdasarkan umpan balik dan persepsi pengguna.

## DAFTAR PUSTAKA

- [1] D. Toresa, S. R. F. Sitorus, I. Muzdalifah, F. Wiza and R. Syelly, "Analisis Sentimen Terhadap Ulasan Penggunaan Dompot Digital Dana Menggunakan Metode Klasifikasi Support Vector Machine," *Technologica*, vol. 3, no. 2, pp. 64-74, 2024.
- [2] R. Novyantri and M. Setiawardani, "The Effect Of E-Service Quality On Customer Loyalty With Customer Satisfaction As A Mediation Variable On Dana's Digital Wallet (Study On Dana Users) Pengaruh E-Service Quality Terhadap Loyalitas Pelanggan Dengan Kepuasan Pelanggan Sebagai Variabel Medias," *International Journal Administration, Business and Organization (IJABO)*, vol. 2, no. 3, pp. 49-58, 2021.
- [3] Z. Azindhani, "Penggunaan Aplikasi Dana Sebagai Media Dompot Digital dan Transaksi di Indonesia," *Jurnal Institut Bisnis dan Teknologi Indonesia*, vol. 1, pp. 1-5, 2021.
- [4] D. S. Pratiwi and I. K. D. Nuryana, "Analisis Tingkat Penerimaan dan Kepercayaan Pengguna Teknologi Terhadap Penggunaan Dompot Digital DANA," *Journal of Emerging Information Systems and Business Intelligence*, vol. 02, no. 04, p. 2021, 2021.
- [5] Normah, B. Rifai, S. Vambudi and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *Jurnal Teknik Komputer AMIK BSI*, vol. 8, no. 2, pp. 174-180, 2022.
- [6] D. A. Agustina, S. Subanti and E. Zukhrona, "Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Marketplace di Indonesia Menggunakan Algoritma Support Vector Machine," *Indonesian Journal of Applied Statistics*, vol. 3, no. 2, p. 109, 2021.
- [7] S. M. K. Azzahra, N. C. Santi and S. Wahyudi, "Implementasi Naïve Bayes Classifier Untuk Analisis Sentimen Aplikasi Pinjaman Online Pada Media Sosial Twitter," *Jurnal Multidisciplinary Applications of Quantum Information Science (al-mantiq)*, vol. 04, no. 02, pp. 1-6, 2023.
- [8] D. T. Novitasari, M. A. Barata and P. E. Yuwita, "Analisis Sentimen Pengguna Twitter Terhadap Skincare Dengan Metode Support Vector Machine (SVM)," *INTI Nusa Mandiri*, vol. 19, no. 2, pp. 325-332, 2025.
- [9] F. F. Irfani, M. Triyanto, A. D. Hartanto and Kusnawi, "Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine," *JBMI (Jurnal Bisnis, Manajemen, dan Informatika)*, vol. 16, no. 3, pp. 258-266, 2020.
- [10] I. S. K. Idris, Y. A. Mustofa and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura Journal of Electrical and Electronics Engineering*, vol. 5, no. 1, pp. 32-35, 2023.
- [11] A. M. Yolanda and R. T. Mulya, "Implementasi Metode Support Vector Machine untuk Analisis Sentimen pada Ulasan Aplikasi Sayurbox di Google Play Store," *Journal of Statistics and Its Application on Teaching and Research*, vol. 6, no. 2, pp. 76-83, 2024.
- [12] Hendiana, A. I. Purnamasari and I. Ali, "Analisis Sentimen Komentar Berita Detik.Com Menggunakan Algoritma Suport Vektor Machine (SVM)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 3175-3181, 2024.
- [13] W. E. Saputro, H. Yuana and W. D. Puspitasari, "Analisis Sentimen Pengguna Dompot Digital Dana Pada Kolom Komentar Google Play Store Dengan Metode Klasifikasi Support Vector Machine," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 2, pp. 1151-1156, 2023.
- [14] A. A. Muhammad, Ermatita and D. S. Prasvita, "Analisis Sentimen Pengguna Aplikasi Dana Berdasarkan Ulasan pada Google Play Menggunakan Metode Support Vector Machine," *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, pp. 194-204, 2022.
- [15] N. G. Ramadhan, "Comparative Analysis of ADASYN-SVM and SMOTE-SVM Methods on the Detection of Type 2 Diabetes Mellitus," *Scientific Journal of Informatics*, vol. 8, no. 2, pp. 276-282, 2021.
- [16] T. A. Dewi and E. Mailoa, "Perbandingan Implementasi Metode SMOTE Pada Algoritma Support Vector Machine (SVM) Dalam Analisis Sentimen Opini Masyarakat Tentang Mixue," *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, vol. 4, no. 3, pp. 849-855, 2023.
- [17] P. Ahirwar, A. J. Deen and M. K. Ahirwar, "Analysis of Machine Learning Algorithm: SMOTE with SVM," *International Journal of Novel Research and Development*, vol. 8, no. 4, pp. 256-264, 2023.
- [18] F. R. Rakhman, R. W. Ramadhani and Y. A. Kuncoroyakti, "Analisis Sentimen dan Opini Digital Kampanye 3M di Masa Covid-19 Melalui Media Sosial Twitter," *Komunikologi: Jurnal Ilmiah Ilmu Komunikasi*, vol. 18, no. 1, pp. 8-20, 2021.
- [19] M. I. Putra and I. Kharisudin, "Analisis Sentimen Pengguna Aplikasi Marketplace Tokopedia Pada Situs Google Play Menggunakan Metode Support Vector Machine (SVM), Naïve Bayes, dan Logistic Regression," *PRISMA*, vol. 5, pp. 759-766, 2022.

- 
- [20] M. Ilmiyah, M. A. Barata and P. E. Yuwita, "Implementation of ANN Optimization with SMOTE and Backward Elimination for PCOS Prediction," *Scientific Journal of Informatics*, vol. 12, no. 1, pp. 133-144, 2025.
- [21] B. P. Pamungkas, M. J. Vikri and I. A. Sa'ida, "Application of SMOTE-ENN Method in Data Balancing for Classification of Diabetes Health Indicators with C4.5 Algorithm," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 14, no. 2, pp. 183-188, 2025.
- [22] A. Z. Praghakusma and N. Charibaldi, "Komparasi Fungsi Kernel Metode Support Vector Machine untuk Analisis Sentimen Instagram dan Twitter (Studi Kasus: Komisi Pemberantasan Korupsi)," *JSTIE (Jurnal Sarjana Teknik Informatika) (E-Journal)*, vol. 9, no. 2, p. 88, 2021.
- [23] F. R. A. Harianto, Z. Alawi and I. A. Sa'ida, "Pengaruh Komposisi Split Data Pada Akurasi Klasifikasi Penderita Diabetes Menggunakan Algoritma Machine Learning," *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 8, no. 1, pp. 36-44, 2025.
- [24] Fahrudin, S. Gantar Fitra and S. Soim, "Pengembangan Model Support Vector Machine untuk Meningkatkan Akurasi Klasifikasi Diagnosis Penyakit Jantung," *Jurnal Teknologi Sistem Informasi dan Aplikasi*, vol. 7, no. 4, pp. 1418-1428, 2024.
- [25] Trianto, A. Muliawati and H. N. Irmanda, "Penerapan Borderline-SMOTE dan Grid Search pada Bagging-SVM untuk Klasifikasi Penyakit Diabetes," *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya*, vol. 3, no. 2, pp. 102-113, 2022.
- [26] E. F. Laili, Z. Alawi, R. Rohmah and M. A. Barata, "Komparasi Algoritma Decision Tree dan Support Vector Machine (SVM) Dalam Klasifikasi Serangan Jantung," *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 8, no. 1, pp. 67-76, 2025.