

Analisis Sentimen Opini Publik Terhadap Kebijakan Pemerintah di Media Sosial X Menggunakan Algoritma Multinomial Naive Bayes

Fransiskus Gsutianto Jaya^{1)*}, Aldy Saputra Rajagukguk^[2], Doni Kristiawan^[3]

Universitas Bina Sarana Informatika^{[1], [2], [3]}

Tangerang, Indonesia

fransiskusgustianto2@gmail.com

Abstract

This study aims to analyze public sentiment toward government policies on the social media platform X using the Multinomial Naive Bayes algorithm. Data were collected from user-generated tweets and analyzed to identify public opinion trends regarding government policies. The results indicate that this method achieves strong accuracy in classifying public sentiment related to governmental decisions. Based on these findings, it can be concluded that Multinomial Naive Bayes is effective for analyzing public opinion on social media X and can serve as a valuable reference for the government in understanding public perceptions and evaluating the effectiveness of public policies

Keywords— *Sentiment Analysis, Multinomial Naive Bayes, Social Media X, Public Opinion, Government Policy.*

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen opini publik terhadap kebijakan pemerintah pada media sosial X menggunakan algoritma Multinomial Naive Bayes. Data dikumpulkan dari unggahan *tweet* pengguna dan dianalisis untuk mengidentifikasi kecenderungan opini publik terhadap kebijakan pemerintah. Hasil penelitian menunjukkan bahwa metode ini mampu menghasilkan akurasi yang baik dalam mengklasifikasikan sentimen publik terhadap kebijakan pemerintah. Berdasarkan temuan tersebut, dapat disimpulkan bahwa Multinomial Naive Bayes efektif digunakan untuk menganalisis opini publik pada media sosial X dan dapat menjadi acuan bagi pemerintah dalam memahami persepsi masyarakat serta mengevaluasi efektivitas kebijakan publik.

Kata Kunci— *Analisis Sentimen, Multinomial Naive Bayes, Media Sosial X, Opini Publik, Kebijakan Pemerintah.*

I. PENDAHULUAN

Perkembangan teknologi telah mendorong meningkatnya aktivitas komunikasi masyarakat melalui media sosial. Platform seperti X (Twitter) kini menjadi ruang diskusi publik untuk menyampaikan opini, kritik, maupun dukungan terhadap berbagai kebijakan pemerintah. Banyaknya unggahan yang bersifat real-time menjadikan media sosial sebagai sumber data penting dalam memahami persepsi publik terhadap isu pemerintahan. Namun, volume data yang besar membuat analisis manual tidak lagi efektif sehingga dibutuhkan pendekatan komputasional yang mampu mengolah data dalam jumlah besar secara cepat dan akurat.

Dalam penelitian ini, data dikumpulkan dari unggahan *tweet* kemudian diproses melalui tahapan *cleansing*, tokenisasi, *stopword removal*, dan *stemming*. Data yang telah diproses kemudian diklasifikasikan ke dalam kategori positif, negatif, dan netral. Algoritma Multinomial Naive Bayes dipilih karena dikenal memiliki performa yang baik dalam klasifikasi teks berbasis frekuensi kata. Berbagai penelitian menunjukkan bahwa algoritma ini memberikan hasil kompetitif pada data media sosial berbahasa Indonesia dan mampu bekerja secara efisien pada *dataset* besar (Aprilia & Isnain, 2024; Naraswati et al., 2021; Salim & Solichin, 2022; Setyawan et al., 2023).

Walaupun analisis sentiment telah diterapkan pada berbagai isu seperti kampanye politik, pelayanan publik, dan pemilihan umum, penelitian yang secara spesifik membahas opini publik terhadap kebijakan pemerintah Indonesia melalui platform X masih terbatas. Penelitian sebelumnya umumnya berfokus pada isu tertentu atau lingkup analisis yang sempit (Perdana et al., 2022; Sulaiman et al., 2024; Via Laurenz & Fajar Ratnawati, 2025; Zharifa & Ujianto, 2024) Oleh karena itu, penelitian ini bertujuan menerapkan Multinomial Naive Bayes untuk mengklasifikasikan sentimen publik terhadap kebijakan pemerintah Indonesia berdasarkan data unggahan *tweet*. Hasil penelitian diharapkan dapat memberikan gambaran empiris mengenai persepsi masyarakat dan mendukung proses pengambilan kebijakan berbasis data.

II. STUDI PUSTAKA

Analisis sentimen merupakan cabang dari *Natural Language Processing* (NLP) yang bertujuan mengidentifikasi kecenderungan opini publik berdasarkan teks. Pendekatan ini semakin penting seiring dengan meningkatnya penggunaan media sosial untuk membahas isu politik maupun kebijakan publik. Berbagai studi menunjukkan bahwa analisis sentimen efektif dalam memetakan opini masyarakat secara cepat dan sistematis. (Aprilia & Isnain, 2024).

Dalam NLP, metode berbasis *machine learning* banyak digunakan untuk memproses kumpulan data teks berskala besar. Naive Bayes merupakan salah satu algoritma probabilistik yang banyak digunakan karena kesederhanaannya serta kemampuan komputasi yang efisien. Varian yang paling umum diterapkan adalah Multinomial Naive Bayes, yang memanfaatkan frekuensi kemunculan kata sebagai dasar perhitungan probabilitas. (Salim & Solichin, 2022)

Sejumlah penelitian telah membuktikan efektivitas Multinomial Naive Bayes dalam analisis sentimen media sosial berbahasa Indonesia. (Setyawan et al., 2023) menunjukkan bahwa algoritma ini mampu memberikan performa tinggi dalam klasifikasi opini publik terhadap isu pemerintahan. Temuan serupa disampaikan oleh (Hidayat et al., 2025; Prayudani et al., 2024) □, yang menyatakan bahwa metode ini tetap kompetitif dibandingkan algoritma lain seperti SVM.

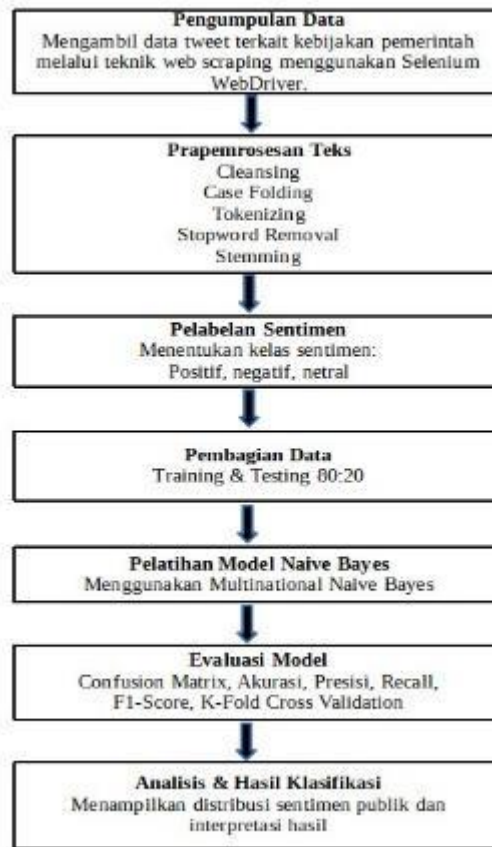
Selain itu, (Perdana et al., 2022; Sulaiman et al., 2024) menemukan bahwa Multinomial Naive Bayes cocok untuk menganalisis opini politik pada skala besar, sementara (Zharifa & Ujianto, 2024) membuktikan keandalannya dalam memetakan sentimen masyarakat pasca debat politik nasional. Berdasarkan temuan-temuan tersebut, algoritma ini dinilai relevan untuk digunakan dalam penelitian mengenai sentimen publik terhadap kebijakan pemerintah di platform X.

III. METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan pendekatan deskriptif. Tujuan penelitian adalah menganalisis sentimen opini publik terhadap kebijakan pemerintah berdasarkan *tweet* pada media sosial X. Analisis dilakukan dengan menggunakan algoritma Multinomial Naive Bayes sebagai metode klasifikasi sentimen.

Objek pada penelitian ini adalah *tweet* pengguna publik. Jenis data yang digunakan adalah data primer berupa teks yang diperoleh melalui *web scraping*. Pemilihan kata kunci ini bertujuan untuk memastikan bahwa *tweet* yang diperoleh memiliki relevansi langsung dengan topik kebijakan pemerintah, sehingga data yang dianalisis lebih fokus, representatif, dan sesuai dengan tujuan penelitian.

Alur penelitian meliputi tahap pengumpulan data, prapemrosesan teks, pelabelan sentimen, pembagian data, pelatihan model Naive Bayes, evaluasi model, dan analisis dan hasil klasifikasi.



Gambar 1 Alur Metode Penelitian Analisis Sentimen Opini Publik Menggunakan Algoritma Naive Bayes

1. Pengumpulan Data

Data dikumpulkan dari unggahan *tweet* pengguna X terkait kebijakan pemerintah menggunakan teknik *web scraping* dengan Selenium WebDriver. Kata kunci yang digunakan antara lain: subsidi, BBM, kebijakan sosial, bansos, ekonomi rakyat, pendidikan, dan isu lain terkait kebijakan pemerintah. Sebanyak 5.000 *tweet* berhasil dikumpulkan dan disimpan dalam format .xlsx.

2. Prapemrosesan Teks

Dataset hasil *scraping* kemudian melalui proses prapemrosesan untuk menyiapkan data teks agar dapat diproses algoritma klasifikasi. Tahapan prapemrosesan meliputi:

2.1. *Cleansing*

Menghapus karakter atau konten yang tidak diperlukan pada teks.

2.2. *Case folding*

Mengubah seluruh teks menjadi huruf kecil.

2.3. *Tokenization*

Memecah kalimat menjadi unit kata / token.

2.4. *Stopword removal*

Menghilangkan kata umum yang tidak berpengaruh pada polaritas sentimen.

2.5. *Stemming*

Mengubah kata ke bentuk dasarnya menggunakan algoritma stemmer bahasa Indonesia (Sastrawi), misalnya “meningkatkan” → “tingkat”.

3. Pelabelan Data

Setiap *tweet* diberikan label sentimen berdasarkan makna isi teks. Label sentimen terdiri dari tiga kategori:

- 3.1. Positif
- 3.2. Negatif
- 3.3. Netral

Pelabelan dilakukan secara otomatis sehingga *dataset* dapat langsung digunakan sebagai data latih (*training data*) untuk membangun model klasifikasi.

4. Pembagian Data

Dataset yang telah dilabeli dibagi menjadi dua bagian dengan porsi:

- 4.1. 80% data latih (*training data*)
- 4.2. 20% data uji (*testing data*)

Pembagian ini bertujuan untuk memastikan bahwa proses pelatihan dan pengujian model dapat dilakukan secara terpisah, sehingga hasil evaluasi model lebih objektif.

5. Pelatihan Model

Proses pelatihan model dilakukan dengan menggunakan algoritma Multinomial Naive Bayes, yang merupakan model probabilistik berbasis frekuensi kata dengan tahapan:

- 5.1. Konversi teks ke bentuk numerik menggunakan *CountVectorizer* yang menghasilkan representasi *bag-of-words* berdasarkan frekuensi kata.
- 5.2. Pelatihan model dengan data latih menggunakan rumus probabilistik:

$$P(c|x) \propto P(c) \times \prod_{i=1}^n P(x_i|c)$$

5.2.1. $P(c|x)$: Probabilitas bahwa data x termasuk kelas c .

5.2.2. Tanda “ \propto ” (*proportional*): Artinya sebanding dengan. Kita tidak menghitung nilai persis, tetapi cukup nilai relatif untuk dibandingkan antar kelas.

5.2.3. $P(c)$: Probabilitas prior dari kelas c .

5.2.4. $\prod P(x_i/c)$

Simbol \prod berarti perkalian.

Ini adalah *likelihood*, probabilitas setiap token x_i terhadap kelas c

Naive Bayes menghitung probabilitas kelas berdasarkan seberapa sering kelas itu muncul $P(c)$ dan seberapa mungkin fitur-fitur muncul di kelas tersebut $P(x_i/c)$. Lalu kelas yang probabilitasnya paling besar dipilih sebagai hasil prediksi.

5.3. Setelah proses pelatihan selesai, model dan *vectorizer* disimpan dalam bentuk format .pkl dan model dapat digunakan untuk memprediksi sentimen pada data uji.

6. Evaluasi Model

Evaluasi model dilakukan menggunakan *confusion matrix*, dengan menghitung metrik:

- 6.1. Akurasi
- 6.2. Presisi
- 6.3. Recall
- 6.4. F1-Score

Setelah itu dilakukan *K-Fold Cross-Validation* untuk mengukur stabilitas dan keandalan model dalam berbagai pembagian data. Teknik ini membantu memastikan bahwa model

tidak hanya bekerja baik pada satu pembagian data saja, tetapi juga konsisten pada berbagai skenario.

7. Analisis dan Hasil Klasifikasi

Tahap akhir penelitian ini adalah menampilkan dan menginterpretasikan hasil prediksi model. Hasil disajikan dalam bentuk:

7.1. Distribusi sentimen (positif, negatif, netral)

7.2. Visualisasi grafik

7.3. Tabel perbandingan

7.4. Interpretasi kecenderungan opini publik terhadap kebijakan pemerintah.

Tahap ini memberikan gambaran deskriptif mengenai bagaimana opini masyarakat terbagi berdasarkan sentimen terhadap isu kebijakan pemerintah.

IV. PEMBAHASAN DAN HASIL

Hasil penelitian ini menunjukkan bahwa algoritma Multinomial Naive Bayes mampu mengklasifikasikan sentimen opini publik terhadap kebijakan pemerintah di media sosial dengan tingkat akurasi yang tinggi. *Dataset* yang digunakan terdiri dari 5.000 *tweet* yang berkaitan dengan subsidi, BBM, kebijakan pemerintah, bansos, pajak, pilpres dan BPJS. Setelah melalui tahapan Prapemrosesan teks meliputi *cleansing*, *case folding*, *tokenization*, *stopword removal* dan *stemming* dianalisis menggunakan pendekatan *probabilistic classification* untuk mengidentifikasi arah sentimen public.

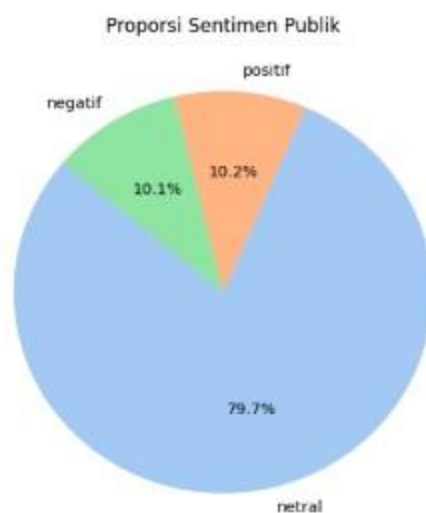
Tabel 1. Metrik Evaluasi	
Metrik	Nilai (%)
Akurasi	89,7
Presisi	82,61
Recall	75,69
F1-Score	78,46

Berdasarkan hasil pengujian, diperoleh nilai akurasi sebesar 89,7%, presisi 82,61%, *recall* 75,69%, dan *F1-score* 78,46%. Hasil ini menunjukkan bahwa algoritma Multinomial Naive Bayes efektif dalam mengenali pola bahasa dan menentukan kategori sentimen publik di media sosial. Evaluasi dilakukan menggunakan *confusion matrix* dan metode *k-fold cross-validation* yang menunjukkan konsistensi hasil klasifikasi di berbagai subset data.

Tabel 2. Distribusi Sentimen		
Kategori Sentimen	Jumlah Tweet	Persentase (%)
Positif	510	10,2
Negatif	506	10,12

Netral	3984	79,68
Total	5.000	100%

Distribusi sentimen publik menunjukkan bahwa opini netral mendominasi sebesar 79,68%, diikuti oleh negatif 10,12%, dan positif 10,2%. Dominasi sentimen netral menandakan bahwa sebagian besar pengguna media sosial menyampaikan opini secara informatif tanpa kecenderungan emosi yang kuat.



Gambar 2 Diagram Lingkaran Proporsi Sentimen Publik

Hasil penelitian ini sejalan dengan temuan (Salim & Solichin, 2022) yang menyatakan bahwa algoritma Naive Bayes mampu memberikan performa yang baik dalam analisis sentimen teks berbahasa Indonesia. Selain itu, (Setyawan et al., 2023) juga membuktikan bahwa Multinomial Naive Bayes memiliki tingkat akurasi yang kompetitif dalam mengklasifikasikan opini publik terhadap kebijakan pemerintah dan cenderung lebih efisien dibandingkan algoritma Support Vector Machine (SVM) pada data teks berskala besar.

Penelitian lain oleh (Perdana et al., 2022; Prayudani et al., 2024) menunjukkan bahwa Multinomial Naive Bayes unggul dalam menangani data teks pendek seperti tweet karena mampu memanfaatkan frekuensi kemunculan kata secara optimal. Hal ini memperkuat hasil penelitian ini, mengingat karakteristik data yang digunakan juga berupa tweet dengan panjang teks yang relatif singkat.

Dari sisi kebijakan publik, hasil analisis memperlihatkan bahwa persepsi masyarakat sangat dipengaruhi oleh konteks kebijakan dan dampak langsung terhadap kehidupan sehari-hari. Kebijakan yang dinilai memberikan beban ekonomi, seperti kenaikan harga atau pengurangan subsidi, cenderung menimbulkan sentimen negatif. Sebaliknya, kebijakan dengan manfaat langsung bagi masyarakat seperti bantuan sosial dan layanan publik memunculkan lebih banyak sentimen positif.

Kebaruan penelitian ini terletak pada penerapan algoritma Multinomial Naive Bayes secara spesifik untuk analisis sentimen terhadap kebijakan pemerintah Indonesia terkini, dengan menggunakan data *real-time* dari media sosial. Pendekatan ini membedakan penelitian ini dari studi sebelumnya yang lebih bersifat umum terhadap isu sosial. Oleh karena itu, hasil penelitian ini dapat dijadikan dasar dalam pengembangan sistem pemantauan opini publik (*public sentiment monitoring system*) berbasis kecerdasan buatan yang membantu pemerintah dalam mengevaluasi efektivitas kebijakan publik secara adaptif dan berbasis data.

V. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa penerapan algoritma Multinomial Naive Bayes dalam analisis sentimen opini publik terhadap kebijakan pemerintah di media sosial mampu memberikan hasil klasifikasi yang akurat dan informatif. Model Multinomial Naive Bayes menunjukkan performa yang baik dalam mengelompokkan opini ke dalam kategori positif, negatif, dan netral dengan tingkat akurasi sebesar 89,7%. Hasil analisis mengungkapkan bahwa mayoritas opini publik bersifat netral, yang menunjukkan bahwa masyarakat cenderung bersikap rasional dalam menanggapi kebijakan pemerintah di media sosial. Penerapan metode ini terbukti efektif dalam mengidentifikasi persepsi masyarakat secara otomatis berbasis data teks, serta berpotensi digunakan sebagai alat pendukung dalam proses pengambilan keputusan pemerintah untuk meningkatkan kualitas kebijakan publik.

SUMBER PUSTAKA

- Aprilia, N. W. A. S., & Isnain, A. R. (2024). Analisis sentimen terhadap media sosial Twitter dengan kasus kampanye anti-korupsi di Indonesia menggunakan Naive Bayes. *Jurnal Media Informatika Budidharma*, 8(2), 695. <https://doi.org/10.30865/mib.v8i2.7582>
- Hidayat, T., Siddiq, M. J., Jayasri, S., Suhendi, A., & Rizky, R. (2025). ANALISIS SENTIMEN OPINI MASYARAKAT TERHADAP PILKADA 2024 DI MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(2). <https://doi.org/10.23960/jitet.v13i2.6280>
- Naraswati, N. P. G., Nooraeni, R., Rosmilda, D. C., Desinta, D., Khairi, F., & Damaiyanti, R. (2021). Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification. *SISTEMASI*, 10(1), 222. <https://doi.org/10.32520/stmsi.v10i1.1179>
- Perdana, A., Hermawan, A., & Avianto, D. (2022). Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Clasifier. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 11(2), 195–200. <https://doi.org/10.32736/sisfokom.v11i2.1412>
- Prayudani, S., Situmorang, D. R. B., Hidayah, R., & Ginting, H. S. (2024). Sentiment analysis of social media X in the 2024 Indonesian presidential election using the Naive Bayes algorithm: candidates' backgrounds and political promises. *Journal of Applied Informatics and Computing*, 8(2), 291–295. <https://doi.org/10.30871/jaic.v8i2.7580>
- Salim, E., & Solichin, A. (2022). Analisis sentimen pada media sosial Twitter terhadap pelayanan dinas kependudukan dan pencatatan sipil menggunakan algoritma Naive Bayes. *IDEALIS: Indonesia Journal Information System*, 5(2), 79–86. <https://doi.org/10.36080/idealism.v5i2.2961>
- Setyawan, H., Azizah, L. M., & Pradani, A. Y. (2023). Sentiment analysis of public responses on Indonesia government using Naive Bayes and Support Vector Machine. *Emerging Information Science and Technology*, 4(1), 1–7. <https://doi.org/10.18196/eist.v4i1.18681>
- Sulaiman, C., Salamah, U. G., Oktaviati, R., & Faizah, S. (2024). Sentiment analysis and classification of public opinion on Prabowo Subianto using Naive Bayes on Twitter. *Journal of Global Engineering Research and Science*, 3(2), 56–62. <https://doi.org/10.56904/j-gers.v3i2.98>
- Via Laurenz, & Fajar Ratnawati. (2025). Analisis Sentimen Komentar Netizen di Twitter terhadap Keputusan Mahkamah Konstitusi pada Hasil Pemilihan Presiden 2024 Menggunakan Algoritma Naive Bayes. *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer*, 5(2), 636–643. <https://doi.org/10.55606/juisik.v5i2.1491>
- Zharifa, A. H. A., & Ujianto, E. I. H. (2024). Analisis sentimen publik di Twitter pasca debat kelima Pilpres 2024 dengan Naive Bayes. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 754–763. <https://doi.org/10.29408/edumatic.v8i2.28048>