

Penerapan SVM dan Word2Vec untuk Analisis Sentimen Ulasan Pengguna Aplikasi DANA

Acuan Supian, Bagus Tri Revaldo, Nanda Marhadi, Rahmaddeni, dan Lusiana Efrizoni

Program Studi S1 Teknik Informatika, Universitas Sains Dan Teknologi Indonesia, Pekanbaru
Jln. Purwodadi Indah No.km, Kota Pekanbaru, 28294, Indonesia

E-mail : acuansupian123@gmail.com, bagustrirevaldo05@gmail.com, nanda.marhadi02@gmail.com
rahmaddeni@sar.ac.id*), dan lusiana@stmik-amik-riau.co.id

Abstrak

Dengan meningkatnya penggunaan aplikasi mobile, analisis sentimen terhadap ulasan pengguna menjadi sangat penting untuk memahami persepsi dan kepuasan pelanggan. Aplikasi DANA, sebagai salah satu platform keuangan digital populer di Indonesia, memiliki ribuan ulasan pengguna di Google Playstore yang dapat memberikan wawasan berharga mengenai pengalaman pengguna. Banyaknya ulasan membuat analisis manual tidak efisien dan rentan terhadap bias. Penelitian ini bertujuan mengembangkan model analisis sentimen menggunakan metode Support Vector Machine (SVM) dengan representasi fitur berbasis Word2Vec untuk mengklasifikasikan ulasan pengguna aplikasi DANA di Google Playstore. Tahapan penelitian meliputi pengumpulan data ulasan, preprocessing data, pelatihan model Word2Vec untuk mendapatkan representasi vektor dari teks ulasan, dan penerapan algoritma SVM untuk klasifikasi sentimen. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa model SVM dengan fitur Word2Vec mampu mengklasifikasikan sentimen ulasan pengguna dengan tingkat akurasi mencapai 88%, efektif dalam mengidentifikasi sentimen positif, negatif, dan netral. Hasil penelitian ini dapat digunakan untuk meningkatkan kualitas layanan dan kepuasan pengguna.

Kata kunci : Analisis Sentimen; Ulasan Pengguna; Support Vector Machine; Word2Vec; Aplikasi DANA

Pendahuluan

Di era digital sekarang, aplikasi seluler sudah berwujud alat penting bagi eksistensi keseharian. Dana adalah aplikasi yang sangat populer di Indonesia, DANA adalah salah satu aplikasi pembayaran dengan pertumbuhan tercepat di Indonesia [1]. DANA adalah dompet digital dan platform pembayaran digital terbuka yang banyak digunakan di aplikasi terdaftar dan outlet online tradisional [2]. Aplikasi dompet digital DANA pertama kali dipublikasikan ke publik pada 5 Desember 2018 dan didirikan oleh Elang Sejahtera Mandiri yang mencakup anak perusahaannya PT Elang Mahkota Teknologi Tbk (EMTEK) dan BersamaAnt Financial Partnership. Aplikasi DANA dirancang mengikuti perkembangan masa kini yang menjadikan transaksi cashless dan cardless menjadi digital, cepat dan nyaman [3]. Saat ini E-Wallet DANA menjadi salah satu dompet digital baru di Indonesia. Sebelum E-Wallet DANA, ada e-wallet lain seperti OVO, GoPay, dan LinkAja [4]. Dengan banyaknya aplikasi dompet digital yang ada, pengguna aplikasi akan memberikan reaksi positif dan

negatif terhadap DANA baik dari segi fungsi aplikasi, keamanan, kenyamanan, dan lain-lain. Salah satunya ada di Google PlayStore [3]. Karena aplikasi Dana memiliki jumlah pengguna yang besar, banyak ulasan positif, negatif, dan netral yang tidak bergantung pada rating Google PlayStore [5].

Penelitian ini dilakukan dengan menggunakan data review pengguna aplikasi DANA melalui aplikasi Google Play Store. Google Play Store adalah platform layanan digital yang bertindak sebagai toko resmi dan memungkinkan pengguna untuk mencari dan mengunduh aplikasi [6]. Selain menawarkan aplikasi, Google Play Store juga menawarkan beragam konten digital, antara lain buku, film, musik, dan permainan. Pengguna dapat menemukan aplikasi berbayar dan gratis, memeriksa ulasan dan penilaian dari pengguna lain, dan memilih salah satu yang sesuai dengan kebutuhan mereka.

Google Play Store juga menawarkan pembaruan aplikasi otomatis, sehingga pengguna selalu memiliki versi terbaru dengan fitur terkini dan perbaikan bug. Sebagai platform yang terintegrasi dalam sis-

tem operasi Android, Google Play Store menjadi sumber utama bagi pengguna Android untuk memperoleh beragam aplikasi dan konten digital yang aman dan andal.

Penelitian opini adalah ilmu menganalisis opini, perasaan, penilaian, survei, penilaian, sikap, dan sentimen sejumlah besar orang terhadap suatu entitas seperti produk, layanan, organisasi, orang, isu, peristiwa, topik, atau atribut lainnya. sedang dianalisis. Itu adalah sebuah bidang [7]. Analisis manual tidak efisien dan rentan bias karena banyaknya ulasan yang kami terima setiap hari. Oleh karena itu, diperlukan pendekatan otomatis yang dapat memproses dan menganalisis data penilaian secara efektif.

Tujuan dari penelitian ini adalah mengembangkan model analisis sentimen menggunakan metode support vector machine (SVM) dengan representasi fitur berbasis Word2Vec untuk mengklasifikasikan review pengguna aplikasi DANA di Google Play Store.

Dalam penelitian ini, SVM dipilih sebagai metode klasifikasi utama. Pasalnya keduanya terbukti efektif untuk analisis sentimen data teks [8]. Support Vector Machine (SVM) adalah pembelajaran yang diawasi. SVM pertama kali diperkenalkan oleh Vapnik. SVM bertujuan untuk mencari hyperplane dengan jarak maksimum antara hyperplane dengan support vector (jarak terdekat dengan hyperplane) [9]. Algoritma Support Vector Machine (SVM) diterapkan untuk mengklasifikasikan data, dengan tujuan mengevaluasi keakuratan pemahaman emosi yang terkandung dalam dataset yang dianalisis [10]. Support Vector Machines (SVM) adalah cara yang andal untuk memecahkan masalah klasifikasi data [11]. Pelatihan model Word2Vec digunakan dalam penelitian ini. Word2Vec digunakan untuk mendapatkan representasi vektor yang kaya dari teks ulasan, yang dapat meningkatkan kinerja model klasifikasi. Word2Vec adalah metode penyematan kata yang mempelajari representasi kata dalam ruang vektor berdimensi tinggi [12]. Metode Word2Vec bertujuan untuk menemukan interaksi tersembunyi di dalam kata-kata. Setiap kata mewakili distribusi bobot dalam elemen tersebut [12]. Word2vec adalah metode penyematan kata yang membantu merepresentasikan kata sebagai vector [13]. Dalam penelitian [14] menggunakan metode Support Vector Machine (SVM) dengan kernel linear untuk analisis sentimen terhadap pengguna e-wallet DANA dan GoPay di Twitter menunjukkan bahwa SVM merupakan metode klasifikasi yang robust dalam mengklasifikasikan dua kelas data dengan hasil akurasi 92% untuk DANA dan 90% untuk GoPay.

Pada penelitian [3] analisis sentimen terhadap evaluasi pengguna aplikasi dompet digital DANA di layanan google play dilakukan menerapkan metode Support Vector Machine (SVM). Hasil penelitian menunjukkan bahwa sebanyak 35% pengguna aplikasi DANA memiliki sentimen positif, sementara

65% lainnya memiliki sentimen negatif. Model SVM yang digunakan berhasil mencapai akurasi sebesar 80%, dengan precision sebesar 84,06% untuk sentimen negatif dan 74,08% untuk sentimen positif, serta recall sebesar 87,02% untuk sentimen negatif dan 69,21% untuk sentimen positif.

Dalam penelitian [2] menunjukkan bahwa evaluasi pengguna aplikasi dompet digital DANA di layanan Google Play dapat dianalisis menerapkan teknik Support Vector Machine (SVM) dengan ekstraksi fitur chi-square. Hasil evaluasi penelitian ini menunjukkan bahwa model SVM memiliki akurasi sebesar 87,58%, presisi 91,20%, dan recall 90,21%. Sedangkan model SVM dengan seleksi fitur chi-square memiliki akurasi sebesar 89,41%, presisi 93,29%, dan recall 90,76%.

Penelitian ini bertujuan untuk mengembangkan system evaluasi perasaan yang efektif menggunakan teknik SVM dan Word2Vec untuk tinjauan pelanggan software DANA di layanan google play, menganalisis pola sentimen dalam ulasan tersebut, memberikan wawasan pengembang untuk meningkatkan kualitas layanan., mengevaluasi kinerja model dengan berbagai metrik, dan mengembangkan metode yang dapat diaplikasikan pada ulasan aplikasi lain untuk analisis sentimen.

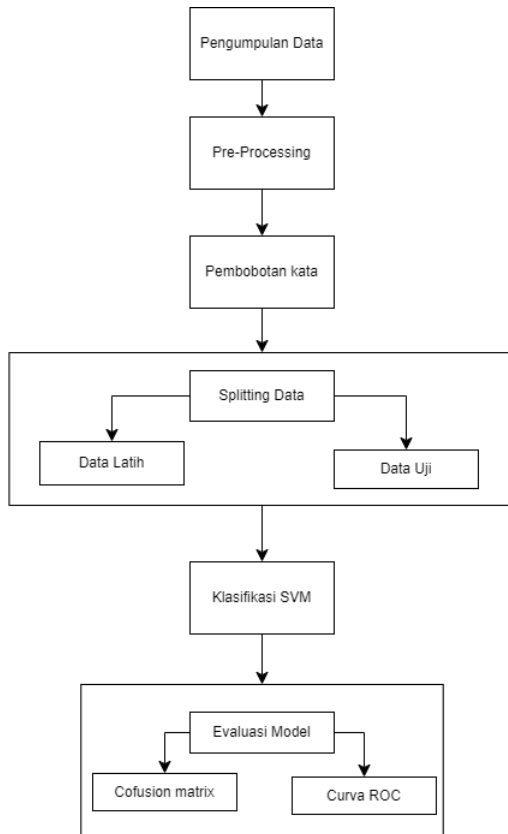
Berdasarkan beberapa studi yang telah di kutip diatas, peneliti melakukan studi yang menggunakan metode Super Vector Machine (SVM) sebagai mengklasifikasi sentiment dari ulasan pengguna aplikasi dana.

Oleh karena itu, penelitian ini diharapkan dapat menyediakan pemahaman tentang bagaimana pandangan pengguna terhadap aplikasi DANA di Google Play Store. Memahami pergeseran pandangan dan pendapat masyarakat terhadap aplikasi DANA dapat mendukung pengembang dan pihak berwenang untuk mengambil keputusan yang lebih adaptif terhadap kebutuhan pengguna.

Metode Penelitian

Kerangka acuan dibuat untuk memudahkan pelaksanaan penelitian. tahap penelitian yang terstruktur dan sistematis. Kerangka fase ini dimaksudkan supaya studi ini bisa dilaksanakan Secara berturut-turut, efektif dan efisien. Kerangka kerja ini memungkinkan setiap tahap penelitian dipantau dan dievaluasi dengan baik. Hal ini memungkinkan Anda melakukan penyesuaian yang diperlukan untuk mencapai hasil yang optimal. Kerangka kerja ini demikian pula membantu mengidentifikasi dan menyelesaikan Masalah apa pun yang mungkin muncul selama proses tersebut penelitian, memastikan bahwa semua langkah yang diperlukan diikuti dengan benar. Dengan cara ini, Anda akan dapat melakukan penelitian secara lebih terorganisir dan sistematis, sehingga lebih efisien dan produktif dari waktu ke waktu. Selain itu, kerangka kerja ini memfasilitasi komunikasi antar anggota tim peneliti

dan memeriksa bahwa tiap orang memahami dengan cara yang serupa tentang langkah-langkah yang harus diambil dengan tujuan yang ingin dicapai. Kerangka alur analisis ditunjukkan pada Gambar 1. Mulailah dengan pengumpulan data, prapemrosesan, pembobotan kata, pemisahan data pengujian dan pelatihan, klasifikasi menggunakan SVM, dan penilaian menggunakan matriks konfusi dan kurva ROC untuk visualisasi model. Dengan mengikuti kerangka ini, diharapkan penelitian dapat berjalan lebih lancar dan memberikan hasil yang lebih efektif.



Gambar 1: Langkah Penelitian

Pengumpulan Data

Kegiatan ini merupakan langkah awal yang penting dalam melakukan penelitian dan menganalisis data untuk memperoleh informasi yang relevan dan bermakna. Pengumpulan data merupakan langkah penting dalam siklus penelitian dan tidak boleh diabaikan. Hal ini dikarenakan kualitas data yang diperoleh sangat mempengaruhi validitas dan ketepatan hasil penelitian. Proses ini seringkali memerlukan perencanaan yang matang dan strategi yang tepat untuk melakukannya memastikan bahwa data yang dikumpulkan memenuhi tujuan penelitian dan berkualitas tinggi. Sebagai bagian dari penelitian ini, kami memilih untuk menggunakan kumpulan data dari Kaggle, sebuah platform yang terkenal dengan beragam kumpulan data berkualitas tinggi yang tersedia untuk penelitian. Dataset yang kami gunakan terdiri dari 50.000 data dan di-

harapkan cukup representatif untuk memperkuat analisis yang kami lakukan.

Dataset ini Anda dapat mengaksesnya dengan menggunakan tautan berikut:

<https://www.kaggle.com/datasets/alexmariosi/manjuntak/dana-app-sentiment-review-on-playstore-indonesia> .

Dataset komprehensif ini bertujuan untuk melakukan penelitian lebih tepat sasaran dan memberikan hasil yang akurat dan bermakna.

PreProcessing

Langkah preprocessing termasuk langkah pemurnian data. Data GitHub yang dikumpulkan adalah data tidak terstruktur [15]. Preprocessing yang dilakukan pada penelitian ini meliputi perubahan huruf menjadi huruf kecil (lowercase Conversion), pembersihan data (cleaning), penyaringan kata (filtering) atau menghilangkan stop word, kata dasar yang meliputi beberapa tahapan yaitu: mengembalikan kata (lemmatization), dan memecah menjadi unit-token. (tokenisasi),

1. Langkah konvolusi meliputi pengerjaan perubahan huruf besar menjadi huruf kecil secara keseluruhan [16]. Menerapkan sensitivitas huruf besar-kecil juga memberikan manfaat berikut ketika menangani inkonsistensi penulisan pada data Anda: B. Kapitalisasi non-standar atau potensi masalah penulisan huruf [17].
2. Fase pembersihan membuang kata-kata yang tidak perlu [16]. Seperti, angka (1), pemisah kata, khususnya koma (,), titik (.), tanda seru (!) dan tanda baca lainnya. Tokenisasi melibatkan pemisahan sepotong teks menjadi kata atau kalimat [18].
3. Berikutnya, dilakukan proses filtering di mana kata-kata penting dipilih dari hasil tokenisasi [19]. Kata-kata yang tidak informatif seperti itu, dan, di, atau, dengan, dll. akan dihapus.
4. Selanjutnya, kosakata tersebut akan diterapkan pada ulasan yang difilter. Lemmatisasi mengacu pada mengubah kata ke bentuk aslinya untuk meningkatkan kinerja model [20].
5. Token adalah kata unik yang akan digunakan sebagai pengidentifikasi emosi kelompok. Sebelum enkripsi

Pembobotan kata Word2Vec

Word2vec adalah alat baru yang dikembangkan melalui Thomas Mikolov. Word2vec dapat memproses kalimat dari kumpulan data yang sangat besar dalam keadaan yang sangat singkat dan hasil yang tinggi dibandingkan alat sebelumnya [21]. Cara kerja alat ini adalah menangkap hasil tulisan

sebagai masukan dan menciptakan penyajian vektoran atas semua kata dalam hasil teks selaku keluaran [21]. Word2vec diterapkan untuk menata fitur setiap kata ke dalam format vector [22]. Vektorisasi adalah tahap mengubah struktur teks menjadi format vektor berisi angka. Sebuah metode vektorisasi yang mungkin Anda gunakan adalah Word2Vec. Word2Vec dapat mewakili sebuah kata dengan vektor dengan ukuran penyematan tertentu [23]. Tahapan word2vec bisa di perhatikan dalam Gambar 2:



Gambar 2: Tahapan Word2Vec

Splitting Data

Sebelum pelatihan model machine learning dilakukan, dataset terlebih dahulu dibagi menjadi dua bagian. Dataset dibagi menjadi data pelatihan dan data pengujian. Data uji ini digunakan untuk mengetahui berapa persentase algoritma klasifikasi yang bekerja dengan benar. Data dibagi menjadi dua kelompok: data pelatihan dan data pengujian. Percobaan pemisahan data dilakukan sebanyak tiga kali, yaitu:

1. dalam pengujian pertama, data pelatihan menyumbang 70% dan 30% data pengujian.
2. dalam pengujian kedua, data pelatihan menyumbang 80% dan 20% data pengujian.
3. kemudian dalam pengujian terakhir data pelatihan menyumbang 90% dan 10% data uji

Data Latih (Data Training)

Penggunaan data latih untuk menginstruksikan sistem penelitian ini. Data pelatihan digunakan untuk mengajar metode klasifikasi SVM agar dapat memahami cara mengklasifikasikan komentar ke dalam kategori negatif, netral, atau positif.

Data Uji (Data Testing)

Setelah model klasifikasi dilatih, langkah berikutnya adalah mengevaluasi kinerja metode tersebut

dengan menggunakan data uji. Proses pengujian bertujuan untuk menguji keefektifan metode klasifikasi SVM dengan memasukkan data baru, di mana metode tersebut akan mengklasifikasikan data baru tersebut ke dalam kategori negatif, netral, atau positif secara akurat.

Klasifikasi SVM

Mesin vektor pendukung merupakan Sistem pembelajaran ini menggunakan hipotesis berupa fungsi linier dalam ruang fitur yang memiliki dimensi tinggi, dan menerapkan kesalahan pembelajaran yang diperoleh dari teori statistik untuk algoritma pembelajarannya [24]. Teori yang mendasari SVM telah dikembangkan. Pengklasifikasian dengan SVM ini melibatkan penerapan metode kernel linier dan menampilkan nilai akurasi yang diperoleh.

Mesin vektor pendukung awal sekali dipublikasikan oleh Vapnik pada tahun 1992 sebagai serangkaian konsep luar biasa yang harmonis di bidang pengenalan pola [25].

Menurut ciri-ciri nya teknik SVM dikelompokkan kedalam dua kelas yaitu linier dan nonlinier. Linear SVM memisahkan data secara linier, yaitu dua kelas pada hyperplane yang dipisahkan oleh soft edge. Di sisi lain, nonlinier adalah fungsi tipu daya kernel dalam ruang berdimensi tinggi [26]. Kami melakukan uji Akurasi didasarkan pada perhitungan yang tepat yang dilakukan dan model mampu memprediksi secara akurat 88%. Algoritma Support Vector Machine (SVM) adalah teknik pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. Berikut adalah langkah-langkah umum SVM:

1. Transformasi Data:
 - Jika data tidak dapat dipisahkan secara linear, SVM menggunakan fungsi kernel untuk mentransformasikan data ke dimensi yang lebih tinggi sehingga dapat dipisahkan secara linear.
2. Menemukan Hyperplane:
 - SVM mencari hyperplane yang memisahkan data ke dalam kelas-kelas dengan margin maksimum.
3. Identifikasi Support Vectors:
 - Titik-titik data terdekat ke hyperplane disebut support vectors. Mereka membantu menentukan hyperplane yang optimal.
4. Optimisasi Margin:
 - SVM memaksimalkan margin, yaitu jarak antara hyperplane dan titik data terdekat dari kedua kelas.
5. Fungsi Kernel: Beberapa fungsi kernel yang umum digunakan adalah:

- Linear: (xi, xj)
- Polynomial: $(y(xi, xj) + r)d$
- Radial Basis Function (RBF): $exp(-y//xi - xj//2)$
- Sigmoid: $tanh(y(xi, xj) + r)$

6. Pelatihan Model:

- SVM menggunakan data pelatihan untuk menyelesaikan masalah optimasi dan menemukan hyperplane yang optimal.

7. Klasifikasi:

- Setelah model dilatih, SVM mengklasifikasikan data baru berdasarkan sisi hyperplane tempat data tersebut berada.

Evaluasi Model

Uji kinerja diperlukan Untuk menilai performa metode support vector machine (SVM). Studi penilaian berdasarkan skor ketepatan. Skor ketepatan merupakan matriks penilaian Akurasi adalah ukuran yang menunjukkan seberapa baik model klasifikasi atau prediksi dapat memperoleh hasil yang tepat atau sesuai dengan data yang ada. Ini mencerminkan tingkat kesuksesan model dalam memprediksi label atau kelas yang benar. Akurasi menggambarkan seberapa baik model mampu mengklasifikasikan keseluruhan data. benar.

$$Akurasi = \frac{JumlahPrediksiBenar}{TotalJumlahData} \quad (1)$$

Confusion Matrix

Matriks konfusi Ini digunakan untuk menilai hasil dari model yang telah dibuat prediksi dari penerapan klasifikasi. Matriks konfusi menggunakan dua perintah untuk membandingkan dua algoritma

Curva ROC

Grafik yang menunjukkan kinerja (ROC) Ini adalah grafik yang menggambarkan kinerja model klasifikasi dalam berbagai ambang atau threshold batas yang berbeda. Kurva ROC memplot tingkat positif sebenarnya (TPR) dan tingkat positif palsu False Positive Rate (FPR) adalah ukuran yang menunjukkan seberapa sering model klasifikasi salah mengklasifikasikan kelas negatif sebagai positif pada berbagai pengaturan ambang batas yang berbeda. Tingkat positif sebenarnya (TPR), juga disebut tingkat deteksi positif atau perolehan kembali, adalah rasio prediksi positif sebenarnya terhadap jumlah total prediksi. Contoh positif, Tingkat positif palsu (FPR) adalah rasio prediksi positif palsu terhadap jumlah total kasus negatif. Garis diagonal pada diagram ROC mewakili performa model yang membuat prediksi acak. Semakin mendekati sudut kiri atas, performa model akan semakin baik

menurut kurva ROC. Area Under the Curve (AUC) adalah metrik kinerja yang merangkum kurva ROC menjadi satu nilai, dengan nilai yang mendekati 1 menunjukkan kinerja yang lebih baik. Pada grafik ini terlihat bahwa AUC model SVM adalah 0,06 yang menunjukkan performa model. dalam membedakan antara kategori atau kelompok yang dianggap positif dan negatif

Visualisasi Model

Visualisasi model adalah proses membuat representasi grafis dari model pembelajaran mesin atau data analitis untuk memahami performa dan properti model tersebut. Visualisasi memberikan wawasan tentang cara kerja model, cara data diinterpretasikan oleh model, dan cara hasil prediksi diinterpretasikan. Dalam penelitian ini, kami memvisualisasikan dan menyajikan matriks fusi dan model kurva Roc.

Hasil dan Pembahasan

Komponen “Hasil dan Pembahasan” ini menampilkan hasil utama analisis sentimen yang dilakukan terhadap review pengguna aplikasi DANA. Bagian ini memberikan penjelasan rinci dari setiap hasil, termasuk interpretasi analisis sentimen positif dan negatif. Maka dari itu, bagian ini diperkirakan bisa memberikan wawasan secara detail mengenai sentimen pengguna terhadap aplikasi DANA.

Deskriptif Data

Data yang dipergunakan dalam penelitian ini diberi label sentimen yang menunjukkan apakah review pengguna terhadap aplikasi DANA di Google Play Store positif atau negatif. Kumpulan data ini terdiri dari beberapa kolom, namun fokusnya adalah pada kolom peringkat dan sentimen. Sebelum memasuki tahap preprocessing, data diperiksa untuk memastikan tidak ada data kosong atau NaN. Setelah pemeriksaan ini, data kolom review mengalami proses preprocessing untuk membersihkan dan menormalkan teks serta mempersiapkannya untuk analisis lebih lanjut.

Pada bagian ini, kami menyajikan statistik deskriptif dari dataset yang digunakan dalam penelitian ini. Dataset ini memuat pendapat atau komentar pengguna aplikasi Dana yang diambil dari Google Play Store. Ulasan-ulasan ini kemudian dianalisis untuk menentukan sentimen yang terkandung di dalamnya, yaitu positif, negatif, atau netral. Berikut adalah distribusi masing-masing sentimen yang bisa dilihat pada Gambar 3:

username	score	dt	content	sentimen	
0	Eliya Kati	5	2024-02-15 11:24:56	Bagus	POSITIVE
1	Rizwan Man	2	2024-02-15 11:24:53	Data ming keun kurang	POSITIVE
2	Chia Saepi	1	2024-02-15 11:22:46	Saya ngapain oggapan dana paman kina ku lag	NEGATIVE
3	Rizki Niki	3	2024-02-15 11:22:46	Kocak mana daban nye ni malah enor tegala lag	NEGATIVE
4	Fil Alifiah	1	2024-02-15 11:21:34	Sabtu siang karena no sama hilang gant no sa	NEGATIVE
49995	Maria Maha	5	2023-12-31 16:34:55	Bagus	POSITIVE
49996	Unique Lady	1	2023-12-31 16:33:47	emr. tidak bisa masuk	NEGATIVE
49997	superid superid	5	2023-12-31 16:30:08	Cukup membantu	POSITIVE
49998	hasanandam	1	2023-12-31 16:28:51	Isong pihak dana trip sabtu saya hilang,ada yg...	NEUTRAL
49999	Maria Aerdau	1	2023-12-31 16:28:46	good	NEGATIVE

Gambar 3: Data Asli

- Jumlah Ulasan Positif : 26.555
- Jumlah Ulasan Negatif : 17.073
- Jumlah Ulasan Netral : 6.372

Untuk memberikan gambaran lebih jelas mengenai ulasan-ulasan ini, berikut adalah beberapa contoh ulasan dari masing-masing kategori sentimen:

- Contoh Ulasan Positif :
 1. Saya sangat senang sekali aplikasi dana ini membantu sekali masalah saya, saya kasih bintang 5.
 2. Aplikasi ini sangat bermanfaat bagi penggunanya, Aplikasi yg bagus
- Contoh Ulasan Negatif :
 1. Proses update ke premium sangat lama, mana udah saya top up jadi gak bisa kirim uang.
 2. Sering banget gangguan
- Contoh Ulasan Netral :
 1. . Maaf boleh nanya gak, bisa top up gak di dana?
 2. Saya baru mau mencobanya

Statistik deskriptif ini memberikan gambaran awal mengenai karakteristik dari Data yang digunakan dalam penelitian ini. Tindakan berikutnya adalah melakukan preprocessing data dan menerapkan metode SVM dengan word2vec untuk analisis sentimen yang lebih mendalam.

Pre-Processing Data

Tahap preprocessing data adalah langkah penting dalam analisis sentimen, karena membantu mempersiapkan data mentah menjadi bentuk yang lebih siap untuk dianalisis oleh model. Dalam penelitian ini, terdapat 50.000 data ulasan, beberapa teknik preprocessing telah diterapkan untuk memastikan kualitas data yang lebih baik dan untuk mengoptimalkan kinerja model SVM dengan word2vec. Berikut code yang kami gunakan dalam melakukan tahapan preprocessing yang bisa di lihat pada Gambar 4:

```
def preprocess_text(text):
    if not isinstance(text, str):
        return ""
    # Ubah ke huruf kecil
    text = text.lower()

    # Hapus angka
    text = re.sub(r'\d+', '', text)

    # Hapus tanda baca
    text = re.sub(r'[^\w\s]', '', text)

    # Hapus spasi berlebih
    text = re.sub(r'\s+', ' ', text).strip()

    # Terapkan penggantian slang dan penghapusan perpanjangan kata
    text = replace_slang(text)
    text = replace_word_elongation(text)

    # Tokenisasi
    tokens = nltk.word_tokenize(text)

    # Hapus kata-kata yang bukan alfabet
    tokens = [token for token in tokens if token.isalpha()]

    # Stemming (contoh saja, bisa dimasukkan jika diperlukan)
    # factory = StemmerFactory()
    # stemmer = factory.create_stemmer()
    # tokens = [stemmer.stem(token) for token in tokens]

    # Print setiap 100 kata yang diproses
    if len(tokens) % 100 == 0:
        print(f"Processing words: {len(tokens)}")

    return tokens

# Terapkan fungsi pembersihan teks pada subset dataset
df['cleaned_content'] = df['content'].apply(preprocess_text)

# Simpan dataset yang telah diproses
df.to_csv('preprocessed_dana.csv', index=False)

# Tampilkan beberapa baris pertama dari dataset yang telah diproses
print("\nDataset yang telah diproses:")
print(df.head())
```

Gambar 4: Code Preprocessing

Kode pada Gambar 4 merupakan fungsi preprocessing teks untuk membersihkan dan mempersiapkan teks sebelum analisis lebih lanjut. Langkah-langkahnya meliputi: memastikan input adalah string, mengubah semua huruf menjadi kecil, menghapus angka, tanda baca, dan spasi berlebihan, mengganti slang dan kata yang diperpanjang, melakukan tokenisasi, menghapus kata non-alfabet, menerapkan stemming (opsional), menampilkan pesan setiap 100 kata yang diproses, menerapkan fungsi pada kolom 'content' dari dataframe, menyimpan hasil preprocessing ke file CSV, dan menampilkan beberapa baris pertama dari dataset yang telah diproses.

Berikut adalah langkah-langkah preprocessing yang telah dilakukan serta pengaruhnya terhadap data:

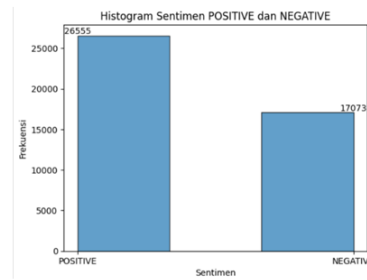
Tabel 1: Hasil Pre-processing

Ulasan Asli	Ulasan Yang Sudah di Pre-Processing
Dana mng keren mantap.	[dana, memang, keren, mantap]
Saya sangat senang sekali aplikasi dana ini membantu sekali masalah saya .saya kasih bintang 5	[saya, sangat, senang, sekali, aplikasi, dana, bantu, masalah, kasih, bintang]
smoga sukses dan terima kasih	[semoga, sukses, terimakasih]
nyaman bngt smga tambah baik lagi	[nyaman, banget, semoga, tambah, baik, lagi]
Saya suka aplikasi DANA sungguh sangat membantu	[saya, suka, aplikasi, dana, sungguh, sangat, bantu]

Pengujian dilakukan terhadap Dataset Sentimen Review Pengguna Aplikasi DANA dari Google Play Store. Label sentimen positif dan negatif dimasukkan secara manual ke dalam dataset ini. Data ini akan dibagi menjadi data latih dan data uji dengan perbandingan 90: 10. 90% digunakan untuk melatih model dan 10% digunakan untuk menguji model.

Distribusi data dilakukan secara acak untuk menjamin keabsahan hasil evaluasi. Perhitungan menunjukkan bahwa data sentimen terdiri dari 45.000 ulasan positif dan 5.000 ulasan negatif dalam kumpulan data.

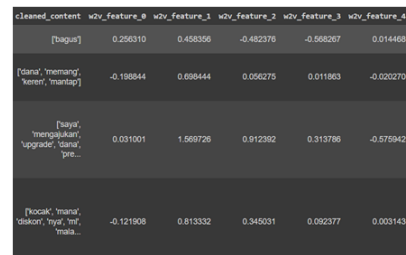
Proses pelatihan dan pengujian dilakukan dengan sangat teliti untuk mengoptimalkan performa model dalam mengklasifikasikan ulasan pengguna berdasarkan pendapat mereka. Dengan menggunakan rasio pembagian data yang telah ditentukan sebelumnya (misalnya 80:20 untuk pelatihan dan pengujian), penelitian ini memastikan bahwa model SVM dapat mempelajari pola dalam data pelatihan dan kemudian menguji akurasi dengan data yang belum pernah dilihat sebelumnya. Ini memberikan gambaran yang lebih objektif tentang kemampuan model dalam mengkategorikan ulasan sebagai positif atau negatif. Selain itu, penelitian ini juga menyertakan analisis mendalam mengenai tanggapan pengguna terhadap aplikasi DANA, sehingga dapat memberikan wawasan yang berharga untuk pengembangan aplikasi lebih lanjut. Hasil perbandingan klasifikasi dan akurasi model ditampilkan dalam bentuk diagram batang, seperti yang ditunjukkan pada Gambar 5, yang memudahkan pemahaman performa model secara visual:



Gambar 5: Kelas Sentimen

Pembobotan Kata

Pada bagian ini, penelitian ini menggunakan teknik Word2Vec untuk membuat representasi vektor kata. Word2Vec mengubah setiap kata dalam korpus teks menjadi vektor berdimensi tetap yang menangkap makna semantik kata tersebut. Implementasi ini dilakukan dengan menggunakan pustaka ‘Gensim’ dan memungkinkan pengelompokan kata-kata yang memiliki arti serupa secara lebih ketat dalam ruang vektor. Visualisasi hasil pembobotan dilakukan dengan menggunakan teknik reduksi dimensi seperti PCA dan t-SNE untuk memudahkan interpretasi. Hasilnya, Word2Vec berhasil menangkap hubungan semantik antar kata dalam kumpulan data, sehingga meningkatkan akurasi model pembelajaran mesin dalam analisis sentimen. Singkatnya, Word2Vec memberikan representasi kata yang lebih kaya dan bermakna, yang berdampak positif terhadap performa model. Berikut beberapa contoh hasil pembobotan katanya:



Gambar 6: Hasil Word2Vec

Pada Gambar 6, bisa dilihat hasil dari pembobotan kata menggunakan Word2Vec. Untuk keterangan dari setiap kolomnya yaitu sebagai berikut :

- ‘cleaned_content’ : Kolom ini berisi daftar kata-kata yang telah dibersihkan dari setiap kalimat atau dokumen yang dianalisis.
- ‘w2v_feature_0’ sampai ‘w2v_feature_4’ : Kolom-kolom ini mewakili vektor fitur yang dihasilkan oleh model Word2Vec untuk setiap kata yang ada di kolom ‘cleaned_content’.

Splitting Data

Pada langkah ini, dikerjakan pembagian dataset berperan dua subset utama: data latih (training

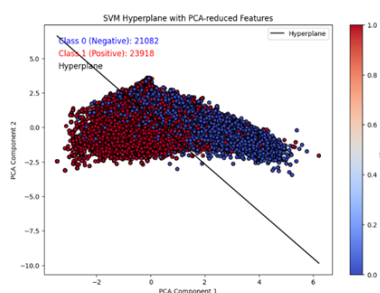
data) dan data uji (testing data) dengan rasio yang berbeda-beda. Pembagian ini bertujuan untuk mengevaluasi performa model pada berbagai proporsi data uji. Splitting data dengan beberapa rasio yang berbeda dengan random state = 0:

- Rasio 90:10 Pembagian dengan rasio 90:10 bermakna 90% dari dataset diterapkan untuk data latih dan 10% diterapkan untuk data uji. Rasio ini cocok digunakan ketika dataset cukup besar dan kita ingin memaksimalkan penggunaan data latih.
- Rasio 80:20 Pembagian dengan rasio 80:20 adalah yang paling umum digunakan. Rasio ini membagi dataset menjadi 80% data latih dan 20% data uji. Ini memberikan keseimbangan antara jumlah data latih yang cukup besar dan data uji yang mencukupi untuk evaluasi model.
- Rasio 70:30 Pembagian dengan rasio 70:30 mengalokasikan 70% dataset untuk data latih dan 30% untuk data uji. Rasio ini cocok digunakan ketika dataset tidak terlalu besar dan kita ingin memastikan model dapat diuji dengan baik pada berbagai data uji.

Klasifikasi SVM

Setelah melalui tahapan-tahapan di atas, selanjutnya kami melakukan visualisasi model SVM untuk memahami bagaimana model memisahkan antara kelas positif dan negatif dalam analisis sentimen ulasan pengguna aplikasi DANA.

Hasil klasifikasi ini membantu kami untuk menunjukkan bagaimana SVM memisahkan kata-kata yang dianggap positif dan negatif berdasarkan ulasan pengguna. Garis yang terlihat adalah hyperplane yang dibangun oleh SVM untuk memaksimalkan margin antara kedua kelas. Bisa di lihat pada Gambar sekian di bawah ini:



Gambar 7: Hasil Word2Vec

Berikut adalah beberapa kesimpulan yang dapat diambil dari visualisasi SVM pada Gambar 7 dengan fitur yang direduksi menggunakan PCA:

1. Distribusi Kelas:

- Kelas 0 (Negative): Terdapat 21,082 data yang termasuk dalam Kelas 0 (biru).

- Kelas 1 (Positive): Terdapat 23,918 data yang termasuk dalam Kelas 1 (merah).
- Kelas 1 memiliki sedikit lebih banyak data dibandingkan dengan Kelas 0.

2. Pemisahan oleh Hyperplane:

- Garis hitam yang disebut hyperplane mencoba memisahkan titik data dari dua kelas yang berbeda.
- Meskipun ada beberapa tumpang tindih antara dua kelas, hyperplane memberikan garis batas yang cukup baik antara dua kelas tersebut.

3. Efektivitas PCA:

- PCA telah berhasil mereduksi dimensi data ke dalam dua komponen utama, memungkinkan visualisasi yang lebih mudah dari data yang mungkin memiliki dimensi yang jauh lebih tinggi.
- Data dalam dua dimensi ini masih mempertahankan karakteristik yang cukup untuk memungkinkan SVM memisahkan dua kelas.

4. Visualisasi Gradien:

- Gradien warna di sisi kanan menunjukkan bahwa ada spektrum nilai prediksi kelas, dengan nilai yang lebih tinggi (merah) menunjukkan Kelas 1 dan nilai yang lebih rendah (biru) menunjukkan Kelas 0.
- Gradien ini juga membantu memahami bagaimana model SVM memprediksi probabilitas atau margin untuk setiap kelas.

5. Tumpang Tindih Kelas:

- Ada beberapa tumpang tindih antara dua kelas, yang dapat dilihat dari titik-titik merah dan biru yang bercampur di sekitar hyperplane.
- Ini menunjukkan bahwa meskipun SVM mencoba untuk memisahkan kelas-kelas ini, ada beberapa data yang sulit untuk diklasifikasikan dengan jelas.

6. Kualitas Model:

- Meskipun terdapat tumpang tindih, garis hyperplane tampaknya cukup baik dalam memisahkan sebagian besar data dari dua kelas.
- Ini menunjukkan bahwa model SVM mampu belajar pola yang ada dalam data yang direduksi menggunakan PCA.

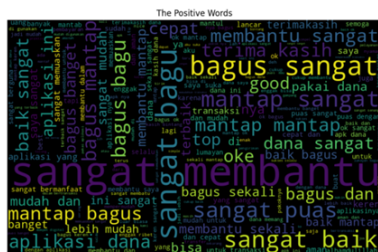
Kesimpulan Perbandingan:

- Ulasan Positif Lebih Banyak: Terdapat 2,836 lebih banyak ulasan positif dibandingkan dengan ulasan negatif.
- Persentase Ulasan:
 - Ulasan Negatif :
 $21,08221,082+23,918 \approx 46.85\%$
 $21,082+23,91821,082 \approx 46.85\%$
 - Ulasan Positif :
 $23,91821,082+23,918 \approx 53.15\%$
 $21,082+23,91823,918 \approx 53.15\%$

Dari perbandingan ini, dapat disimpulkan bahwa aplikasi Dana di Playstore menerima lebih banyak ulasan positif dibandingkan dengan ulasan negatif. Hal ini menunjukkan bahwa secara keseluruhan, pengguna cenderung memiliki pengalaman yang lebih baik dengan aplikasi tersebut.

Sentimen Positif

Data respon positif yang telah dilabeli dimasukkan ke dalam kelas positif dengan memanfaatkan analisis sentimen, di mana sentimen positif ditetapkan berdasarkan frekuensi kata-kata yang menunjukkan evaluasi yang menguntungkan dalam keseluruhan dataset. Metode ini memungkinkan identifikasi dan pengelompokan ulasan atau teks yang mengandung ungkapan positif, memungkinkan pemahaman yang lebih dalam tentang pendekatan tinjauan sentimen yang dipakai dalam pengkajian ini.



Gambar 8: Hasil Word2Vec

Sentimen Negatif

Data sentiment negatif hasil pelabelan dimasukkan ke dalam kelas negatif menerapkan analisis sentiment. Data sentiment negatif tersebut ditentukan berlandaskan jumlah frekuensi kata dalam Kumpulan data.



Gambar 9: Hasil Word2Vec

Evaluasi Model

Pada tahap ini, kami melakukan evaluasi rinci terhadap model klasifikasi SVM untuk analisis sentimen ulasan pengguna aplikasi DANA. Berdasarkan hasil eksperimen dari beberapa ratio Evaluasi dilakukan dengan memeriksa matriks konfusi dan kurva ROC untuk memahami kinerja model secara lebih detail.

Pada studi ini, kami mengimplementasikan penerapan Support Vector Machine (SVM) untuk mengklasifikasikan ulasan pengguna aplikasi DANA. SVM dipilih karena kemampuannya dalam menangani dataset berdimensi tinggi dan kemampuannya untuk menangani masalah klasifikasi dengan baik, terutama pada data teks. Berikut hasil eksperimen pada setiap rasio yang telah dilakukan:

1. Rasio 90:10

```

Accuracy: 88.26
Precision: 88.2892606864501
Recall: 88.26
F1 Score: 88.24388526328909
Classification Report:

```

	precision	recall	f1-score	support
0	0.89	0.86	0.87	2363
1	0.88	0.91	0.89	2637
accuracy			0.88	5000
macro avg	0.88	0.88	0.88	5000
weighted avg	0.88	0.88	0.88	5000

Gambar 10: Hasil Rasio 90:10

Gambar 10 menunjukkan hasil evaluasi model klasifikasi dengan akurasi 88,26%, presisi 88,29%, recall 88,26%, dan skor F1 88,24%. Dari laporan klasifikasi, untuk kelas 0 presisi 0,89, recall 0,86, dan skor F1 0,87, sedangkan untuk kelas 1 presisi 0,88, recall 0,91, dan nilai F1 0,89. Akurasi keseluruhan adalah 0,88 karena dukungan kelas 0 adalah 2363 dan dukungan kelas 1 adalah 2637. Rata-rata makro dan tertimbang skor presisi, perolehan, dan F1 adalah 0,88, yang menunjukkan bahwa performa kedua kelas model seimbang.

2. Rasio 80: 20

```

Accuracy: 88.03
Precision: 88.04313807007212
Recall: 88.03
F1 Score: 88.01526958859563
Classification Report:

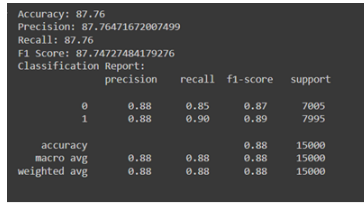
```

	precision	recall	f1-score	support
0	0.88	0.86	0.87	4683
1	0.88	0.90	0.89	5317
accuracy			0.88	10000
macro avg	0.88	0.88	0.88	10000
weighted avg	0.88	0.88	0.88	10000

Gambar 11: Hasil Rasio 80:20

Gambar 11 menunjukkan hasil evaluasi model klasifikasi dengan akurasi 88,03%, presisi 88,04%, recall 88,03%, dan skor F1 88,02%. Dari laporan klasifikasi, untuk kelas 0 presisi 0,88, recall 0,86, dan skor F1 0,87, sedangkan untuk kelas 1 presisi 0,88, recall 0,90, dan nilai F1 0,89. Dukungan kelas 0 adalah 4683, dukungan kelas 1 adalah 5317, dan presisi keseluruhan adalah 0,88. Rata-rata makro dan tertimbang skor presisi, perolehan, dan F1 adalah 0,88, yang menunjukkan bahwa performa kedua kelas model seimbang.

3. Rasio 70:30



Gambar 12: Hasil Rasio 70:30

Gambar 12 menunjukkan hasil evaluasi model klasifikasi dengan akurasi 87.76%, precision 87.76%, recall 87.76%, dan F1 score 87.75%. Laporan klasifikasi mengindikasikan bahwa untuk kelas 0, precision adalah 0.88, recall 0.85, dan F1 score 0.87, sementara untuk kelas 1, precision adalah 0.88, recall 0.90, dan F1 score 0.89. Dengan support masing-masing 7005 untuk kelas 0 dan 7995 untuk kelas 1, akurasi keseluruhan adalah 0.88. Rata-rata macro dan weighted untuk precision, recall, dan F1 score adalah 0.88, menunjukkan performa model yang seimbang untuk kedua kelas.

Confusion Matrix

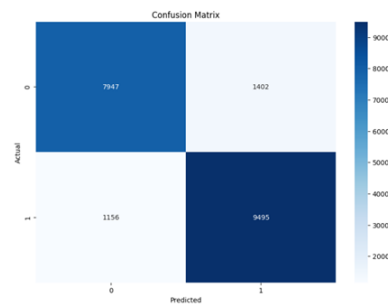
Matriks konfusi adalah sebuah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi dalam pembelajaran mesin. Tabel ini menampilkan hasil prediksi dari model, yang mengelompokkan data ke dalam kategori atau kelas tertentu, serta menunjukkan seberapa akurat model tersebut dalam memprediksi setiap kelas. Dengan matriks konfusi, kita dapat melihat secara jelas jumlah prediksi yang benar (true positives dan true negatives) serta yang salah (false positives dan false negatives) untuk setiap kelas yang diprediksi oleh model. Ini memberikan gambaran yang komprehensif tentang seberapa baik model dapat memisahkan dan mengklasifikasikan data yang diberikan.

Tabel 2: Hasil Confusion Matrix

		Nilai Prediksi	
		Positif	Negatif
Nilai Aktual	Positif	9495	7497
	Negatif	1402	1156

Berdasarkan Tabel 2, yaitu hasil Confusion Matrix dari rasio 90:10 menunjukkan perhitungan hasil pengujian model dimana data uji (nyata) pada ulasan berlabel positif diprediksi oleh model dengan akurasi 9495 dan probabilitas kesalahan 7497. Sedangkan untuk ulasan berlabel negatif, probabilitas benar adalah 1402 dan probabilitas salah sebesar 1156. Di bawah ini adalah hasil matriks ke-

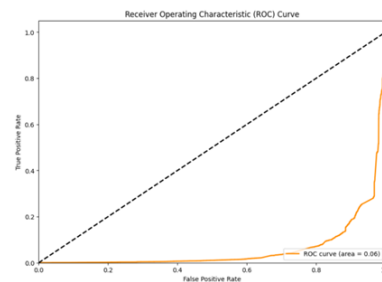
bingungan klasifikasi sentimen Dana, terlihat pada gambar 13 :



Gambar 13: Hasil Evaluasi Confusion Matrix

Curva Roc

Kurva ROC adalah alat visualisasi yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan keseimbangan antara tingkat positif sebenarnya (TPR) dan tingkat positif palsu (FPR) pada nilai ambang batas yang berbeda. Untuk hasil kurva ROC bisa di lihat pada Gambar 14:



Gambar 14: Hasil Curva Roc

Penjelasan Kurva ROC:

Kurva ROC (Receiver Operating Characteristic) adalah grafik yang menampilkan kinerja model klasifikasi pada tingkat ambang batas yang berbeda.

Kurva ROC menggambarkan tingkat positif sebenarnya (TPR) versus tingkat positif palsu (FPR) pada pengaturan ambang batas yang berbeda. Tingkat positif sebenarnya (TPR), juga dikenal sebagai sensitivitas atau perolehan, adalah sebuah rasio prediksi benar positif terhadap jumlah total contoh positif.

False Positive Rate (FPR) adalah rasio prediksi salah positif terhadap jumlah total contoh negatif.

Garis diagonal pada plot ROC merepresentasikan kinerja model yang membuat prediksi acak. Semakin dekat kurva ROC ke sudut kiri atas, semakin baik performanya.

Area di bawah kurva (AUC) adalah metrik kinerja yang merangkum kurva ROC menjadi satu nilai, dengan nilai lebih dekat ke 1 menunjukkan kinerja yang lebih baik. Pada plot ini, kita dapat melihat bahwa model SVM kita memiliki AUC sebesar 0.06, yang menunjukkan kinerja Pemodelan dalam

mengidentifikasi perbedaan antara kategori yang dianggap positif dan negatif.

Penutup

Pada penelitian ini, kami berhasil mengembangkan dan mengevaluasi model analisis sentimen menggunakan metode support vector machine (SVM) dengan representasi fitur berbasis Word2Vec untuk menganalisis ulasan pengguna terhadap aplikasi DANA di Google Playstore. Model yang dikembangkan mencapai akurasi 88% dengan rasio data 90:10, menunjukkan kemampuan yang baik dalam mengklasifikasikan sentimen ulasan. Teknologi Word2Vec mengubah kata-kata dalam ulasan menjadi vektor numerik, memungkinkan model mendeteksi dan membedakan secara efektif antara sentimen positif, negatif, dan netral. Penerapan ini tidak hanya membuktikan efektivitasnya dalam menangani penilaian yang beragam dan kompleks secara otomatis, namun juga mengatasi tantangan analisis manual yang cenderung subjektif.

Hasil penelitian ini berkontribusi secara signifikan terhadap pemahaman yang lebih baik tentang reaksi pengguna terhadap aplikasi DANA dan memberikan landasan yang kuat untuk mengembangkan strategi manajemen layanan yang lebih baik di masa depan yang dapat meningkatkan pengalaman pengguna secara keseluruhan.

Penelitian di masa depan akan mencakup pengoptimalan parameter SVM, mempertimbangkan teknik ekstraksi fitur lainnya, menggabungkan data dari platform berbeda, mengembangkan analisis sentimen multibahasa, mempertimbangkan faktor temporal, dan melakukan studi perbandingan dengan metode klasifikasi lain analisis sentimen pengguna.

Ucapan Terimakasih

Penulis ingin mengucapkan terima kasih kepada Universitas Sains dan Teknologi Indonesia atas dukungan yang diberikan dalam pelaksanaan penelitian ini. Terima kasih Bapak Rahmaddeni M.Kom dan Ibu Dr. Lusiana M.Kom atas bimbingan dan kontribusinya yang berharga. Selain itu, penulis juga mengucapkan terima kasih kepada Universitas Sains dan Teknologi Indonesia, para responden, dan seluruh pihak yang telah mendukung penelitian ini. Semua kontribusi dan kerja sama yang diberikan sangat berarti dalam suksesnya pengujian ini.

Daftar Pustaka

- [1] Z. Azindhani, "Penggunaan Aplikasi Dana Sebagai Media Dompot Digital dan Transaksi di Indonesia", *J. Inst. Bisnis dan Teknol. Indones.*, vol. 1, pp. 1–5, 2021.
- [2] A. A. Muhammad, Ermatita, dan D. S. Prasvita, "Analisis Sentimen Pengguna Aplikasi DANA Berdasarkan Ulasan pada Google Play Menggunakan Metode Support Vector Machine", *Prodi S1 Informatika / Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta*, Semin. Nas. Mhs. Ilmu Komput. dan Apl., pp. 194–204, 2022.
- [3] W. Eko Saputro, H. Yuana, dan W. Dwi Puspitarsari, "Analisis Sentimen Pengguna Dompot Digital Dana Pada Kolom Komentar Google Play Store Dengan Metode Klasifikasi Support Vector Machine", *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1151–1156, doi: 10.36040/jati.v7i2.6842, 2023.
- [4] N. D. Abrilia dan S. Tri, "Pengaruh Persepsi Kemudahan Dan Fitur Layanan Terhadap Minat Menggunakan E-Wallet Pada Aplikasi Dana Di Surabaya", *J. Pendidik. Tata Niaga*, vol. 8, no. 3, pp. 1006–1012, 2020.
- [5] F. A. Larasati, D. E. Ratnawati, dan B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest", *Jurnal Teknol. Inf. dan Komputer*, vol. 6, no. 9, pp. 4305–4313, 2022.
- [6] R. D. Wahyuni dan A. N. Utomo, "Penggunaan Metode Lexicon Untuk Analisis Sentimen pada Ulasan Aplikasi KAI Access di Google Play Store", *J. Rekayasa Inf.*, vol. 11, no. 2, pp. 134–145, 2022.
- [7] W. C. Widyaningtyas dan S. Al Faraby, "Klasifikasi Sentiment Analysis pada Review Film Berbahasa Inggris dengan Menggunakan Metode Doc2Vec dan Support Vector Machine (SVM) ", *e-Proceeding of Engineering*, vol. 5, no. 1, pp. 1570–1578, 2018.
- [8] A. Supian, B. T. Revaldo, N. Marhadi, L. Efrizoni, dan Rahmadani, "Perbandingan Kinerja Naïve Bayes dan SVM pada Analisis Sentimen Twitter Ibukota Nusantara", *Jurnal Ilmiah Informatika*, Vol. 12, No.01, pp: 15-21, DOI: 10.33884/jif.v12i01.8721, 2024.
- [9] I. M. Yulietta dan S. Al Faraby, "Klasifikasi Sentimen Review Film Menggunakan Algoritma Support Vector Machine", *e-Proceeding Eng.* vol. 4, no. 3, pp. 4740–4750, 2017.
- [10] A. Ardiansyah, E. A. Pratama, dan N. I. Fadlilah, "Analisis Sentimen Pengguna Terhadap Aplikasi ChatGPT Di Google Play Store: Penerapan Algoritma Support Vector Machine", *J. Tek. Inform. dan Sist. Inf.*, vol. 11, no. 2, pp. 247–254, 2024.
- [11] S. N. Cahyani dan G. W. Saraswati, "Penerapan Metode Support Vector Machine Pada

- Klasifikasi Buku Perpustakaan Sekolah Dengan Kombinasi TF-IDF Dan Word2Vec”, *J. Tek. Inform.*, vol. 4, no. 6, pp. 1555–1566, 2023.
- [12] K. F. Wahyu dan M. Warih Maharani, “Analisis Sentimen Twitter Bahasa Indonesia Menggunakan”, *e-Proceeding Eng.*, vol. 6, no. 2, pp. 7821–7829, 2022.
- [13] D. T. Hermanto, A. Setyanto, dan E. T. Luthfi, “Algoritma LSTM-CNN untuk Binary Klasifikasi dengan Word2vec pada Media Online”, *Creat. Inf. Technol. J.*, vol. 8, no. 1, p. 64, doi: 10.24076/citec.2021v8i1.264, 2021.
- [14] M. K. Rifa, M. H. Totohendarto, dan M. R. Muttaqin,, “Analisis Sentimen Pengguna E-Wallet Dana Dan Gopay Pada Twitter Menggunakan Metode Support Vector Machine (SVM)”, *J. Tek.*, vol. 17, no. x, pp. 323–332, 1978.
- [15] J. Ipmawati, S. Saifulloh, dan K. Kusnawi, “Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine”, *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 247–256, doi: 10.57152/malcom.v4i1.1066, 2024.
- [16] F. Sidik, I. Suhada, A. H. Anwar, dan F. N. Hasan, “Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier”, *J. Linguist. Komputasional*, vol. 5, no. 1, p. 34, doi: 10.26418/jlk.v5i1.79, 2022.
- [17] R. R. Salam, M. F. Jamil, Y. Ibrahim, R. Rahmadden, S. Soni, dan H. Herianto, “Analisis Sentimen Terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan Support Vector Machine”, *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 27–35, doi: 10.57152/malcom.v3i1.590, 2023.
- [18] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, dan M. K. Anam, “Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM”, *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 153–160, doi: 10.57152/malcom.v3i2.897, 2023.
- [19] U. Khaira, R. Johanda, P. E. P. Utomo, and T. Suratno, “Sentiment Analysis Of Cyberbullying On Twitter Using SentiStrength”, *Indones. J. Artif. Intell. Data Min.*, vol. 3, no. 1, p. 21, doi: 10.24014/ijaidm.v3i1.9145, 2020.
- [20] L. Hermawan dan M. Bellanier Ismiati, “Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval”, *J. Transform.*, vol. 17, no. 2, p. 188, doi: 10.26623/ttransformatika.v17i2.1705, 2020.
- [21] N. N. Widyastuti, A. Bijaksana, dan I. L. Sardi, “Analisis Word2vec untuk Perhitungan Kesamaan Semantik antar Kata Pendahuluan Kajian Pustaka”, *e-Proceeding Eng.*, vol. 5, no. 3, pp. 7603–7612, 2018.
- [22] L. A. H. Raudhoti, A. Herdiani, dan A. Romadhony, “Identifikasi Cyberbullying pada Kolom Komentar Instagram dengan Metode Support Vector Machine dan Semantic Similarity”, *J. Comput. Sci. Informatics Eng.*, vol. 4, no. 1, pp. 1–8, doi: 10.29303/jcosine.v4i1.318, 2020.
- [23] P. Ayuningtyas dan H. Tantyoko, “Perbandingan Metode Word2vec Model Skipgram pada Ulasan Aplikasi Linkaja menggunakan Algoritma Bidirectional LSTM dan Support Vector Machine”, *JUSTIN (Jurnal Sist. dan Teknol. Informasi)*, vol. 12, no. 1, pp. 189–196, doi: 10.26418/justin.v12i1.72530, 2024.
- [24] F. S. Jumeilah, “Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian”, *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 1, pp. 19–25, doi: 10.29207/resti.v1i1.11, 2017.
- [25] E. Susilowati, M. K. Sabariah, and A. A. Gozali, “Implementasi Metode Support Vector Machine untuk Melakukan Klasifikasi Kemacetan Lalu Lintas Pada Twitter”, *E-Proceeding Eng.*, vol. 2, no. 1, pp. 1478–1484, 2015.
- [26] Auliya Rahman Isnain, Adam Indra Sakti, Debby Alita, dan Nurman Satya Marga, “Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM”, *Jurnal Data Mining dan Sistem Informasi (JDMSI)*, vol. 2, no. 1, pp. 31–37, DOI: <https://doi.org/10.33365/jdmsi.v2i1.1021>, 2021.