

Klasifikasi Sentimen Masyarakat terhadap Presiden Indonesia Menggunakan Metode *Naive Bayes*

Muhammad Affan Irtiza Pramono^{*}, Aditya Haryanto Ramdani, Ade Gagas Sudirjo, Eneng Dalpa Pahira, Rosyida Aurelia Fitri Riayah

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

affanir1@email.com, adityaharyanto3@gmail.com, gagasade123@gmail.com, enengdalpapahira@gmail.com, rosyida.aurelia@gmail.com

Abstract. Social media platform X has become an important platform for expressing public opinion, particularly in the political context, including the 2024 Presidential Election in Indonesia. This study analyzes public sentiment toward the elected presidential pair, Prabowo Subianto and Gibran Rakabuming Raka, using the Naive Bayes algorithm. Research data was obtained through a scraping process on social media X, covering 1,478 tweets relevant to the research topic. The analysis process began with text preprocessing, which included cleansing, stemming, tokenizing, and stopword removal, followed by sentiment labeling into three categories: positive, neutral, and negative. The Naive Bayes algorithm was chosen for its simplicity, efficiency in processing large datasets, and ability to provide reliable results despite the assumption of feature independence. Additionally, the SMOTE method was applied to address data imbalance. The research results show that positive sentiment dominates with a percentage of 43.30%, followed by neutral sentiment at 36.47%, and negative sentiment at 20.23%. The application of SMOTE improved the algorithm's accuracy from 55.41% to 75.52%. These findings confirm that the Naive Bayes algorithm, with the support of data balancing techniques, is capable of providing reliable results in large-scale sentiment analysis. This research makes an important contribution to understanding public opinion on social media and serves as a guide in designing evidence-based policies.

Keywords: *Public Sentiment, President, Indonesia.*

Abstrak. Media sosial X telah menjadi platform penting untuk mengekspresikan opini publik, khususnya dalam konteks politik, termasuk Pemilihan Presiden 2024 di Indonesia. Penelitian ini menganalisis sentimen masyarakat terhadap pasangan presiden terpilih, Prabowo Subianto dan Gibran Rakabuming Raka, dengan menggunakan algoritma Naive Bayes. Data penelitian diperoleh melalui proses scrapping pada media sosial X, mencakup 1.478 cuitan yang relevan dengan topik penelitian. Proses analisis dimulai dari text preprocessing yang meliputi cleansing, stemming, tokenizing, dan stopword removal, diikuti oleh pelabelan sentimen menjadi tiga kategori: positif, netral, dan negatif. Algoritma Naive Bayes dipilih karena keunggulannya dalam kesederhanaan, efisiensi pemrosesan data besar, dan kemampuannya memberikan hasil yang andal meskipun dengan asumsi independensi antarfitur. Selain itu, metode SMOTE diterapkan untuk mengatasi ketidakseimbangan data. Hasil penelitian menunjukkan bahwa sentimen positif mendominasi dengan persentase 43,30%, diikuti sentimen netral sebesar 36,47%, dan sentimen negatif sebesar 20,23%. Penerapan SMOTE meningkatkan akurasi algoritma dari 55,41% menjadi 75,52%. Temuan ini menegaskan bahwa algoritma Naive Bayes, dengan dukungan teknik penyeimbangan data, mampu memberikan hasil yang andal dalam analisis sentimen volume data besar. Penelitian ini memberikan kontribusi penting dalam memahami opini publik di media sosial serta menjadi panduan dalam merancang kebijakan berbasis data yang lebih responsif.

Kata Kunci: *Sentimen Publik, Presiden, Indonesia.*

A. Pendahuluan

Pada era digital saat ini, teknologi informasi dan komunikasi menjadi bagian yang tidak terpisahkan dari kehidupan masyarakat. Media mengambil peran besar dalam perubahan tahapan kehidupan yang berfungsi sebagai penyampaian pesan dan informasi yang berkembang dengan begitu dahsyat. Media sosial X adalah salah satu *platform microblogging* paling populer di dunia, termasuk di Indonesia. Berdasarkan laporan *We Are Social dan Meltwater*, terdapat 611,3 juta pengguna X di dunia pada April 2024, dan Indonesia menempati urutan keempat pengguna X terbanyak dengan 24,85 juta pengguna. Hal ini karena X memiliki fitur dan karakteristik yang unik serta sesuai dengan kebutuhan informasi dan interaksi modern. Dengan karakteristiknya yang memungkinkan berbagi informasi secara singkat dan cepat, X telah menjadi ruang diskusi publik yang dinamis, terutama untuk isu-isu sosial, ekonomi, dan politik (Jha & Verma, 2024).

Di Indonesia, X berada di peringkat lima besar aplikasi media sosial yang paling banyak diunduh, sehingga memperkuat statusnya sebagai sumber informasi utama bagi pengguna internet (Abidin & Herawati, 2024). Selama kampanye Pilpres 2024, penggunaan media sosial X mengarah pada peningkatan keterlibatan pemilih yang lebih besar, baik dalam bentuk diskusi maupun partisipasi dalam debat publik secara daring. Fenomena ini membuktikan bahwa X tidak hanya berfungsi sebagai saluran komunikasi satu arah, melainkan juga sebagai wadah interaktif yang memfasilitasi pembentukan opini politik secara kolektif (Sunata et al., 2024).

Pemilihan Presiden 2024 menjadi momen penting yang disambut dengan antusiasme serta respon dari berbagai kalangan. Tingginya perspektif masyarakat di *platform* X mencerminkan harapan serta kekhawatiran terhadap proses dan hasil pemilu pasangan presiden terpilih (Fauzan & Hikmah, 2022). Oleh karena itu, memahami sentimen publik terhadap pasangan presiden terpilih sangat penting karena dapat memberikan gambaran tentang apa yang diinginkan dan dikhawatirkan masyarakat.

Komunikasi yang efektif dan pengumpulan umpan balik dari masyarakat menjadi sangat penting bagi seorang pemimpin untuk dapat mengambil kebijakan yang adil dan tepat sasaran. Dengan semakin banyaknya masyarakat yang mengungkapkan pendapat mereka di platform seperti X, analisis sentimen dapat menjadi alat yang sangat berguna bagi pemimpin untuk memahami pandangan publik secara lebih mendalam. Analisis sentimen adalah pendekatan untuk mengeksplorasi dan mengidentifikasi nuansa emosional atau pandangan yang tersirat dalam sebuah teks. Proses ini biasanya diterapkan pada komentar, ulasan, atau percakapan di platform digital. Tujuan utamanya adalah untuk mengungkap apakah teks tersebut mencerminkan perasaan positif, negatif, atau netral terhadap suatu produk, isu, atau layanan (Pramastuti, 2016). Analisis ini memungkinkan pemahaman lebih dalam terhadap teks, dari tingkat kalimat hingga keseluruhan dokumen. Penerapan analisis sentimen sangat beragam, mulai dari dunia pemasaran untuk menggali persepsi konsumen, politik untuk mengukur opini masyarakat, hingga media sosial untuk mengetahui reaksi terhadap berbagai kejadian atau topik (Liu, 2023).

Labelling sentimen pada penelitian ini akan menggunakan mesin bernama *TextBlob*. Cuitan dikategorikan sesuai sentimen positif, netral, dan negatif (Macfud et al., 2023). *Labelling* bertujuan meningkatkan efisiensi dan kualitas dalam berbagai tugas. Dalam konteks analisis sentimen, *labelling* bertujuan untuk mengelompokkan teks ke dalam kategori yang sudah ditentukan. *Labelling* merupakan tahap yang penting karena secara langsung mempengaruhi kinerja model analisis. *Labelling* yang tidak tepat dapat menghasilkan model yang kurang akurat dalam melakukan klasifikasi.

Salah satu metode yang efektif untuk melakukan analisis sentimen adalah dengan menggunakan *Naive Bayes*, sebuah algoritma pembelajaran mesin yang sering digunakan untuk klasifikasi teks. *Naive Bayes* merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yaitu memprediksi peluang berdasarkan pengalaman di masa sebelumnya (Teorema Bayes) dengan ciri utamanya adalah asumsi yang sangat kuat (naif) akan ketergantungan dari masing-masing kondisi/kejadian (Rish, 2014). Metode klasifikasi ini cocok digunakan ketika jumlah masukan yang sangat besar. Klasifikasi ini lebih disukai karena kecepatan dan kesederhanaannya (Goel et al., 2016). Meskipun dalam kenyataannya fitur-fitur tersebut mungkin saling terkait, asumsi ini tidak mengurangi efektivitasnya dalam berbagai tugas klasifikasi, terutama dalam analisis teks seperti identifikasi sentimen, pemisahan pesan spam, dan pengelompokan konten.

Melalui penelitian ini, analisis sentimen di media sosial diharapkan dapat membantu pemimpin memahami persepsi masyarakat dan mendukung terciptanya kepemimpinan yang

bertanggung jawab serta berpihak kepada rakyat.

Tujuannya adalah mengungkap sentimen publik terhadap pasangan presiden terpilih periode 2024-2029, baik yang bersifat positif, negatif, maupun netral, melalui analisis data teks di media sosial X. Serta untuk mengetahui keakuratan algoritma Naive Bayes dalam menangani volume data besar dari media sosial X terkait opini publik mengenai pasangan presiden terpilih periode 2024-2029.

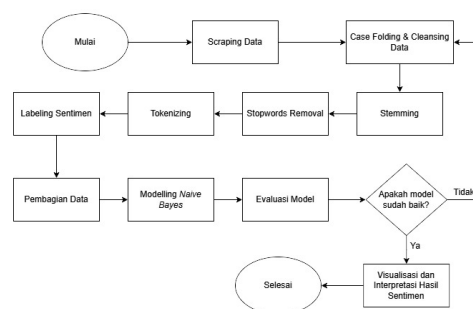
Adapun batasan penelitian yang dilakukan adalah data cuitan diambil dari media sosial X dengan menggunakan keyword "Presiden", "Prabowo", dan "Gibran". Data pengguna X yang digunakan adalah cuitan yang menggunakan bahasa Indonesia. Jumlah data yang diambil sebanyak 1478 cuitan. Klasifikasi sentimen dibagi menjadi tiga, yaitu sentimen positif, netral, dan negatif dengan menggunakan metode klasifikasi Naive Bayes.

B. Metode

Data pada penelitian ini menggunakan data cuitan (teks), labelling *TextBlob*, *oversampling* data dengan *SMOTE*, dan klasifikasi sentimen menggunakan Naive Bayes. Analisis ini dilakukan menggunakan bahasa pemrograman Python dan dijalankan pada platform Google Colab untuk proses yang efisien. Tahapan analisis dilakukan sebagai berikut

1. Pengumpulan data dengan scrapping melalui platform media sosial X yang berupa komentar yang relevan dengan pasangan presiden terpilih tahun 2024 dikumpulkan.
2. Preprocessing data atau text preprocessing : Melakukan pembersihan teks, yang mencakup penghapusan karakter khusus, tautan, tanda baca, dan simbol. Melakukan tokenisasi, yaitu memecah teks menjadi kata-kata atau token. Menghapus "stopwords" (kata-kata umum yang tidak memberikan informasi sentimen, seperti "dan", "atau"). Dan melakukan stemming untuk menyederhanakan kata ke bentuk dasarnya.
3. Labelling Sentimen setiap teks diberi label sentimen, yaitu "positif", "negatif", atau "netral" dengan menggunakan mesin *TextBlib*.
4. Vektorisasi data, mengubah data teks menjadi data numerik sehingga dapat dilakukan analisis data lebih lanjut menggunakan *TF – IDF*
5. Membagi data latih (training set) dan data uji (test set) sesuai dengan kebutuhan proporsi penelitian, untuk memastikan model terlatih dengan data yang memadai serta diuji keakuratannya.
6. Pemodelan analisis sentimen menggunakan algoritma Naive Bayes berdasarkan data training.
7. Klasifikasi hasil pemodelan Naive Bayes menggunakan data uji dan melihat seberapa besar akurasi yang dihasilkan.

Adapun tahapan penelitian dapat digambarkan dengan *flowchart* berikut :



Gambar 1. *Flowchart* Analisis

Proses klasifikasi *Naive Bayes* secara umum dapat dilihat pada Persamaan 1 (C.Aggarwal, 2015).

$$P(c_j|w_i) = \frac{P(c_j) \times P(w_i|c_j)}{P(w_i)} \quad (1)$$

Keterangan:

$P(c_j|w_i)$: *Posterior* merupakan peluang kategori j ketika terdapat kemunculan kata i

$P(w_i|c_j)$: *Conditional probability* merupakan peluang sebuah kata i masuk ke dalam kategori j

$P(c_j)$: *Prior* merupakan peluang kemunculan sebuah kategori j

$P(w_i)$: Peluang kemunculan sebuah kata

i : Indeks kata yang dimulai dari 1 hingga kata ke- k

j : Indeks kategori yang dimulai dari 1 hingga kategori ke- n

Peluang kemunculan kata sesungguhnya dapat dihilangkan pada proses perhitungan klasifikasi karena peluang tersebut tidak akan berpengaruh pada perbandingan hasil klasifikasi dari setiap kategori. Proses pada klasifikasi dapat disederhanakan dengan Persamaan 2 (Manning et al., 2009).

$$P(c_j|w_i) = P(c_j) \times P(w_i|c_j) \quad (2)$$

Untuk menghitung *prior* atau peluang kemunculan suatu kategori pada semua dokumen dapat dilakukan dengan menggunakan Persamaan 3 (Manning, et al., 2009)

$$P(c_j) = \frac{N_{c_j}}{N} \quad (3)$$

Keterangan:

N_{c_j} : Dokumen yang masuk kategori c_j

N : Jumlah keseluruhan dokumen latih yang digunakan

Pada umumnya data uji memiliki banyak kata yang diproses mulai indeks ke-1 hingga ke- k , dalam hal ini conditional probability kata w_i pada kategori c_j dilakukan perkalian dari $i=1$ sampai $i=k$ sehingga untuk mengetahui nilai posterior dapat dihitung dengan menggunakan Persamaan 4.

$$P(c_j|w_i) = P(c_j) \times P(w_1|c_j) \times P(w_2|c_j) \times P(w_3|c_j) \dots \times P(w_i|c_j) \quad (4)$$

C. Hasil Penelitian dan Pembahasan

Data yang digunakan adalah data primer berupa teks komentar, *tweet*, atau *retweet* dari media sosial X yang memuat *keyword* Prabowo, Gibran, Presiden. Data yang diperoleh dari hasil *scrapping* sebanyak 1478 yang dilakukan pada tanggal 21-30 Oktober 2024 dimana waktu tersebut adalah saat pasangan presiden dan wakil presiden terpilih periode 2024-2029 dilantik. Berikut data yang digunakan pada penelitian ini.

Data Penelitian

Tabel 1. Data Penelitian

No	Cuitan X
1	Rencana pemerintahan Prabowo Subianto menggelar lagi program transmigrasi ke Papua memicu kecemasan sekaligus kemarahan di kalangan https://t.co/Ts4FzKwzIH
2	@msaid_didu Mungkin saja itu strategi Prabowo utk mengetahui keterlibatan menteri lainnya....operasi senyap bawah tanah sedang dimainkan...
⋮	⋮
1478	Selama 10 tahun kepemimpinannya Presiden Jokowi berhasil membangun Indonesia#Jokowihebat#TerimaKasihPakJokowi#PengabdianPenuhKepedulian https://t.co/kygCnw4ld9

Text Preprocessing

Sebelum melakukan analisis sentimen, dilakukan preprocessing data untuk membersihkan dan menyiapkan data. Berikut adalah perbandingan antara sesudah dan sebelum dilakukan *text preprocessing* disajikan pada tabel 2.

Tabel 2. Text Preprocessing

No	Full Text	Cleasning
1	Rencana pemerintahan Prabowo Subianto menggelar lagi program transmigrasi ke	['rencana', 'perintah', 'prabowo', 'subianto', 'gelar', 'program',

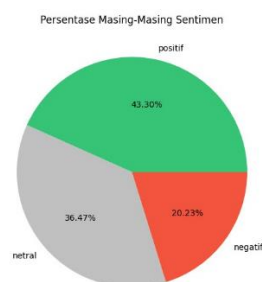
Papua memicu kecemasan sekaligus ‘transmigrasi’, ‘papua’, ‘picu’, ‘cemas’, kemarahan di kalangan orang asli papua ‘sekalius’, ‘marah’, ‘kalangan’]
<https://t.co/Ts4FzKwzIH>

Labelling Sentimen

Selanjutnya dilakukan proses *labelling* pada data cuitan sehingga memiliki label sesuai dengan sentimennya. Proses *labelling* memanfaatkan mesin *TextBlob* yang diaplikasikan menggunakan bahasa python. Berikut merupakan hasil dari proses *labelling* disajikan pada tabel 3 dan gambar 1.

Tabel 3. Hasil *Labelling*

No	Full text	Labelling
1	Rencana pemerintahan Prabowo Subianto menggelar lagi program transmigrasi ke Papua memicu kecemasan sekaligus kemarahan di kalangan orang asli Papua. https://t.co/Ts4FzKwzIH	Negatif
2	@msaid_didu Mungkin saja itu strategi Prabowo utk mengetahui keterlibatan menteri lainnya...operasi senyap bawah tanah sedang dimainkan...	Negatif
:	:	:
1477	@DivHumas_Polri kok kalian lemot ya kasus yg beginian apa Krn libatkan anak presiden dan calon wapres 2024?	Netral
1478	Selama 10 tahun kepemimpinannya Presiden Jokowi berhasil membangun Indonesia #Jokowihebat #TerimaKasihPakJokowi #PengabdianPenuhKepedulian https://t.co/kygCnw4ld9	Positif



Gambar 2. Pie Chart *Labelling* Sentimen

Split Data Training - Testing

Sebanyak 1.478 *tweet* yang telah diberi label dibagi menjadi dua kelompok, yaitu data *training* dan data *testing*. Data *training* digunakan untuk membangun dan melatih model klasifikasi, sedangkan data *testing* dimanfaatkan untuk mengevaluasi kinerja model tersebut. Penelitian ini menerapkan rasio pembagian data sebesar 80% untuk *training* dan 20% untuk *testing*, karena rasio tersebut dianggap optimal. Pembagian data dilakukan secara acak menggunakan fungsi *train_test_split* dalam Python. Berikut hasil pembagian data disajikan pada tabel 4.

Tabel 4. Data *Training* dan Data *Testing*

Keterangan	Data <i>Training</i>	Data <i>Testing</i>	Total
Persentase	80%	20%	100%
Jumlah	1.182	296	1.478

Evaluasi Model Naive Bayes**Tabel 5.** Evaluasi Model *Naive Bayes*

	<i>Before SMOTE</i>	SMOTE	<i>Increase</i>
<i>Accuracy</i>	55,41%	75,52%	20,11%
<i>Recall</i>	75,99%	78,42%	2,43%
<i>Precision</i>	45,43%	75,52%	30,09%
<i>F1-Score</i>	41,35%	75,09%	33,74%

Pada percobaan awal, model *Naive Bayes* menunjukkan performa yang kurang optimal saat diterapkan pada data yang tidak seimbang. Hal ini terlihat dari nilai akurasi sebesar 55,41%, *recall* sebesar 75,99%, *precision* sebesar 45,43%, dan *F1-score* sebesar 41,35%. Hasil tersebut mengindikasikan bahwa model memiliki kesulitan dalam mengenali kelas minoritas akibat ketidakseimbangan data, yang menyebabkan prediksi terhadap kelas tertentu menjadi kurang akurat.

Untuk mengatasi masalah *class imbalance*, metode SMOTE (*Synthetic Minority Oversampling Technique*) diterapkan. SMOTE menghasilkan data sintetik untuk menyeimbangkan jumlah sampel di setiap. Dengan menggunakan data hasil *oversampling*, performa model meningkat secara signifikan. Akurasi model naik menjadi 75,52%, *recall* meningkat menjadi 78,42%, *precision* menjadi 75,52%, dan *F1-score* mencapai 75,09%. Peningkatan ini menunjukkan bahwa model mampu mengenali lebih baik semua kelas, terutama kelas minoritas, setelah distribusi data menjadi lebih seimbang.

Tabel 6. Classification Report

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Sentimen Negatif	67%	93%	78%
Sentimen Netral	87%	68%	76%
Sentimen Positif	79%	66%	71%

Berdasarkan analisis classification report, sentimen netral memiliki precision tertinggi sebesar 87%, yang menunjukkan bahwa prediksi untuk sentimen netral sangat akurat. Sebaliknya, sentimen negatif memiliki recall tertinggi sebesar 93%, yang berarti hampir semua sampel dengan sentimen negatif berhasil dikenali dengan baik oleh model. Namun, terdapat kelemahan pada precision untuk sentimen negatif sebesar 67%, yang mengindikasikan banyak prediksi negatif yang salah, serta recall untuk sentimen positif sebesar 66%, yang menunjukkan bahwa banyak sampel positif tidak terdeteksi. Dari sisi F1-Score, sentimen negatif memiliki performa terbaik dengan nilai 78%, diikuti oleh sentimen netral (76%) dan positif (71%). Secara keseluruhan, penerapan SMOTE terbukti efektif dalam meningkatkan performa model *Naive Bayes* dengan mengatasi ketidakseimbangan data. Langkah ini penting untuk memastikan model tidak hanya fokus pada sentimen mayoritas, tetapi juga mampu mengenali sentimen minoritas secara lebih akurat, sebagaimana ditunjukkan oleh peningkatan kinerja pada tabel di bawah.

Tabel 7. Confusion Matrix

		Nilai Prediksi		
		Sentimen Positif	Sentimen Netral	Sentimen Negatif
Nilai Akurat	Sentimen Positif	84	10	34
	Sentimen Netral	17	87	24
	Sentimen Negatif	6	3	119

Pada tabel *confusion matrix* menunjukkan bahwa model *Naive Bayes* mampu mengklasifikasikan *tweet* ke dalam tiga kategori sentimen, yaitu negatif, netral, dan positif. Dari total 128 *tweet* sentimen negatif, sebanyak 119 *tweet* berhasil diklasifikasikan dengan benar sebagai negatif, sementara 3 *tweet* salah diklasifikasikan sebagai netral dan 6 *tweet* salah diklasifikasikan sebagai positif. Untuk 128 *tweet* sentimen netral, sebanyak 87 *tweet* diklasifikasikan dengan benar sebagai netral, sedangkan 17 *tweet* salah diklasifikasikan sebagai negatif dan 24 *tweet* salah diklasifikasikan sebagai positif. Sementara itu, dari 128 *tweet* sentimen positif, sebanyak 84 *tweet* berhasil diklasifikasikan dengan benar sebagai positif, 10 *tweet* salah diklasifikasikan sebagai netral, dan 34 *tweet* positif yang salah diklasifikasikan sebagai negatif. Hal ini menunjukkan bahwa model memiliki performa yang baik dalam mengidentifikasi *tweet* sentimen negatif dengan *recall* yang tinggi, meskipun terdapat beberapa kesalahan pada prediksi sentimen netral dan positif.

D. Kesimpulan

Berdasarkan hasil analisis sentimen masyarakat terhadap pasangan presiden terpilih periode 2024-2029 di media sosial X menggunakan algoritma *Naive Bayes*, diperoleh bahwa:

Analisis menunjukkan bahwa opini masyarakat di media sosial X terkait pasangan presiden terpilih Prabowo Subianto dan Gibran Rakabuming Raka terdiri dari tiga kategori sentimen: positif, negatif, dan netral. Sentimen positif mendominasi dengan persentase 43,30%, diikuti sentimen netral sebesar 36,47%, dan sentimen negatif yang mencapai 20,23%.

Algoritma *Naive Bayes* terbukti efektif dalam mengklasifikasikan sentimen publik dengan akurasi meningkat setelah penerapan teknik SMOTE, yang menunjukkan kemampuannya dalam menangani ketidakseimbangan data dengan lebih baik setelah data disesuaikan. Model *Naive Bayes* yang digunakan mencapai akurasi sebesar 75,52% setelah penerapan SMOTE. Hal ini membuktikan keandalannya dalam menangani volume data besar dari media sosial X dan memberikan hasil klasifikasi cukup akurat dalam klasifikasi sentimen.. Analisis sentimen terhadap pasangan presiden terpilih periode 2024-2029 menggunakan algoritma *Naive Bayes* memberikan gambaran persepsi publik di media sosial X yang mayoritas bersifat positif. Meskipun awalnya model menunjukkan performa kurang optimal akibat ketidakseimbangan data, penerapan metode SMOTE berhasil meningkatkan *accuracy*, *precision*, dan *F1-score* secara signifikan, terutama dalam mengenali sentimen minoritas. Penelitian ini menyoroti pentingnya *preprocessing* data, seperti *cleansing*, *stemming*, dan *tokenizing*, untuk memastikan data siap digunakan dalam analisis. Hasil yang diperoleh tidak hanya relevan bagi pengambil kebijakan untuk memahami opini publik, tetapi juga membuka peluang untuk penelitian lebih lanjut dengan integrasi metode lain atau analisis sentimen berbasis waktu untuk memahami dinamika perubahan persepsi masyarakat.

Daftar Pustaka

- Abidin, R., & Herawati, A. (2024). Analisis Sentimen Publik Terhadap Kebijakan Program Tabungan Perumahan Rakyat (Tapera). *Journal of Information System and Computer*, 13–19.
- C.Aggarwal, C. (2015). *Data Classification: Algorithms and Applications*. Berilustrasi penyunt. CRC Press.

- Diyandini, S. M., & Kurniati, E. (2024). Analisis Maksimalisasi Utilitas Konsumen Melalui Grafik dan Persamaan Rumus. *DataMath: Journal of Statistics and Mathematics*, 2(1), 1–10.
- Fauzan, Abd. C., & Hikmah, K. (2022). IMPLEMENTASI ALGORITMA NAIVE BAYES DALAM ANALISIS POLARISASI OPINI MASYARAKAT TERKAIT VAKSIN COVID-19. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 7(2), 122–128. <https://doi.org/10.36341/rabit.v7i2.2403>
- Goel, A., Gautam, J., & Kumar, S. (2016). Real time sentiment analysis of tweets using Naive Bayes. *2nd International Conference on Next Generation Computing Technologies (NGCT)*, 257–261.
- Jha, A. K., & Verma, N. K. (2024). Social Media Platforms and User Engagement: A Multi-Platform Study on One-way Firm Sustainability Communication. *Information Systems Frontiers*, 26(1), 177–194. <https://doi.org/10.1007/s10796-023-10376-8>
- Liu, B. (2023). Sentiment Analysis and Opinion Mining: A Survey of New Methods and Applications. *IEEE Xplore*.
- Macfud, A. Z., Kusuma, A. P., & Puspitasari, W. D. (2023). ANALISIS ALGORITMA NAIVE BAYES CLASSIFIER (NBC) PADA KLASIFIKASI TINGKAT MINAT BARANG DI TOKO VIOLET CELL. *Jurnal Mahasiswa Teknik Informatika*, 7(1).
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge University Press.
- Noor, F. R., & Sirodj, D. A. N. (2024). Penerapan Support Vector Machine untuk Klasifikasi Opini Masyarakat Terhadap Isu Bullying. *Jurnal Riset Statistika*, 57–66. <https://doi.org/10.29313/jrs.v4i1.3877>
- Pramastuti, J. I. R. (2016). Sentimen Analisis Review Aplikasi Mobile Menggunakan Metode Ensemble Classifier Berbasis Average Of Probabilities. *Doro*, 7(6).
- Refi Maulana Yusuf, & Dadi Ahmadi. (2022). Kampanye Budaya Beberes di Media Sosial Instagram. *Jurnal Riset Jurnalistik Dan Media Digital*, 151–158. <https://doi.org/10.29313/jrjmd.v2i2.1530>
- Rish, I. (2014). *An Empirical Study of the Naïve Bayes Classifier An empirical study of the naive Bayes classifier*. <https://www.researchgate.net/publication/228845263>
- Sunata, M. H. A., Irwiensyah, F., & Hasan, F. N. (2024). Analisis Sentimen Calon Presiden 2024 di Media Sosial X Menggunakan Naive Bayes dan SMOTE. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(3), 1313. <https://doi.org/10.30865/mib.v8i3.7708>