

SENTIMENT ANALYSIS OF ONLINE DATING APPS USING SUPPORT VECTOR MACHINE AND NAÏVE BAYES ALGORITHMS

Urip Hadi Laksono^{*1}, Ryan Randy Suryono^{*2}

^{2,3}Information System, Universitas Teknokrat Indonesia, Indonesia
Email: ¹urip_hadi_laksono@teknokrat.ac.id, ²ryan@teknokrat.ac.id

(Article received: May 17, 2024; Revision: May 30, 2024; published: February 20, 2025)

Abstract

In daily life, the use of digital applications is increasingly widespread, making dating apps increasingly popular and an important part of modern social interaction. This research aims to analyze user sentiment towards online dating apps, specifically Tinder, using Support Vector Machine (SVM) and Naïve Bayes algorithms. The problem underlying the importance of this research is the lack of balance between positive and negative sentiments in Tinder app users, which can affect user experience and the quality of service provided by Tinder. Utilizing the CRISP-DM framework, this research involves six stages, from data collection to evaluation. The results showed a significant imbalance between the number of positive and negative sentiments before optimization, but after the application of the SMOTE technique, there was a balancing between the two sentiment categories. SVM achieved 85% accuracy, while Naïve Bayes achieved 84%, with similar performance in identifying positive and negative sentiments. While both models performed satisfactorily, SVM appeared more stable in recognizing both positive and negative sentiments, suggesting the potential to be a superior choice in the context of dating apps. As such, this research makes an important contribution to the understanding of users' views on Tinder apps and provides a basis for further development.

Keywords: Dating Apps, Naïve Bayes, Sentimen Analysis, Support Vector Machine, SMOTE Optimization.

ANALISIS SENTIMEN APLIKASI KENCAN ONLINE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN NAÏVE BAYES

Abstrak

Dalam kehidupan sehari-hari, penggunaan aplikasi digital semakin meluas, membuat aplikasi kencan menjadi semakin populer dan menjadi bagian penting dari interaksi sosial modern. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi kencan *online*, khususnya Tinder, menggunakan algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes*. Permasalahan yang mendasari pentingnya penelitian ini adalah kurangnya keseimbangan antara sentimen positif dan negatif dalam pengguna aplikasi Tinder, yang dapat memengaruhi pengalaman pengguna dan kualitas layanan yang disediakan oleh Tinder. Dengan memanfaatkan kerangka kerja CRISP-DM, penelitian ini melibatkan enam tahapan, mulai dari pengumpulan data hingga evaluasi. Hasil penelitian menunjukkan adanya ketidakseimbangan yang signifikan antara jumlah sentimen positif dan negatif sebelum optimasi, namun setelah penerapan teknik SMOTE, terjadi penyeimbangan antara kedua kategori sentimen. SVM mencapai akurasi sebesar 85%, sementara *Naïve Bayes* mencapai 84%, dengan performa serupa dalam mengidentifikasi sentimen positif dan negatif. Meskipun kedua model menunjukkan kinerja yang memuaskan, SVM terlihat lebih stabil dalam mengenali sentimen positif dan negatif, menunjukkan potensi untuk menjadi pilihan yang lebih unggul dalam konteks aplikasi kencan. Dengan demikian, penelitian ini memberikan kontribusi penting dalam pemahaman terhadap pandangan pengguna terhadap aplikasi Tinder serta memberikan dasar bagi pengembangan lebih lanjut dalam meningkatkan kualitas layanan Tinder dan memperkuat daya saing aplikasi kencan di pasar yang semakin kompetitif.

Kata kunci: Analisis sentimen, Dating Apps, Naïve Bayes, Optimasi SMOTE, Support Vector Machine.

1. PENDAHULUAN

Dalam beberapa tahun terakhir, ada peningkatan yang signifikan dalam penggunaan platform *online* untuk kencan [1]. Kencan *online* menggambarkan

cara membangun hubungan romantis melalui internet. Segala sesuatu kini dilakukan dengan cepat, mudah, dan efisien, mempersingkat waktu dan mengakibatkan terjadinya akulturasi budaya dalam

praktik kencan [2]. Motivasi di balik penggunaan kencan *online* sangat beragam, dari mencari pasangan hidup untuk hubungan jangka panjang, hingga mencari teman obrolan daring atau hanya untuk *flirting* [3]. Sebelum kemunculan aplikasi kencan, orang sering bergantung pada jaringan sosial tradisional atau perantara seperti teman atau keluarga untuk menemukan pasangan hidup.

Namun, dengan adanya aplikasi kencan *online*, strategi dalam mencari pasangan mengalami perubahan signifikan ke arah yang lebih praktis dan efisien di era digital. Hal ini memungkinkan individu untuk mengidentifikasi calon pasangan potensial dengan cepat melalui perangkat seluler mereka, tanpa adanya kendala waktu dan tempat serta tanpa perlu melakukan pertemuan langsung [4]. Sejak popularitas pencarian jodoh secara *online* meningkat, aplikasi seperti Tinder telah menjadi pilihan utama bagi banyak individu, terutama remaja, untuk menjalin hubungan dengan orang baru dan mencari pasangan [5].

Tinder adalah platform media sosial khusus untuk mencari pasangan, yang menjadi fitur utama aplikasinya [6]. Diluncurkan pada tahun 2012 oleh Sean Rad, Justin Mateen, dan Jonathan Badeen, Tinder memperkenalkan cara inovatif untuk membantu masyarakat menemukan pasangan atau teman baru secara online, tanpa perlu bertemu langsung [7]. Menurut hasil survei, Tinder merupakan aplikasi kencan terbaik dengan tingkat keberhasilan 16,5%. *Plenty of Fish* berada di posisi kedua dengan tingkat keberhasilan sebesar 11,18%, diikuti oleh *Bumble* di posisi ketiga dengan tingkat keberhasilan sebesar 5,75%. *Match.com* berada di posisi keempat dengan tingkat keberhasilan sebesar 5,64%, dan *Badoo* menempati posisi kelima dengan tingkat keberhasilan sebesar 4,91% [8]. Oleh karena itu aplikasi Tinder menjadi fokus utama pada penelitian ini.

Ulasan-ulasan yang diberikan oleh pelanggan dapat dikumpulkan dan diinterpretasikan sebagai sentimen [9]. Dalam konteks bisnis, analisis ulasan tersebut dapat memberikan panduan untuk meningkatkan kualitas layanan [10]. Analisis sentimen adalah proses memproses data teks untuk mengenali dan memahami sentimen yang terkandung di dalamnya [11]. Penggunaannya meliputi analisis ulasan pengguna di *Google Play Store* untuk mendapatkan wawasan yang berguna. Metode ini, yang juga dikenal sebagai penambangan opini, secara otomatis mengambil informasi dari teks untuk menentukan apakah opini tersebut bersifat positif atau negatif [12]. Penelitian ini menganalisis sentimen penggunaan aplikasi Tinder menggunakan *Support Vector Machine* (SVM) dan algoritma *Naive Bayes*. Kedua algoritma tersebut telah terbukti dapat mengevaluasi sentimen teks, termasuk penelusuran sentimen terhadap ulasan atau umpan balik pengguna terhadap aplikasi [13].

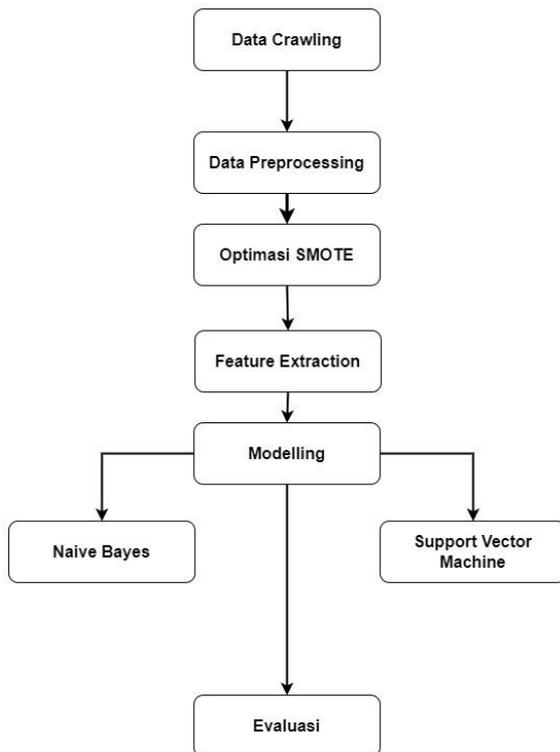
Penelitian terkait analisis sentimen sebelumnya telah dilakukan menunjukkan bahwa Tik Tok, meskipun populer, sering kali digunakan oleh anak-anak di bawah umur, dengan dampak negatif yang mungkin terjadi. Namun, beberapa orang juga melihat manfaatnya dalam dunia bisnis. Untuk menggali lebih dalam tentang pendapat masyarakat, dilakukan analisis sentimen terhadap ulasan pengguna. Hasilnya, 76,7% dari 2000 data diklasifikasikan sebagai ulasan positif dan 23,3% sebagai ulasan negatif. Metode *Naive Bayes* mencapai akurasi 79%, sementara SVM mencapai 84%, menunjukkan keunggulan SVM dalam mengklasifikasikan sentimen dengan lebih tepat [14].

Analisis sentimen telah digunakan dalam penelitian sebelumnya untuk memahami opini masyarakat terhadap Shopee. Dalam penelitian ini, algoritma *Naive Bayes* dan *Support Vector Machine* digunakan untuk mengklasifikasikan sentimen. Dari pengujian dengan 2000 data, *Naive Bayes* mencapai akurasi 85%, sedangkan *Support Vector Machine* mencapai 81%. Temuan ini konsisten dengan penelitian sebelumnya yang menunjukkan keunggulan *Naive Bayes* dalam mengklasifikasikan sentimen dibandingkan dengan *Support Vector Machine* [15].

Adapun tujuan dari penelitian ini yaitu untuk mengevaluasi efektivitas dua metode klasifikasi dalam mengidentifikasi sentimen pengguna terhadap aplikasi Tinder. Melalui hal ini, diharapkan akan tercapai pemahaman yang lebih mendalam tentang pandangan pengguna terhadap aplikasi *dating* ini. Selain itu, hasil penelitian ini diharapkan dapat memberikan masukan berharga bagi pengembang aplikasi untuk meningkatkan kualitas layanan mereka.

2. METODE PENELITIAN

Penelitian ini memulai prosesnya dengan menerapkan pendekatan metodologi yang mengadopsi kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang terdiri dari enam tahapan yang telah ditetapkan. Tahap awal adalah pengumpulan data (data crawling), langkah selanjutnya adalah tahap pra-pemrosesan data (data *preprocessing*). Kemudian, diterapkanlah teknik optimasi SMOTE untuk menangani ketidakseimbangan kelas dalam data. Proses ekstraksi fitur (*feature extraction*) berikutnya dilakukan untuk mengekstraksi fitur-fitur yang penting dari data yang telah diproses. Selanjutnya, dua model klasifikasi, yaitu *Naive Bayes* dan SVM, diterapkan untuk melakukan analisis sentimen. Pada tahap akhir, dilakukan evaluasi untuk mengevaluasi kinerja dari model-model yang telah dikembangkan. Gambar 1 merupakan alur dari tahapan penelitian.



Gambar 1. Tahapan Penelitian

2.1. Data Crawling

Tahap awal dalam eksplorasi ini dimulai dengan pengumpulan data atau proses *crawling*, yang berperan penting dalam menghimpun kumpulan data yang relevan untuk penelitian [15]. Data yang berhasil dihimpun terdiri dari sejumlah ulasan pengguna aplikasi Tinder di *Google Play Store*, yang mengungkapkan pandangan mereka terhadap pengalaman menggunakan aplikasi tersebut dengan menggunakan kata kunci "Tinder". Proses pengambilan data ulasan ini dilakukan dalam rentang waktu tertentu, mulai dari September 2018 hingga April 2024. Jumlah total ulasan yang berhasil terkumpul mencapai 2787. Data ini akan diolah lebih lanjut untuk memfasilitasi analisis sentimen selanjutnya.

2.2. Data Preprocessing

Preprocessing merupakan langkah yang umum dilakukan sebelum melakukan pengelompokan dokumen. Salah satu tujuan dari *preprocessing* adalah untuk menghasilkan dataset yang seragam agar dapat diproses dengan lebih mudah [16]. Tahapan *preprocessing* melibatkan beberapa tahap yakni *cleansing*, *casefolding*, *filtering and stopword*, *tokenizing*, *stemming*, dan *labeling*.

2.3. Optimasi SMOTE

Synthetic Minority Oversampling Technique (SMOTE) adalah salah satu implementasi dari metode *oversampling*. Keunggulan utama dari metode ini adalah tidak menghasilkan informasi yang

hilang. SMOTE dapat meningkatkan akurasi prediksi data dengan menciptakan representasi baru dari kelas minoritas, memperluas pemahaman dan kemampuan model dalam memprediksi kelas minoritas. SMOTE mengatasi ketidakseimbangan data dengan menciptakan data sintesis dari kelas minoritas, yang sering kali memiliki jumlah data yang terbatas [17].

2.4. Feature Extraction

Pada tahap ekstraksi fitur, setiap token kata yang telah diproses melalui tahap *preprocessing* akan diubah menjadi vektor. Vektor ini akan memvisualisasikan kata-kata yang ada setelah melewati proses *preprocessing* [18]. Teknik yang digunakan pada penelitian ini adalah Tf-IDF. Tf-IDF memberikan informasi tentang seberapa penting setiap kata dalam dokumen saat mengekstrak fitur dari dokumen. Pada dasarnya, kata atau frasa dianggap sebagai penanda yang baik untuk klasifikasi jika sering muncul dalam satu kelas dokumen tetapi jarang muncul dalam kelas lain [19]. Persamaan 1, 2, dan 3 merupakan perhitungan Tf-IDF:

- Frekuensi Term (TF)

$$TF(t, d) = \frac{\text{Jumlah kemunculan term } t \text{ dalam dokumen } d}{\text{Total jumlah term dalam dokumen } d} \quad (1)$$

- Invers Dokumen Frekuensi (IDF)

$$IDF(t, D) =$$

$$\log\left(\frac{\text{Total jumlah dokumen dalam } d}{\text{Jumlah dokumen dalam } D \text{ yang mengandung term } t} + 1\right) \quad (2)$$

- TF-IDF

$$TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (3)$$

2.4.1. Modeling

Dalam tahap pemodelan ini, menggunakan dua algoritma klasifikasi, yaitu *Support Vector Machine* (SVM) dan *Naïve Bayes*, untuk menganalisis sentimen dari ulasan pengguna aplikasi Tinder.

2.4.2. Support Vector Machine

Support Vector Machine (SVM) adalah metode prediksi untuk klasifikasi dan regresi yang termasuk dalam kategori *supervised learning*. Proses implementasinya melibatkan tahap pelatihan menggunakan *sequential training* SVM dan dilanjutkan dengan pengujian. SVM mencari *hyperplane* optimal untuk memisahkan kelas data, dapat menangani dimensi tinggi dengan kernel trik, dan menggunakan titik data terpilih sebagai vektor pendukung untuk model klasifikasi [20]. Performa model yang dihasilkan oleh proses transformasi dengan SVM sangat dipengaruhi oleh pilihan fungsi kernel dan pengaturan parameter yang digunakan [21].

2.4.3. Naïve Bayes

Naïve Bayes, merupakan sebuah metode statistik untuk mengestimasi probabilitas bahwa suatu objek akan masuk ke dalam kelas tertentu. Pendekatan ini berakar pada Teorema Bayes, yang memiliki kemampuan klasifikasi yang sebanding dengan *decision tree* dan jaringan saraf [22]. Rumus Naïve bayes dijelaskan pada persamaan 4 berikut ini.

$$P(H|X) = \frac{P(H|X) P(H)}{P(X)} \quad (4)$$

Persamaan diatas merupakan rumus dari Teorema Bayes, yang merupakan dasar dari *Bayesian classification*. Dalam konteks ini, X merupakan data class yang belum diketahui, sedangkan H adalah hipotesa class yang spesifik terkait dengan X. $P(H | X)$ adalah probabilitas bahwa hipotesa H benar berdasarkan kondisi X, sementara $P(H)$ adalah probabilitas prior dari hipotesa H sebelum melihat data X. $P(X | H)$ adalah probabilitas dari X berdasarkan kondisi H. Rumus ini menyajikan cara untuk memperbarui atau menyesuaikan probabilitas prior $P(H)$ dengan informasi baru yang diberikan oleh data X, sehingga memungkinkan kita untuk membuat estimasi yang lebih akurat tentang kelas atau hipotesa yang mungkin benar berdasarkan data yang ada.

2.5. Evaluation

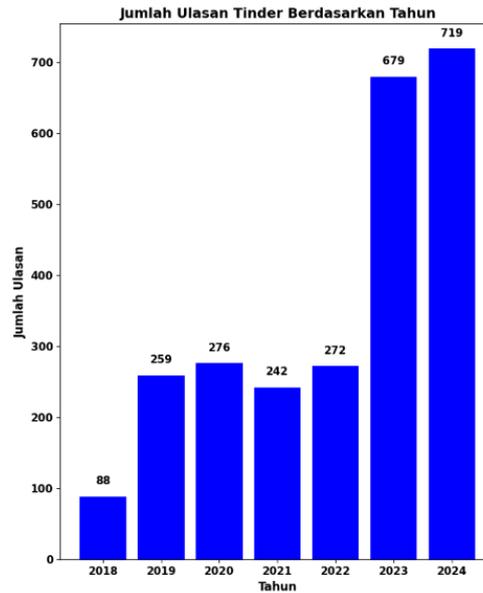
Pada tahap evaluasi, performa algoritma klasifikasi yang digunakan dalam penelitian dievaluasi untuk mengetahui seberapa baiknya. Evaluasi ini melibatkan penggunaan metrik seperti akurasi, *recall*, *precision*, dan *f1-score*. Untuk menghitung metrik-metrik tersebut, digunakan metode *confusion matrix* [23].

3. HASIL DAN PEMBAHASAN

3.1. Dataset

Total ulasan yang berhasil terkumpul dari rentang waktu September 2018 hingga April 2024 sebanyak 2787. Gambar 2 merupakan hasil visualisasi dari jumlah ulasan pengguna tinder berdasarkan gambar 2.

Tren pertumbuhan jumlah ulasan pengguna aplikasi Tinder dari *Google Play Store*, terjadi peningkatan yang signifikan dari tahun 2018 hingga 2024. Dari jumlah ulasan yang relatif rendah pada tahun 2018 sebanyak 88, terjadi lonjakan yang konsisten hingga mencapai puncaknya pada tahun 2024 dengan 719 ulasan. Fenomena ini mencerminkan adanya pertumbuhan yang substansial dalam penggunaan dan popularitas aplikasi Tinder seiring berjalannya waktu, serta menunjukkan bahwa aplikasi tersebut mempertahankan ketertarikannya di kalangan pengguna.



Gambar 2. Jumlah ulasan pengguna Tinder berdasarkan tahun

3.2. Data Preprocessing

Data yang dihasilkan dari proses *crawling* masih dalam bentuk yang belum diolah. Untuk mengatasinya, diperlukan langkah-langkah *preprocessing* guna mengubah data mentah tersebut menjadi format yang lebih terstruktur. Berikut ini merupakan langkah-langkah yang dilakukan dalam proses *preprocessing* data:

3.2.1. Cleansing

Dalam tahap ini, segala karakter kecuali huruf akan dihapus dari data, termasuk tanda baca dan angka [24].

3.2.2. Casefolding

Case folding adalah proses mengubah huruf-huruf alfabet dalam teks menjadi huruf kecil. Hal ini penting untuk memastikan konsistensi dalam analisis teks, sehingga kata-kata yang sama diakui tanpa memperhatikan huruf besar atau kecilnya [25].

3.2.3. Tokenize

Tokenisasi merupakan Langkah dalam memisahkan kata-kata dalam sebuah teks. Pada tahap ini, setiap bagian kata dipisahkan dan disebut sebagai token [26].

3.2.4. Filtering

Filtering merupakan tahap penting dalam persiapan data yang melibatkan penghapusan elemen-elemen yang tidak diinginkan serta pemberian parameter berdasarkan analisis. Beberapa langkah yang dilakukan dalam proses filtering data termasuk penghapusan URL, mencatat, dan tag, penghilangan kata-kata yang tidak relevan, serta transformasi huruf besar menjadi huruf kecil [27].

3.2.5. Stemming

Stemming merupakan kelanjutan dari proses seleksi hasil. Pada tahap ini, berbagai teknik diaplikasikan untuk mengubah kata-kata ke dalam bentuk yang seragam. Namun, perlu diingat bahwa penerapan stemming pada kata-kata berbahasa Inggris mungkin lebih tepat daripada pada kata-kata berbahasa Indonesia [28]. Seperti yang ditunjukkan pada Tabel 1 yang merupakan hasil dari tahapan preprocessing.

Tabel 1. Hasil preprocessing

Process	Result
Dataset	aplikasi ini sangat bagus ..kalau kalian mau download aplikasi ini ..kalian harus download juga auto like swipe☺□
Cleansing	aplikasi ini sangat bagus kalau kalian mau download aplikasi ini kalian harus download juga auto like swipe
Casefolding	aplikasi ini sangat bagus kalau kalian mau download aplikasi ini kalian harus download juga auto like swipe
Tokenize	['aplikasi', 'ini', 'sangat', 'bagus', 'kalau', 'kalian', 'mau', 'download', 'aplikasi', 'ini', 'kalian', 'harus', 'download', 'juga', 'auto', 'like', 'swipe']
Filtering	['aplikasi', 'bagus', 'download', 'aplikasi', 'download', 'auto', 'like', 'swipe']
Stemming	aplikasi bagus download aplikasi download auto like swipe

3.2.6. Labeling

Tahap berikutnya adalah labeling, dimana data diberi tanda atau penilaian emosional yang sesuai dengan konteksnya. Penilaian ini bisa ulasan yang mencerminkan pendapat positif atau negatif, berdasarkan skala bintang yang diberikan oleh pengguna. Tabel 2 memperlihatkan hasil dari proses hasil dari tahapan labeling.

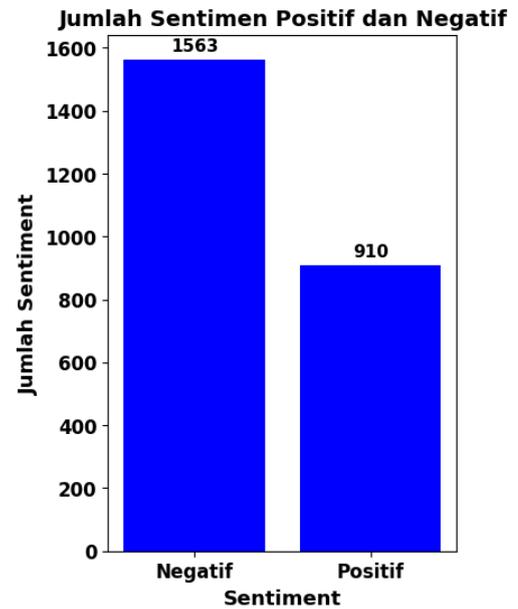
Tabel 2. Hasil labeling

Ulasan	Label
aplikasi ini sangat bagus ..kalau kalian mau download aplikasi ini ..kalian harus download juga auto like swipe☺□	Positif
setiap mau login gabisa, tulisannya malah ada masalah saat masuk tinder, silahkan coba lagi. udah coba beberapa kali tapi tetep gabisa, padahal ga lupa password kok:')	Negatif

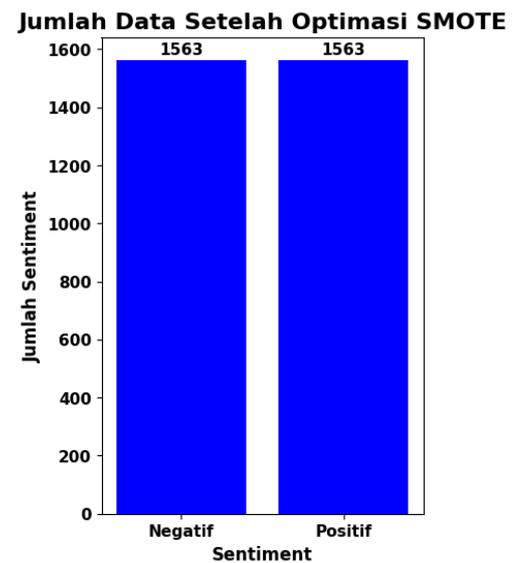
3.3. Tahap pengujian dan Evaluasi

Hasil dari pengujian menunjukkan adanya ketimpangan yang signifikan antara jumlah sentimen positif dan negatif, dengan 910 ulasan yang bersifat positif dan 1563 ulasan yang bersifat negatif. Untuk mengatasi ketidakseimbangan ini, dilakukan optimasi dengan teknik SMOTE (*Synthetic Minority Over-sampling Technique*). Setelah dilakukan optimasi menggunakan metode SMOTE, terjadi penyeimbangan antara jumlah ulasan dengan sentimen positif dan negatif. Saat ini, masing-masing kategori memiliki 1563 ulasan, menciptakan kesetimbangan yang lebih baik dalam dataset. Gambar 3 dan Gambar 4 menunjukkan representasi

data sebelum dan sesudah dilakukan optimasi menggunakan SMOTE.



Gambar 3. Jumlah sentimen positif dan negatif



Gambar 4. Jumlah sentimen positif dan negatif setelah optimasi SMOTE

Dalam penelitian ini menggunakan sekitar 1978 dari total 2473 dataset sebagai data pelatihan, yang setara dengan sekitar 80% dari total. Sementara itu, sekitar 495 sisanya, yang setara dengan sekitar 20%, kami alokasikan sebagai data uji. Keputusan untuk memprioritaskan jumlah data latih lebih besar daripada data uji didasarkan pada strategi untuk memberikan model kesempatan yang lebih besar untuk belajar dari beragam pola dan tren dalam data. Dengan memberikan lebih banyak data untuk pelatihan, diharapkan model dapat mengembangkan pemahaman yang lebih mendalam dan akurat tentang karakteristik sentimen yang tersembunyi dalam dataset, sehingga meningkatkan kinerjanya dalam mengklasifikasikan data baru.

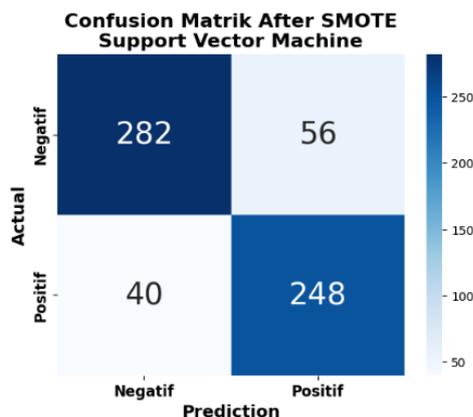
3.3.1. Support Vector Machine (SVM)

Hasil *classification report* SVM setelah penerapan SMOTE menunjukkan akurasi sebesar 85%, di mana klasifikasi sentimen negatif memiliki tingkat presisi sebesar 88%, *recall* sebesar 83%, dan *F1-score* sebesar 85%, menunjukkan kemampuan model dalam mengidentifikasi sentimen negatif dengan baik namun dengan beberapa sampel yang salah diklasifikasikan. Sementara itu, klasifikasi sentimen positif menampilkan presisi sebesar 82%, *recall* sebesar 86%, dan *F1-score* sebesar 84%, menunjukkan keseimbangan yang baik antara presisi dan *recall* untuk klasifikasi ini. Secara keseluruhan, hasil ini menggambarkan kinerja yang solid dari model SVM setelah penerapan SMOTE dalam mengklasifikasikan sentimen positif dan negatif, dengan akurasi yang tinggi dan nilai *F1-score* yang seimbang untuk kedua klasifikasi. Hasil klasifikasi algoritma SVM dapat dilihat pada Tabel 3.

Tabel 3. Hasil *classification report* SVM

Support Vector Machine after SMOTE				
	Accuracy	Precision	recall	F1-score
Negatif	85%	88 %	83%	85%
Positif		82 %	86%	84%

Gambar 5 merupakan dari hasil *confusion matrix* untuk model SVM, tampak bahwa model mampu dengan akurat mengklasifikasikan 248 sampel sebagai *True Positives* (TP) yang merupakan sentimen positif, dan 282 sampel sebagai *True Negatives* (TN) yang merupakan sentimen negatif. Namun, terdapat 40 sampel yang salah diklasifikasikan sebagai *False Positives* (FP), dan 56 sampel yang seharusnya positif namun salah diklasifikasikan sebagai *False Negatives* (FN). Hal ini mengindikasikan bahwa model memiliki kecenderungan untuk mengalami kesulitan dalam mengenali sentimen positif, yang tercermin dari jumlah FN yang signifikan, meskipun mampu mempertahankan tingkat akurasi yang relatif tinggi. Evaluasi ini memberikan wawasan mendalam tentang area-area di mana model dapat ditingkatkan, seperti meningkatkan kemampuan untuk mengenali sentimen positif yang lebih baik guna meningkatkan performa keseluruhan model.



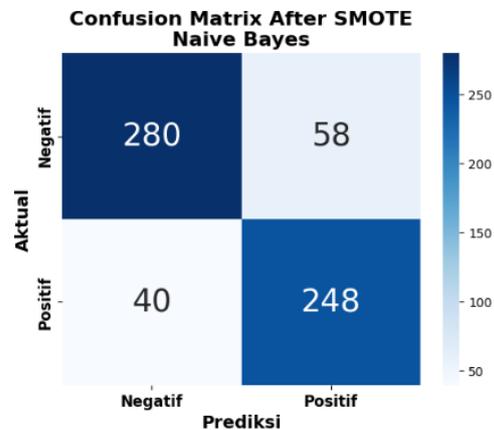
Gambar 5. Hasil *Confusion Matrix* SVM

3.3.2. Naïve Bayes

Tabel 4 yang memberikan hasil *classification report* pasca penerapan SMOTE, dengan tingkat akurasi mencapai 84%. Klasifikasi sentimen negatif, model menunjukkan tingkat presisi sebesar 88%, *recall* sebesar 83%, dan *F1-score* sebesar 85%. Temuan ini mengindikasikan bahwa model mampu dengan baik mengidentifikasi sentimen negatif, dengan sebagian besar prediksi yang tepat dan kemampuan untuk menangkap sebagian besar sampel yang sebenarnya adalah sentimen negatif. Namun, pada klasifikasi sentimen positif, presisi mencapai 81%, dengan *recall* sebesar 86%, dan *F1-score* sebesar 84%. Meskipun presisi dalam mengklasifikasikan sentimen positif menunjukkan kinerja yang baik, *recall* yang lebih tinggi menunjukkan bahwa model cenderung lebih baik dalam mendeteksi sampel yang sebenarnya positif. Hasil evaluasi ini memberikan gambaran mendalam tentang kinerja model *Naïve Bayes* setelah penerapan SMOTE, menyoroti kelebihan dan kekurangan dalam mengenali sentimen positif dan negatif, serta memberikan dasar bagi peningkatan lebih lanjut dalam penelitian.

Tabel 4. Hasil *classification report* Naïve Bayes

Naïve Bayes after SMOTE				
	Accuracy	Precision	recall	F1-score
Negatif	84%	88 %	83%	85%
Positif		81 %	86%	84%



Gambar 6. Hasil *Confusion Matrix* Naive Bayes

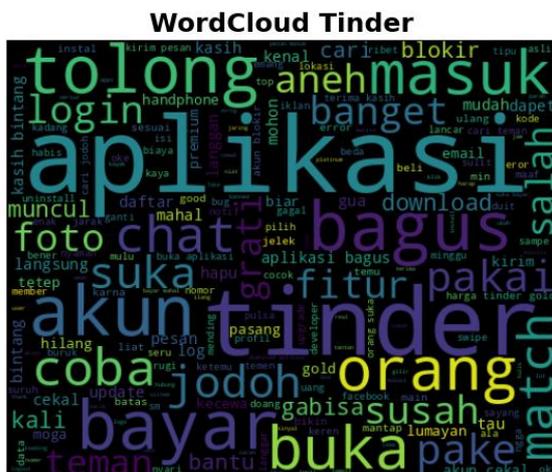
Gambar 6 menunjukkan bahwa dalam *confusion matrix* untuk model Naïve Bayes, terlihat model berhasil mengklasifikasikan 248 sampel sebagai *True Positives* (TP), menunjukkan kemampuannya dalam mengidentifikasi sentimen positif dengan tepat. Namun, terdapat 40 sampel yang salah diklasifikasikan sebagai *False Positives* (FP), mengindikasikan adanya kesalahan dalam mengklasifikasikan sampel yang seharusnya negatif sebagai positif. Sementara itu, model juga berhasil mengklasifikasikan 280 sampel sebagai *True Negatives* (TN), menunjukkan kemampuannya dalam mengenali sentimen negatif dengan benar. Namun, terdapat 58 sampel yang salah diklasifikasikan

sebagai *False Negatives* (FN), menunjukkan kecenderungan model untuk mengabaikan atau kesulitan dalam mengenali sampel yang seharusnya positif.

3.4. Visualisasi Data

3.4.1. WordCloud

Dari hasil *WordCloud* aplikasi Tinder, terlihat bahwa kata "aplikasi" dan "Tinder" mendominasi, menandakan fokus yang kuat pada platform tersebut. Kehadiran kata "akun" menyoroti perhatian terhadap aspek teknis seperti pembuatan dan pengelolaan akun. Ukuran yang signifikan dari kata "bagus" mengindikasikan bahwa banyak pengguna memiliki persepsi positif terhadap Tinder, mungkin merujuk pada pengalaman yang memuaskan atau fitur-fitur yang disukai. Kemudian, keberadaan kata "orang" menunjukkan bahwa percakapan cenderung berkesan pada interaksi dengan individu yang ditemui melalui aplikasi tersebut. Kata "jodoh" menyoroti tujuan utama pengguna untuk menemukan pasangan hidup melalui Tinder, sementara kata "bayar" menunjukkan kesadaran terhadap aspek keuangan seperti model pembayaran atau fitur berbayar. Secara keseluruhan, analisis ini menggambarkan beragam topik yang dibicarakan oleh pengguna Tinder, dari pengalaman pengguna hingga harapan untuk menemukan hubungan yang berarti. Visualisasi wordcloud Aplikasi Tinder dapat dilihat pada Gambar 7.



Gambar 7. Hasil *WordCloud* Aplikasi Tinder

Dalam analisis *WordCloud* sentimen positif terhadap aplikasi Tinder, kata-kata yang paling menonjol adalah "aplikasi", "Tinder", "tolong", "bagus", "jodoh", "suka", dan "bantu". Ukuran kata-kata tersebut menunjukkan tingkat frekuensi kemunculannya dalam teks yang dianalisis. Dari kata-kata tersebut, dapat disimpulkan bahwa pengguna merasa positif terhadap pengalaman mereka dengan aplikasi Tinder, menganggapnya sebagai alat yang bermanfaat dalam mencari jodoh. Kemunculan kata "tolong" dan "bantu" mungkin menunjukkan bahwa pengguna merasa aplikasi ini memberikan dukungan

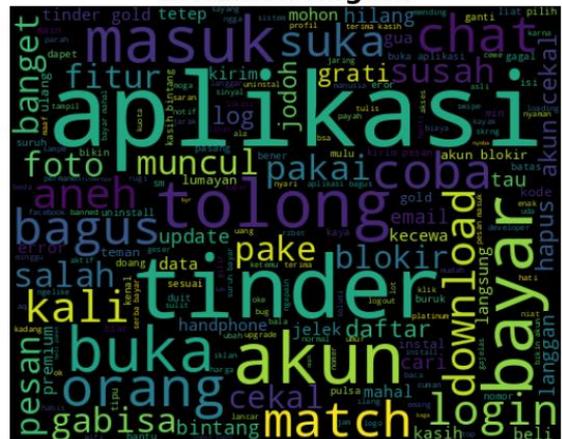
atau bantuan dalam perjalanannya mencari pasangan. Adanya kata "bagus" juga menunjukkan bahwa pengguna menyukai fitur atau pengalaman yang diberikan oleh aplikasi. Namun, untuk pemahaman yang lebih mendalam, perlu diperhatikan konteks di mana kata-kata ini muncul, seperti dalam ulasan pengguna atau percakapan, untuk memahami lebih lanjut tentang pengalaman individual dan harapan pengguna terhadap aplikasi ini. Gambar 8 menunjukkan visualisasi wordcloud aplikasi tinder sentimen positif.

WordCloud Aplikasi Tinder Sentimen Positif



Gambar 8. Hasil *WordCloud* Aplikasi Tinder sentimen positif

WordCloud Aplikasi Tinder Sentimen Negatif



Gambar 9. Hasil *WordCloud* Aplikasi Tinder sentimen negatif.

Analisis *WordCloud* menyoroti pengalaman negatif pengguna Tinder, dengan kata-kata seperti "gabisa", "susah", "hilang", dan "blokir" menonjol. Hal ini menunjukkan adanya kesulitan teknis dan interaksi sosial yang dihadapi pengguna, mungkin terkait dengan kehilangan koneksi atau diblokir oleh pengguna lain. Selain itu, kata "tolong" mencerminkan permintaan bantuan atau ketidakpuasan terhadap berbagai masalah yang mungkin mereka hadapi, sementara kata "bayar" menyoroti kemungkinan ketidakpuasan terkait dengan model pembayaran atau biaya terkait.

Pengguna juga menyebutkan pengalaman "aneh", mungkin merujuk pada interaksi yang tidak diinginkan atau tidak biasa. Analisis ini menggambarkan kompleksitas tantangan teknis dan sosial yang dihadapi pengguna Tinder, menekankan perlunya perbaikan dan peningkatan dalam pengalaman pengguna untuk mengatasi masalah yang teridentifikasi. Gambar 9 memberikan visualisasi yang jelas tentang aplikasi tinder sentimen negative.

4. DISKUSI

4.1. Penelitian Sebelumnya

Penelitian sebelumnya yang serupa juga mengulas analisis sentimen Tik Tok menggunakan algoritma yang sama, hasilnya menunjukkan pola yang serupa, dengan SVM unggul dengan akurasi 84% dibandingkan dengan *Naïve Bayes* yang mencapai 79%. Perbedaan ini menyoroti pentingnya pemilihan model dalam konteks aplikasi yang spesifik. Meskipun demikian, keputusan akhir harus mempertimbangkan bukan hanya tingkat akurasi, tetapi juga stabilitas dan kebutuhan analisis yang terkait dengan platform media sosial [12].

Selain itu penelitian terdahulu mengenai Media berita online mengungguli media tradisional seperti media elektronik dan cetak dalam popularitasnya karena kepraktisan, kecepatan *real-time*, dan kelengkapan informasinya. Penilaian masyarakat terhadap layanan dan berita yang disajikan oleh media online sangat penting untuk meningkatkan kualitasnya, yang dapat dilihat melalui ulasan pengguna di Google Play. Analisis sentimen terhadap ulasan tersebut dilakukan untuk memahami pandangan publik terhadap aplikasi media berita online, dengan menerapkan algoritma *Naïve Bayes* dan SVM. Hasil eksperimen menunjukkan bahwa SVM memiliki akurasi lebih tinggi (88%) dibandingkan *Naïve Bayes* (87%), dan mayoritas opini publik cenderung positif terhadap aplikasi media berita online, dengan jumlah ulasan positif yang jauh lebih banyak daripada ulasan negatif [29].

Penelitian sebelumnya menunjukkan bahwa kedatangan pengungsi Rohingya di Indonesia menjadi isu kontroversial yang menimbulkan berbagai reaksi dari masyarakat. Analisis sentimen publik terhadap kedatangan pengungsi ini penting untuk memahami dinamika opini dan sikap masyarakat Indonesia. Untuk mencapai hasil yang akurat, pemilihan metode analisis sentimen menjadi krusial. Penelitian ini bertujuan untuk menganalisis sentimen publik terkait kedatangan pengungsi Rohingya dengan menggunakan metode Support Vector Machine (SVM) dan *Naïve Bayes*. Penelitian ini melibatkan analisis terhadap 3350 tweet menggunakan pembagian data 70:30, di mana 70% data digunakan untuk pelatihan dan 30% untuk pengujian model. Hasilnya menunjukkan bahwa SVM memiliki akurasi 76%, sedangkan *Naïve Bayes* mencapai akurasi 70%, menunjukkan bahwa SVM

lebih efektif dalam memprediksi sentimen publik dengan kesalahan yang lebih rendah dibandingkan *Naïve Bayes* [30].

Penelitian terdahulu menyoroti transformasi signifikan dalam interaksi manusia dengan dunia digital melalui fenomena *metaverse*. Pemerintah dan industri kini berusaha memahami arah pengembangan *metaverse* untuk menerapkan digitalisasi dalam berbagai sektor. Studi ini menggunakan data dari media sosial X untuk melakukan analisis sentimen publik terhadap *metaverse*. Dua model klasifikasi teks, yaitu Support Vector Machine dan Random Forest, digunakan dalam analisis ini, dan optimasi SMOTE diterapkan untuk meningkatkan kinerja model. Hasil penelitian menunjukkan bahwa model Random Forest mencapai akurasi 91%, sedangkan Support Vector Machine mencapai akurasi 90%. Ini menunjukkan bahwa Random Forest lebih unggul dibandingkan Support Vector Machine. Selain itu, penerapan SMOTE terbukti meningkatkan kemampuan model dalam mengenali sentimen positif pada kedua algoritma, meskipun terdapat *trade-off* antara *recall* dan *precision* [31].

4.2. Interpretasi Hasil

Dari hasil eksperimen ini, kedua model *Support Vector Machine* (SVM) dan *Naïve Bayes* menunjukkan kinerja yang memuaskan setelah penerapan teknik SMOTE, dengan SVM mencapai akurasi 85% dan *Naïve Bayes* mencapai akurasi 84%. SVM memiliki nilai presisi, *recall*, dan *F1-score* yang seimbang untuk kedua klasifikasi sentimen, sedangkan *Naïve Bayes* memiliki performa yang serupa dalam mengidentifikasi sentimen positif dan negatif, meskipun sedikit lebih rendah dari SVM dalam beberapa metrik evaluasi. Dalam memilih model yang lebih unggul, perlu dipertimbangkan bahwa SVM mungkin menjadi pilihan yang lebih baik karena memiliki performa yang sedikit lebih tinggi dan terlihat lebih stabil dalam mengenali sentimen positif dan negatif. Hasil ini sejalan dengan penelitian sebelumnya yang juga menunjukkan keunggulan SVM dalam analisis sentimen, baik dalam konteks aplikasi TikTok maupun media berita online, namun keputusan akhir harus disesuaikan dengan kebutuhan analisis yang spesifik dari aplikasi tersebut.

5. KESIMPULAN

Dari hasil penelitian ini, meskipun kedua model, yaitu *Support Vector Machine* (SVM) dan *Naïve Bayes*, menunjukkan kinerja yang memuaskan setelah penerapan teknik SMOTE, terdapat perbedaan yang dapat diamati dalam evaluasi performa. SVM mencapai akurasi sebesar 85% dengan nilai presisi, *recall*, dan *F1-score* yang seimbang untuk kedua klasifikasi sentimen. Sementara itu, *Naïve Bayes* mencapai akurasi sebesar

84% dengan performa yang serupa dalam mengidentifikasi sentimen positif dan negatif, meskipun sedikit lebih rendah dari SVM dalam beberapa metrik evaluasi. Dalam memilih model yang lebih unggul, perlu dipertimbangkan bahwa SVM mungkin menjadi pilihan yang lebih baik karena memiliki performa yang sedikit lebih tinggi dan terlihat lebih stabil dalam mengenali sentimen positif dan negatif. Namun, keputusan ini harus disesuaikan dengan konteks aplikasi dan kebutuhan analisis yang spesifik.

DAFTAR PUSTAKA

- [1] A. P. Meiliani and I. Fuady, "Pengaruh Motivasi Pengguna Aplikasi Kencan Terhadap Kesehatan Mental Pada Mahasiswa," *Hum. J. Sosiohumaniora Nusant.*, vol. 1, no. 1, pp. 43–55, 2023, doi: 10.62180/yp2gp136.
- [2] V. A. Shania, M. F. Tri Palupi, and H. Kusumaningrum, "2747-File Utama Naskah-5038-1-10-20230718," pp. 176–185, 2021, [Online]. Available: <http://repository.untag-sby.ac.id/id/eprint/25766>
- [3] M. Mafruh and Q. N. Wijayani, "Pendekatan Terhadap Interaksi Pengguna Tinder Dalam Konteks Kencan Digital," *Madani J. Ilm. Multidisiplin*, vol. 1, no. 11, pp. 842–845, 2023.
- [4] F. A. Madarina, "Budaya Hook-Up pada Online Dating Tinder," *Asketik*, vol. 4, no. 2, pp. 187–196, 2020, doi: 10.30762/ask.v4i2.2420.
- [5] A. J. Pendidikan, S. Humaniora, N. Februari, Y. Sarah, A. Agustang, and M. R. S. Ahmad, "Aplikasi Tinder Sebagai Media Mencari Pasangan Dalam Membangun Interaksi Pada Kalangan Remaja Di Kota Makassar," vol. 2, no. 1, 2024.
- [6] D. Kusumo and R. Afandi, "Table Of Content information Article Rechtsidee," *Indones. J. Innov. Stud.*, vol. 13, no. 1, pp. 1–12, 2020, doi: 10.21070/acopen.9.2024.5394.
- [7] M. O. R. Nakano and S. Sumardijjati, "Resepsi Perempuan Surabaya terhadap Pemberitaan Negatif Aplikasi Kencan Online Tinder," *JHIP - J. Ilm. Ilmu Pendidik.*, vol. 7, no. 1, pp. 320–326, 2024, doi: 10.54371/jiip.v7i1.3642.
- [8] C. T. Saputri, S. Nursanti, and F. O. Lubis, "Proses Keberhasilan Hubungan Pengguna Aplikasi Kencan OnlineTinder Generasi Z," *Pendidik. Tambusai*, vol. 7, no. 3, pp. 23081–23087, 2023, [Online]. Available: <https://jptam.org/index.php/jptam/article/view/10258/8246>
- [9] A. W. V. Hutabarat, N. L. S. S. Adnyani, and K. Suryadi, "Analisis Sentimen Data Ulasan Pengguna MyPertamina di Twitter dengan Metode Text Mining," *J. Rekayasa Sist. Ind.*, vol. 13, no. 1, pp. 145–154, 2024.
- [10] F. N. Hasan and M. Dwijayanti, "Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naïve Bayes Classifier," *J. Linguist. Komputasional*, vol. 4, no. 2, pp. 52–58, 2021, doi: <https://doi.org/10.26418/jlk.v4i2.61>.
- [11] Y. Ardian Pradana, I. Cholissodin, and D. Kurnianingtyas, "Analisis Sentimen Pemindahan Ibu Kota Indonesia pada Media Sosial Twitter menggunakan Metode LSTM dan Word2Vec," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, pp. 2389–2397, 2023, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [12] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.
- [13] I. Sugiyarto *et al.*, "Sentimen Analisis Pengguna Aplikasi Grab Menggunakan Algoritma Naive Bayes Classifier dan Support Vector Machine," vol. 18, no. x, pp. 331–341, 1978.
- [14] I. P. Rahayu, A. Fauzi, and J. Indra, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 296, 2022, doi: 10.30865/json.v4i2.5381.
- [15] Tania Puspa Rahayu Sanjaya, Ahmad Fauzi, and Anis Fitri Nur Masruriyah, "Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naïve bayes dan support vector machine," *INFOTECH J. Inform. Teknol.*, vol. 4, no. 1, pp. 16–26, 2023, doi: 10.37373/infotech.v4i1.422.
- [16] A. Amalia, M. S. Lydia, S. D. Fadilla, and M. Huda, "Perbandingan Metode Klaster dan Preprocessing Untuk Dokumen Berbahasa Indonesia," *J. Rekayasa Elektr.*, vol. 14, no. 1, pp. 35–42, 2018, doi: 10.17529/jre.v14i1.9027.
- [17] F. Rahman, "WAKTU MENGGUNAKAN BINNING DAN SYNTHETIC MINORITY," vol. 4, no. 1, pp. 29–35, 2024.
- [18] J. Ipmawati, S. Saifulloh, and 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, 2024, doi: 10.57152/malcom.v4i1.1066.
- [19] P. Widyantara *et al.*, “Analisis Sentimen pada Teks Berbahasa Bali Menggunakan Metode Multinomial Naive Bayes dengan TF-IDF dan BoW,” *Jnatia*, vol. 2, no. 1, pp. 37–46, 2023.
- [20] Y. Julianto, D. H. Setiabudi, and S. Rostianingsih, “Analisis Sentimen Ulasan Restoran Menggunakan Metode SVM,” *J. Infra*, vol. 10, no. 1, 2022.
- [21] W. Eko Saputro, H. Yuana, and W. Dwi Puspitasari, “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, 2023, doi: 10.36040/jati.v7i2.6842.
- [22] A. M. Nur, N. Nurhidayati, and I. Fathurrahman, “Penerapan Metode Naive Bayes Untuk Penentuan Penerima Beasiswa Program Indonesia Pintar (PIP),” *Infotek J. Inform. dan Teknol.*, vol. 7, no. 1, pp. 93–102, 2024, doi: 10.29408/jit.v7i1.23995.
- [23] A. I. Tangraeni and M. N. N. Sitokdana, “Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naive Bayes,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.
- [24] D. Wijaya, R. A. Saputra, and F. Irwiensyah, “KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Pada Google Playstore Menggunakan Algoritma Naive Bayes,” *Media Online*, vol. 4, no. 4, pp. 2369–2380, 2024, doi: 10.30865/klik.v4i4.1738.
- [25] O. I. Gifari, M. Adha, F. Freddy, and F. F. S. Durrand, “Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine,” *J. Inf. Technol.*, vol. 2, no. 1, pp. 36–40, 2022, doi: 10.46229/jifotech.v2i1.330.
- [26] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, and A. Mulyani, “Comparison of Classification Algorithm and Feature Selection in Bitcoin Sentiment Analysis,” *J. Tek. Inform.*, vol. 3, no. 3, pp. 739–744, 2022, [Online]. Available: <https://doi.org/10.20884/1.jutif.2022.3.3.343%0Ahttp://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/343>
- [27] N. Suarna and W. Prihartono, “PENGUNAAN NAIVE BAYES DALAM MENGANALISIS SENTIMEN ULASAN APLIKASI MCDONALD ’ S DI INDONESIA,” vol. 8, no. 2, pp. 1949–1956, 2024.
- [28] K. D. Indarwati and H. Februariyanti, “Analisis Sentimen Terhadap Kualitas Pelayanan Aplikasi Gojek Menggunakan Metode Naive Bayes Classifier,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 10, no. 1, 2023, doi: 10.35957/jatisi.v10i1.2643.
- [29] U. Kusnia, F. Kurniawan, and S. Artikel, “Analisis Sentimen Review Aplikasi Media Berita Online Pada Google Play menggunakan Metode Algoritma Support Vector Machines (SVM) Dan Naive Bayes INFO ARTIKEL ABSTRAK,” *Explor. IT*, vol. 4, no. 36, pp. 24–28, 2022, [Online]. Available: <https://doi.org/10.35891/explorit>
- [30] D. Ananda and R. R. Suryono, “Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan Naive Bayes,” vol. 8, no. April, pp. 748–757, 2024, doi: 10.30865/mib.v8i2.7517.
- [31] P. K. Sari and R. R. Suryono, “Komparasi Algoritma Support Vector Machine Dan Random Forest Untuk Analisis Sentimen Metaverse,” *J. Mnemon.*, vol. 7, no. 1, pp. 31–39, 2024, doi: 10.36040/mnemonic.v7i1.8977.