



Analisis Sentimen Publik Terhadap Danantara di Media Sosial X Menggunakan Naïve Bayes dan Support Vector Machine

Fabian Firmando¹, Ryan Randy Suryono^{2*}

¹Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

²Fakultas Teknik dan Ilmu Komputer, Magister Ilmu Komputer, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ¹fabian_firmando@teknokrat.ac.id, ^{2,*}ryan@teknokrat.ac.id

Email Penulis Korespondensi: ryan@teknokrat.ac.id

Submitted: 30/04/2025; Accepted: 01/09/2025; Published: 02/09/2025

Abstrak—Danantara sebagai lembaga pengelola investasi milik negara, telah menjadi topik perbincangan hangat di kalangan masyarakat, khususnya di media sosial X yang menampilkan berbagai opini publik. Penelitian ini bertujuan untuk mengevaluasi sentimen publik terhadap Danantara melalui analisis sentimen berbasis machine learning. Data yang digunakan terdiri dari 10.108 tweet, yang setelah melalui proses pembersihan (preprocessing), menghasilkan 9.790 tweet siap analisis. Metode yang digunakan dalam penelitian ini adalah pembobotan kata dengan TF-IDF, serta penerapan dua algoritma klasifikasi, yaitu Naïve Bayes dan Support Vector Machine (SVM). Untuk menangani ketidakseimbangan kelas dalam data sentimen, digunakan teknik Synthetic Minority Over-sampling Technique (SMOTE). Evaluasi kinerja model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil awal menunjukkan bahwa sebelum penerapan SMOTE, algoritma Naïve Bayes memiliki akurasi sebesar 64%, sementara SVM menunjukkan performa lebih baik dengan akurasi sebesar 80%. Setelah SMOTE diterapkan, akurasi Naïve Bayes meningkat menjadi 72% dan SVM meningkat signifikan hingga mencapai 89%. Hasil ini menunjukkan bahwa SMOTE efektif dalam menangani ketimpangan data dan meningkatkan performa klasifikasi. Secara keseluruhan, studi ini memberikan gambaran yang lebih jelas mengenai opini publik terhadap Danantara serta menunjukkan bahwa kombinasi teknik preprocessing, TF-IDF, algoritma machine learning, dan penyeimbangan data mampu menghasilkan analisis sentimen yang lebih akurat.

Kata Kunci: Danantara; Analisis Sentimen; Machine Learning; SMOTE; Support Vector Machine (SVM); Naïve Bayes

Abstract—Danantara a state-owned investment management institution, has become a topic of widespread public discussion, particularly on social media platform X, where diverse public opinions are expressed. This study aims to evaluate public sentiment toward Danantara through sentiment analysis using machine learning techniques. The dataset consists of 10,108 tweets, of which 9,790 tweets remained after the preprocessing stage and were ready for analysis. The methodology involves word weighting using Term Frequency-Inverse Document Frequency (TF-IDF) and the implementation of two classification algorithms: Naïve Bayes and Support Vector Machine (SVM). To address the class imbalance in sentiment data, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. Model performance was evaluated using metrics such as accuracy, precision, recall, and F1-score. Initial results show that before applying SMOTE, the Naïve Bayes algorithm achieved an accuracy of 64%, while SVM performed better with an accuracy of 80%. After applying SMOTE, Naïve Bayes accuracy improved to 72%, and SVM increased significantly to 89%. These results indicate that SMOTE is effective in handling data imbalance and enhancing classification performance. Overall, this study provides a clearer picture of public opinion toward Danantara and demonstrates that the combination of preprocessing, TF-IDF, machine learning algorithms, and data balancing techniques can produce more accurate sentiment analysis.

Keywords: Danantara; Sentiment Analysis; Machine Learning; SMOTE; Support Vector Machine (SVM); Naïve Bayes

1. PENDAHULUAN

Kemajuan teknologi informasi telah berubah dalam hal menyampaikan pandangan mereka, terutama melalui media sosial seperti X, yang kini telah menjadi ruang diskusi publik yang utama. salah satu tema yang paling sering dibahas diskusikan di platform tersebut merupakan Danantara, sebuah dana kekayaan negara yang didirikan untuk mengelola kekayaan negara secara strategis. Nama Danantara sendiri mengandung filosofi yang mendalam, Daya melambangkan kekuatan, Anagata melambangkan masa depan, dan Nusantara melambangkan Indonesia, sehingga menjadi simbol vitalitas ekonomi masa depan negara ini. Selain memainkan peran yang signifikan, Danantara juga mendapat beragam tanggapan publik di media sosial. Sebagian pihak menyambut positif kebijakan ini dan meyakini akan membawa angin segar bagi pengelolaan aset negara. Namun, banyak pula yang mengkritik kebijakan ini karena dinilai kurang transparan dan berpotensi menimbulkan risiko politik. Untuk memahami persepsi publik tersebut secara sistematis, analisis sentimen menjadi pendekatan yang relevan [1]. Penelitian ini memanfaatkan dua algoritma machine learning, yaitu Naïve Bayes dan Support Vector Machine (SVM), untuk mengklasifikasikan sentimen publik dalam tweet tentang Danantara menjadi kelompok positif, negatif, atau netral. Data diambil dari platform media X dan kemudian melalui tahap praproses, ekstraksi fitur, pelatihan model, dan evaluasi kinerja algoritma [2]. Hasil analisis diharapkan dapat memberikan gambaran menyeluruh mengenai opini publik dan isu-isu penting, sekaligus memberikan masukan strategis bagi Danantara untuk merancang komunikasi publik yang lebih efektif dan membangun kepercayaan publik terhadap lembaga yang dimaksud.

Penelitian ini bertujuan untuk mengkaji reaksi masyarakat terhadap kehadiran Danantara dengan menganalisis sentimen yang terdapat dalam tweet di media sosial X. Tujuannya adalah untuk memperoleh pemahaman yang komprehensif tentang persepsi masyarakat terhadap Danantara, baik positif maupun negatif. Dalam proses ini, dua algoritma pembelajaran mesin, Naïve Bayes dan Support Vector Machine (SVM), akan digunakan untuk mengklasifikasikan emosi menjadi tiga kategori: positif, negatif, dan netral. Studi ini juga difokuskan pada



mengidentifikasi isu-isu yang sering dibahas publik seperti transparansi, kinerja investasi, dan faktor politik yang terkait dengan Danantara. Diharapkan hasil analisis ini akan memberikan wawasan berharga untuk mengembangkan strategi komunikasi publik yang lebih responsif dan berbasis data, sehingga meningkatkan kepercayaan publik terhadap lembaga tersebut [3]. Untuk mengatasi tantangan dalam memahami lebih jauh tentang perspektif publik, studi ini mengusulkan penggunaan analisis sentimen berbasis data untuk memetakan persepsi publik secara efektif. Studi ini mengandalkan algoritma Naïve Bayes dan SVM untuk menyediakan metode efektif dalam memproses data teks dalam jumlah besar secara otomatis kedua metode ini digunakan sebab memiliki kapabilitas yang baik pada mengelompokkan data berbasis teks dengan akurasi tinggi, terutama dalam analisis bahasa alami [4]. Hasil klasifikasi sentimen akan divisualisasikan dan dianalisis lebih lanjut untuk dijadikan dasar keputusan strategis Danantara. Selain itu, penelitian ini merekomendasikan penerapan pemantauan sentimen berkelanjutan untuk memungkinkan Danantara dengan cepat mendeteksi perubahan opini publik dan mengambil tindakan proaktif yang diperlukan.

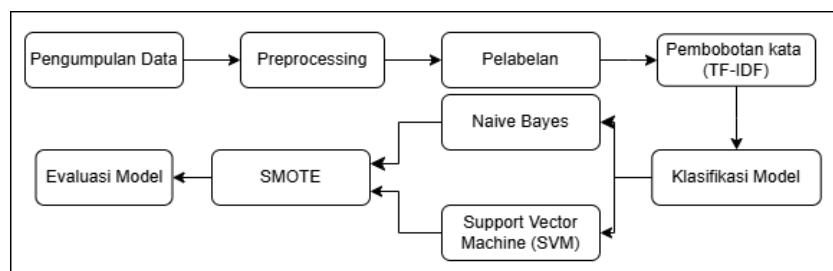
Berbagai penelitian telah dilakukan pada penerapan algoritma pengklasifikasi seperti Naïve Bayes dan Support Vector Machine (SVM) pada berbagai situasi. Salah satunya adalah penelitian Ginabila (2023) yang menggunakan Naïve Bayes untuk menganalisis sentimen 84,73% pemutar musik daring Spotify. Pada saat yang sama, akurasi algoritma Support Vector Machine adalah 82,42% [5]. Thanty Tawaqalia (2020) juga menggunakan Naïve Bayes dan Support Vector Machine dalam rangka menganalisis pandangan terhadap figur publik, dengan Naïve Bayes mencapai akurasi sebesar 91,81% dan SVM mencapai akurasi sebesar 85,47%. Tingkat akurasi Naïve Bayes sebesar 89,28%, dan tingkat akurasi SVM sebesar 90,95%. Sementara itu, Naïve Bayes memiliki recall sebesar 91,58% sedangkan SVM memiliki recall sebesar 76,18% [6]. Freddy Andrew Dkk (2025) terhadap Kemenkes didapatkan bahwa algoritma SVM mendapat angka akurasi yang lebih unggul berdasarkan perbandingan dengan Naïve Bayes, dengan SVM mendapatkan akurasi sebesar 79%, sementara Naïve Bayes mendapatkan nilai akurasi sebesar 77% [7]. Acuan Supian (2024) Hasil penelitian tentang analisis sentimen Twitter Ibukota Nusantara mengindikasikan bahwa metode SVM terus-menerus menunjukkan akurasi yang lebih tinggi dalam analisis sentimen, yaitu sebesar 94%, dibandingkan dengan Naïve Bayes yang mencatat akurasi sebesar 91% [8]. Altolyto Sitanggang dkk (2024) juga menggunakan Naïve Bayes Pada program makan siang gratis, algoritma Naïve Bayes digunakan serta menghasilkan akurasi sebesar 72,2%, dengan presisi 63,2%, recall 66,1%, serta F1-Score mencapai 64% [9].

Pendekatan komparatif yang digunakan dalam studi ini memberikan wawasan yang lebih menyeluruh. Penelitian ini tidak sekadar menitikberatkan pada aspek akurasi semata, tetapi juga mengevaluasi aspek-aspek seperti kecepatan pemrosesan, kemampuan algoritma dalam menangani ketidakseimbangan data, dan akurasi dalam menangkap kompleksitas sentimen publik. Dengan demikian, studi ini tidak hanya mengidentifikasi algoritma yang optimal tetapi juga menguraikan berbagai faktor yang memengaruhi kinerjanya dalam konteks analisis sentimen informasi baru. Temuan ini diharapkan dapat memperdalam persepsi publik terhadap Danantara dan memberikan wawasan berharga bagi penyedia layanan dalam menilai penerimaan publik.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tujuan dari studi yang dimaksud adalah untuk mengevaluasi sentimen yang muncul terkait Danantara di platform media sosial X, pertama-tama mengumpulkan data opini yang diberi label positif, negatif, dan netral. Kemudian data ini diperoleh dilakukan tahap preprocessing yang meliputi pembersihan teks dengan menghilangkan kata henti, stemming, dan menghilangkan karakter yang tidak relevan[10]. Selanjutnya, teknik SMOTE (*Synthetic Minority Over-sampling Technique*), dilanjutkan dengan representasi data menggunakan TF-IDF serta pembagian data menjadi data latih dan uji seperti pada Gambar 1.



Gambar 1. Tahapan Penelitian

Proses klasifikasi teks dalam diagram ini mencakup tahapan yang sistematis mulai dari pengumpulan data hingga evaluasi model. Pembobotan kata dengan metode TF-IDF berperan penting dalam mengubah data teks menjadi bentuk numerik yang dapat diproses oleh algoritma klasifikasi. Penggunaan algoritma Naïve Bayes dan Support Vector Machine (SVM) memungkinkan model mempelajari pola dari data teks yang telah diproses. Penerapan teknik SMOTE membantu mengatasi ketidakseimbangan data untuk meningkatkan performa model. Evaluasi akhir menunjukkan efektivitas model dalam mengklasifikasikan data teks secara akurat dan seimbang. Alur ini menjadi pendekatan yang kuat dalam pengolahan dan analisis teks berbasis machine learning.



2.2 Pengumpulan Data

Tahap awal dalam proses analisis ini adalah pengumpulan data, yang dilakukan dengan mengambil cuitan dari media sosial X. Proses ini menggunakan kata kunci tertentu, seperti "Danantara", untuk menyaring cuitan yang relevan dengan topik penelitian. Data yang diperoleh berbentuk teks mentah dari unggahan pengguna, yang mencerminkan opini, persepsi, maupun respons publik terhadap Danantara. Data ini menjadi fondasi utama dalam analisis sentimen karena mewakili suara masyarakat Dengan segera dan dalam waktu yang bersamaan [11]. Untuk Memastikan keakuratan data yang diperoleh memiliki kualitas dan relevansi yang tinggi, proses pengumpulan dilakukan dalam kurun waktu tertentu, serta memperhatikan faktor-faktor seperti jumlah pengikut akun, bahasa yang digunakan, dan frekuensi cuitan. Penggunaan *Application Programming Interface* (API) dari X atau tools crawling lainnya memungkinkan pengambilan data secara otomatis dan dalam jumlah besar [12]. Dengan cara ini, peneliti dapat memperoleh representasi opini publik yang luas dan beragam sebagai bahan analisis lanjutan.

2.3 Preprocessing

Pada fase ini, data mentah yang dikumpulkan menjalani proses pembersihan dan persiapan untuk diproses Secara lebih terperinci . Beberapa tindakan yang dilakukan termasuk menghilangkan simbol atau karakter yang tidak relevan, memisahkan kata-kata (tokenisasi), menyamakan bentuk kata (normalisasi), Mengurangi penggunaan kata-kata standar yang tidak mempunyai arti utama (*stop word removal*), serta mengembalikan kata ke bentuk dasarnya (stemming) [13]. Fokus utama dari fase bertujuan untuk mengatur serta menyederhanakan struktur data teks sehingga akan lebih mudah dan lebih akurat dalam proses analisis selanjutnya.

2.4 Pelabelan Data

Setelah melalui proses pembersihan, setiap tweet kemudian diklasifikasikan berdasarkan jenis sentimennya, seperti sentimen positif, negatif, atau netral[14] . Jumlah tweet yang dapat digunakan untuk analisis berkurang dari 10108 menjadi 9790 tweet. Tahapan berikutnya pelabelan sentiment yg bertujuan untuk mengklasifikasi masing-masing tweet kedalam kategori.

2.5 Pembobotan Kata (TF-IDF)

Menyiapkan data teks sehingga algoritma machine learning dapat memprosesnya., setiap kata dalam dataset perlu dikonversi ke dalam bentuk angka. Salah satu teknik yang digunakan adalah TF-IDF (*Term Frequency - Inverse Document Frequency*), yaitu cara pembobotan kata untuk mengukur sejauh mana sering kata tertentu muncul pada satu dokumen dibandingkan dengan seluruh dokumen dalam kumpulan data [15]. Dengan pendekatan ini, sistem dapat mengenali kata-kata mana yang memiliki bobot informasi lebih tinggi, sehingga mampu membedakan istilah umum dari kata-kata yang relevan dan khas terhadap topik tertentu [16]. Proses ini penting agar model dapat memahami konteks dan makna relatif dari kata dalam analisis sentimen.

2.6 Klasifikasi Model

Setelah data teks dikonversi menjadi bentuk numerik melalui pembobotan, langkah selanjutnya adalah menerapkan model klasifikasi untuk memprediksi jenis sentimen yang terkandung dalam setiap cuitan [17]. Dua pendekatan algoritma *machine learning*, yaitu Naïve Bayes dan Support Vector Machine (SVM), digunakan dalam penelitian ini. Kedua Pemilihan ini didasarkan pada efektivitas yang telah terbukti dalam memproses data Kalimat dan melakukan klasifikasi sentimen. Naive Bayes didasarkan pada prinsip probabilitas, sedangkan SVM mencari hiperbidang terbaik yang memisahkan kelas emosi dengan margin maksimum. Keduanya mengambil pendekatan yang berbeda tetapi saling melengkapi untuk menilai hasil perkiraan yang akurat dan dapat diandalkan [18].

2.6.1 Naive Bayes

Naive Bayes yaitu Model klasifikasi yang mengaplikasikan prinsip Teorema Bayes, melalui anggapan bahwa setiap fitur dan atribut data tidak bergantung satu sama lain [19]. Meskipun asumsi independensi ini tidak selalu sesuai dengan kondisi data nyata, metode ini tetap sangat efektif, terutama dalam klasifikasi teks seperti analisis sentimen, deteksi spam, atau penyaringan email [20]. Keunggulan algoritma ini terletak pada kemudahannya dalam implementasi, kecepatan dalam proses pelatihan, serta kemampuannya menangani data dalam skala besar. Dalam analisis sentimen, Naïve Bayes digunakan Untuk mengukur kemungkinan bahwa sebuah dokumen Bagian dari kategori sentimen tertentu dengan melihat pola kemunculan kata-kata yang termasuk di dalamnya [21].

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

Dalam rumus tersebut, $P(C|X)$ adalah probabilitas posterior, yaitu peluang data X termasuk ke dalam kelas C setelah mempertimbangkan bukti yang diberikan oleh fitur-fitur X. Nilai $P(X|C) P(C)$ disebut likelihood, yaitu kemungkinan fitur X muncul apabila data tersebut memang berasal dari kelas C. Selanjutnya, $P(C)$ merupakan prior, yaitu peluang awal kemunculan kelas C tanpa mempertimbangkan data. Sedangkan $P(X)$ adalah evidence, yaitu total probabilitas dari fitur X secara keseluruhan tanpa memperhatikan kelasnya. Karena



nilai $P(X)P(X)$ sama untuk semua kelas, dalam proses klasifikasi kita hanya perlu fokus pada hasil perkalian $P(X|C) \cdot P(C)P(X|C)P(C)$, dan kelas dengan nilai tertinggi dari hasil ini dipilih sebagai prediksi akhir.

2.6.2 Support Vector Machine (SVM)

Salah satu teknik *machine learning* yang diterapkan dalam menyelesaikan tantangan dalam klasifikasi dan regresi terdiri dari adalah *Support Vector Machine* (SVM). SVM menentukan bidang pemisah ideal atau *hyperplane* yang dapat secara jelas memisahkan data ke dalam dua kelas berbeda.[22]. *Hyperplane* yang paling dekat, atau jarak terdekat antara titik data terdekat dari masing-masing kelas terhadap garis pemisah, adalah yang dipilih. Support vector adalah nama titik-titik ini. Data dapat dipetakan ke ruang fitur yang lebih besar dengan kernel. memudahkan pemisahan kelas. SVM dilatih dengan optimasi untuk menemukan hyperplane terbaik yang memaksimalkan margin pemisah, menghasilkan klasifikasi yang lebih akurat. SVM sangat cocok digunakan pada data dengan dimensi tinggi serta mampu mengatasi klasifikasi non-linear melalui penggunaan fungsi kernel, sehingga sering digunakan dalam pengolahan data teks seperti analisis sentimen [23].

$$f(x) = w \cdot x + b \quad (2)$$

Pada persamaan $f(x) = w \cdot x + b$, vektor x merupakan representasi fitur dari data input yang diperoleh melalui proses ekstraksi fitur. Vektor w adalah bobot yang dipelajari selama proses pelatihan dan digunakan untuk menentukan kontribusi masing-masing fitur terhadap hasil prediksi. Sementara itu, b adalah nilai bias yang berfungsi untuk menggeser posisi hyperplane agar model dapat melakukan pemisahan kelas dengan lebih fleksibel, tanpa harus selalu melewati titik asal.

2.7 Smote

SMOTE (*Synthetic Minority Over-sampling Technique*) digunakan sebagai solusi untuk mengurangi ketidakseimbangan jumlah data antar kelompok sentimen dalam kumpulan data [24]. Ketika jumlah data pada salah satu kelas, seperti sentimen negatif atau netral, Model cenderung bias terhadap kelas yang dominan karena jumlah populasinya relatif kecil. Untuk mencegah hal ini, SMOTE interpolasi sampel yang ada untuk menghasilkan data sintetis baru tentang kelas minoritas. [25]. Dengan menyeimbangkan distribusi data, metode ini membantu meningkatkan kemampuan model dalam mengenali semua jenis sentimen secara lebih adil dan akurat.

2.8 Evaluasi Model

Tahap terakhir dalam proses analisis ini adalah Merupakan analisis model, yang dimaksudkan untuk menilai seberapa baik sistem klasifikasi dalam memperkirakan sentimen secara akurat. Evaluasi dilaksanakan dengan menggunakan sejumlah metrik kinerja, termasuk akurasi, presisi, *recall*, dan F1-score [26]. Memberikan penjelasan gambaran menyeluruh tentang ketepatan krangka untuk mengkategorikan sentimen, termasuk kemampuannya dalam mengidentifikasi sentimen yang benar (presisi), menangkap seluruh sentimen yang relevan (*recall*), serta keseimbangan antara keduanya (F1-score). Dengan evaluasi ini, peneliti dapat menentukan apakah model yang digunakan sudah layak diterapkan atau perlu dilakukan penyempurnaan lebih lanjut.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (3)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

$$f1 - Score = \frac{2 \times \text{presisi} \times \text{recall}}{\text{presisi} + \text{recall}} \quad (6)$$

Akurasi digunakan untuk mengukur seberapa besar proporsi prediksi yang dilakukan model sesuai dengan nilai sebenarnya. Nilainya diperoleh dari jumlah prediksi benar, yaitu True Positive (TP) dan True Negative (TN), membagi dengan seluruh jumlah prediksi ($TP + TN + False Positive (FP) + False Negative (FN)$), lalu dikalikan 100%. Sementara itu, presisi menunjukkan Tingkat ketepatan model dalam memperkirakan kelas positif, dengan cara membagi jumlah TP dengan total prediksi positif ($TP + FP$). *Recall*, di sisi lain, Menentukan sejauh mana model berhasil mengenali seluruh data untuk benar-benar positif, Diperoleh Dengan cara membagi total *True Positive* (TP) dengan keseluruhan data yang benar-benar termasuk kategori positif ($TP + FN$).

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data dikumpulkan menggunakan metode crawling dengan memanfaatkan library Tweet Harvest pada platform Google Colab, serta menggunakan API Token Twitter sebagai mekanisme autentikasi untuk aplikasi atau skrip yang mengakses data melalui API resmi Twitter. Data yang diperoleh disusun dan disajikan dalam Tabel 1.

**Tabel 1.** Hasil Pengumpulan Data

Username	Tweet
@Sinergi	Membuka Peluang Baru bagi Pengusaha! Dengan dukungan infrastruktur yang kuat para pelaku usaha akan mendapatkan manfaat besar dari proyek ini. https://t.co/hQUYL1XDY4
@jatamnas	Danantara itu ya bro salah satu Langkah besar dan emang harus punya sikap buat ambil Langkah ini jadi proyek-proyek strategis nasional bisa dipakai nantinya gaperlu hutang keluar negri

Berdasarkan data sampel yang ditampilkan pada Tabel 1, tahap selanjutnya adalah melakukan *preprocessing* guna mempersiapkan data agar layak dan optimal untuk dianalisis lebih lanjut.

3.2 Preprocessing

Setelah data berhasil dikumpulkan, tahap berikutnya adalah *preprocessing*, yaitu langkah awal Pada tahapan analisis terhadap data teks guna untuk membersihkan serta mempersiapkan data secara sistematis. Proses ini mencakup beberapa tahapan, yaitu *cleaning* untuk menghilangkan elemen-elemen yang kurang berkaitan, *case folding* untuk mengonversi seluruh isi dalam bentuk teks berhuruf kecil, dan *tokenization* guna memecah teks menjadi satuan kata. Tahapan berikutnya adalah *stopword removal* guna mengeliminasi Kata-kata biasa yang tidak memiliki kontribusi inti, *normalisasi* untuk menyeragamkan bentuk variasi kata, serta *stemming* yang bertujuan mengembalikan kata kebentuk aslinya. Keseluruhan proses ini bertujuan agar Data teks menjadi lebih terstruktur dan siap dianalisis. digunakan dalam tahap analisis selanjutnya. Tabel 2 memperlihatkan hasil yang diperoleh dari preprocessing.

Tabel 2. Hasil *Preprocessing*

Tahapan	Tweet
<i>Tweet</i>	Membuka Peluang Baru bagi Pengusaha! Dengan dukungan infrastruktur yang kuat para pelaku usaha akan mendapatkan manfaat besar dari proyek ini. https://t.co/hQUYL1XDY4
<i>Cleansing</i>	Membuka Peluang Baru bagi Pengusaha Dengan dukungan infrastruktur yang kuat para pelaku usaha akan mendapatkan manfaat besar dari proyek ini.
<i>Case Folding</i>	membuka peluang baru bagi Pengusaha dengan dukungan infrastruktur yang kuat para pelaku usaha mendapatkan manfaat besar dari proyek ini.
<i>Tokenizing</i>	['membuka', 'peluang', 'baru', 'bagi', 'pengusaha', 'dengan', 'dukungan', 'infrastruktur', 'yang', 'kuat', 'para', 'pelaku', 'usaha', 'akan', 'mendapatkan', 'manfaat', 'besar', 'dari', 'proyek', 'ini']
<i>Stopword</i>	['membuka', 'peluang', 'pengusaha', 'dukungan', 'infrastruktur', 'kuat', 'pelaku', 'usaha', 'manfaat', 'proyek']
<i>Stemming</i>	buka peluang usaha dukung infrastruktur kuat laku usaha manfaat proyek

Merujuk pada Tabel 2, hasil proses preprocessing menunjukkan bahwa data tweet telah dibersihkan dan siap digunakan untuk tahap evaluasi. Misalnya, Teks " usaha akan mendapatkan manfaat besar dari proyek ini. <https://t.co/hQUYL1XDY4>" Diubah menjadi "usaha akan mendapatkan manfaat besar dari proyek ini". Case folding Mengubah setiap karakter menjadi ukuran yang lebih Minim untuk konsistensi, seperti "Pengusaha" menjadi "pengusaha". Pada tokenizing, teks dipecah menjadi kata-kata individual, misalnya "Pengusaha dengan dukungan" menjadi ['Pengusaha', 'dengan', 'dukungan']. Stopword removal Proses penghilangan kata-kata yang tidak relevan bermakna, contohnya "baru" maupun "bagi", meninggalkan kata-kata relevan. Pada akhirnya, stemming menyederhanakan Kata-kata yang disusun menjadi, seperti "pelaku" menjadi "laku". Semua tahapan ini membersihkan, menyelaraskan, dan menyederhanakan teks agar siap untuk di analisis lebih lanjut. Langkah-langkah ini sangat penting untuk memastikan bahwa data masukan yang digunakan oleh model analisis sentimen bersifat konsisten, terstruktur, dan bebas dari elemen-elemen yang dapat mengganggu akurasi prediksi.

3.3 Pelabelan Data

Setelah melalui tahap pemrosesan data, sebanyak 9.791 tweet berhasil disaring menjadi data yang bersih dan siap untuk dianalisis. Selanjutnya, setiap tweet dilabeli menggunakan *library TextBlob* berdasarkan analisis sentimen hasil pelabelan tertera pada Tabel 3 berikut ini:

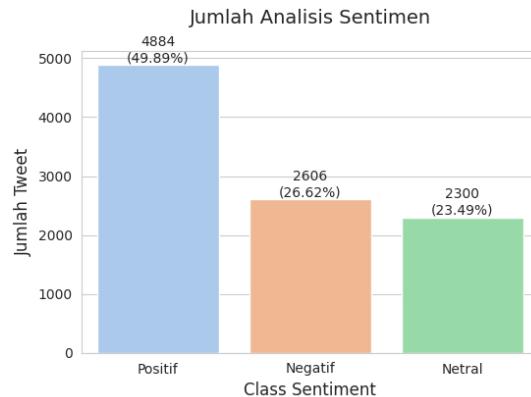
Tabel 3. Hasil Proses Pelabelan

Stemming	Score	Label
buka peluang usaha dukung infrastruktur kuat laku usaha manfaat proyek	2	Positif
fokus proyek dorong tumbuh ekonomi	0	Netral
danantara celengan rakksasa negara simpan uang untung Perusahaan perusahaan negara bumn uang simpan pakai bangun proyek-proyek ekonomi indonesia maju	-1	Negatif

Dari tahapan pelabelan tersebut didapati data hasil pengkategorianya adalah 4884 positif, 2300 Netral dan 2606 Negatif. Hasil tersebut diperoleh melalui perhitungan nilai polaritas, di mana teks diklasifikasikan sebagai



positif bila nilai polaritasnya Di atas 0, negatif Bila nilai lebih rendah dari 0, dan netral apabila bernilai tepat 0. Gambar 2 di bawah ini merupakan hasil pelabelan dalam bentuk Visualisasi.

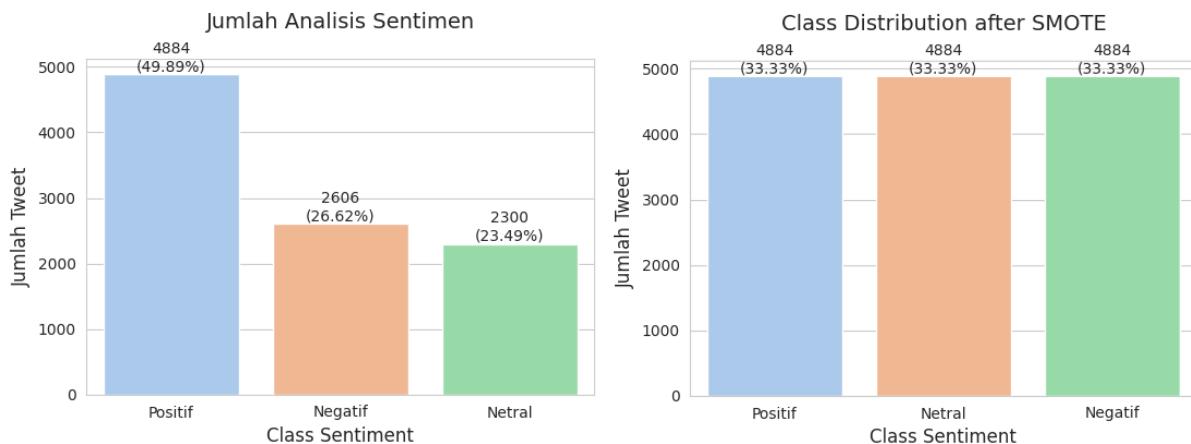


Gambar 2 Visualiasi Hasil Pelabelan Data

Hasil pelabelan sentimen menunjukkan bahwa sebagian besar komentar bersifat positif, dengan persentase mencapai 49,89%. Sementara itu, sentimen netral tercatat sebesar 23,49%, dan sentimen negatif sebesar 26,62%. Temuan ini mengindikasikan bahwa pandangan publik terhadap Danantara umumnya bernada positif. Tingginya proporsi sentimen positif ini kemungkinan dipengaruhi oleh sejumlah faktor, seperti dominasi kata-kata bernada positif yang terdeteksi dalam kamus sentimen, atau karena Danantara secara umum memperoleh tanggapan baik dari masyarakat. Meski demikian, pola distribusi sentimen ini juga bisa dipengaruhi oleh karakteristik data Twitter yang cenderung singkat dan padat.

3.4 SMOTE

Pada Hasil awal analisis menunjukkan bahwa terdapat 4.884 tweet positif, 2.606 tweet negatif dan 2.300 netral. Yang menunjukkan ketidakseimbangan data mayoritas dan minoritas, Untuk mengatasi ketidakseimbangan data ini, digunakan teknik SMOTE. Pada Hasil awal analisis menunjukkan bahwa terdapat 4.884 tweet positif, 2.606 tweet negatif dan 2.300 netral. Yang menunjukkan ketidakseimbangan data mayoritas dan minoritas, Untuk mengatasi ketidakseimbangan data ini, digunakan teknik SMOTE. Ketidakseimbangan ini berpotensi menyebabkan model lebih memihak pada kelas mayoritas, yaitu sentimen positif. Akibatnya, akurasi prediksi untuk kelas negatif dan netral bisa menurun. Untuk itu, digunakan teknik SMOTE yang bertujuan meningkatkan jumlah sampel pada kelas minoritas secara sintetik. SMOTE menghasilkan data baru dengan membuat interpolasi antara sampel-sampel yang mirip dalam kelas yang sama. Setelah penerapan teknik ini, jumlah tweet negatif dan netral meningkat sehingga mendekati jumlah tweet positif. Penerapan SMOTE tidak mengubah isi data yang ada, melainkan menyeimbangkan jumlahnya untuk kebutuhan pelatihan model. Dengan distribusi yang lebih seimbang, model diharapkan dapat mengenali pola dari ketiga kelas sentimen secara lebih adil. Hal ini penting untuk meningkatkan akurasi dan generalisasi model dalam tugas klasifikasi sentimen. Dengan demikian, proses pelatihan menjadi lebih optimal dan tidak biasa terhadap kelas tertentu. Hasil analisis setelah SMOTE menunjukkan peningkatan kinerja model dalam mendeteksi sentimen negatif dan netral. Hasil perbandingan distribusi sentimen dan Ditampilkan pada Gambar 3 di bawah ini:



Gambar 3. Visualisasi Sebelum SMOTE dan Sesudah SMOTE

Gambar 3 menunjukkan perbandingan kondisi data pada tahap awal dan akhir penerapan metode SMOTE. Sebelum SMOTE diterapkan, dataset cenderung didominasi oleh komentar dengan sentimen positif, sementara jumlah



data untuk sentimen negatif dan netral jauh lebih sedikit. Ketidakseimbangan ini menyebabkan model cenderung mengklasifikasikan komentar ke dalam kategori positif, sehingga mengurangi tingkat presisi dalam mendeteksi sentimen negatif dan netral. Setelah penerapan SMOTE, data pada kelas minoritas (negatif dan netral) ditingkatkan dengan menambahkan data sintetis, sehingga distribusi antar kelas menjadi lebih seimbang. Kondisi ini membuat model lebih mampu mengenali dan mengklasifikasikan sentimen negatif maupun netral dengan lebih akurat.

3.5 Hasil Pengujian Model

Dalam bagian ini, hasil dari pengujian model ini menggambarkan kinerja masing-masing algoritma dalam menentukan sentimen berdasarkan kategori positif, negatif, dan netral berdasarkan data *Twitter* yang telah diproses. Proses pengujian dikerjakan dengan menggunakan 20% data latih dan 80% data uji. Metriks Yang diterapkan untuk evaluasi adalah akurasi, presisi, recall, dan F1-score. Berikut adalah hasil pengujian sebelum dan setelah penerapan SMOTE.

3.5.1 Hasil Sebelum Penerapan Smote

Sebelum SMOTE diterapkan, ketiga model menunjukkan variasi performa yang cukup mencolok. Tabel 4 berikut menyajikan hasil evaluasi kinerja untuk model Naïve Bayes dan SVM.

Tabel 4. Hasil Proses Pelabelan

Model	Accuracy	Precision (Negatif)	Recall (Negatif)	Precision (Positif)	Recall (Positif)	F1-Score (Negatif)	F1-Score (Positif)
Naïve Bayes	64%	74%	59%	61%	95%	65%	74%
SVM	80%	80%	79%	87%	90%	79%	88%

Berdasarkan Tabel 4, dapat dilihat bahwa meskipun akurasi keseluruhan kedua model tergolong tinggi (Naïve Bayes 64% dan SVM 80%), performa dalam mengklasifikasikan sentimen negatif masih rendah, terutama pada nilai presisi dan recall. Hal ini mengindikasikan adanya ketidakseimbangan data, di mana model cenderung lebih akurat dalam mengenali sentimen positif. Ketimpangan ini disebabkan oleh dominasi jumlah data sentimen positif dalam dataset, sehingga model lebih terlatih untuk mengenali pola-pola positif dibandingkan sentimen negatif.

3.5.2 Hasil Sesudah Penerapan Smote

Setelah teknik SMOTE diterapkan untuk mengatasi ketidakseimbangan data, hasil pengujian menunjukkan peningkatan yang signifikan pada semua model. SMOTE menghasilkan data gabungan mengenai kelompok untuk kelas minoritas, yang meningkatkan representasi sentimen negatif dan netral dalam dataset. Tambahan data sintetis dari SMOTE juga membantu model dalam mengenali pola yang sebelumnya kurang terwakili. Akurasi, presisi, dan recall model meningkat secara konsisten, terutama dalam mendeteksi sentimen negatif. Hal ini menunjukkan bahwa SMOTE efektif dalam meningkatkan performa model pada data yang tidak seimbang. Hasil setelah penerapan SMOTE disajikan pada Tabel 5 di bawah ini

Tabel 5. Hasil Proses Pelabelan

Model	Accuracy	Precision (Negatif)	Recall (Negatif)	Precision (Positif)	Recall (Positif)	F1-Score (Negatif)	F1-Score (Positif)
Naïve Bayes	72%	70%	91%	70%	71%	79%	71%
SVM	89%	93%	91%	96%	82%	92%	88%

Setelah SMOTE diterapkan, ada peningkatan yang signifikan pada performa semua model, sebagaimana ditunjukkan pada Tabel 5. Akurasi meningkat pada seluruh algoritma, di mana SVM mencatat akurasi tertinggi sebesar 89%. Selain itu Presisi, recall, dan skor F1 untuk kategori sentimen negatif juga Mengalami kenaikan yang signifikan. Temuan ini menunjukkan bahwa SMOTE dapat menyelesaikan masalah Memperbaiki ketidakseimbangan data dengan Membuat data sintetis untuk kelas minoritas, yang mengembalikan keseimbangan model. dan mampu mengklasifikasikan semua kelas sentimen baik negatif maupun positif dengan lebih akurat.

3.6 Hasil Visualisasi Wordcloud

Visualisasi word cloud menampilkan hasil Analisis sentimen yang mencakup positif dan negatif terhadap Danantara. Word cloud merupakan gambaran visual Dari istilah Yang paling sering ditemui dalam teks, di mana ukuran suatu kata yang lebih besar, semakin tinggi frekuensi kemunculannya dalam data. Visualisasi ini membantu dalam mengidentifikasi topik-topik yang paling banyak dibicarakan oleh masyarakat. Analisis ini memberikan pemahaman tentang aspek-aspek yang mempengaruhi persepsi masyarakat terhadap Danantara, dengan cara ini, word cloud menjadi alat eksploratif Yang efisien untuk memahami pandangan publik secara cepat dan intuitif. Visualisasi ini turut membantu peneliti dalam mengidentifikasi kata-kata utama yang sering muncul dalam percakapan publik. Selain menjadi langkah awal dalam analisis yang lebih mendalam terhadap opini dan isu yang beredar, word cloud juga berfungsi sebagai cara yang efektif untuk menyajikan data secara ringkas dan mudah dimengerti oleh berbagai lapisan masyarakat. Dengan desain visual yang menarik, hasil word cloud ini dapat dimanfaatkan dalam penyampaian



informasi untuk keperluan kebijakan publik. Secara keseluruhan, visualisasi ini menjadi alat yang penting dalam mendukung pengambilan keputusan yang didasarkan pada data yang tepat. Hal ini dapat dilihat pada Gambar 4, 5, dan 6.



Gambar 4. WordCloud Negatif

Gambar 4 menunjukkan bahwa, word cloud ini menyajikan istilah yang paling sering digunakan pada sentimen negatif terhadap Danantara. Ukuran kata yang lebih besar mencerminkan jumlah kemunculan yang lebih banyak dalam percakapan atau opini publik. Kata-kata seperti "danantara", "apbn", "negara", "rakyat", dan "korupsi" mendominasi, menunjukkan kekhawatiran utama masyarakat terkait pengelolaan dana negara dan potensi penyalahgunaan kekuasaan. Selain itu, munculnya kata-kata seperti "proyek", "gagal", "rugi", "pajak", dan "beban" mengindikasikan ketidakpuasan terhadap efektivitas dan dampak ekonomi dari proyek investasi yang dilakukan. Kritik lainnya tercermin dari kata-kata seperti "jabatan", "bumn", "utang", dan "efisiensi", yang memperkuat persepsi negatif terhadap tata kelola, transparansi, dan akuntabilitas dalam pengelolaan Danantara. Secara keseluruhan, word cloud ini mencerminkan persepsi publik yang kritis terhadap pengelolaan dana publik dan arah kebijakan investasi strategis. Dominasi kata-kata bernada negatif ini menunjukkan adanya keresahan publik yang cukup kuat terhadap kebijakan dan implementasi proyek Danantara. Word cloud ini berfungsi sebagai representasi visual yang efektif untuk memahami sentimen masyarakat secara cepat dan menyeluruh.

Temuan ini dapat menjadi dasar penting bagi evaluasi komunikasi publik dan strategi transparansi Danantara ke depan. Dengan memahami kata-kata yang sering muncul, pemangku kepentingan dapat merancang pendekatan yang lebih responsif terhadap isu yang menjadi perhatian publik, Hal ini juga membuka peluang bagi perbaikan kebijakan agar lebih selaras dengan harapan dan kebutuhan masyarakat.



Gambar 5. WordCloud Positif

Gambar 5 menunjukkan bahwa, word cloud ini Memperlihatkan kata-kata yang paling sering muncul digunakan dalam sentimen positif terhadap Danantara. Kata-kata utama seperti "transformasi", "ekonomi", "gerakcepatkonomiehebat", "danantara", dan "global" muncul dengan ukuran besar, menandakan tingginya frekuensi kata-kata tersebut dalam percakapan publik yang bernada positif. Hal ini mencerminkan apresiasi masyarakat terhadap peran Danantara dalam mendorong transformasi ekonomi nasional dan mempercepat pertumbuhan ekonomi.

Selain itu, kata-kata seperti "efisiensi", "tumbuh", "investasi", "transparan", dan "percaya" menyoroti aspek-aspek pengelolaan yang dianggap positif, seperti efisiensi anggaran, transparansi, dan dampak terhadap pertumbuhan ekonomi. Kata-kata seperti "bumn", "daya saing", dan "percaya" juga menunjukkan dukungan publik terhadap strategi Danantara dalam memperkuat peran negara melalui investasi strategis.



Gambar 6. WordCloud Netral

Ekonomi Indonesia berdasarkan Gambar 6, word cloud ini Menunjukkan kata-kata yang paling umum dalam sentimen netral terhadap Danantara. Kata-kata utama seperti "danantara", "apbn", "transformasi ekonomi", "gerakcepatkonomiehebat", dan "ekonomi" muncul dengan ukuran besar, menunjukkan bahwa kata-kata tersebut banyak digunakan dalam perbincangan netral. Kata-kata seperti "investasi", "bumn", "kelola", "negara", dan "rakyat" juga banyak muncul, mencerminkan pembahasan objektif terkait peran Danantara dalam pengelolaan dana negara dan transformasi ekonomi. Selain itu, kemunculan kata seperti "efisiensi", "tumbuh", "usaha", dan "proyek" menandakan fokus masyarakat terhadap aspek operasional dan potensi dampak kebijakan Danantara tanpa mengandung muatan emosional yang kuat, baik positif maupun negatif.

3.7 Feature Extraction

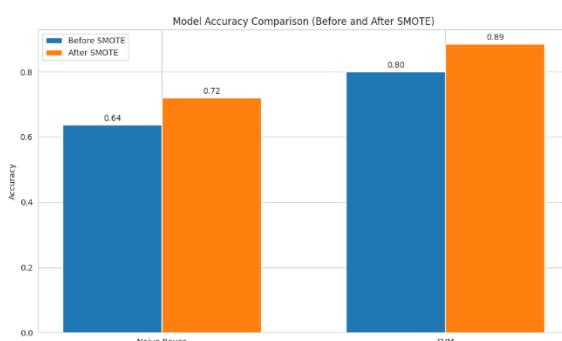
Pada langkah Feature Extraction, analisis ini menerapkan Metode TF-IDF (Term Frequency-Inverse Document Frequency) digunakan untuk mengubah komentar terkait menjadi format numerik yang dapat diproses oleh algoritma machine learning. TF-IDF berfungsi sebagai mengukur relevansi Kata-kata dalam suatu komentar sebagai mempertimbangkan frekuensinya di dokumen tersebut dan di keseluruhan kumpulan data. Kata-kata yang sering muncul dan kurang informatif akan diberi bobot rendah, sedangkan kata-kata yang jarang tetapi mengandung makna penting seperti istilah yang berkaitan dengan penipuan akan mendapatkan bobot lebih tinggi. Hasil dari proses ini adalah sebuah matriks berdimensi jarang (sparse matrix), di mana setiap kolom merepresentasikan kata unik, dan setiap baris menggambarkan satu komentar.

Tabel 6. Hasil Proses Pelabelan

	Term	TFIDF	SUM
1973	Danantara	618.179623	
2783	Ekonomi	576.372617	
13962	Transformasi	364.589364	
7716	Investasi	353.812650	
2765	Efisiensi	260.973627	

3.8 Visualisasi Perbandingan Algoritma

Untuk memahami dampak Implementasi metode Synthetic Minority Over-sampling Technique (SMOTE) terhadap kinerja model, dilakukan visualisasi perbandingan akurasi Antara dua metode klasifikasi, yaitu Naive Bayes dan Support Vector Machine (SVM). Gambar di bawah ini menyajikan perbandingan akurasi model sebelum dan sesudah penerapan SMOTE, memberikan gambaran yang lebih jelas mengenai peningkatan performa yang diperoleh setelah penyeimbangan data. Dari visualisasi terlihat bahwa kedua algoritma mengalami peningkatan akurasi setelah diterapkan SMOTE. Peningkatan tertinggi terjadi pada algoritma SVM, yang menunjukan bahwa Teknik penyeimbang data ini sangat efektif dalam menangani permasalahan kelas tidak seimbang pada model tersebut.



Gambar 7. Perbandingan Naïve Bayes dan SVM



Gambar 7 merupakan grafik batang yang menampilkan perbandingan akurasi Perbandingan model Naïve Bayes dan SVM sebelum serta setelah menggunakan SMOTE (Synthetic Minority Over-sampling Technique). Dari grafik terlihat bahwa kedua model mengalami peningkatan akurasi setelah dilakukan proses SMOTE. Model Naïve Bayes menunjukkan peningkatan akurasi dari 64% menjadi 72%, sementara model SVM meningkat dari 80% menjadi 89%. Peningkatan ini mengindikasikan bahwa SMOTE berhasil memperbaiki ketidakseimbangan kelas dalam dataset, sehingga model dapat mengklasifikasikan data dengan lebih baik. Secara keseluruhan, grafik ini mengonfirmasi bahwa penggunaan SMOTE berdampak positif terhadap Efektivitas model klasifikasi dalam melakukan analisis sentimen. Peningkatan akurasi ini juga mencerminkan kemampuan model yang lebih baik dalam mengenali pola pada data minoritas, seperti sentimen negatif dan netral. Dengan distribusi data yang lebih seimbang, model menjadi lebih robust dan tidak terlalu bias terhadap kelas mayoritas. Hal ini menunjukkan pentingnya penanganan ketidakseimbangan data sebagai langkah krusial dalam meningkatkan performa model klasifikasi.

4. KESIMPULAN

Hasil yang diperoleh dalam penelitian ini menunjukkan bahwa, diperoleh pemahaman tentang performa algoritma klasifikasi Naïve Bayes dan Support Vector Machine (SVM) selama melakukan analisis Respons terhadap tweet yang berkaitan dengan topik Danantara di platform media sosial X. Sebelum diterapkannya metode SMOTE, Naïve bayes mencatatkan kinerja paling rendah dengan akurasi sebesar 64%, dan SVM yang memiliki akurasi tertinggi yaitu 80%. Setelah SMOTE diterapkan, seluruh model mengalami peningkatan kinerja. Naïve Bayes menunjukkan performa terbaik dengan akurasi mencapai 72%, dan SVM yang naik menjadi 89%. Temuan ini menegaskan bahwa SVM merupakan algoritma paling andal dalam klasifikasi sentimen, baik sebelum maupun setelah penyeimbangan data dilakukan. Namun demikian, Penelitian ini memiliki beberapa kekurangan yang perlu menjadi pertimbangan. Pertama, jumlah data yang dianalisis masih relatif kecil dan cenderung didominasi oleh tweet dengan sentimen positif. Hal ini kemungkinan disebabkan oleh keterbatasan dalam proses pelabelan selama tahap pra-pemrosesan data, di mana beberapa kata atau frasa bermakna negatif tidak berhasil dikenali secara akurat, sehingga menyebabkan ketidakseimbangan klasifikasi. Untuk mengatasi hal ini, peningkatan kualitas kamus sentimen atau penerapan pendekatan berbasis machine learning dapat menjadi solusi strategis dalam studi selanjutnya. Selain itu, ketergantungan terhadap data yang hanya berasal dari media sosial X dapat membatasi representasi opini publik secara keseluruhan. Oleh karena itu, penelitian mendatang disarankan untuk menggabungkan data dari berbagai platform media sosial dan menerapkan pendekatan lanjutan seperti *deep learning* guna menggali aspek emosional yang lebih kompleks dan beragam dari opini publik.

REFERENCES

- [1] R. R. Putri and N. Cahyono, “Analisis Sentimen Komentar Masyarakat Terhadap Pelayanan Publik Pemerintah Dki Jakarta Dengan Algoritma Super Vector Machine Dan Naïve Bayes,” *J. JATI*, vol. 8, no. 2, pp. 2363–2371, 2024, doi: <https://doi.org/10.36040/jati.v8i2.9472>.
- [2] Hidayatunnisa, Kusrini, and Kusnawi, “Perbandingan Kinerja Metode Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Ulasan Pengguna Aplikasi Pintu,” *J. Fasilkom*, vol. 13, no. 2, pp. 173–180, 2023, doi: <http://dx.doi.org/10.23960/jitet.v13i2.6220>.
- [3] D. D. Saputra *et al.*, “Analisis Sentimen Terhadap Twitter Direktorat Jenderal Bea dan Cukai Menggunakan komparasi Algoritma Naïve Bayes dan Support Vector Machine,” *J-INTECH*, Vol 12, No 2, pp. 285–296, 2024, doi: <https://doi.org/10.32664/j-intech.v12i02.1274>.
- [4] D. I. Efendi, D. Solihudin, C. L. Rohmat, and S. E. Permana, “Bpjks Kesehatan Dengan Pendekatan Support Vector Machine (Svm) Dalam Analisis Sentimen,” *J. JATI*, vol. 8, no. 1, pp. 88–94, 2024, doi: <https://doi.org/10.36040/jati.v8i1.8304>.
- [5] A. Fauzi, P. Studi, S. Informasi, U. Bina, and S. Informatika, “Analisis Sentimen Terhadap Pemutar Musik Online Spotify Dengan Algoritma Naïve Bayes dan Support Vector Machine,” *J. Ilk.*, vol. 6, no. 2, pp. 111–122, 2023, doi: <https://doi.org/10.47324/ilkominfo.v6i2.180>.
- [6] T. T. Widowati *et al.*, “Analisis Sentimen Twitter Terhadap Tokoh Publik Dengan Algoritma Naïve Bayes Dan Support Vector,” *J. SIMETRIS*, vol. 11, no. 2, 2020, doi: <https://doi.org/10.24176/simet.v11i2.4568>.
- [7] F. A. Ryandi, D. Pratiwi, and S. Sari, “Analisis Sentimen Masyarakat Di Media Sosial X Terhadap Kemenkes Dengan Naïve Bayes dan SVM,” *J. Sains Dan Teknol.*, vol. 7, no. 1, pp. 1–6, 2025, [Online]. Available: <https://ejournal.sisfokomtek.org/index.php/saintek/article/view/4615>
- [8] A. Supian, B. T. Revaldo, N. Marhadi, and L. Efizoni, “Perbandingan Kinerja Naïve Bayes dan SVM pada Analisis Sentimen Twitter Ibukota Nusantara,” *J. Ilm. Inform.*, vol. 12, No 1, 2024, doi: <https://doi.org/10.33884/jif.v12i01.8721>.
- [9] A. Sitanggang, Y. Umaidah, R. I. Adam, U. S. Karawang, and T. Timur, “Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naïve Bayes,” *J. Ilmu dan Teknol. Komput.*, vol. 12, no. 3, 2024, doi: <http://dx.doi.org/10.23960/jitet.v12i3.4902>.
- [10] T. Maulana, Bintang Adhiyaksa, Rifqi Arul Fauzi, Rissa Ilmia Agustin, Siti Alia Azhaar and Rohana, “Komparasi Algoritma Naïve Bayes Dan Svm Untuk Analisis Sentimen Twitter Korupsi Bansos Beras Masa Pandemi,” *J. Ilmu dan Teknol. Komput.*, vol. 12, no. 2, pp. 912–918, 2024, doi: <http://dx.doi.org/10.23960/jitet.v12i2.4020>.
- [11] A. M. Rizqiyah and I. K. D. Nuryana, “Analisis Sentimen Masyarakat terhadap Kebijakan Iuran Tabungan Perumahan Rakyat (Tapera) pada Platform X Menggunakan Algoritma Naïve Bayes Classifier dan Support Vector Machine,” *JEISBI*, vol. 05, no. 03, pp. 298–306, 2024, doi: <https://doi.org/10.26740/jeisbi.v5i3.64074>.
- [12] E. Yuniar and N. Hendrastuty, “Perbandingan Metode Naïve Bayes , Random Forest dan SVM Untuk Analisis Sentimen



Pada Twitter Tentang Kenaikan Gaji Guru," *Build. Informatics, Technol. Sci.*, vol. 6, no. 4, pp. 2480–2490, 2025, doi: 10.47065/bits.v6i4.6970.

- [13] D. R. Firmansyah and E. Lestariningsih, "Analisis Sentimen Ulasan Aplikasi Smart Campus Unisbank di Google Playstore Menggunakan Algoritma Naïve Bayes," *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, 2024, doi: <https://doi.org/10.35870/jtik.v8i2.1882>.
- [14] D. N. Agustia, R. R. Suryono, U. T. Indonesia, L. Ratu, and K. B. Lampung, "Comparison Of Naïve Bayes , Random Forest , And Logistic Regression Algorithms For Sentiment Analysis Online Gambling Komparasi Algoritma Naïve Bayes , Random Forest , Dan Logistic Regresion Untuk Analisis," *INOVTEK Polbeng*, vol. 10, no. 1, pp. 284–295, 2025, doi: <https://doi.org/10.35314/prk93630>.
- [15] and N. A. V. I. A. Fahrezi, Rudiman, "Analisis Sentimen Twitter Atas Isu Hak Angket Menggunakan Pembobotan TF-IDF dan Algoritma SVM," *Sci-Tech J.*, vol. 3, pp. 179–192, 2024, doi: <https://doi.org/10.56709/stj.v3i2.526>.
- [16] P. I. Purnama and R. R. Suryono, "Analisis Sentimen Acara Clash of Champions dengan Algoritma Naïve Bayes dan Support Vector Machine," *INOVTEK Polbeng*, vol. 6, no. 4, pp. 2277–2287, 2025, doi: 10.47065/bits.v6i4.6575.
- [17] N. S. Ramadan and D. Darwis, "Perbandingan Metode Naïve Bayes Dan Svm Untuk Sentimen Analisis Masyarakat Terhadap Serangan Ransomware Pada Data Kip-K," *J. Simika*, vol. 8, no. 1, pp. 12–23, 2025, doi: <https://doi.org/10.47080/simika.v8i1.3621>.
- [18] A. M. Ndapamuri, D. Manongga, and A. Iriani, "Analisis Sentimen Ulasan Aplikasi Tripadvisor Dengan Metode Support Vector Machine, K-Nearest Neighbor, Dan Naive Bayes," *INOVTEK Polbeng - Seri Inform.*, vol. 8, no. 1, p. 127, 2023, doi: 10.35314/isi.v8i1.3260.
- [19] I. Iin, R. Supriatna, M. Mulyawan, and D. Rohman, "Penerapan Natural Language Processing Dalam Analisis Sentimen Cawapres 2024 Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 1109–1115, 2024, doi: 10.36040/jati.v8i1.8572.
- [20] M. A. Hermawan, A. Faqih, G. Dwilestari, T. Informatika, and S. Informasi, "Implementasi Akurasi Model Naïve Bayes Menggunakan Smote Dalam Analisis Sentimen Pengguna Aplikasi Brimo," *J. Ilmu dan Teknol. Komput.*, vol. 13, no. 1, 2025, doi: <http://dx.doi.org/10.23960/jitec.v13i1.5748>.
- [21] L. Ode *et al.*, "Analisis Sentimen Aplikasi Peminjaman Online Berdasarkan Ulasan Pada Play Store Menggunakan Metode Naïve Bayes Dan Support Vector Machine (Studi Kasus : Adakami Dan Easycash)," *AnoatIK J. Teknol. Inf. Dan Komput.*, vol. 2, no. 2, 2024, doi: <https://doi.org/10.33772/anoatik.v2i2.71>.
- [22] J. Anggraini and D. Alita, "Implementasi Metode SVM Pada Sentimen Analisis Terhadap Pemilihan Presiden (Pilpres) 2024 Di Twitter," *J. Inform. J. Pengemb. IT*, vol. 9, no. 2, pp. 102–111, 2024, doi: 10.30591/jpit.v9i2.6560.
- [23] M. Samantri and Afriyati, "Perbandingan Algoritma Support Vector Machine dan Random Forest untuk Analisis Sentimen Terhadap Kebijakan Pemerintah Indonesia Terkait Kenaikan Harga BBM Tahun 2022," *J. JTIK (Jurnal Teknol. Inf. Dan Komunikasi)*, vol. 8, no. 1, pp. 1–9, 2024, doi: <https://doi.org/10.35870/jtik.v8i1.1202>.
- [24] "Eskiyaturrofikoh" and R. R. Suryono, "Analisis Sentimen Aplikasi X Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine (Svm)," *JIPI(Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 3, pp. 1408–1419, 2024, doi: <https://doi.org/10.29100/jipi.v9i3.5392>.
- [25] P. M. Rafli Saputra, "Analisis sentimen twitter terhadap konflik di papua menggunakan perbandingan naive bayes dan svm," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 10, no. 2, pp. 1197–1208, 2025, doi: <https://doi.org/10.29100/jipi.v10i2.6180>.
- [26] M. A. Java, Mohammad Syafrullah, W. Windarto, and P. Painem, "Analisis Sentimen Ulasan Pengguna Aplikasi Threads pada Google Play Store Menggunakan Multinomial Naïve Bayes dan Support Vector Machine," *J. Ticom Technol. Inf. Commun.*, vol. 12, no. 2, pp. 75–80, 2024, doi: 10.70309/ticom.v12i2.112.