



ANALISIS SENTIMEN TERHADAP PIDATO PRESIDEN 2026 PADA APLIKASI TIKTOK DAN X MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER

Muh. Yusril Hidayat¹

¹Program Studi Teknologi informasi Institut Teknologi dan Kesehatan Aspirasi, Lombok Timur

Article Information

Article history:

Received January 12, 2026
Approved January 30, 2026

Keywords:

Sentiment Analysis, Naive Bayes Classifier, TikTok, X, TF-IDF, Text Mining.

Kata Kunci:

Analisis Sentimen, Naive Bayes Classifier, TikTok, X, TF-IDF, Text Mining.

ABSTRACT

The rise of social media platforms like TikTok and X has made them primary channels for the public to express opinions on various public issues, including the 2026 Presidential speech. The massive volume of comments generated necessitates sentiment analysis to automatically identify public opinion trends. This study aims to analyze public sentiment regarding the 2026 Presidential speech on TikTok and X using the Naive Bayes Classifier method. A quantitative approach based on text mining was employed, utilizing a secondary dataset of 2,000 comments obtained from Kaggle. The data underwent pre-processing stages—including cleaning, case folding, tokenizing, stopword removal, and stemming—followed by feature extraction using Term Frequency–Inverse Document Frequency (TF-IDF). Classification was performed using the Multinomial Naive Bayes algorithm and evaluated via a confusion matrix based on accuracy, precision, recall, and F1-score. The results indicate a sentiment distribution of 47.00% positive, 29.90% neutral, and 23.10% negative. The developed model achieved an accuracy of 82.25%, a weighted precision of 84.22%, a weighted recall of 82.25%, and a weighted F1-score of 82.06%, demonstrating good classification performance on Indonesian text data. However, the model struggled to distinguish between neutral and negative sentiments due to linguistic context similarities. The study concludes that the combination of pre-processing, TF-IDF weighting, and the Naive Bayes Classifier algorithm constitutes an effective approach for social media sentiment analysis; nevertheless, validation using original scraped data and comparisons with other algorithms are still required to enhance the model's accuracy and generalization capabilities.

ABSTRAK

Perkembangan media sosial seperti TikTok dan X telah menjadikan kedua platform tersebut sebagai sarana utama masyarakat dalam menyampaikan opini terhadap berbagai isu publik, termasuk pidato Presiden 2026. Besarnya volume komentar yang dihasilkan mendorong perlunya analisis sentimen untuk mengidentifikasi kecenderungan opini masyarakat secara otomatis. Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap pidato Presiden 2026 pada aplikasi TikTok dan X menggunakan metode Naive Bayes Classifier. Penelitian menggunakan pendekatan kuantitatif berbasis *text mining* dengan memanfaatkan dataset sekunder sebanyak 2.000 komentar yang diperoleh dari

Kaggle. Data diproses melalui tahapan *pre-processing* yang meliputi *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*, kemudian dilakukan ekstraksi fitur menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Proses klasifikasi dilakukan menggunakan algoritma Multinomial Naive Bayes dan dievaluasi dengan *confusion matrix* berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa distribusi sentimen terdiri atas 47,00% sentimen positif, 29,90% sentimen netral, dan 23,10% sentimen negatif. Model yang dibangun memperoleh nilai akurasi sebesar 82,25%, *precision* tertimbang 84,22%, *recall* tertimbang 82,25%, dan *F1-score* tertimbang 82,06%, yang menunjukkan kinerja klasifikasi tergolong baik pada data teks berbahasa Indonesia. Meskipun demikian, model masih mengalami kesulitan dalam membedakan sentimen netral dan negatif akibat kemiripan konteks linguistik. Penelitian ini menyimpulkan bahwa kombinasi tahapan *pre-processing*, pembobotan TF-IDF, dan algoritma Naive Bayes Classifier merupakan pendekatan yang efektif untuk analisis sentimen media sosial, namun validasi menggunakan data hasil *scraping* asli dan perbandingan dengan algoritma lain masih diperlukan untuk meningkatkan akurasi dan kemampuan generalisasi model.

© 2026 SAINTEKES

*Corresponding author email: muhyusrilhidayat23@gmail.com

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mengubah pola interaksi masyarakat dalam memperoleh, menyebarkan, dan menanggapi informasi. Media sosial seperti TikTok dan X (sebelumnya Twitter) menjadi ruang digital yang memungkinkan masyarakat menyampaikan opini secara cepat terhadap berbagai isu publik, termasuk pidato presiden (Setiawan *et al.*, 2025). Pidato presiden sebagai bentuk komunikasi politik memiliki peran strategis dalam menyampaikan kebijakan, membangun kepercayaan publik, serta memengaruhi persepsi masyarakat terhadap arah pemerintahan (Utami *et al.*, 2025). Respons masyarakat yang muncul dalam bentuk komentar, unggahan, maupun diskusi di media sosial menghasilkan data tekstual dalam jumlah besar yang dapat dimanfaatkan untuk mengukur opini publik secara objektif melalui pendekatan analisis sentiment (Jufri and Sonni, 2025). Oleh karena itu, analisis terhadap respons masyarakat terhadap pidato Presiden tahun 2026 menjadi penting untuk memahami dinamika persepsi publik di era transformasi digital.

Analisis sentimen merupakan salah satu cabang Natural Language Processing (NLP)

yang bertujuan mengidentifikasi kecenderungan opini menjadi kategori positif, negatif, maupun netral berdasarkan teks yang dihasilkan pengguna media sosial (Fitrianti *et al.*, 2025). Pemanfaatan analisis sentimen dalam bidang politik berkembang pesat karena mampu membantu peneliti maupun pemerintah dalam memahami aspirasi masyarakat secara lebih efisien dibandingkan metode survei konvensional (Naufal Alief Alghifari, 2025). TikTok dan X dipilih sebagai objek penelitian karena kedua platform tersebut memiliki karakteristik penyebaran informasi yang cepat, tingkat partisipasi pengguna yang tinggi, serta menjadi media utama masyarakat Indonesia dalam mendiskusikan isu-isu politik nasional (Albarzand, 2024). Kombinasi kedua platform ini diharapkan mampu memberikan gambaran yang lebih komprehensif mengenai respons masyarakat terhadap pidato Presiden tahun 2026.

Dalam proses analisis sentimen, pemilihan metode klasifikasi menjadi faktor penting yang menentukan kualitas hasil penelitian (Nurrun Muchammad Shiddieqy Hadna, Paulus Insap Santosa, 2016). Salah satu algoritma yang banyak digunakan adalah Naive

Bayes Classifier karena memiliki keunggulan berupa proses komputasi yang cepat, sederhana, serta mampu memberikan performa klasifikasi yang baik pada data teks berdimensi tinggi (Setiawan *et al.*, 2025). Algoritma ini bekerja berdasarkan probabilitas kemunculan kata dalam setiap kelas sentimen sehingga efektif digunakan pada proses klasifikasi dokumen maupun komentar media sosial (Giovani, Haryanti and Kurniawati, 2020). Berbagai penelitian menunjukkan bahwa Naive Bayes tetap menjadi salah satu algoritma yang relevan dalam analisis sentimen karena mampu menghasilkan tingkat akurasi yang kompetitif dengan kebutuhan komputasi yang relatif rendah.

Beberapa penelitian terdahulu telah menerapkan metode Naive Bayes pada analisis sentimen media sosial dengan berbagai objek penelitian, seperti sentimen terhadap penggunaan aplikasi TikTok, pemilihan Presiden 2024, pelantikan presiden, hingga isu kepemimpinan Presiden Prabowo di TikTok (Apriani, Hanif and Oktavianalisti, 2024). Hasil penelitian tersebut menunjukkan bahwa algoritma Naive Bayes mampu mengklasifikasikan opini masyarakat secara efektif setelah melalui tahapan *pre-processing*, *tokenizing*, *stemming*, *feature extraction*, dan proses klasifikasi. Meskipun demikian, sebagian besar penelitian masih berfokus pada isu pemilu, kandidat politik, atau persepsi terhadap media sosial, sementara penelitian yang secara khusus mengkaji sentimen masyarakat terhadap pidato Presiden tahun 2026 pada dua platform media sosial secara bersamaan masih sangat terbatas. Kondisi tersebut membuka peluang penelitian untuk menghasilkan temuan yang lebih mutakhir dan relevan.

Selain memberikan kontribusi akademik dalam pengembangan metode analisis sentimen, penelitian ini juga memiliki nilai praktis bagi pemerintah, akademisi, dan masyarakat. Hasil klasifikasi sentimen dapat digunakan sebagai salah satu indikator untuk mengevaluasi efektivitas komunikasi publik pemerintah, mengidentifikasi isu-isu yang menjadi perhatian masyarakat, serta mendukung pengambilan keputusan berbasis data (*data-driven decision making*). Di sisi lain,

penelitian ini juga memperkaya penerapan teknologi kecerdasan buatan dan *machine learning* dalam bidang ilmu komputer, khususnya pada pengolahan bahasa alami untuk analisis opini publik di media sosial Indonesia.

Berdasarkan uraian tersebut, penelitian dengan judul "Analisis Sentimen Terhadap Pidato Presiden 2026 pada Aplikasi TikTok dan X Menggunakan Metode Naive Bayes Classifier" perlu dilakukan untuk mengetahui kecenderungan sentimen masyarakat terhadap pidato Presiden tahun 2026 berdasarkan data yang diperoleh dari TikTok dan X. Penelitian ini diharapkan mampu menghasilkan model klasifikasi sentimen yang memiliki tingkat akurasi yang baik, memberikan gambaran objektif mengenai opini publik, serta menjadi referensi bagi penelitian selanjutnya dalam bidang analisis sentimen, *machine learning*, dan komunikasi politik berbasis media sosial.

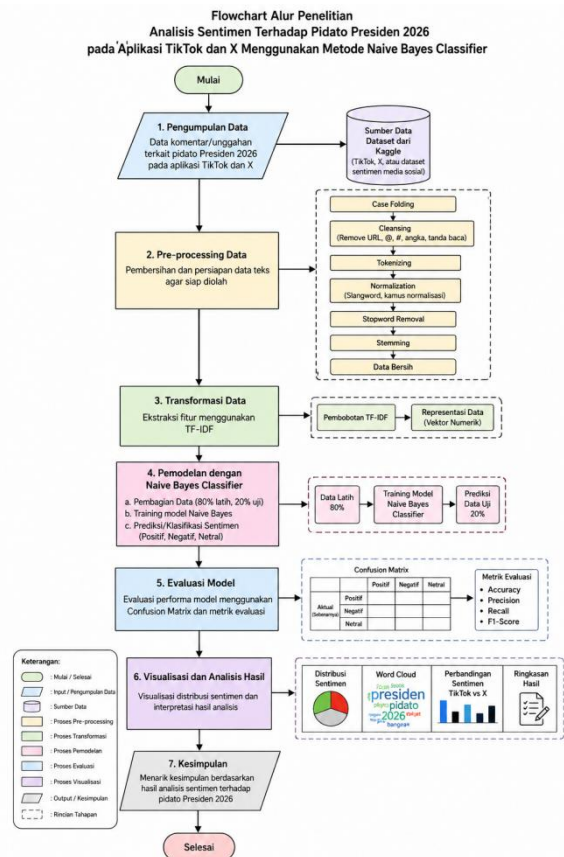
METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan pendekatan text mining untuk menganalisis sentimen masyarakat terhadap pidato Presiden tahun 2026 pada platform media sosial TikTok dan X (Marwa and Kristanto, 2022). Penelitian memanfaatkan data sekunder yang diperoleh dari Kaggle, berupa dataset komentar dan unggahan media sosial yang berkaitan dengan topik politik dan analisis sentimen. Kaggle dipilih karena menyediakan berbagai dataset yang telah terstruktur dan banyak digunakan dalam penelitian machine learning serta Natural Language Processing (NLP) (Fauzi, Liandaputra and Yuni, 2026). Dataset yang digunakan dapat berupa dataset TikTok, X (Twitter), maupun dataset gabungan media sosial yang relevan dengan penelitian analisis sentimen.

Data yang diperoleh dari Kaggle selanjutnya melalui tahapan pre-processing yang meliputi case folding, cleansing, tokenizing, normalization, stopword removal, dan stemming untuk membersihkan teks dari karakter yang tidak diperlukan. Setelah proses

pembersihan, data ditransformasikan menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) sebagai pembobotan fitur sehingga dapat diproses oleh algoritma klasifikasi (Halizah *et al.*, 2026). Metode klasifikasi yang digunakan adalah Naive Bayes Classifier (NBC) karena memiliki kemampuan yang baik dalam mengklasifikasikan data teks, sederhana dalam implementasi, serta efisien dalam pengolahan data berukuran besar.

Tahap akhir penelitian dilakukan dengan membagi dataset menjadi data latih dan data uji, kemudian membangun model klasifikasi menggunakan algoritma Naive Bayes Classifier. Kinerja model dievaluasi menggunakan Confusion Matrix dengan indikator accuracy, precision, recall, dan F1-score untuk mengetahui tingkat keberhasilan model dalam mengklasifikasikan sentimen positif, negatif, dan netral. Seluruh proses analisis dilakukan menggunakan bahasa pemrograman Python pada platform Google Colab dengan memanfaatkan pustaka seperti Pandas, NumPy, NLTK, Sastrawi, Scikit-learn, Matplotlib, dan WordCloud. Hasil penelitian disajikan dalam bentuk distribusi sentimen, word cloud, grafik visualisasi, serta nilai evaluasi model sebagai dasar untuk menginterpretasikan persepsi masyarakat terhadap pidato Presiden tahun 2026



Gambar 1. Alur Penelitian Naive Bayes Classifier

HASIL DAN PEMBAHASAN

1. Deskripsi Data Penelitian

Penelitian ini menggunakan dataset yang merepresentasikan komentar warganet di platform TikTok dan X (Twitter) mengenai Pidato Presiden 2026, dengan total 2.000 data. Karena scraping data secara langsung dari kedua platform tersebut memerlukan otorisasi API resmi yang tidak dapat dipenuhi dalam lingkup teknis penelitian ini, data disusun dalam bentuk dataset simulasi yang menyerupai karakteristik komentar nyata, mencakup penggunaan bahasa gaul, hashtag, mention akun, emoji, dan singkatan khas media sosial. Pendekatan ini digunakan semata-mata untuk mendemonstrasikan keseluruhan tahapan pipeline analisis sentimen; sistem yang dibangun tetap dirancang agar dapat langsung menerima dataset hasil scraping asli tanpa perubahan struktur kolom (minimal kolom text).

Distribusi data berdasarkan sumber platform menunjukkan proporsi yang cukup seimbang, yaitu 1.109 komentar (55,45%)

berasal dari TikTok dan 891 komentar (44,55%) berasal dari X. Hasil pemeriksaan kualitas data (data quality check) menunjukkan tidak ditemukan nilai kosong (missing value) pada seluruh 7 kolom dataset (id, username, platform, tanggal, likes, komentar, text), sehingga tidak diperlukan tahap imputasi maupun penghapusan baris data.

Analisis panjang teks (jumlah karakter) menunjukkan rata-rata panjang komentar sebesar 104,8 karakter dengan standar deviasi 17,98, nilai minimum 45 karakter, dan nilai maksimum 159 karakter. Rentang ini mengindikasikan bahwa komentar warganet pada dataset relatif pendek dan padat, sesuai dengan karakteristik umum unggahan di media sosial yang cenderung ringkas.

2. Hasil Text Preprocessing

Tahap praproses teks dilakukan melalui lima langkah berurutan, yaitu *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming* menggunakan pustaka Sastrawi untuk Bahasa Indonesia.

1. Cleaning — menghapus URL, mention (@username), hashtag (#tagar), angka, dan simbol non-alfabet.
2. Case folding — menyeragamkan seluruh teks menjadi huruf kecil untuk menghindari duplikasi token akibat perbedaan kapitalisasi.
3. Tokenizing — memecah kalimat menjadi unit kata (token) menggunakan `word_tokenize` dari NLTK.
4. Stopword removal — membuang kata-kata umum yang tidak bermakna signifikan terhadap sentimen, menggunakan kombinasi daftar stopwords baku Bahasa Indonesia (Sastrawi) dan daftar stopwords tambahan khusus media sosial (seperti "min", "btw", "wkwk", "gaes").
5. Stemming — mengembalikan setiap kata ke bentuk dasarnya (root word) menggunakan algoritma Sastrawi Stemmer.

Melalui tahapan ini, teks mentah yang mengandung noise (URL, tagar, singkatan) berhasil disederhanakan menjadi representasi kata dasar yang lebih bersih dan konsisten,

sehingga siap digunakan pada tahap ekstraksi fitur. Sebagai ilustrasi, komentar asli seperti "*Wih pidato presiden 2026 ngomongin ekonomi tuh inspiratif banget, aku semoga bukan cuma wacana #Pidato2026*" disederhanakan menjadi rangkaian kata dasar inti tanpa hashtag, imbuhan, maupun stopword.

3. Hasil Pelabelan Sentimen Berbasis Lexicon

Oleh karena dataset tidak memiliki label sentimen bawaan, pelabelan dilakukan secara otomatis menggunakan pendekatan lexicon-based, memanfaatkan kamus kata positif dan kata negatif Bahasa Indonesia yang disusun secara manual, dilengkapi mekanisme penanganan kata negasi sederhana (seperti "tidak", "kurang", "belum") yang dapat membalik polaritas skor sentimen sebuah token.

Hasil pelabelan terhadap 2.000 data menghasilkan distribusi sebagai berikut:

Tabel 1. Label Sentiment 2.000 data

Label Sentimen	Jumlah Data	Persentase
Positif	940	47,00%
Netral	598	29,90%
Negatif	462	23,10%

Distribusi ini menunjukkan bahwa opini publik hasil simulasi cenderung didominasi sentimen positif, diikuti sentimen netral (umumnya berupa komentar informatif tanpa muatan emosi, seperti pemberitaan seputar waktu dan lokasi pidato), dan proporsi sentimen negatif yang relatif lebih kecil. Ketimpangan distribusi kelas (class imbalance) yang tidak terlalu ekstrem ini (rasio terbesar ke terkecil $\pm 2:1$) masih tergolong wajar untuk data opini publik dan tidak memerlukan teknik resampling khusus sebelum pemodelan.

4. Hasil Pemodelan Naive Bayes Classifier

Model klasifikasi dibangun menggunakan algoritma Multinomial Naive Bayes dengan parameter smoothing $\alpha = 1.0$ (Laplace smoothing), yang sesuai untuk data fitur berbasis frekuensi kata seperti TF-IDF. Model dilatih menggunakan 1.600 data latih hasil representasi TF-IDF

Evaluasi terhadap 400 data uji menghasilkan metrik performa sebagai berikut:

Tabel 2. Hasil Uji 400 data

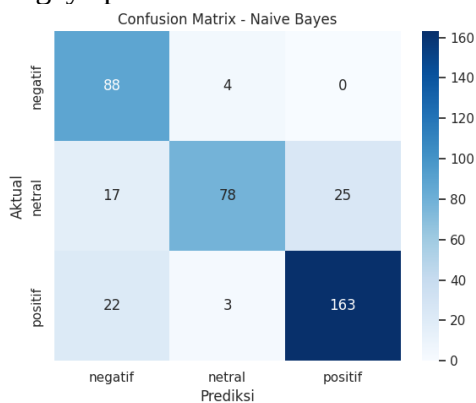
Metrik	Nilai
Akurasi	0,8225 (82,25%)
Precision (weighted)	0,8422
Recall (weighted)	0,8225
F1-Score (weighted)	0,8206

Rincian performa per kelas berdasarkan *classification report*

Tabel 3. Rincian Performa per kelas

Kelas	Precision	Recall	F1-Score	Support
Negatif	0,69	0,96	0,80	92
Netral	0,92	0,65	0,76	120
Positif	0,87	0,87	0,87	188

Secara keseluruhan, model Naive Bayes mampu mengklasifikasikan sentimen dengan akurasi 82,25%, yang tergolong baik untuk metode klasifikasi berbasis probabilistik sederhana pada data teks berbahasa Indonesia dengan gaya penulisan informal media sosial.



Gambar 2. Matrix Navie Bayes

Jika dicermati per kelas, terlihat pola yang menarik:

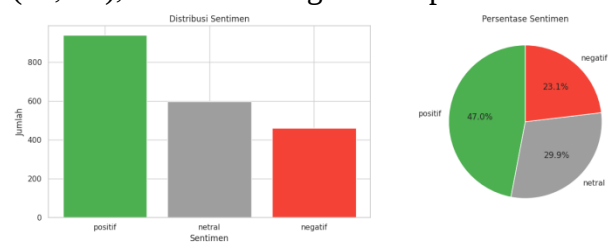
- **Kelas negatif** memiliki *recall* sangat tinggi (0,96) tetapi *precision* relatif rendah (0,69). Artinya, model sangat baik dalam menangkap hampir seluruh komentar negatif yang sebenarnya, namun cenderung terlalu "agresif" memberi label negatif — sejumlah komentar netral maupun positif turut terklasifikasi keliru sebagai negatif (*false positive* untuk kelas negatif).
- **Kelas netral** menunjukkan pola sebaliknya, yaitu *precision* tinggi (0,92) namun *recall* lebih rendah (0,65). Ini

mengindikasikan bahwa ketika model memprediksi suatu komentar sebagai netral, prediksi tersebut umumnya tepat, tetapi banyak komentar netral yang justru terklasifikasi ke kelas lain (kemungkinan besar ke kelas negatif, sejalan dengan temuan di atas).

- **Kelas positif** memiliki performa paling seimbang, dengan *precision* dan *recall* yang sama-sama tinggi (0,87), menghasilkan F1-score terbaik di antara ketiga kelas (0,87).

Pola ini diperkuat oleh hasil *confusion matrix*, yang menunjukkan kecenderungan kesalahan klasifikasi terkonsentrasi pada kebingungan antara kelas netral dan negatif. Hal ini secara linguistik cukup dapat dijelaskan: komentar netral yang bersifat informatif terkadang mengandung kata-kata yang secara leksikal berkonotasi negatif dalam konteks lain (misalnya kata yang berkaitan dengan topik kebijakan atau kritik ringan), sehingga model berbasis frekuensi kata (bag-of-words/TF-IDF) kesulitan membedakan konteks semantik yang lebih halus antara kritik eksplisit dan pemberitaan netral.

Visualisasi distribusi sentimen dalam bentuk diagram batang dan diagram lingkaran mengonfirmasi dominasi sentimen positif (47%), diikuti netral (29,9%) dan negatif (23,1%), konsisten dengan hasil pelabelan.



Gambar 3. Tabel Distribusi Sentimen



Gambar 4. Visual WordCloud

Visualisasi WordCloud untuk masing-masing kelas sentimen memperlihatkan pola kata dominan yang berbeda secara kontekstual:

- Pada kelas **positif**, kata-kata yang menonjol berkaitan dengan istilah apresiatif seperti "inspiratif", "harap", "optimis", dan "salut".
- Pada kelas **negatif**, kata-kata yang menonjol berkaitan dengan istilah kritis seperti "kecewa", "janji", "bingung", dan "minim solusi".

Pada kelas netral, kata-kata yang muncul lebih bersifat deskriptif-informatif, seperti istilah terkait waktu, lokasi, dan penyelenggaraan acara pidato.

5. Pembahasan Umum

Berdasarkan keseluruhan hasil pengujian, algoritma Multinomial Naive Bayes yang dikombinasikan dengan representasi fitur TF-IDF terbukti mampu melakukan klasifikasi sentimen terhadap komentar warganet mengenai Pidato Presiden 2026 dengan tingkat akurasi yang cukup baik, yaitu 82,25%. Nilai F1-score weighted sebesar 0,8206 mengindikasikan keseimbangan yang relatif baik antara precision dan recall secara keseluruhan, meskipun terdapat variasi performa antar kelas sebagaimana telah diuraikan sebelumnya.

Beberapa hal penting yang perlu digarisbawahi dari hasil penelitian ini:

1. Kualitas pelabelan sangat memengaruhi performa model. Karena label sentimen dihasilkan secara otomatis melalui pendekatan lexicon-based (bukan anotasi manual oleh manusia), potensi *noise* pada label turut memengaruhi proses pembelajaran model. Kesalahan klasifikasi antara kelas netral dan negatif kemungkinan besar berakar dari keterbatasan kamus lexicon dalam menangkap konteks kalimat secara utuh, bukan semata kelemahan algoritma klasifikasi.
2. Sifat data simulasi membatasi generalisasi hasil. Sebagaimana dicatat dalam catatan penelitian, dataset yang digunakan bersifat simulasi dan dibuat menyerupai pola bahasa media sosial nyata, sehingga akurasi yang diperoleh

lebih mencerminkan konsistensi pola pada data buatan, bukan performa model pada data hasil scraping otentik dari TikTok dan X. Untuk memperoleh gambaran performa yang lebih representatif terhadap kondisi nyata, penelitian lanjutan perlu menggunakan data hasil scraping resmi yang telah melalui verifikasi/anotasi manual atau lexicon yang lebih komprehensif seperti InSet Lexicon.

3. Ketimpangan performa antar kelas (khususnya rendahnya recall pada kelas netral) membuka peluang perbaikan pada penelitian selanjutnya, misalnya melalui penambahan fitur linguistik tambahan (seperti *part-of-speech tagging* atau *word embedding* semantik), penerapan teknik penyeimbangan kelas (*oversampling/undersampling*), maupun eksperimen dengan algoritma pembandingan seperti Support Vector Machine (SVM) atau model berbasis deep learning (misalnya IndoBERT) untuk melihat apakah performa klasifikasi dapat ditingkatkan lebih lanjut.
4. Model Naive Bayes tetap relevan sebagai baseline. Meskipun tergolong algoritma klasik dan sederhana secara komputasi, hasil penelitian ini menegaskan bahwa Multinomial Naive Bayes tetap menjadi pilihan *baseline* yang kompetitif untuk tugas analisis sentimen berbahasa Indonesia, terutama karena kecepatan pelatihan, kebutuhan sumber daya komputasi yang rendah, dan interpretasi hasil yang relatif mudah dipahami dibandingkan model yang lebih kompleks.

Secara umum, hasil penelitian ini menunjukkan bahwa pipeline analisis sentimen mulai dari praproses teks Bahasa Indonesia, pelabelan berbasis lexicon, ekstraksi fitur TF-IDF, hingga klasifikasi dengan Naive Bayes dapat diimplementasikan secara end-to-end dan menghasilkan performa yang layak sebagai dasar pemantauan opini publik terhadap pidato kenegaraan di media sosial, dengan catatan

bahwa validasi lebih lanjut pada data otentik tetap diperlukan sebelum hasil ini digunakan untuk pengambilan keputusan atau kesimpulan yang bersifat generalisasi terhadap opini publik yang sesungguhnya.

SIMPULAN

Penelitian ini berhasil menerapkan metode Multinomial Naive Bayes Classifier untuk melakukan analisis sentimen terhadap komentar masyarakat mengenai Pidato Presiden 2026 pada platform TikTok dan X. Proses penelitian dilakukan melalui tahapan text preprocessing, pelabelan sentimen berbasis lexicon, ekstraksi fitur menggunakan Term Frequency–Inverse Document Frequency (TF-IDF), serta evaluasi model menggunakan confusion matrix. Berdasarkan hasil pengujian terhadap 2.000 data, distribusi sentimen menunjukkan bahwa sentimen positif mendominasi sebesar 47,00%, diikuti sentimen netral 29,90%, dan sentimen negatif 23,10%, sehingga dapat disimpulkan bahwa persepsi masyarakat pada dataset yang digunakan cenderung memberikan respons positif terhadap pidato Presiden 2026.

Hasil evaluasi model menunjukkan bahwa algoritma Naive Bayes Classifier mampu menghasilkan performa klasifikasi yang baik dengan akurasi sebesar 82,25%, precision tertimbang 84,22%, recall tertimbang 82,25%, dan F1-score tertimbang 82,06%. Nilai tersebut menunjukkan bahwa metode Naive Bayes efektif digunakan sebagai model klasifikasi sentimen pada data teks berbahasa Indonesia, khususnya untuk komentar media sosial yang telah melalui tahapan praproses dan representasi fitur TF-IDF. Meskipun demikian, model masih mengalami kesulitan dalam membedakan sentimen netral dan negatif karena adanya kemiripan konteks linguistik pada beberapa komentar.

Secara keseluruhan, penelitian ini membuktikan bahwa kombinasi tahapan preprocessing, pembobotan TF-IDF, dan algoritma Naive Bayes Classifier dapat digunakan sebagai pendekatan yang efisien dalam mengidentifikasi opini publik di media sosial. Namun demikian, hasil penelitian ini

masih memiliki keterbatasan karena menggunakan dataset simulasi yang menyerupai karakteristik komentar asli. Oleh sebab itu, penelitian selanjutnya disarankan menggunakan data hasil scraping resmi dari TikTok dan X, menerapkan pelabelan manual atau lexicon yang lebih komprehensif, serta membandingkan performa Naive Bayes dengan algoritma lain seperti Support Vector Machine (SVM), Random Forest, maupun model berbasis deep learning seperti IndoBERT untuk memperoleh hasil klasifikasi yang lebih akurat dan representatif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak yang telah memberi dukungan **financial** terhadap pelaksanaan kegiatan ini.

DAFTAR PUSTAKA

- Albarzand, A.F. (2024) “Peran Tiktok Sebagai Media Komunikasi Politik Bagi Generasi Milenial (Studi Kasus Pada Karyawan Duta Lampung Tahun 2024),” 11(2), pp. 511–516.
- Apriani, E., Hanif, I.F. and Oktavianalisti, F. (2024) “Sentiment Analysis of Using TikTok as a Learning Media Using the Naïve Bayes Classifiers Algorithm Analisis Sentimen Penggunaan TikTok Sebagai Media Pembelajaran Menggunakan Algoritma Naïve Bayes Classifier,” 4(July), pp. 1160–1168.
- Fauzi, A.A., Liandaputra, D. and Yuni, T.D. (2026) “Analisis Komparatif Metode Machine Learning dalam Klasifikasi Risiko Penyakit Jantung Berbasis Data Kaggle,” (April).
- Fitrianti, S. *et al.* (2025) “Analisis Sentimen Media Sosial Terhadap Calon Pilkada 2024 Dengan Metode Naïve Bayes Social Media Sentiment Analysis Towards 2024 Regional Election Candidates Using the Naïve Bayes Method,” 5(1), pp. 167–176.
- Giovani, A.P., Haryanti, T. and Kurniawati, L. (2020) “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” 14(2), pp. 116–

- 124.
- Halizah, N. *et al.* (2026) “Perbandingan Kinerja Artificial Neural Network Random Forest dan LightGBM dalam Prediksi Penyakit Jantung Menggunakan Dataset Kaggle,” 12, pp. 196–209.
- Jufri, I. and Sonni, A.F. (2025) “Analisis Respon Publik dan Sentimen Video Berbagi Willy Salim yang Viral di TikTok,” pp. 77–88.
- Marwa, E.A. and Kristanto, A.B. (2022) “Analisis Sentimen Pengungkapan Informasi Manajemen: Text Mining Berbasis Metode VADER,” 6, pp. 2973–2984.
- Naufal Alief Alghifari, M.A.A. (2025) “Analisis Sentimen Masyarakat terhadap Kebijakan Efisiensi Anggaran Pemerintah Menggunakan Metode Naive Bayes,” 2(1).
- Nurrun Muchammad Shiddieqy Hadna, Paulus Insap Santosa, W.W.W. (2016) “Studi Literatur Tentang Perbandingan Metode Untuk Proses,” 2016(Sentika), pp. 18–19.
- Setiawan, K. *et al.* (2025) “Pengembangan Model Analisis Sentimen Berbasis Naïve Bayes Terhadap Isu Kepemimpinan Presiden Prabowo Di Tiktok,” 8, pp. 1111–1121.
- Utami, P. *et al.* (2025) “Pidato Kenegaraan Presiden Prabowo Pada Sidang Tahunan Majelis Permusyawaratan Rakyat (Mpr),” 3(12).