

ENHANCING SENTIMENT ANALYSIS OF THE 2024 INDONESIAN PRESIDENTIAL INAUGURATION ON X USING SMOTE-OPTIMIZED NAIVE BAYES CLASSIFIER

Lasmedi Afuan^{*1}, Muthia Khanza², Adila Zahira Hasyati³

^{1,2,3}Informatics, Universitas Jenderal Soedirman, Indonesia

Email: ¹lasmedi.afuan@unsoed.ac.id, ²muthia.khanza@mhs.unsoed.ac.id, ³adila.hasyati@mhs.unsoed.ac.id

(Article received: January 05, 2025; Revision: January 08, 2025; published: February 20, 2025)

Abstract

The inauguration of the President and Vice President of Indonesia for the 2024-2029 period has drawn significant public attention, reflecting widespread political and societal interest. This study aims to optimize sentiment analysis of public opinion on X (formerly Twitter) regarding the inauguration by enhancing the Naïve Bayes Classifier (NBC) with the Synthetic Minority Over-sampling Technique (SMOTE). Addressing the issue of class imbalance in sentiment data, the research demonstrates how SMOTE improves classification robustness. The methodology includes data crawling from X, preprocessing involving tokenization, stemming, and TF-IDF feature extraction, and sentiment labeling using TextBlob. Sentiment classification is conducted with NBC, evaluated under conditions with and without SMOTE. Metrics such as accuracy, precision, recall, and F1-score are utilized to assess performance. Results indicate that the application of SMOTE increases the accuracy of NBC from 98% to 99%, with precision improving from 0.98 to 1 and recall maintaining high levels (0.99). This 1% accuracy enhancement underscores the significance of addressing class imbalance for reliable sentiment analysis. The findings contribute to a better understanding of public sentiment during critical political events and highlight the effectiveness of SMOTE in improving text classification tasks. This research provides valuable insights into leveraging machine learning techniques for analyzing imbalanced datasets, offering implications for both academic and practical applications in sentiment analysis and political studies.

Keywords: 2024 Indonesian Presidential Inauguration, Naïve Bayes Classifier, Sentiment Analysis, Synthetic Minority Over-sampling Technique (SMOTE), TF-IDF, TextBlob, Twitter.

PENINGKATAN ANALISIS SENTIMEN PELANTIKAN PRESIDEN RI TAHUN 2024 PADA X MENGGUNAKAN NAIVE BAYES CLASSIFIER YANG DIOPTIMALKAN SMOTE

Abstrak

Pelantikan Presiden dan Wakil Presiden Indonesia untuk periode 2024-2029 telah menarik perhatian publik yang signifikan, yang mencerminkan minat politik dan sosial yang luas. Penelitian ini bertujuan untuk mengoptimalkan analisis sentimen opini publik di X (sebelumnya Twitter) mengenai pelantikan dengan meningkatkan Naïve Bayes Classifier (NBC) dengan Synthetic Minority Over-sampling Technique (SMOTE). Mengatasi masalah ketidakseimbangan kelas dalam data sentimen, penelitian ini menunjukkan bagaimana SMOTE meningkatkan ketahanan klasifikasi. Metodologi tersebut mencakup perayapan data dari X, praproses yang melibatkan tokenisasi, stemming, dan ekstraksi fitur TF-IDF, dan pelabelan sentimen menggunakan TextBlob. Klasifikasi sentimen dilakukan dengan NBC, dievaluasi dalam kondisi dengan dan tanpa SMOTE. Metrik seperti akurasi, presisi, recall, dan skor F1 digunakan untuk menilai kinerja. Hasil penelitian menunjukkan bahwa penerapan SMOTE meningkatkan akurasi NBC dari 98% menjadi 99%, dengan presisi meningkat dari 0,98 menjadi 1 dan perolehan kembali tetap pada level tinggi (0,99). Peningkatan akurasi sebesar 1% ini menggarisbawahi pentingnya mengatasi ketidakseimbangan kelas untuk analisis sentimen yang andal. Temuan ini berkontribusi pada pemahaman yang lebih baik tentang sentimen publik selama peristiwa politik kritis dan menyoroti efektivitas SMOTE dalam meningkatkan tugas klasifikasi teks. Penelitian ini memberikan wawasan berharga dalam memanfaatkan teknik pembelajaran mesin untuk menganalisis kumpulan data yang tidak seimbang, yang menawarkan implikasi untuk aplikasi akademis dan praktis dalam analisis sentimen dan studi politik.

Kata kunci: Analisis Sentimen, Pelantikan Presiden RI 2024, Twitter, Naïve Bayes Classifier, Synthetic Minority Over-sampling Technique (SMOTE), TF-IDF, TextBlob.

1. PENDAHULUAN

Pemilihan presiden 2024 menghadirkan pasangan calon Prabowo Subianto dan Gibran Rakabuming Raka, yang tidak hanya menarik perhatian luas dari masyarakat Indonesia, tetapi juga menjadi topik diskusi hangat di berbagai kalangan, mulai dari pengamat politik hingga masyarakat umum[1]. Pasangan ini menjadi sorotan utama dalam pemilu yang menandai momen transisi kepemimpinan negara menuju era baru. Prabowo Subianto, yang sudah lama dikenal dalam dunia politik Indonesia, tampil sebagai calon presiden, sementara Gibran Rakabuming Raka, yang merupakan putra Presiden Joko Widodo, menjadi calon wakil presiden[2]. Setelah pelantikan pada 20 Oktober 2024, Presiden Prabowo menyampaikan komitmennya untuk memperkuat sektor-sektor vital, seperti ketahanan pangan dan energi, guna memastikan Indonesia lebih mandiri. Proses pelantikan ini menjadi titik balik penting bagi politik Indonesia, yang disambut dengan berbagai harapan dari masyarakat[3].

Saat ini, media sosial X merupakan salah satu platform jejaring sosial yang paling populer di dunia. Pengguna dapat memposting pesan yang dikenal sebagai "tweet," yang dapat berisi teks, foto, video, dan tautan. Dengan popularitasnya yang tinggi, X menjadi salah satu sumber data utama untuk berbagai analisis, termasuk analisis sentimen. Pada Maret 2024, platform ini memiliki 250 juta pengguna aktif harian[5]. Perubahan nama dari Twitter menjadi "X" terjadi pada 22 Juli 2023 sebagai bagian dari kebijakan baru yang diperkenalkan oleh Elon Musk[6].

Sebagai platform yang sering digunakan untuk diskusi publik, media sosial X telah menjadi sumber utama dalam memahami opini masyarakat terhadap isu-isu tertentu. Analisis sentimen adalah teknik yang digunakan untuk mengevaluasi teks dengan tujuan mendeteksi emosi atau opini yang mendasarinya. Namun, karena prevalensi bahasa gaul, singkatan, dan salah eja dalam teks yang diposting di X, klasifikasi sentimen memerlukan metode yang robust untuk memberikan hasil yang akurat. Oleh karena itu, analisis data dari X menjadi semakin penting untuk memahami sentimen masyarakat terhadap berbagai isu, termasuk peristiwa besar seperti pelantikan Presiden. Dalam penelitian ini, algoritma Naïve Bayes Classifier (NBC) digunakan untuk klasifikasi sentimen. Algoritma ini dikenal dengan kesederhanaan dan efisiensinya dalam menganalisis data berbasis teks. Dengan menggabungkan algoritma ini dengan teknik seperti SMOTE, penelitian ini bertujuan untuk meningkatkan akurasi analisis sentimen terhadap data yang tidak seimbang, sehingga dapat memberikan wawasan yang lebih baik terhadap opini publik[7][8].

Algoritma NBC menawarkan keunggulan yang berbeda dalam klasifikasi sentimen, didukung oleh

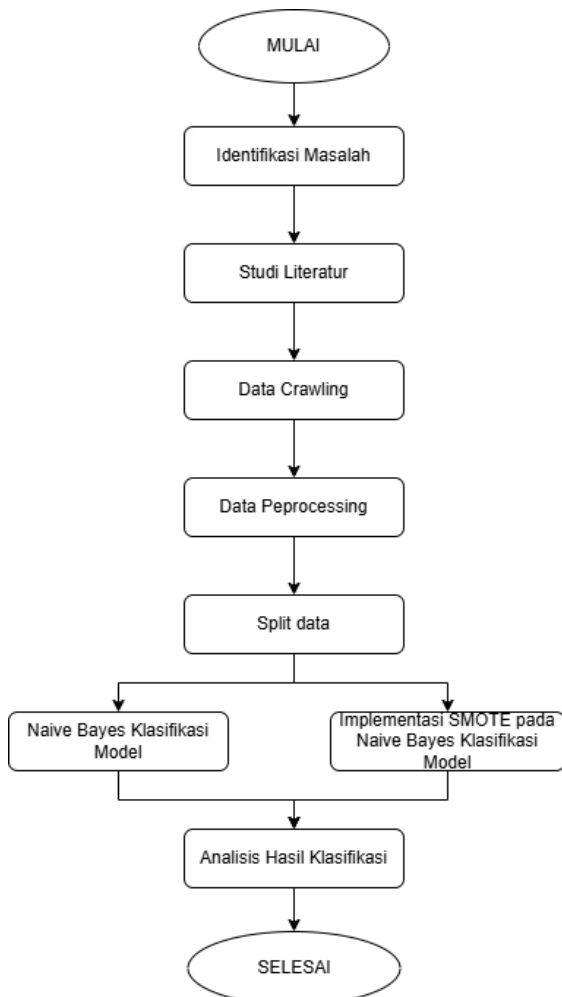
beberapa tahapan yang diperlukan [9]. Skema pembelajaran NBC digunakan dengan baik pada sebagian besar tugas klasifikasi karena kesederhanaan, efektivitas, kemampuan interpretasi, dan efisiensi komputasi. Namun, pendekatan ini membuat asumsi utama bahwa semua atribut sepenuhnya independen dari satu sama lain dalam kelas yang diberikan, yang jarang akurat dalam aplikasi dunia nyata dan akan merusak kinerjanya dengan ketergantungan atribut [10]. Dengan bantuan Naïve Bayes Classifier, kami akan mengkategorikan ulasan ke dalam sentimen positif dan negatif dan menggunakan visualisasi komparatif untuk menyajikan temuan dalam format yang mudah dimengerti[11]. Salah satu pendekatan yang menonjol untuk mengatasi ketidakseimbangan kelas adalah Synthetic Minority Over-sampling Technique (SMOTE), yang menghasilkan sampel sintetis untuk kelas minoritas berdasarkan contoh yang ada. Dengan memahami dataset yang tidak seimbang, para peneliti dan praktisi dapat memperoleh perspektif yang komprehensif tentang tantangan unik dalam masalah klasifikasi yang tidak seimbang. Wawasan ini sangat penting untuk mengembangkan teknik yang efektif seperti SMOTE untuk meningkatkan kinerja model klasifikasi[12]. Dalam SMOTE, sampel sintetis dihasilkan alih-alih menggunakan oversampling dengan penggantian [13].

Metode oversampling adalah salah satu metode yang paling representatif untuk memecahkan masalah klasifikasi data yang tidak seimbang, yang dapat mewujudkan keseimbangan antara sampel dari kelas yang berbeda dengan mensintesis sampel kelas minoritas yang baru[14]. Bias terhadap sampel kelas mayoritas membuat pengklasifikasi kurang mampu mengidentifikasi sampel kelas minoritas, dan bahkan mungkin sepenuhnya mengabaikan pentingnya kelas minoritas, yang dapat menyebabkan kesalahan klasifikasi dalam aplikasi praktis [15], [16]. Sebagai contoh, untuk kumpulan data di mana hanya 1% dari contoh yang termasuk dalam kelas minoritas, meskipun model mengklasifikasikan semua contoh sebagai kelas mayoritas, model tersebut masih mencapai akurasi keseluruhan sebesar 99%. Namun, contoh kelas minoritas, yang ingin kami klasifikasikan secara akurat, semuanya salah diklasifikasikan oleh model ini meskipun model ini mencapai akurasi yang sangat tinggi [17]. Kinerja algoritma dievaluasi menggunakan operator SMOTE dan non-SMOTE. Analisis komparatif ini memungkinkan untuk menilai efektivitas SMOTE dalam mengatasi ketidakseimbangan kelas dan meningkatkan generalisasi model. Dengan menguji algoritma dalam kedua kondisi tersebut, penelitian ini menentukan apakah penerapan SMOTE meningkatkan akurasi klasifikasi dan ketangguhan model [18].

Tujuan dari penelitian ini adalah untuk mengoptimalkan performa NBC dalam analisis

sentimen pada data pelantikan Presiden 2024 di platform X dengan menggunakan SMOTE. Penelitian ini juga bertujuan untuk memberikan wawasan tentang bagaimana mengelola dataset tidak seimbang dalam analisis sentimen dan memberikan kontribusi signifikan bagi penelitian di bidang ilmu komputer dan teknologi informasi.

2. METODE PENELITIAN



Gambar 1. Metode Penelitian

2.1. Identifikasi Masalah

Permasalahan yang diidentifikasi dalam penelitian ini adalah bagaimana mengoptimalkan algoritma NBC dengan menggunakan metode SMOTE untuk meningkatkan akurasi analisis sentimen mengenai Pelantikan Presiden Indonesia tahun 2024 di Twitter.

2.2. Studi Literatur

Penelitian ini menggunakan setidaknya 29 karya literatur yang berbeda. Studi literatur tersebut dipilih dengan parameter penelitian berdasarkan judul berita, database jurnal, tahun publikasi, dan indeks jurnal. Informasi tersebut berisi tentang analisis sentimen,

SMOTE, NBC, dan pelantikan Presiden Indonesia Tahun 2024.

2.3. Data Crawling

Tujuan utama dari pengumpulan data adalah untuk mendapatkan data penelitian. Data tweet akan diambil dari aplikasi X dengan menggunakan bahasa pemrograman Python. Pengumpulan data penelitian ini memanfaatkan Google Collaboratory dan Chrome sebagai tools utama dalam proses pengumpulan data.

2.4. Data Preprocessing

Preprocessing sangat penting untuk membersihkan data agar siap digunakan. Pada langkah ini, beberapa prosedur dilakukan untuk memproses data sebelum digunakan untuk klasifikasi menggunakan algoritma NBC. Prosedur-prosedur tersebut antara lain [19], [20]:

- Pembersihan data adalah proses yang bertujuan untuk memastikan bahwa data dalam kumpulan data memiliki konsistensi dan kegunaan yang optimal.
- Case Folding digunakan untuk menyeragamkan atau mengubah semua huruf dari “a” hingga “z” dalam dataset menjadi huruf kecil
- Normalisasi adalah prosedur untuk menyamakan atau menyeragamkan kata-kata yang ditulis dengan cara yang berbeda tetapi memiliki arti yang sama serta mengubah kata-kata yang tidak baku menjadi kata-kata baku.
- Stopword removal* adalah proses penghapusan kata-kata yang sering muncul tetapi tidak penting dan tidak relevan, seperti kata penghubung, kata ganti orang, dan kata ganti kepemilikan..
- Tokenisasi adalah aksi memecah kalimat menjadi kata-kata atau token. Metode ini memecah kata-kata dari setiap kalimat menjadi token.
- Stemming, pada tahap ini, kata-kata dengan imbuhan diubah menjadi kata dasar.

2.5. TF-IDF

TF-IDF digunakan sebagai metode ekstraksi fitur untuk mengubah teks dari dokumen menjadi vektor numerik yang dapat digunakan oleh algoritma klasifikasi. Dengan menggunakan TF-IDF, kita dapat mengukur tingkat kepentingan kata dalam sebuah dokumen dan menghasilkan representasi fitur yang berguna untuk membedakan kelas-kelas yang berbeda dalam klasifikasi teks. Ekstraksi fitur dalam data mining adalah proses yang melibatkan langkah-langkah untuk mengurangi jumlah data yang tersedia untuk mendeskripsikan kumpulan data yang besar. Ketika menganalisis sentimen teks yang kompleks, salah satu masalah utama yang muncul adalah banyaknya variabel. Menerapkan teknik ekstraksi fitur pada data input sebelum meneruskannya ke

algoritma klasifikasi akan meningkatkan akurasi model pengklasifikasi.

Bagian ini memperkenalkan teknik Ekstrak Elemen baru untuk mengekstrak elemen dari teks yang diberikan. Menggunakan kata-kata konteks untuk mewakili atau mengekstrak makna dari teks yang besar dikenal sebagai “vektorisasi balik.” Beberapa batasan timbal balik diberlakukan pada setiap kata untuk menentukan korespondensi kata yang mungkin. Data teks juga dapat digunakan untuk mendapatkan metrik yang berguna seperti Term Frequency (TF) dan Inverse Document Frequency (IDF). Persamaan (1) dan (2) digunakan oleh TF dan IDF untuk menghitung frekuensi kemunculan sebuah istilah dalam sebuah dokumen. Berikut ini adalah persamaan TF-IDF yang ditunjukkan pada persamaan (1) - (3) [8], [21]:

$$TF_{ij} = \frac{n_{ij}}{d_j} \quad (1)$$

$$IDF_i = \lg\left(\frac{D}{N_i}\right) \quad (2)$$

$$TF - IDF = TF_{ij} \cdot IDF_i \quad (3)$$

2.6. Data Split

Proses pemisahan data membagi dataset menjadi data pelatihan dan pengujian. Awalnya, data pengujian diekstrak dari dataset. Data yang tersisa kemudian digunakan untuk tujuan pelatihan dan disusun ke dalam kelompok pengujian yang terpisah.

2.7. Labeling

Dalam proses pelabelan ini, digunakan pustaka TextBlob. TextBlob adalah sebuah library Python yang digunakan untuk pemrosesan teks dan natural language processing (NLP). TextBlob cukup baik dalam mengidentifikasi dan mengkategorikan emosi tekstual. Dalam penggunaan TextBlob, polarity digunakan untuk menentukan apakah sebuah sentimen masuk ke dalam kategori positif atau negatif berdasarkan nilai polarity yang dihasilkan [22], [23], [24], [25].

2.8. Model Klasifikasi

Dalam pembuatan model klasifikasi ini, ada beberapa metode yang digunakan yaitu klasifikasi NBC dan Synthetic Minority Over-sampling Technique (SMOTE). Setelah ekstraksi fitur, akan dilakukan pengujian klasifikasi dengan pengklasifikasi NBC, yang ditunjukkan pada persamaan (4)[8]:

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

Pengujian selanjutnya adalah SMOTE yang digunakan untuk mengatasi oversampling data dan meningkatkan akurasi dari penelitian ini. Persamaan dari SMOTE (5)[8]:

$$X_{syn} = X_i + rand(0,1) \times |X_i - X_{neighbour}| \quad (5)$$

2.9. Analisis Hasil Klasifikasi

Langkah terakhir dalam penelitian ini adalah analisis hasil klasifikasi. Analisis hasil klasifikasi dilakukan setelah kedua skenario perbandingan data melewati penerapan ketiga varian NBC. Pada penelitian ini, evaluasi dilakukan dengan menggunakan confusion matrix untuk ketiga varian NBC pada seluruh skenario perbandingan, yang terdiri dari klasifikasi pro dan kontra. Tahap evaluasi menggunakan confusion matrix dilakukan untuk menghitung nilai akurasi, presisi, recall, dan F1-score, yang digunakan untuk mengevaluasi performa kedua varian NBC pada tiga skenario perbandingan data yang berbeda [26]. Perhitungan nilai akurasi pada Tabel 1 ditunjukkan pada Persamaan (6).[8], [27]:

Tabel 1. *Confusion Matrix*

<i>Actual Class</i>	<i>Predicted Class</i>	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	<i>True Positive</i>	<i>False Positive</i>
<i>Negative</i>	<i>False Negative</i>	<i>True Negative</i>

$$precision = \frac{TP}{TP+FP} \quad (6)$$

$$recall = \frac{TP}{TP+FN} \quad (7)$$

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

$$f1 - score = 2 \times \frac{precision+recall}{precision \times recall} \quad (9)$$

Pada persamaan akurasi di atas TP berarti True Positive, TN berarti True Negative, FP berarti False Positive, dan FN berarti False Negative. Setelah proses perhitungan akurasi, maka akan didapatkan nilai akurasi dan dapat melakukan analisis untuk menyimpulkan hasil penelitian.

3. HASIL DAN PEMBAHASAN

3.1. Data Crawling

Pengumpulan data dilakukan melalui tiga langkah utama. Pertama, menentukan kata kunci dan rentang pencarian. Kata kunci yang digunakan adalah “Fufufafa”, “Pelantikan Prabowo”, “Prabowo Gibran”, “Presiden RI 2024”, “Pelantikan Presiden 2024”, dan “Kabinet Mera. Data tweet akan diambil dari aplikasi X dengan menggunakan Python. Pengumpulan data penelitian ini menggunakan

Google Collaboratory dan Chrome sebagai alat utama dalam proses pengumpulan data. Langkah terakhir yang dilakukan adalah mengumpulkan data selama periode 1 Oktober 2024 hingga 11 November 2024.

Setelah mengumpulkan data, analisis dilakukan terhadap 1127 baris tweet dengan 15 kolom tabel. Data ini dianalisis untuk mengidentifikasi tren, sentimen, dan topik utama dari diskusi online selama periode tersebut. Hasil pengumpulan data dirinci dalam Tabel 2, yang menunjukkan distribusi tweet per kata kunci dan berbagai metrik lain yang relevan.

Table 2. Detailed Data Collection Results

Keywords	Total
Fufufafa	549
Pelantikan Prabowo	114
Prabowo Gibran	44
Presiden RI 2024	179
Pelantikan Presiden 2024	121
Kabinet Merah Putih	120
Total	1.127

3.2. Data Preprocessing

Data Preprocessing dimulai dengan langkah awal yaitu menghapus 14 kolom dari dataset, menyisakan hanya 1 kolom yang berfungsi sebagai tempat untuk komentar teks dari hasil perayapan. Hal ini memungkinkan fokus pada informasi teks yang relevan untuk analisis lebih lanjut. Hasil dari preprocessing awal dapat dilihat pada Tabel 3.

Tabel 3. Data Set

No	Teks
1	Presiden China Xi Jinping menegaskan dukungannya kepada pemerintahan Presiden RI Prabowo Subianto ... percaya bahwa https://t.co/yvg11f0Gi
2	Presiden terpilih RI tampak berjalan kurang stabil wakilnya yang sangat muda. Ia bisa kejatuhan tanggung jawab memimpin negeri sebesar RI. https://t.co/EMjH7GWXCI
...
1127	Poin ² menarik yg aku highlight dari Bu Connie: ... mention Mulyono wkwk. - potensi proxy war (I think belio udah tahu sesuatu cuma nggak mau eksplisit aja) https://t.co/jQq2w3n4eN

Preprocessing lebih lanjut meliputi tahap Cleansing, Case Folding, Normalisasi, Stopword, dan Tokenisasi. Langkah selanjutnya akan mengeksekusi dokumen keluaran dari setiap tahap *preprocessing* lanjutan. Tabel 4 menampilkan hasil *preprocessing* lanjutan untuk satu kalimat atau dokumen.

Tabel 4. Tahap Preprocessing

No	Tahap Preprocessing	Dokumen
1	-	Bismillah Aksi 411 Besok 4 Nopember 2024 Bersama Menuju Indonesia yg baik Adil Makmur utk seluruh Rakyat Indonesia Agar Keadilan & amp
2	Cleansing	Bismillah Aksi 411 Besok 4 Nopember 2024 Bersama Menuju Indonesia yg baik Adil Makmur utk seluruh Rakyat Indonesia Agar Keadilan
3	Case Folding	bismillah aksi 411 besok 4 nopember 2024 bersama menuju indonesia yg

4	Normalization	baik adil makmur utk seluruh rakyat indonesia agar keadilan bismillah aksi 411 besok 4 nopember 2024 bersama menuju indonesia baik adil makmur untuk seluruh rakyat indonesia agar keadilan
5	Stopword	bismillah aksi 411 besok 4 nopember 2024 bersama menuju indonesia baik adil makmur untuk seluruh rakyat indonesia agar keadilan
6	Tokenization	bismillah aksi 411 besok 4 nopember 2024 bersama menuju indonesia baik adil makmur untuk seluruh rakyat indonesia agar keadilan
7	Stemming	bismillah aksi 411 besok 4 nopember 2024 bersama menuju indonesia baik adil makmur untuk seluruh rakyat indonesia agar keadilan

3.3. TF-IDF

Data tersebut memerlukan ekstraksi fitur menggunakan metode TF-IDF sebelum dioperasikan. Hasil uji coba TF-IDF untuk satu contoh kalimat atau dokumen data latih dapat dilihat pada Gambar 2.

(0, 173)	0.1982670841368174
(0, 782)	0.293846470288588
(0, 1056)	0.08604017522437768
(0, 1425)	0.1547690432079409
(0, 1643)	0.27777845570438403
(0, 2496)	0.33760334304269696
(0, 2533)	0.33760334304269696
(0, 3019)	0.20990109952093272
(0, 3097)	0.20671629595780355
(0, 3427)	0.15581863171856655
(0, 3444)	0.27777845570438403
(0, 3563)	0.33760334304269696
(0, 3567)	0.33760334304269696
(0, 3573)	0.142883173039273
(0, 3582)	0.3182406183915033

Gambar 2. TF-IDF

3.4. Split Data

Tahap selanjutnya setelah persiapan data adalah pemisahan dataset. Dataset telah melalui proses preprocessing dan perhitungan TF-IDF, menghasilkan 1080 data yang bersih dan siap digunakan. Data ini kemudian dibagi menjadi dua subset: 216 data untuk pengujian dan 864 data untuk pelatihan. Data training digunakan untuk memodelkan setiap kelompok uji menggunakan metode Naïve Bayes, baik dengan maupun tanpa penerapan SMOTE.

3.5. Labeling

Program ini menggunakan pustaka TextBlob untuk menganalisis sentimen dataset menggunakan 1080 data untuk pelatihan. Program ini mengklasifikasikan sentimen tweet sebagai positif atau negatif. Hasilnya disimpan dalam sebuah bingkai data yang mencakup teks tweet dan klasifikasinya. Program mencetak hasil pelabelan, yaitu jumlah sentimen positif 846 dan 18 sentimen negatif. Hasil labeling dapat dilihat pada Gambar 3.

Results of Data Analysis:
 Positive = 846
 Negative = 18

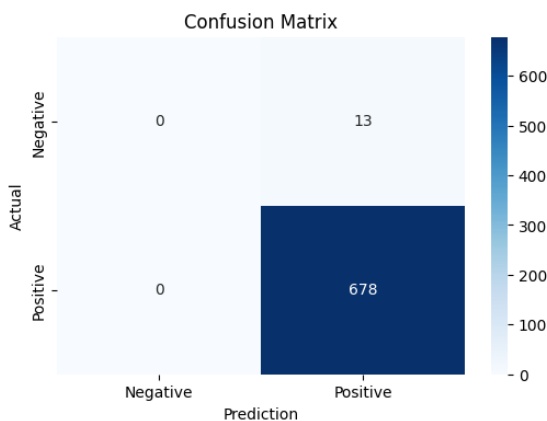
Gambar 3. Labeling

3.6. Model Klasifikasi

Setelah memberi label pada data pelatihan, klasifikasi model dilakukan untuk menentukan sentimen teks yang dianalisis dengan menggunakan beberapa metode.

3.6.1. Implementasi Naïve Bayes

Langkah selanjutnya adalah penerapan Naïve Bayes untuk mengevaluasi setiap kelompok uji. Proses ini akan menghasilkan Confusion Matrix yang berguna untuk menentukan nilai akurasi dari pemodelan. Gambar 4 menampilkan data confusion matrix Naïve Bayes pada pengujian.



Gambar 4. Naïve Bayes Confusion Matrix Result

Dari hasil Confusion Matrix pada Gambar 4, dapat dilihat bahwa pada pengujian, Naïve Bayes berhasil memprediksi dengan benar sebanyak 678 kali, namun mengalami 13 kali prediksi yang salah. Untuk mengukur performa dari model, terdapat empat pengukuran dalam confusion matrix yaitu precision, recall, overall accuracy, dan f1-score [27].

Pertama, untuk mengukur rasio observasi positif yang diprediksi dengan benar, persamaan presisi digunakan seperti yang dijelaskan dalam persamaan (6). Kedua, rasio observasi positif yang diprediksi dengan benar dibandingkan dengan semua pernyataan di kelas yang sebenarnya digunakan untuk mengukur recall. Recall menilai seberapa efektif model dalam mengidentifikasi sampel positif. Untuk mengukur recall, persamaan seperti yang dijelaskan dalam persamaan (7) digunakan. Selanjutnya, kinerja model diukur dengan menggunakan akurasi. Akurasi menunjukkan kemampuan model untuk memprediksi dengan benar baik positif maupun negatif dari semua hasil. Untuk mengukur akurasi, digunakan persamaan seperti pada persamaan (8). Dari perhitungan akurasi, diperoleh nilai akurasi sebesar 0,98. Terakhir, keseimbangan kinerja dinilai dengan perhitungan f1-

score, yang merupakan rata-rata dari nilai precision dan recall. Untuk mengukur f1-score, digunakan persamaan seperti yang dijelaskan pada persamaan (9).

Dari hasil analisis performa menggunakan confusion matrix setelah menerapkan NBC didapatkan nilai precision sebesar 0,98, recall sebesar 1, akurasi sebesar 0,98, dan nilai f1-score sebesar 0,99. Hasil pengukuran confusion matrix dapat dilihat pada Gambar 5.

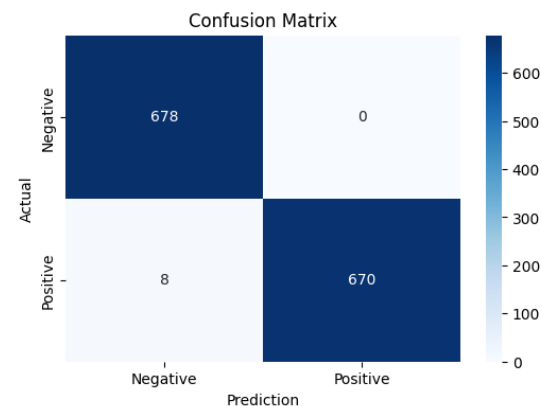
Classification Report:

	precision	recall	f1-score	support
Negatif	0.00	0.00	0.00	13
Positif	0.98	1.00	0.99	678
accuracy			0.98	691

Gambar 5. Hasil NBC

3.6.2. Implementasi SMOTE dalam Naïve Bayes Classifier

Selanjutnya, SMOTE diterapkan untuk menangani ketidakseimbangan data. Proses ini menggunakan data yang sama dengan pengujian Naïve Bayes sebelumnya. Model dibangun dengan menggunakan metode NBC pada data yang telah melalui ekstraksi fitur.



Gambar 6. Implementasi SMOTE pada NBC

Gambar 6 menunjukkan *confusion matrix*, yang digunakan untuk mengevaluasi kemampuan SMOTE. Menurut gambar tersebut, penerapan SMOTE pada Naïve Bayes menghasilkan 1348 prediksi yang benar dan 8 prediksi yang salah dalam pengujian ini. Untuk mengukur performa model, terdapat empat pengukuran dalam confusion matrix, yaitu *precision*, *recall*, *overall accuracy*, dan *f1-score*. Pengukuran performa dilakukan dengan cara yang sama seperti yang telah dilakukan pada pengukuran performa model NBC tanpa SMOTE sebelumnya.

Dari hasil analisis performa menggunakan confusion matrix setelah menerapkan SMOTE, didapatkan nilai precision sebesar 1, recall sebesar 0,99, akurasi sebesar 0,99, dan nilai f1-score sebesar 0,99. Hasil pengukuran confusion matrix dapat dilihat pada Gambar 7.

Classification Report (Resampled Data):			
	precision	recall	f1-score
Negatif	0.99	1.00	0.99
Positif	1.00	0.99	0.99
accuracy	0.99		

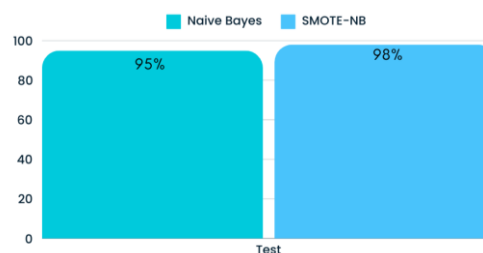
Gambar 7. NBC dengan SMOTE Classification Report

4. DISKUSI

Penelitian ini menunjukkan bahwa penerapan SMOTE pada Naïve Bayes Classifier berhasil meningkatkan kinerja model dalam analisis sentimen, khususnya pada data yang memiliki ketidakseimbangan kelas signifikan. Data yang digunakan dalam penelitian ini terdiri dari 1.127 baris tweet terkait Pelantikan Presiden Indonesia tahun 2024. Setelah proses preprocessing, data yang siap digunakan berjumlah 1.080 baris, dengan distribusi kelas yang tidak seimbang. Dengan implementasi SMOTE, data kelas minoritas berhasil diperluas sehingga model mampu mempelajari pola dari kelas tersebut dengan lebih baik. Hasil analisis performa menunjukkan peningkatan akurasi sebesar 1% (dari 98% menjadi 99%), dengan precision meningkat dari 0.98 menjadi 1.00 dan recall tetap tinggi di angka 0.99. Peningkatan ini mengindikasikan bahwa model mampu menangkap pola dari data minoritas tanpa mengorbankan generalisasi terhadap data mayoritas. Dalam analisis lebih lanjut, penggunaan confusion matrix memberikan wawasan mendalam tentang kontribusi SMOTE dalam mengurangi jumlah prediksi salah pada kelas minoritas. Perbandingan dengan penelitian sebelumnya menunjukkan bahwa SMOTE secara konsisten efektif dalam menangani dataset tidak seimbang. Sebagai contoh, penelitian serupa pada analisis sentimen terhadap kandidat pemilu 2024 menggunakan SMOTE dan menunjukkan peningkatan akurasi dari 57.50% menjadi 73.17%. Penelitian ini memperkuat bukti bahwa SMOTE dapat digunakan pada berbagai skenario klasifikasi dengan ketidakseimbangan data, baik dalam konteks politik maupun domain lainnya. Namun, terdapat beberapa keterbatasan dalam penelitian ini. Salah satunya adalah penggunaan satu algoritma klasifikasi, yaitu Naïve Bayes Classifier, yang meskipun sederhana dan efisien, memiliki asumsi independensi antar fitur yang mungkin tidak selalu terpenuhi dalam data dunia nyata. Selain itu, penelitian ini hanya menggunakan data dari satu platform media sosial (X/Twitter), sehingga hasilnya belum dapat digeneralisasi ke platform lain. Implikasi praktis dari penelitian ini mencakup penerapan metode SMOTE untuk meningkatkan kinerja analisis sentimen pada acara politik besar lainnya atau aplikasi serupa. Selain itu, hasil penelitian ini dapat menjadi acuan bagi praktisi dan peneliti yang ingin mengatasi masalah ketidakseimbangan data pada klasifikasi teks. Penelitian mendatang disarankan untuk mengeksplorasi kombinasi SMOTE dengan

algoritma klasifikasi lain, seperti Random Forest atau Support Vector Machine, untuk mengevaluasi potensi peningkatan kinerja. Selain itu, mengaplikasikan metode ini pada platform media sosial lain atau topik yang berbeda dapat memberikan wawasan tambahan tentang efektivitas metode balancing data. Hasil uji akurasi dari kedua metode dapat dilihat pada Gambar 8.

MODEL ACCURACY RESULT



Gambar 8. Hasil akurasi model

5. KESIMPULAN

Pada penelitian ini, implementasi SMOTE pada algoritma Naïve Bayes Classifier diterapkan untuk mengatasi ketidakseimbangan kelas pada analisis sentimen terkait Pelantikan Presiden Indonesia tahun 2024. Penelitian ini membandingkan kinerja model sebelum dan sesudah penerapan SMOTE untuk meningkatkan akurasi dan ketangguhan klasifikasi sentimen. Hasil penelitian menunjukkan bahwa penggunaan SMOTE berhasil meningkatkan akurasi model sebesar 1%, dari 98% menjadi 99%. Selain itu, precision dan recall juga mengalami peningkatan yang signifikan, dengan precision meningkat dari 0.98 menjadi 1.00 dan recall tetap tinggi di angka 0.99. Penerapan SMOTE tidak hanya membantu mengatasi masalah ketidakseimbangan data, tetapi juga memberikan kontribusi yang signifikan dalam meningkatkan kualitas hasil analisis sentimen, terutama dalam konteks data sosial media yang sangat dinamis. Hal ini menunjukkan bahwa pendekatan oversampling berbasis SMOTE dapat menjadi solusi efektif untuk meningkatkan performa model dalam skenario klasifikasi yang kompleks. Hasil ini memperkuat temuan dari penelitian sebelumnya dan memberikan wawasan baru tentang pentingnya pengelolaan dataset tidak seimbang dalam analisis sentimen politik. Sebagai implikasi praktis, metode yang digunakan dalam penelitian ini dapat diterapkan pada platform media sosial lain untuk analisis sentimen yang serupa, khususnya dalam konteks acara politik besar. Penelitian ini memiliki beberapa keterbatasan, seperti terbatasnya data yang digunakan serta fokus pada satu algoritma klasifikasi. Penelitian mendatang disarankan untuk mengeksplorasi pengaruh metode balancing lain atau mengombinasikan beberapa algoritma klasifikasi untuk meningkatkan kinerja analisis sentimen lebih lanjut. Selain itu, analisis lebih mendalam pada

platform media sosial lain juga dapat memberikan generalisasi yang lebih luas terhadap efektivitas metode ini.

UCAPAN TERIMA KASIH

Artikel ini adalah hasil dari proyek penelitian yang didanai oleh LPPM Universitas Jenderal Soedirman dengan nomor kontrak 26.827/UN23.35.5/PT.01/II/2024. Kami dengan tulus mengucapkan terima kasih atas dukungan dan sumber daya yang diberikan oleh Universitas. Para penulis juga ingin menyampaikan rasa terima kasih kepada semua pihak yang terlibat dalam penyelesaian karya ini.

DAFTAR PUSTAKA

- [1] J. Jtik *et al.*, "Analisis Semiotika Pierce terhadap Pasangan Prabowo-Gibran dalam Pilpres 2024," vol. 8, no. 4, 2024.
- [2] DDTC News, "Sah! Prabowo-Gibran Resmi Jadi Presiden dan Wakil Presiden 2024-2029," DDTC News, 3 Desember 2024. [Online]. Tersedia: <https://news.ddtc.co.id/berita/nasional/1806298/sah-prabowo-gibran-resmi-jadi-presiden-dan-wakil-presiden-2024-2029>.
- [3] Kementerian Pertahanan Republik Indonesia, "Pelantikan Prabowo Subianto dan Gibran Rakabuming Raka sebagai Presiden dan Wakil Presiden RI 2024-2029," Kementerian Pertahanan RI, 21 Oktober 2024. [Online]. Tersedia: <https://www.kemhan.go.id/2024/10/21/pelantikan-prabowo-subianto-dan-gibran-rakabuming-raka-sebagai-presiden-dan-wakil-presiden-ri-2024-2029.html>.
- [4] E. Memi, "applied sciences Comparative Study for Sentiment Analysis of Financial Tweets with Deep Learning Methods," 2024.
- [5] J. L. Kent K. Lee, Marilyn L. Abbott, Shiran Wang, "The Uses of X/Twitter by Members of the TESOL Community," *CJLT RCAT*, vol. 49, no. 3, 2023, [Online]. Available: <https://cjlt.ca/index.php/cjlt/article/view/28495/20792>
- [6] K. P. Widiatmika, "Seeking Answers In The Shadowsthe Use of Auto-Based Accounts on Twitter/X As A Platform for Social Q&A," *J. Sos. Hmaniora*, vol. 15, 2024, [Online]. Available: <https://ojs.unida.ac.id/JSH/article/view/12884/5066>
- [7] D. B. R. Krishna, "Twitter Sentiment Analysis," *Int. J. Innov. Eng. Manag. Res.*, vol. 24, no. 05, pp. 811–825, 2022, doi: 10.48047/ijiemr/v11/i06/53.
- [8] S. Setianingsih *et al.*, "Implementation of SMOTE to Overcome Class Imbalance in the Naive Bayes Classifier in Sentiment Analysis of Presidential Candidates in the 2024 General Election," *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 99, no. 1, pp. 1–1, 2019, doi: 10.12928/TELKOMNIKA.v99i1.paperID.
- [9] Y. Afrianto Singgalen, "Sentiment Classification of Over-Tourism Issues in Responsible Tourism Content using Naïve Bayes Classifier," *J. Comput. Syst. Informatics*, vol. 5, no. 2, pp. 275–285, 2024, doi: 10.47065/josyc.v5i2.4904.
- [10] Z. Gao, S. Wu, Z. Wan, and S. Agaian, "A Hybrid Method for Implicit Intention Inference Based on Punished-Weighted Naïve Bayes," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1826–1836, 2023, doi: 10.1109/TNSRE.2023.3259550.
- [11] A. V. D. Sano, A. A. Stefanus, E. D. Madyatmadja, H. Nindito, A. Purnomo, and C. P. M. Sianipar, "Proposing a visualized comparative review analysis model on tourism domain using Naïve Bayes classifier," *Procedia Comput. Sci.*, vol. 227, pp. 482–489, 2023, doi: 10.1016/j.procs.2023.10.549.
- [12] B. Nemade, V. Bharadi, S. S. Alegavi, and B. Marakarkandy, "A Comprehensive Review: SMOTE-Based Oversampling Methods for Imbalanced Classification Techniques, Evaluation, and Result Comparisons," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 9s, pp. 790–803, 2023.
- [13] T. Watthaisong and N. M. , Khamron Sunat, "Comparative Evaluation of Imbalanced Data Management Techniques for Solving Classification Problems on Imbalanced Datasets," *Int. Acad. Press*, vol. 12, pp. 547–570, 2024, [Online]. Available: <http://www.iapress.org/index.php/soic/article/view/1890/1063>
- [14] Y. Zhang, L. Deng, and B. Wei, "Imbalanced Data Classification Based on Improved Random-SMOTE and Feature Standard Deviation," pp. 1–17, 2024.
- [15] M. J. A. Fatemeh Daneshfar, "Enhanced text classification through an improved discrete laying chicken algorithm," *Wiley*, 2024, [Online]. Available: <https://onlinelibrary.wiley.com/doi/epdf/10.1111/exsy.13553>
- [16] V. R. Revathy, A. S. Pillai, and F. Daneshfar, "LyEmoBERT: Classification of lyrics' emotion and recommendation using a pre-trained model," *Procedia Comput. Sci.*, vol. 218, pp. 1196–1208, 2022, doi: 10.1016/j.procs.2023.01.098.

- [17] J. Park, S. Kwon, and S. P. Jeong, "A study on improving turnover intention forecasting by solving imbalanced data problems : focusing on SMOTE and generative adversarial networks," *J. Big Data*, 2023, doi: 10.1186/s40537-023-00715-6.
- [18] Y. A. Singgalen, "Comparative Analysis of DT and SVM Model Performance with SMOTE in Sentiment Classification," vol. 4, no. 5, pp. 2485–2494, 2024, doi: 10.30865/klik.v4i5.1828.
- [19] D. A. Fatah, E. M. S. Rochman, W. Setiawan, A. R. Aulia, F. I. Kamil, and A. Su'ud, "Sentiment Analysis of Public Opinion Towards Tourism in Bangkalan Regency Using Naïve Bayes Method," *E3S Web Conf.*, vol. 499, pp. 1–8, 2024, doi: 10.1051/e3sconf/202449901016.
- [20] A. B. Putra Negara, "The Influence Of Applying Stopword Removal And Smote On Indonesian Sentiment Classification," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 14, no. 3, p. 172, 2023, doi: 10.24843/lkjiti.2023.v14.i03.p05.
- [21] S. Saranya and G. Usha, "A Machine Learning-Based Technique with Intelligent WordNet Lemmatize for Twitter Sentiment Analysis," *Intell. Autom. Soft Comput.*, vol. 36, no. 1, pp. 339–352, 2023, doi: 10.32604/iasc.2023.031987.
- [22] E. D. Ajik, A. B. Suleiman, and M. Ibrahim, "Enhancing User Experience Through Sentiment Analysis for Katsina State Transport Agency: a Textblob Approach," *Fudma J. Sci.*, vol. 7, no. 6, pp. 117–122, 2023, doi: 10.33003/fjs-2023-0706-2057.
- [23] O. Abiola, A. Abayomi-Alli, O. A. Tale, S. Misra, and O. Abayomi-Alli, "Sentiment analysis of COVID-19 tweets from selected hashtags in Nigeria using VADER and Text Blob analyser," *J. Electr. Syst. Inf. Technol.*, vol. 10, no. 1, 2023, doi: 10.1186/s43067-023-00070-9.
- [24] Ilham Firman Ashari, F. A. M. Daffa, and Sekar A, "Sentiment Analysis of Tweets About Allowing Outdoor Mask Wear Using Naïve Bayes and TextBlob," *Indones. J. Comput. Sci.*, vol. 12, no. 3, pp. 1092–1103, 2023, doi: 10.33022/ijcs.v12i3.3238.
- [25] Nehal, D. Jeet, V. Sharma, S. Mishra, C. Iwendi, and J. Osamor, "Twitter Sentiment Analysis and Emotion Detection Using NLTK and TextBlob," *2023 4th Int. Conf. Comput. Autom. Knowl. Manag. ICCAKM 2023*, 2023, doi: 10.1109/ICCAKM58659.2023.10449540.
- [26] B. Wahyu Andrian, F. Adline Twince Tobing, I. Zuhdi Pane, and A. Kusnadi, "Implementation of Naïve Bayes Algorithm in Sentiment Analysis of Twitter Social Media Users Regarding Their Interest to Pay the Tax," *Int. J. Sci. Technol. Manag.*, vol. 4, no. 6, pp. 1733–1742, 2023, doi: 10.46729/ijstm.v4i6.1015.
- [27] A. A. Ajhari, "The Comparison of Sentiment Analysis of Moon Knight Movie Reviews between Multinomial Naive Bayes and Support Vector Machine," *Appl. Inf. Syst. Manag.*, vol. 6, no. 1, pp. 13–20, 2023, doi: 10.15408/aism.v6i1.26045..