JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 3, Juli 2024, Page 1313-1322
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i3.7708

Analisis Sentimen Calon Presiden 2024 di Media Sosial X Menggunakan Naive Bayes dan SMOTE

Muhamad Hafidz Ardian Sunata, Faldy Irwiensyah*, Firman Noor Hasan

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, DKI Jakarta, Indonesia
Email: [1]hafidzard@gmail.com, [2,*]faldy@uhamka.ac.id, [3]firman.noorhasan@uhamka.ac.id
Email Penulis Korespondensi: faldy@uhamka.ac.id

Abstrak−Dalam era digital, penggunaan media sosial semakin meningkat, memungkinkan individu untuk secara terbuka menyuarakan pendapat mereka. penelitian ini menyoroti penggunaan platform media sosial X sebagai wadah utama bagi pengguna untuk berbagi pendapat, terutama terkait topik politik, khususnya dalam konteks pemilihan presiden. Metode analisis sentimen, khususnya menggunakan Algoritma naïve bayes dan teknik Synthetic Minority Oversampling (SMOTE), telah menjadi fokus penelitian untuk menarik kesimpulan masyarakat terhadap kandidat presiden. Meskipun telah ada berbagai penelitian sebelumnya, namun masih terdapat kekurangan dalam akurasi dan ketidakseimbangan data. Penelitian ini bertujuan untuk meningkatkan kinerja analisis sentimen dengan menggunakan metode naïve bayes bersama dengan SMOTE. Melalui analisis tweet di media sosial X dari 12 Desember 2023 hingga 31 Maret 2024, data dibagi menjadi kategori positif dan negatif. Hasil penelitian menunjukkan bahwa dengan SMOTE, akurasi meningkat menjadi 88% pada dataset Ganjar-Mahfud, sementara tanpa SMOTE, akurasi rendah sekitar 69% pada dataset Anies-Imin. Dari total 1589 tweet dengan sentimen positif, sekitar 27.7% mengarah ke Anies-Imin, 28.7% ke Prabowo-Gibran, dan 43.5% ke Ganjar-Mahfud. Mayoritas sentimen negatif ditujukan kepada Anies-Imin (41.5%) dan Prabowo-Gibran (40.8%).

Kata Kunci: Analisis Sentimen; Calon Presiden; Naive Bayes; SMOTE; Media Sosial X

Abstract−In the era of digital advancement, the utilization of social media has surged, enabling individuals to express their viewpoints openly. This research underscores the utilization of social media platform X as the primary avenue for users to express their opinions, particularly on political matters, notably within the framework of the presidential election. Sentiment analysis techniques, specifically employing the Naïve Bayes algorithm and the Synthetic Minority Oversampling (SMOTE) method, have been the central focus of inquiry to infer people's inclinations toward presidential candidates. Despite numerous antecedent studies, deficiencies persist in terms of precision and data imbalance. This study endeavors to enhance the efficacy of sentiment analysis by integrating the Naïve Bayes approach with SMOTE. By scrutinizing tweets on social media X spanning from December 12, 2023, to March 31, 2024, the data is categorized into positive and negative sentiments. The findings reveal that employing SMOTE bolstered accuracy to 88% for the Ganjar-Mahfud dataset, whereas accuracy without SMOTE languished at approximately 69% for the Anies-Imin dataset. Out of 1589 tweets conveying positive sentiments, approximately 27.7% were directed towards Anies-Imin, 28.7% towards Prabowo-Gibran, and 43.5% towards Ganjar-Mahfud. The preponderance of negative sentiments was aimed at Anies-Imin (41.5%) and Prabowo-Gibran (40.8%).

Keywords: Sentiment Analysis; Presidential Candidates; Naive Bayes; SMOTE; Social Media X

1. PENDAHULUAN

Penggunaan media sosial yang semakin diminati dalam era digital ini memungkinkan penggunanya untuk secara terbuka berbagi pendapat mengenai berbagai topik, termasuk ungkapan keresahan yang mereka alami [1]. Dalam evolusi teknologi, aplikasi kini diperkaya dengan versi mobile, menyederhanakan proses bagi pengguna untuk berbagi pengalaman dengan lebih lancar dan adaptif. [2]. Media sosial X memungkinkan pesan yang disebut tweet untuk dibagikan oleh penggunanya. Satu fitur unggulan dimiliki oleh media sosial X, yaitu kemampuan untuk melakukan unggahan yang disebut tweet, di mana apa pun bisa diposting oleh pengguna, baik itu teks, foto, maupun video. Menurut Statista, 500 juta tweet diposting setiap hari dan 350.000 tweet diposting setiap menit. Beragam topik di platform sosial X juga diperbincangkan, termasuk kegemaran, pendapat, dan ulasan terhadap berbagai hal. Topik-topik yang paling sering diperbincangkan di kalangan pengguna Indonesia dihimpun oleh media sosial X Indonesia. Perlu dicatat bahwa politik pada masa ini menjadi salah satu topik yang paling populer dan sering dibicarakan di kalangan pengguna media sosial X Indonesia [3].

Pengguna platform media sosial, terutama media sosial X, kerap membagikan cerita sehari-hari dan sudut pandangnya secara rutin [4]. Hal ini menandakan peran penting media sosial dalam membentuk opini publik terhadap para kandidat. Tanggapan yang berkembang di media sosial dapat menjadi cerminan dari dinamika politik yang sedang berlangsung di masyarakat. Terutama setelah pemilihan umum serentak, seperti yang baru saja terjadi akhir-akhir ini di Indonesia, di mana rakyat memilih presiden beserta Wakil Presiden yang baru. Dalam konteks pemilihan tersebut, tiga calon yang bersaing untuk posisi Presiden dan wakil presiden adalah Anies-Imin, Prabowo-Gibran, dan Ganjar-Mahfud [5].

Setelah proses perdebatan antara kandidat presiden dan kandidat wakil presiden yang dipertemukan oleh Komisi Pemilihan Umum sampai dengan pengumuman hasil pemilihan umum, muncul gelombang diskusi di media sosial X yang semakin intensif mengenai para kandidat untuk periode kepemimpinan 2024-2029. Dengan banyaknya pendapat yang beredar dari warganet Indonesia, terutama di media sosial X, diharapkan bahwa dari tanggapan kolektif warganet Indonesia, analisis sentimen mampu memberikan kesimpulan. Dalam penelitian

Submitted: **14/05/2024**; Accepted: **25/07/2024**; Published: **26/07/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 3, Juli 2024, Page 1313-1322
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i3.7708

opini, prosedur esensial dalam analisis sentimen adalah pengolahan yang memungkinkan pemahaman, ekstraksi, serta pengelolaan teks secara otomatis, dengan tujuan mendapatkan Gambaran sentimen dari kalimat opini [6].

Metode seperti naïve bayes, random forest, dan sejenisnya sering kali menjadi pilihan utama dalam menerapkan analisis sentimen [7]. Alat RapidMiner digunakan, yaitu alat yang berfokus pada pemrosesan data [8]. RapidMiner menggunakan berbagai algoritma dan prinsip penambangan data [9]. RapidMiner studio menawarkan berbagai solusi seperti analisis prediksi, text mining, dan data mining. Berbagai teknik deskriptif digunakan untuk memastikan pengambilan keputusan yang tepat [10]

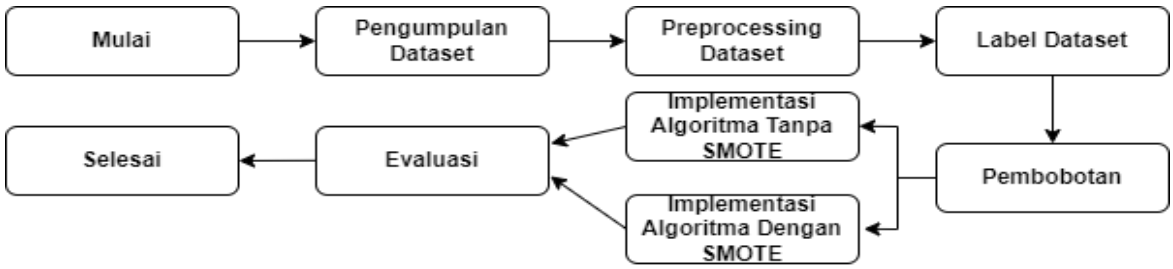
Beberapa penelitian sebelumnya telah berupaya menganalisis sentimen yang terkait dengan calon Presiden 2024. Salah satu penelitian yang dilakukan oleh Fais dkk, dengan menggunakan metode naive bayes. Hasil penelitian tersebut menunjukkan bahwa algoritma naive bayes memiliki tingkat akurasi sebesar 73,68%. Hal ini mengindikasikan bahwa akurasi yang diperoleh dari algoritma naive bayes masih tergolong cukup rendah [11]. Menurut Nardilasari dkk, penelitian terkait algoritma naïve Bayes dengan membandingkannya dengan Algoritma Support Vector Machine (SVM), mengungkapkan bahwa Algoritma naïve Bayes mencatat akurasi sebesar 73,68%, sedangkan Algoritma SVM berhasil mencapai tingkat akurasi sebesar 98,61%[12].

Namun demikian, selain penelitian terkait dengan sentimen terkait calon presiden 2024, beberapa penelitian lain telah mengeksplorasi penggunaan SMOTE dalam konteks yang berbeda. Menurut Saifurrachman dkk, penelitian analisis sentimen pada aplikasi duolingo menyoroti perbandingan kinerja algoritma naive bayes dengan dan tanpa SMOTE. Hasilnya menunjukkan peningkatan yang signifikan dalam akurasi setelah penerapan SMOTE, mencapai kenaikan sebesar 14.81% [13]. Di sisi lain, penelitian oleh Andreyestha dan Azizah terfokus pada penggunaan SMOTE dalam mengoptimalkan data yang tidak seimbang pada kicauan Tokopedia. Integrasi SMOTE dengan naive bayes menghasilkan peningkatan akurasi yang mencolok sebesar 3.4%[14]. Penelitian selanjutnya, dimana Yerik Afrianto menggunakan metode naive bayes digabungkan dengan SMOTE dalam analisis sentimen terhadap konten robot restoran. Dengan SMOTE, algoritma menunjukkan akurasi 70,11%. Sebaliknya, tanpa SMOTE, algoritma menghasilkan akurasi 48,90% [15].

Dari penelitian sebelumnya, jelas terlihat bahwa penggabungan SMOTE dengan algoritma naive bayes dapat meningkatkan akurasi dan presisi dalam analisis sentimen. Penelitian ini bertujuan untuk mengisi gap dengan menerapkan gabungan naive bayes dan SMOTE dalam konteks analisis sentimen calon presiden 2024 di media sosial X, suatu pendekatan yang belum pernah dilakukan sebelumnya. Dengan menggabungkan naive bayes dan SMOTE, diharapkan penelitian ini dapat meningkatkan akurasi analisis sentimen dibandingkan dengan penelitian-penelitian sebelumnya yang dilakukan oleh Fais (2022) dan Nardilasari (2023) hanya menggunakan naive bayes tanpa SMOTE [11][12]. Oleh karena itu, tujuan penelitian ini adalah menggunakan metode naive bayes dalam klasifikasi data, dengan penerapan SMOTE untuk menjaga keseimbangan data. Serta menarik kesimpulan tentang pandangan masyarakat Indonesia terhadap calon presiden 2024. Selain itu, penelitian ini menggunakan dataset dari 12 Desember 2023 hingga 31 Maret 2024.

2. METODOLOGI PENELITIAN

Penelitian ini melibatkan langkah-langkah metodologis seperti pengumpulan data, preprocessing, pelabelan data, pembobotan menggunakan TF-IDF, penerapan naive bayes dengan dan tanpa SMOTE, dan evaluasi menggunakan confusion matrix. Adapun Detail penelitian tersebut tergambar pada Gambar 1.



Gambar 1. Diagram urutan langkah-langkah penelitian

Gambar 1 menunjukkan rangkaian langkah penelitian ini. Mulai dari pengumpulan data komentar atau tweet dari media sosial X melalui tweet-harvest, terkait pandangan publik tentang ketiga pasangan calon presiden dan wakil presiden 2024 dari debat pertama hingga setelah pengumuman pemilu. Langkah berikutnya termasuk preprocessing data, pelabelan manual untuk sentimen, pembobotan kata dengan TF-IDF, analisis menggunakan algoritma naive bayes (dengan atau tanpa SMOTE). dan Evaluasi dilakukan dengan confusion matrix.

2.1 Pengumpulan Dataset

Data penelitian diperoleh oleh peneliti dari platform media sosial X. Peneliti memanfaatkan alat pengumpulan tweet-harvest untuk mengatur data melalui platform Google Colab. Proses pengambilan data dilakukan secara

Submitted: **14/05/2024**; Accepted: **25/07/2024**; Published: **26/07/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 3, Juli 2024, Page 1313-1322
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i3.7708

real-time melalui API media sosial X, setelah itu data disimpan dalam format basis data untuk dilakukan analisis lebih lanjut [16].

2.2 Preprocessing Dataset

Preprocessing merupakan langkah penting dalam menyiapkan analisis sentimen untuk meningkatkan kinerja proses klasifikasi [17]. Dalam penelitian ini, penting untuk memilih metode preprocessing yang tepat. Sejumlah Langkah preprocessing diperlukan untuk mempersiapkan data, seperti menghapus entri ganda, mengganti kata-kata tertentu seperti mention, URL, hashtag, dan simbol, mengubah format huruf, membagi menjadi kata-kata terpisah, menyaring kata penghubung, melakukan pemangkasan kata, dan pada akhirnya, menyaring kata berdasarkan panjangnya. Langkah-langkah ini bertujuan untuk memastikan bahwa dataset telah dibersihkan dan siap untuk digunakan dalam analisis sentimen selanjutnya.

2.3 Memilih Dataset

Secara faktual, data yang diperoleh dari pengumpulan tweet tetap melibatkan berbagai informasi yang tidak mencerminkan dengan jelas pandangan baik yang negatif maupun yang positif. Oleh karena itu, langkah penyaringan atau seleksi data diperlukan agar sesuai dengan keperluan penelitian ini. Untuk memastikan keakuratan refleksi opini atau sentimen yang relevan terhadap fokus penelitian, tujuan utamanya adalah memvalidasi data yang dianalisis.

2.4 Label Dataset

Langkah selanjutnya yang harus diambil adalah melakukan pelabelan. Ini penting karena penelitian ini akan menggunakan algoritma pembelajaran yang diawasi, sehingga dataset harus disiapkan dengan label yang sesuai. Proses pelabelan ini menjadi dasar penting untuk memberikan informasi kepada algoritma dalam mengenali pola dan karakteristik yang berkaitan dengan tujuan penelitian [18]. Sangat penting untuk melakukan pelabelan data dengan akurat dalam rangka membentuk model pembelajaran mesin yang efektif. Hal ini berdampak signifikan terhadap kemampuan model dalam memberikan prediksi berkualitas tinggi atau mengklasifikasikan data yang belum pernah diproses sebelumnya [19]. Dalam penelitian ini, data dilabeli secara manual oleh peneliti, terbagi menjadi kelompok emosi positif dan negatif, memastikan label sentimen yang akurat untuk kumpulan data berkualitas tinggi.

2.5 Pembobotan

Dalam penelitian ini, sebuah metode dalam pemrosesan bahasa alami, yaitu Frequency-Inverse Document Frequency (TF-IDF), dipergunakan dengan tujuan mengevaluasi tingkat signifikansi suatu kata dalam sebuah dokumen atau koleksi dokumen. Dalam konsep ini, perhatian diberikan pada dua faktor utama, yaitu Term Frequency (TF) dan Inverse Document Frequency (IDF). Jumlah term yang muncul meningkat seiring dengan pembobotan yang diberikan, sementara inversi frekuensi dokumen adalah metode untuk mengevaluasi signifikansi kata dalam dokumen tertentu. TF-IDF bekerja dengan mengidentifikasi keberadaan kata-kata tertentu dalam dokumen dan mempertimbangkan seberapa signifikan kata-kata tersebut dalam konteks keseluruhan dokumen yang ada [20].

2.6 SMOTE

Ketidakseimbangan kelas dalam klasifikasi teks menjadi fokus penting dalam penelitian ini. Ketika jumlah data kelas target tidak seimbang maka akan terjadi masalah ketidakseimbangan kelas [21]. Selama proses prediksi, ketidakseimbangan data antar kategori akan mengurangi tingkat akurasi dan perolehan kembali kategori minoritas. Oleh karena itu, Model tersebut memiliki kecenderungan untuk lebih akurat dalam memprediksi kelas mayoritas namun kurang efisien dalam memprediksi kelas minoritas [22]. Untuk mengatasi tantangan tersebut, penelitian akan memanfaatkan pendekatan oversampling, yakni dengan menciptakan kelas tambahan dari kelompok etnis yang minoritas. Oversampling dipilih sebagai metode karena tidak mengurangi jumlah data dalam dataset dan terbukti memberikan hasil yang lebih optimal dalam mengatasi ketidakseimbangan kelas pada dataset yang memiliki jumlah sampel kelas minoritas yang terbatas [23]. Dalam menangani ketidakseimbangan kelas, pendekatan menggunakan metode Synthetic Minority Oversampling Technique (SMOTE) menjadi relevan. Dalam konteks ini, strategi tersebut menghasilkan sampel baru dari kelas yang minoritas dengan cara menciptakan instansi tambahan melalui pembentukan variasi yang luas dari instansi yang berdekatan [24]. Berikut adalah rumus dari SMOTE :

$$X_{syn} = Xi + [X_{knn} - X_i] \times \delta \quad (1)$$

2.7 Algoritma Naive Bayes

Sebuah algoritma yang termasuk dalam kategori metode supervised learning di bidang machine learning. Secara esensial, Algoritma ini memerlukan sampel data yang sudah diberi label untuk keperluan pelatihan [25]. Dalam implementasinya, naïve bayes menerapkan suatu metode di mana kata-kata yang ada dalam dokumen dianggap bermanfaat secara bersama-sama tanpa memperhatikan susunan atau konteks dari kata-kata tersebut. Selain itu,

Submitted: **14/05/2024**; Accepted: **25/07/2024**; Published: **26/07/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 3, Juli 2024, Page 1313-1322
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i3.7708

teknik ini juga mengambil pertimbangan atas frekuensi kemunculan suatu kata dalam dokumen [7]. Hal ini tercermin dalam rumus umum Teorema Bayes yang menjadi dasar naïve bayes, seperti yang ditunjukkan pada rumus berikut :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2)$$

2.8 Evaluasi

Untuk menilai kesuksesan model yang telah dibuat, digunakan confusion matrix sebagai evaluasi sesuai dengan pada Tabel 1 [26]. Pengukuran kinerja model dalam penelitian ini menggunakan metrik accuracy, recall, precision, dan F1-Score.

Tabel 1. Confusion Matrix

Predicted Values	Actual Values	
	Negative	**Positive**
Pred.Negative	TN	FN
Pred. Positive	FP	TP

Dapat dilihat berdasarkan Tabel 1, True Negative (TN) dan True Positive (TP) mengindikasikan keberhasilan classifier dalam klasifikasi, sedangkan False Negative (FN) dan False Positive (FP) menunjukkan kesalahan klasifikasi yang dilakukan oleh classifier. TP dan TN mewakili hasil klasifikasi yang benar, sedangkan FP adalah nilai yang diprediksi sebagai positif namun sebenarnya negatif, dan FN adalah nilai yang terprediksi sebagai negatif namun sebenarnya positif [26]. Hasil dari perhitungan confusion matrix ini memberikan nilai akurasi, presisi (precision), recall, dan F1-Score. Berikut formula metrik evaluasi :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

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

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

$$F1-score = \frac{2\times Precision \times Recall}{Precision + Recall} \quad (6)$$

Metrik accuracy memberikan Gambaran umum tentang efektivitas model dengan mengukur seberapa akurat model dapat memprediksi kelas tertentu. Precision adalah proses penghitungan jumlah prediksi positif yang benar dibagi dengan jumlah total prediksi positif yang dihasilkan, yang memberikan informasi tentang kemampuan model dalam mengidentifikasi kelas positif [26]. Meskipun recall mengukur sejauh mana model dapat mengenali semua contoh positif yang sebenarnya, itu memberikan wawasan tentang keefektifan model dalam menemukan semua contoh positif yang sesungguhnya dari kelas positif. F1-Score, sebuah metrik evaluasi menggabungkan precision dan recall dalam perhitungannya memastikan bahwa kedua nilai tersebut seimbang dalam mengevaluasi kinerja model. Dengan demikian, F1-Score memberikan Gambaran yang lebih lengkap tentang kualitas model dalam melakukan klasifikasi [27].

2.9 Visualisasi WordCloud

Metode WordCloud, yang sering digunakan dalam text mining, membantu visualisasi data teks dengan menarik. Populer karena kemudahannya, teknik ini menggali informasi tentang kata-kata yang paling sering muncul dalam dokumen dan membuat representasi visual dari frekuensinya [28].

3. HASIL DAN PEMBAHASAN

Dalam pengumpulan dataset untuk penelitian ini, peneliti menerapkan metode tweet-harvest dengan kata kunci yang spesifik, seperti nama dari ketiga pasangan contohnya “Anies” dan “Imin”. Pengumpulan dimulai dari debat pertama calon presiden pada 12 Desember 2023, hingga pada 31 Maret 2024.

3.1 Pengumpulan Dataset

Semua data yang diperoleh berasal dari media sosial X. Setiap kali dilakukan pengumpulan data, peneliti berhasil memperoleh 100 data, sehingga total 600 data diperoleh untuk setiap pasangan atau 3 dataset. Untuk memvisualisasikan dataset yang diambil, peneliti menyajikannya dalam Tabel 2.

Tabel 2. Data Awal

No	Tweet
1	@nadyairawan__ @Rzlzl4 @memelord_666666 Ahok jadi ayam sayur pas debat sama anies

Submitted: **14/05/2024**; Accepted: **25/07/2024**; Published: **26/07/2024**

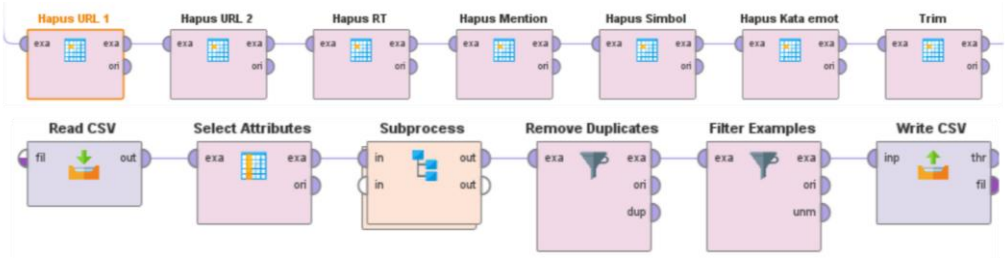
JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 3, Juli 2024, Page 1313-1322
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i3.7708

No	Tweet
2	@leebookmark Itu twtnya bahas debat capres nyindir anies gasih wkwkwkwkw
3	@ainunnajib Seingat saya, jokowi waktu ikut debat pilpres dulu ga kaya anies gini kelakuannya. waktu soal HGB lahan jg tdk frontal merujuk nama ataupun langsung menunjuk lawan debatnya.
…	…
600	@idextratime Plis pak anis dan cak imin udah pakk udah, prepare aja buat 2029

Berdasarkan Tabel 2, Dapat diamati bahwa variasi simbol, angka, dan beragam penulisan kalimat dengan perbedaan huruf besar dan kecil terlihat dalam bentuk awal data yang diambil dari media sosial X.

3.2 Preprocessing Dataset

Langkah berikutnya, dapat dilihat pada Gambar 4 adalah melakukan Preprocessing menggunakan perangkat lunak Rapidminer.



Gambar 2. Proses Preprocessing Dataset

Berdasarkan Gambar 4, langkah pertama dalam preprocessing adalah menggunakan operator “Read CSV” untuk membaca file CSV yang telah disiapkan. Selanjutnya, hubungkan dengan operator “Select Attributes”. Operator ini hanya akan memilih atribut teks. Kemudian, masuk ke dalam operator “Subprocess” yang berisi langkah-langkah untuk menghapus URL, RT, mention, hashtag, simbol, dan emoticon. Langkah ini dilakukan untuk membersihkan teks dari gangguan seperti entitas mention, tautan, hashtag, simbol dan emoticon yang tidak relevan [29]. Proses trim menghilangkan spasi tambahan di awal dan akhir teks dengan tujuan menjaga teks tetap bersih dan rapi. Langkah berikutnya adalah menghapus duplikasi tweet dengan menggunakan operator "Remove Duplicate". Setelah itu, ambil tweet yang sudah bersih dan tidak hilang dengan menggunakan operator "Filter Example". Terakhir, gunakan operator "Write CSV" untuk menyimpan hasilnya kembali dalam bentuk CSV.

3.3 Memilih Dataset

Dalam seleksi tweet untuk penelitian, peneliti memfilter hanya tweet dengan opini berisi sentimen negatif dan positif mengenai Calon presiden dan wakil presiden. Setelah tahap pemilihan dataset, ditemukan total Anies-Imin 559, Prabowo-Gibran 560, Ganjar-Mahfud 470 data dari beragam jenis tweet, termasuk tweet asli, retweet, kutipan, dan balasan.

3.4 Label Dataset

Peneliti memberikan label sentimen secara manual pada data yang telah dipilih. Pengelompokan dilakukan dalam dua klasifikasi, yaitu positif dan negatif. Hasil klasifikasi dapat ditunjukan pada tabel 3.

Tabel 3. Klasifikasi Sentimen

Pasangan Calon	Kategori		
	Negatif	Positif	Total
Anies-Imin	357	202	559
Prabowo-Gibran	351	209	560
Ganjar-Mahfud	153	317	470

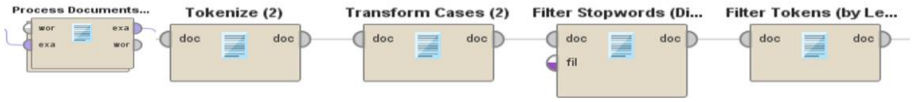
Berdasarkan Tabel 3 dapat dilihat klasifikasi kategori sentimen positif serta negatif. Dalam penilaian sentimen, dataset Anies-Imin, sebagai pasangan calon pertama, menarik perhatian dari 202 pengguna yang dengan tegas menyatakan dukungan positif terhadap visi dan rencananya. Namun, sorotan positif itu diimbangi oleh 357 komentar yang kritis dan menunjukkan ketidaksetujuan terhadap pasangan tersebut. Di sisi lain, dataset Prabowo-Gibran juga mendapat perhatian yang signifikan. Dukungan positif dari 209 komentar menunjukkan keyakinan pada platform politik mereka, meskipun 351 komentar lainnya mencerminkan ketidakpuasan dan ketidakpercayaan terhadap pasangan tersebut. Namun, dalam situasi polarisasi ini, satu pasangan calon, dataset Ganjar-Mahfud, menonjol dengan perbedaan yang mencolok. Meskipun hanya 153 komentar yang menunjukkan ketidaksetujuan terhadap mereka, 317 komentar lainnya dengan penuh antusias menyambut visi dan integritas pasangan ini.

Submitted: **14/05/2024**; Accepted: **25/07/2024**; Published: **26/07/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 3, Juli 2024, Page 1313-1322
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i3.7708

3.5 Pembobotan TF-IDF

Selanjutnya, di dalam operator "Process Document" terdapat beberapa operator yang digunakan, seperti yang disajikan pada Gambar 3, seperti Tokenize, Transform Cases, Filter Stopwords, dan Filter Tokens.



Gambar 3. Operator Yang Digunakan Dalam Pemrosesan Kata

Berdasarkan Gambar 3, Menurut penelitian sebelumnya, Pertama-tama semua huruf harus diubah menjadi huruf kecil sebelum pembobotan untuk memastikan penggunaan huruf yang konsisten dan mencegah kesalahan dalam tokenisasi [30]. Tokenisasi membagi sebuah kalimat menjadi kata-kata yang memiliki arti masing-masing untuk memperjelas struktur dan komposisi kalimat. Selanjutnya Transform cases ini bertujuan untuk menyamakan kapitalisasi setiap kata dalam data. Pilihan "lower case" diterapkan agar semua data diubah menjadi huruf kecil, sehingga konsisten. Kemudian untuk fokus pada kata kunci yang lebih relevan untuk analisis, filter stopword digunakan untuk menghilangkan kata sambung umum [26]. Terakhir, filter token (berdasarkan panjang) digunakan untuk menghilangkan kata-kata dengan jumlah huruf tertentu, seperti kata-kata dengan panjang kurang dari dua karakter atau lebih dari dua puluh lima karakter. Ini digunakan untuk memastikan bahwa kontribusi yang dihapus tidak terlalu besar atau terlalu panjang untuk dievaluasi [29].

Langkah selanjutnya, metode Term Frequency-Inverse Document Frequency (TF-IDF) digunakan untuk memberikan bobot atau nilai tertentu dalam penelitian ini. Teknik ini bekerja dengan menemukan frekuensi kata tertentu sesuai dengan frekuensi kata tersebut di seluruh dokumen dan kemudian membandingkannya. Dalam metode TF-IDF, skor akhir diperoleh dengan mengalikan skor TF dan IDF dari suatu kata, sementara nilai TF (Term Frequency) menunjukkan seberapa sering kata muncul dalam sebuah dokumen, dan IDF (Inverse Document Frequency) menunjukkan bahwa kata yang memiliki skor TF-IDF yang lebih tinggi dalam dokumen semakin terkait dengan dokumen tersebut [26]. Gambar 4 menunjukkan contoh TF-IDF ini.

text	abis	aaamiin	abah
bener abis debat anies cecer buzzer	0.477	0	0

Gambar 4. Pembobotan TF-IDF

Berdasarkan Gambar 4, kata "abis" muncul dengan term frequency (TF) sebesar 0,477 pada dokumen tersebut, contohnya pada kalimat “bener abis debat anies cecer buzzer”. Sebaliknya, IDF menyoroti kata yang jarang muncul dalam seluruh dataset. Kata "aaamin" dan "abah" tidak muncul dalam dokumen tersebut, yang menunjukkan bahwa frekuensi mereka nol,

3.6 SMOTE

Dataset yang digunakan tidak seimbang antara kategori positif dan negatif. Perbedaan yang terlihat setelah penggunaan SMOTE dapat dilihat pada Tabel 4.

Tabel 4. Hasil Penambahan Data Minoritas

Pasangan Calon	SMOTE			
	Sebelum		Sesudah	
	Negatif	Positif	Negatif	Positif
Anies-Imin	357	202	357	357
Prabowo-Gibran	351	209	351	351
Ganjar-Mahfud	153	317	317	317

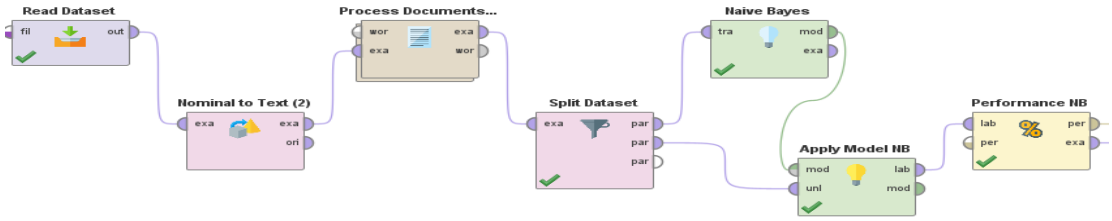
Dapat dilihat dari Tabel 4, sebelum menggunakan SMOTE, jumlah sampel positif untuk Anies-Imin hanya 202. Namun, setelah menerapkan SMOTE, jumlahnya meningkat menjadi 357, sehingga seimbang dengan jumlah data negatif. Peningkatan jumlah data minoritas juga terjadi pada dataset Prabwo-Gibran data minoritas 209 menjadi 351 dan Ganjar-Mahfud 153 menjadi 317.

3.7 Implementasi Algoritma dan SMOTE

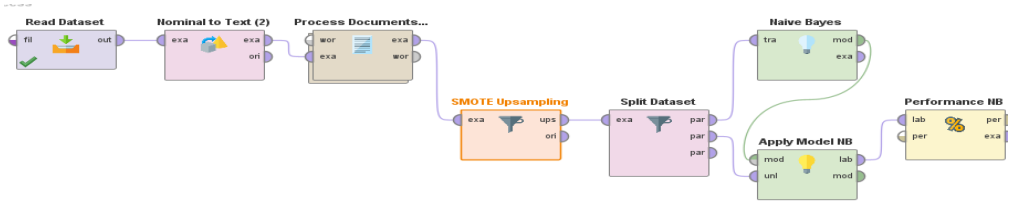
Setelah preprocessing dan pemberian bobot menggunakan metode TF-IDF, peneliti menggunakan metode SMOTE untuk meningkatkan kinerja metode naïve bayes. Gambar 5 mengilustrasikan tahap pertama tanpa menggunakan SMOTE dan langsung ke tahap implementasi metode naïve bayes, sedangkan tahap kedua diolah menggunakan SMOTE, yang diGambarkan pada Gambar 6. Tujuan dari pendekatan ini adalah untuk mendapatkan hasil yang optimal [14].

Submitted: **14/05/2024**; Accepted: **25/07/2024**; Published: **26/07/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 3, Juli 2024, Page 1313-1322
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i3.7708



Gambar 5. Pemodelan Algoritma Naive bayes tidak menggunakan SMOTE



Gambar 6. Pemodelan Algoritma Naive bayes dengan SMOTE

Berdasarkan Gambar 5 dan 6 dapat dilihat beberapa tahapan yang dilakukan. Langkah awal menggunakan operator “Read Dataset” untuk membaca file CSV. Data kemudian diproses melalui operator “Nominal to Text” sebelum dihubungkan dengan "Process Document from Data" agar data nominal dapat diubah menjadi teks. Perbedaan antara Gambar 5 dan 6 terletak pada penambahan operator “SMOTE Upsampling” untuk menyeimbangkan data minoritas. Selanjutnya, dataset dibagi menjadi data latih (80%) dan data uji (20%) menggunakan operator “Split Dataset”. Algoritma dipelajari menggunakan data latih untuk mencari model yang cocok dengan dataset, sedangkan data uji digunakan untuk mengevaluasi model yang telah dilatih [30], Setelah dataset dibagi, operator “Naïve Bayes” memproses data latih, sementara data uji dialirkan ke operator “Apply Model” untuk mengevaluasi performa model yang telah dilatih sebelumnya. Untuk menilai akurasi prediksi terhadap data uji, digunakan operator “Performance” [24].

3.8 Evaluasi

Berikut dapat dilihat pada Gambar 7 hasil pengujian sentimen dataset pertama menggunakan model naïve bayes tanpa menggunakan teknik SMOTE didapatkan dari confusion matrix.

accuracy: 69.37%

	true negatif	true positif
pred. negatif	51	14
pred. positif	20	26

Gambar 7. Confusion matrix Algoritma Naive bayes tanpa SMOTE Anies-Imin

Dari hasil confusion matrix pada Gambar 7. Dapat dilihat total data, terdapat 26 True Positive (TP), 51 True Negative (TN), 14 False Positive (FP), dan 29 False Negative (FN). Model ini mencapai tingkat akurasi sebesar 69,37%, Recall sebesar 56%, Precision sebesar 65%, dan F1-Score sebesar 60%.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{(26+51)}{(26+51+14+20)} = \frac{77}{111} = 0.69$$

$$Precision = \frac{TP}{TP+FP} = \frac{26}{(26+14)} = \frac{26}{40} = 0.65$$

$$Recall = \frac{TP}{TP+FN} = \frac{26}{(26+20)} = \frac{26}{46} = 0.56$$

$$F1-score = \frac{2\times Precision \times Recall}{Precision + Recall} = \frac{2\times(0{,}65 * 0.56)}{(0{,}65 + 0.56)} = \frac{2\times 0.364}{1{,}21} = 0.6$$

Berikut dapat dilihat pada Gambar 8 hasil pengujian sentimen dataset kedua menggunakan model Naïve Bayes dengan teknik SMOTE didapatkan dari confusion matrix.

accuracy: 83.57%

	true negatif	true positif
pred. negatif	53	6
pred. positif	17	64

Gambar 8. Confusion matrix Algoritma Naive bayes dengan SMOTE Prabowo-Gibran

Submitted: **14/05/2024**; Accepted: **25/07/2024**; Published: **26/07/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 3, Juli 2024, Page 1313-1322
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i3.7708

Dari hasil confusion matrix pada Gambar 8. Dari total data, terdapat 64 True Positive (TP), 53 True Negative (TN), 6 False Positive (FP), dan 17 False Negative (FN). Model ini mencapai tingkat akurasi sebesar 83%, Recall sebesar 78%, Precision sebesar 95%, dan F1-Score sebesar 84%.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{(64+53)}{(64+53+6+17)} = \frac{117}{140} = 0.83$$

$$Precision = \frac{TP}{TP+FP} = \frac{64}{(64+6)} = \frac{67}{70} = 0.95$$

$$Recall = \frac{TP}{TP+FN} = \frac{64}{(64+17)} = \frac{64}{82} = 0.78$$

$$F1-score = \frac{2\times Precision \times Recall}{Precision + Recall} = \frac{2\times(0,95 * 0.78)}{(0,95 + 0.78)} = \frac{2\times 0.74}{1,73} = 0.84$$

Berikut dapat dilihat pada Gambar 9 hasil pengujian sentimen dataset ketiga menggunakan model Naïve Bayes dengan teknik SMOTE didapatkan dari confusion matrix.

accuracy: 88.89%

	true negatif	true positif
pred. negatif	58	9
pred. positif	5	54

Gambar 9. Confusion matrix Algoritma Naive bayes dengan SMOTE Ganjar-Mahfud

Dari hasil confusion matrix pada Gambar 9. Dari total data, terdapat 54 True Positive (TP), 58 True Negative (TN), 9 False Positive (FP), dan 5 False Negative (FN). Model ini mencapai tingkat akurasi tertinggi sebesar 88%, Recall sebesar 91%, Precision sebesar 91%, dan F1-Score sebesar 86%.

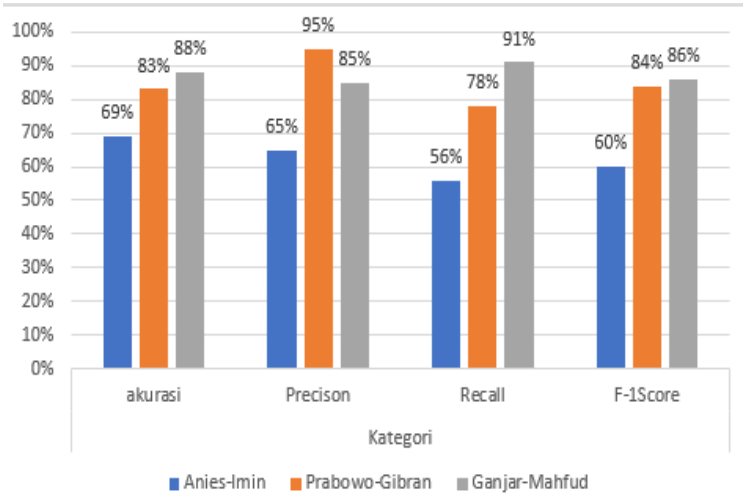
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{(54+58)}{(54+58+9+5)} = \frac{112}{126} = 0.88$$

$$Precision = \frac{TP}{TP+FP} = \frac{54}{(54+9)} = \frac{54}{59} = 0.85$$

$$Recall = \frac{TP}{TP+FN} = \frac{54}{(54+5)} = \frac{54}{59} = 0.91$$

$$F1-score = \frac{2\times Precision \times Recall}{Precision + Recall} = \frac{2\times (0,85 * 0.91)}{(0,85 + 0.91)} = \frac{2\times 0.77}{1,76} = 0.86$$

Berikut dapat dilihat pada Gambar 10 perbandingan hasil evaluasi kinerja antara Naïve Bayes dengan atau tanpa SMOTE. Evaluasi ini mencakup nilai Accuracy, Recall, Precision, dan F1-Score.



Gambar 10. Perbandingan Nilai Evaluasi

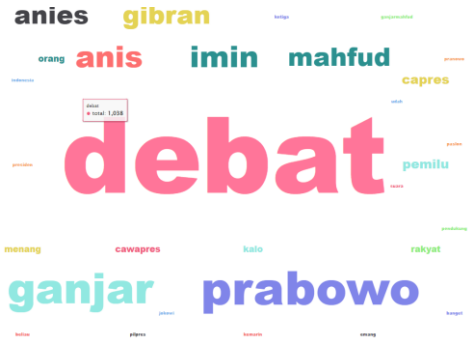
Dari hasil perbandingan pada Gambar diatas, perbedaan dalam penggunaan SMOTE dapat dilihat dengan jelas dalam hasil evaluasi kinerja model untuk setiap pasangan calon. Anies-Imin, yang tidak menggunakan SMOTE, memiliki tingkat akurasi yang rendah sekitar 69%, dengan presisi hanya 65%, recall 56%, dan F-1 Score 60%. Sementara itu, Prabowo-Gibran dan Ganjar-Mahfud, yang memanfaatkan SMOTE, menunjukkan peningkatan yang signifikan dalam kinerja model mereka. Prabowo-Gibran mencapai tingkat akurasi sekitar 83%, dengan presisi mencapai 95%, recall 78%, dan F-1 Score mencapai 84%. Di sisi lain, Ganjar-Mahfud memiliki tingkat akurasi mencapai 88%, presisi 85%, recall 91%, dan F-1 Score 86%.

3.1 Visualisasi WordCloud

Jumlah kemunculan suatu kata dalam dataset menentukan ukuran kata dalam WordCloud, di mana semakin sering suatu kata muncul, ukurannya akan semakin besar [20]. Hasil kemunculan kata dapat dilihat pada Gambar 11.

Submitted: **14/05/2024**; Accepted: **25/07/2024**; Published: **26/07/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 3, Juli 2024, Page 1313-1322
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i3.7708



Gambar 11. Visualisasi WordCloud Semua Pasangan calon presiden

Berdasarkan visualisasi pada Gambar 11 dapat dilihat bahwa kata “Debat” mendominasi. untuk ketiga dataset, kata "debat" muncul sebanyak 1038 kali. Hal ini menunjukkan bahwa masyarakat secara aktif mengikuti debat dan banyak memberikan komentar terkait debat yang diadakan oleh Komisi Pemilihan Umum (KPU).

4. KESIMPULAN

Berdasarkan analisis penelitian terhadap pengolahan dataset yang dikumpulkan antara 12 Desember 2023 hingga 31 Maret 2024, dari total 1589 tweet yang menunjukkan sentimen positif, sekitar 27,7% berkaitan dengan Anies-Imin, sekitar 28,7% berkaitan dengan Prabowo-Gibran, dan sekitar 43,5% berkaitan dengan Ganjar-Mahfud. Hal ini menunjukkan bahwa Ganjar-Mahfud mendapat dukungan positif yang lebih besar dibandingkan dengan dua pasangan calon lainnya. Sedangkan untuk sentimen negatif, mayoritas sentimen negatif tersebut tertuju kepada pasangan calon Anies-Imin dan Prabowo-Gibran. Sekitar 41.5% sentimen negatif mengarah ke Anies-Imin, sementara sekitar 40.8% mengarah ke Prabowo-Gibran. Ini menunjukkan bahwa kedua pasangan calon ini menerima kritik dan ketidaksetujuan yang signifikan dari masyarakat dalam data yang diamati. Sentimen negatif ini disebabkan oleh kurangnya keterampilan dalam berdebat yang terlihat pada saat debat, terutama dari pasangan Anies-Imin dan Prabowo-Gibran. Kemudian hasil penelitian menggunakan metode Naïve Bayes menunjukkan akurasi yang rendah, sekitar 69%, tanpa menggunakan SMOTE pada dataset Anies-Imin. Adapun, akurasi yang tinggi, mencapai 88%, diperoleh dari dataset Ganjar-Mahfud yang menggunakan SMOTE. Selain itu, penggunaan metode SMOTE dapat membantu menyeimbangkan data minoritas dengan menambahkan sampel kelas minoritas sehingga jumlahnya setara dengan kelas mayoritas, yang pada akhirnya akan meningkatkan akurasi klasifikasi sentimen.

REFERENCES

[1] T. Tranggono et al., “Peran Media Sosial Sebagai Wadah Aspirasi Masyarakat,” Bur. J. Indones. J. Law Soc. Gov., vol. 3, no. 2, pp. 2155–2164, 2023, doi: https://doi.org/10.53363/bureau.v3i2.314.
[2] Y. W. S. P. et Al., Pengantar Aplikasi Mobile. CV.Haura Utama, 2023.
[3] K. Arifin and S. I. Al-Idrus, “Klasifikasi Emosi Pengguna Twitter Terhadap Bakal Calon Presiden Pada Pemilu 2024 Menggunakan Algoritma Naïve Bayes,” J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer), vol. 23, no. 1, p. 37, 2024, doi: 10.53513/jis.v23i1.9558.
[4] Alfandi Safira and F. N. Hasan, “Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier,” Zo. J. Sist. Inf., vol. 5, no. 1, pp. 59–70, 2023, doi: 10.31849/zn.v5i1.12856.
[5] N. Fitriani and E. Fitrianti, “Tindak Tutur Asertif dalam Acara Debat Perdana,” vol. 3, no. 3, pp. 120–126, 2024, doi: https://doi.org/10.36057/jips.v7i3.644.
[6] M. Minardi, R. Lasepa, S. Riyadi, S. Ramadhan, and D. D. Saputra, “Sentiment Analysis Terhadap Perspektif Warganet Atas Tragedi Kanjuruhan Malang di Twitter Menggunakan Naïve Bayes Classifier,” J. Inform., vol. 10, no. 1, pp. 45–53, 2023, doi: 10.31294/inf.v10i1.14546.
[7] A. R. Abdillah and F. N. Hasan, “Analisis Sentimen Terhadap Kandidat Calon Presiden Berdasarkan Tweets Di Sosial Media Menggunakan Naive Bayes Classifier,” Smatika J., vol. 13, no. 01, pp. 117–130, 2023, doi: 10.32664/smatika.v13i01.750.
[8] C. Nas, “Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5,” J. Manaj. Inform., vol. 11, no. 2, pp. 131–145, 2021, doi: 10.34010/jamika.v11i2.5506.
[9] K. F. Irnanda, A. P. Windarto, and I. S. Damanik, “Optimasi Particle Swarm Optimization Pada Peningkatan Prediksi dengan Metode Backpropagation Menggunakan Software RapidMiner,” JURIKOM (Jurnal Ris. Komputer), vol. 9, no. 1, p. 122, 2022, doi: 10.30865/jurikom.v9i1.3836.
[10] D. Ardiansyah, “Algoritma C4.5 Untuk Klasifikasi Calon Peserta Lomba Cerdas Cermat Siswa Smp Dengan Menggunakan Aplikasi Rapid Miner,” J. Inkofar, vol. 1, no. 2, pp. 5–12, 2019, doi: 10.46846/jurnalinkofar.v1i2.29.
[11] M. R. Fais Sya’ bani, U. Enri, and T. N. Padilah, “Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritme Naïve Bayes,” JURIKOM (Jurnal Ris. Komputer), vol. 9, no. 2, p. 265, 2022, doi: 10.30865/jurikom.v9i2.3989.
[12] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, T. Tukino, and B. Priyatna, “Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter,” JOINTECS (Journal Inf. Technol. Comput. Sci., vol. 8, no.

Submitted: **14/05/2024**; Accepted: **25/07/2024**; Published: **26/07/2024**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 8, Nomor 3, Juli 2024, Page 1313-1322
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v8i3.7708

1, p. 11, 2023, doi: 10.31328/jointecs.v8i1.4265.
[13] S. Chohan, A. Nugroho, A. M. B. Aji, and W. Gata, "Analisis Sentimen Pengguna Aplikasi Duolingo Menggunakan Metode Naïve Bayes dan Synthetic Minority Over Sampling Technique," Paradig. - J. Komput. dan Inform., vol. 22, no. 2, pp. 139–144, 2020, doi: 10.31294/p.v22i2.8251.
[14] A. Andreyestha and Q. N. Azizah, "Analisa Sentimen Kicauan Twitter Tokopedia Dengan Optimalisasi Data Tidak Seimbang Menggunakan Algoritma SMOTE," Infotek J. Inform. dan Teknol., vol. 5, no. 1, pp. 108–116, 2022, doi: 10.29408/jit.v5i1.4581.
[15] Y. A. Singgalen, "Social Network Analysis and Sentiment Classification of Extended Reality Product Content," KLIK Kaji. Ilm. Inform. dan Komput., vol. 4, no. 4, pp. 2197–2208, 2024, doi: 10.30865/klik.v4i4.1710.
[16] H. Setiawan and I. Zufria, "Analisis Sentimen Pembatalan Indonesia Sebagai Tuan Rumah Piala Dunia FIFA U-20 Menggunakan Naïve Bayes," vol. 7, no. 3, pp. 1003–1012, 2023, doi: 10.30865/mib.v7i3.6144.
[17] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," J. Media Inform. Budidarma, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
[18] Z. Firmansyah and N. F. Puspitasari, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Berdasarkan Opini Pada Twitter MenggunakFirmansyah, Z., & Puspitasari, N. F. (2021). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Berdasarkan Opini Pada Twitter Menggunakan Algoritma Nai," J. Tek. Inform., vol. 14, no. 2, pp. 171–178, 2021, [Online]. Available: https://doi.org/10.15408/jti.v14i2.24024.
[19] O. Manullang, C. Prianto, and N. H. Harani, "Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based Dan Random Forest," J. Ilm. Inform., vol. 11, no. 02, pp. 159–169, 2023, doi: 10.33884/jif.v11i02.7987.
[20] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," Teknika, vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.
[21] J. W. Iskandar and Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 5, no. 6, pp. 1120–1126, 2021, doi: 10.29207/resti.v5i6.3588.
[22] N. Sulistiyowati and M. Jajuli, "Integrasi Naive Bayes Dengan Teknik Sampling Smote Untuk Menangani Data Tidak Seimbang," Nuansa Inform., vol. 14, no. 1, p. 34, 2020, doi: 10.25134/nuansa.v14i1.2411.
[23] L. Qadrini, H. Hikmah, and M. Megasari, "Oversampling, Undersampling, Smote SVM dan Random Forest pada Klasifikasi Penerima Bidikmisi Sejawa Timur Tahun 2017," J. Comput. Syst. Informatics, vol. 3, no. 4, pp. 386–391, 2022, doi: 10.47065/josyc.v3i4.2154.
[24] Hermanto, A. Y. Kuntoro, T. Asra, E. B. Pratama, L. Effendi, and R. Ocanitra, "Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm and Support Vector Machine Based Smote Technique," J. Phys. Conf. Ser., vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012102.
[25] R. Puspita and A. Widodo, "Analisis Sentimen terhadap Layanan Indihome di Twitter dengan Metode Machine Learning," J. Inform. Univ. Pamulang, vol. 6, no. 4, pp. 759–766, 2021, doi: 10.32493/informatika.v6i4.13247.
[26] R. Yunita and M. Kamayani, "Perbandingan Algoritma SVM Dan Naïve Bayes Pada Analisis Sentimen Penghapusan Kewajiban Skripsi," Indones. J. Comput. Sci., vol. 12, no. 5, pp. 2879–2890, 2023, doi: 10.33022/ijcs.v12i5.3415.
[27] M. A. Saddam, E. Kurniawan D, and I. Indra, "Analisis Sentimen Fenomena PHK Massal Menggunakan Naive Bayes dan Support Vector Machine," J. Inform. J. Pengemb. IT, vol. 8, no. 3, pp. 226–233, 2023, doi: 10.30591/jpit.v8i3.4884.
[28] M. G. Pradana, "Penggunaan Fitur Wordcloud dan Document Term Matrix dalam Text Mining," J. Ilm. Inform., vol. 8, no. 1, pp. 38–43, 2020, doi: 10.33884/jif.v8i01.1838.
[29] M. R. Amly, Y. Yusra, and M. Fikry, "Penerapan Metode Naïve Bayes Classifier Pada Klasifikasi Sentimen Terhadap Anies Baswedan Sebagai Bakal Calon Presiden 2024," J. Sist. Komput. dan Inform., vol. 4, no. 4, p. 621, 2023, doi: 10.30865/json.v4i4.6214.
[30] R. Wati, S. Ernawati, and H. Rachmi, "Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH," J. Manaj. Inform., vol. 13, no. 1, pp. 84–93, 2023, doi: 10.34010/jamika.v13i1.9424.

Submitted: **14/05/2024**; Accepted: **25/07/2024**; Published: **26/07/2024**