Jurnal Teknik Informatika dan Teknologi Informasi
Volume 6 Nomor 1 April 2026
E-ISSN .: 2827-9387, P-ISSN .: 2827-9379, Hal 116-129
DOI: https://doi.org/10.55606/jutiti.v6i1.6630
Tersedia: https://journalshub.org/index.php/jutiti

Analisis Sentimen Publik pada Media Sosial X terhadap Program Magang Pemerintah Menggunakan Algoritma *Naïve Bayes*

Cesia Trisani Saragih Garingging[1*], Sasmita Lumban Gaol[2], Sardo Pardingotan Sipayung[3]
[1-3]Teknik Informatika, Universitas Katolik Santo Thomas, Medan, Indonesia
Email: cesiasaragih.@gmail.com[1], sasmitalumbangaol468@gmail.com[2], pinsarsiphom@gmail.com[3]*

**Penulis Korespondensi: cesiasaragih@gmail.com*

Abstract. *Internship programs organized by the government are a strategic effort to enhance human resource competencies while providing employment opportunities for the younger generation. This study aims to examine public perceptions of these programs through sentiment classification on the social media platform X (formerly Twitter). The method employed is Naive Bayes, a probabilistic classification algorithm that is effective for text analysis. Data were collected from tweets containing keywords related to government internship programs within a specific time frame. After the data underwent cleaning and preprocessing, the model was trained and tested to classify sentiments into positive and negative categories. The classification results indicate that positive sentiment dominates with 52.3%, while negative sentiment accounts for 47.7%. This relatively narrow margin suggests that although the majority of users responded positively, a considerable portion expressed criticism or dissatisfaction with the programs. These findings serve as valuable input for the government to improve the quality and implementation of internship initiatives to better align with public expectations. Furthermore, this study demonstrates that the Naive Bayes method is capable of automatically identifying public sentiment using data from social media platform X. The conclusions drawn can be used as a reference for policymakers to enhance communication and execution of government internship programs in the future. It is also recommended that future research expand the data sources by including various other social media platforms to gain a more comprehensive view of public sentiment. This approach would enrich the analysis and contribute to more data-driven and responsive policy development.*

Keywords: *Internship Program; Naive Bayes; Public Sentiment; Social Media X; Text Analysis.*

Abstrak. Program magang yang diselenggarakan oleh pemerintah merupakan langkah strategis untuk meningkatkan kompetensi sumber daya manusia sekaligus membuka peluang kerja bagi generasi muda. Penelitian ini bertujuan mengkaji persepsi masyarakat terhadap program tersebut melalui klasifikasi sentiment di platform media social X (sebelumnya Twitter). Metode yang dipilih adalah Naïve Bayes, sebuah algoritma klasifikasi berbasis probabilitas yang efektif dalam analisis teks. Data dikumpulkan dari tweet yang memuat kata-kata terkait program magang pemerintah dalam rentang waktu tertentu. Setelah data dibersihkan dan diproses, model kemudian dilatih dan diuji untuk mengelompokkan sentimen ke dalam kategori positif dan negatif. Hasil pengklasifikasian menunjukkan bahwa sentimen positif mendominasi sebesar 52,3%, sementara sentimen negatif mencapai 47,7%. Selisih yang relatif tipis ini mengindikasikan bahwa walaupun mayoritas responden memberikan tanggapan positif, ada juga sejumlah besar yang menyatakan kritik atau ketidakpuasan terhadap program. Temuan ini menjadi masukan berharga bagi pemerintah guna meningkatkan kualitas dan implementasi program magang agar lebih sesuai dengan ekspektasi masyarakat. Selain itu, penelitian ini membuktikan bahwa metode Naive Bayes mampu secara otomatis mengidentifikasi sentimen publik berdasarkan data dari media sosial X. Kesimpulan yang diperoleh dapat menjadi referensi penting bagi pembuat kebijakan dalam memperbaiki komunikasi serta pelaksanaan program magang di masa depan. Disarankan pula untuk melakukan studi lanjutan dengan memperluas sumber data dari berbagai platform media sosial agar hasil analisis sentimen lebih menyeluruh.

Kata kunci: Analisis Teks; Media Sosial X; Naive Bayes; Program Magang; Sentimen Publik.

1. LATAR BELAKANG

Perkembangan media sosial telah mengubah pola komunikasi publik dengan menjadikannya sebagai ruang diskursus terbuka antara masyarakat dan pemerintah. Berbagai kebijakan dan program pemerintah kini tidak hanya diimplementasikan, tetapi juga dievaluasi

Naskah Masuk: 18 November 2025; Revisi: 21 Desember 2025; Diterima: 13 Januari 2026; Tersedia: 15 Januari 2026

secara langsung melalui opini yang diekspresikan pengguna di platform digital. Program magang pemerintah merupakan salah satu inisiatif yang memperoleh perhatian luas karena berorientasi pada peningkatan kompetensi dan kesiapan kerja generasi muda. Meskipun program ini memiliki peran strategis dalam penguatan kualitas sumber daya manusia, pemahaman terhadap respons dan sikap masyarakat terhadap pelaksanaannya masih belum terdokumentasi secara komprehensif, khususnya melalui analisis data media sosial yang merepresentasikan opini spontan publik.

Sejumlah penelitian terdahulu menunjukkan bahwa analisis sentimen berbasis pembelajaran mesin mampu mengungkap kecenderungan opini masyarakat terhadap kebijakan publik dengan memanfaatkan data teks tidak terstruktur dari media sosial. Namun, kajian yang secara khusus memetakan sentimen masyarakat terhadap program magang pemerintah secara agregatif, terutama dengan menggunakan data dari platform X setelah perubahan identitas platform, masih relatif terbatas. Kondisi ini menunjukkan adanya kebutuhan akan penelitian yang mengintegrasikan pendekatan klasifikasi sentimen dengan konteks kebijakan ketenagakerjaan pemerintah. Oleh karena itu, penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat terhadap program magang pemerintah menggunakan algoritma Naïve Bayes berbasis data dari platform X, sehingga dapat memberikan gambaran objektif mengenai persepsi publik sekaligus menjadi dasar evaluatif bagi perumusan kebijakan yang lebih adaptif dan berbasis data.

2. KAJIAN TEORITIS

Tinjauan Pustaka

Penelitian mengenai klasifikasi sentimen di media sosial telah mengalami perkembangan pesat dalam dekade terakhir. Studi oleh Siregar et al. (2020) menunjukkan bahwa algoritma Naive Bayes mampu memberikan akurasi yang kompetitif dalam mengklasifikasikan opini publik terhadap kebijakan pemerintah daerah, dengan tingkat akurasi mencapai 91%. Sementara itu, Pramudito dan Lestari (2022) membandingkan performa antara Naive Bayes dan Decision Tree dalam analisis sentimen kampanye vaksinasi COVID-19 di Twitter, di mana Naive Bayes menunjukkan keunggulan dalam kecepatan proses klasifikasi meskipun akurasi sedikit lebih rendah.

Dalam konteks layanan digital, Fauzi dan Hidayat (2021) menerapkan algoritma Naive Bayes untuk mengklasifikasikan ulasan pengguna aplikasi transportasi daring dan menemukan bahwa model tersebut efektif dalam membedakan opini positif dan negatif, terutama untuk teks pendek. Di sisi lain, penelitian oleh Arsyad (2023) yang menggunakan pendekatan text mining

terhadap tanggapan masyarakat mengenai program Kartu Prakerja menemukan bahwa media sosial X (dahulu Twitter) merupakan sumber yang kaya opini, meskipun menghadirkan tantangan dalam praproses data akibat struktur bahasa informal.

Meski banyak studi telah dilakukan, sebagian besar masih fokus pada satu kebijakan spesifik. Penelitian ini mengambil celah tersebut dengan mengkaji seluruh program magang pemerintah secara menyeluruh melalui pendekatan klasifikasi sentimen menggunakan algoritma Naive Bayes. Dengan pendekatan ini, diharapkan dapat memberikan gambaran yang lebih luas dan akurat tentang persepsi masyarakat terhadap kebijakan magang yang dikelola oleh pemerintah.

Media Sosial

Media sosial telah menjadi salah satu sumber utama untuk mengumpulkan data teks dalam berbagai penelitian karena sifatnya yang interaktif, real-time, dan merepresentasikan opini publik secara luas. Dalam konteks penelitian ini, media sosial digunakan sebagai media pengambilan data berbasis teks dari opini masyarakat terkait isu tertentu, dengan cakupan individu yang saling terhubung melalui internet. Platform ini memungkinkan distribusi informasi yang cepat dan terbuka, menjadikannya sumber data penting dalam kajian analisis sentimen dan perilaku digital.

X (Twitter)

Platform X (dulu dikenal sebagai Twitter) digunakan sebagai objek utama dalam pengumpulan data karena karakteristik uniknya, yaitu tweet dengan batas karakter 280 huruf yang mendorong penggunaan bahasa informal, singkatan, dan simbol khas media sosial. Tweet yang dikumpulkan melalui API mengandung ekspresi pengguna dalam bentuk teks pendek, dan sering digunakan sebagai data dalam penelitian sosial digital. Fitur-fitur seperti mention, retweet, hashtag, dan reply memberikan pelang untuk menelusuri percakapan publik secara tematik dan kronologis.

Naïve Bayes

Algoritma Naive Bayes digunakan dalam tahap klasifikasi sentimen karena memiliki kecepatan dan efisiensi tinggi saat diaplikasikan pada data teks berukuran besar. Metode ini bekerja berdasarkan Teorema Bayes, dengan asumsi bahwa fitur (kata) saling bebas secara kondisional. Model menghitung probabilitas posterior $P(Ci|X)$ dari setiap kelas Ci terhadap data masukan X, dan hasil akhir ditentukan berdasarkan nilai probabilitas maksimum. Implementasi dilakukan menggunakan pembagian dataset menjadi data latih dan data uji dengan rasio tertentu, serta evaluasi akurasi dilakukan untuk menilai performa model.

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)}$$

Dengan:

X: data masukan yang akan diklasifikasikan

Ci: kelas ke-i

P(Ci|X): probabilitas posterior

P(X|Ci): probabilitas kemunculan data X jika diketahui kelasnya

P(Ci): probabilitas awal kelas

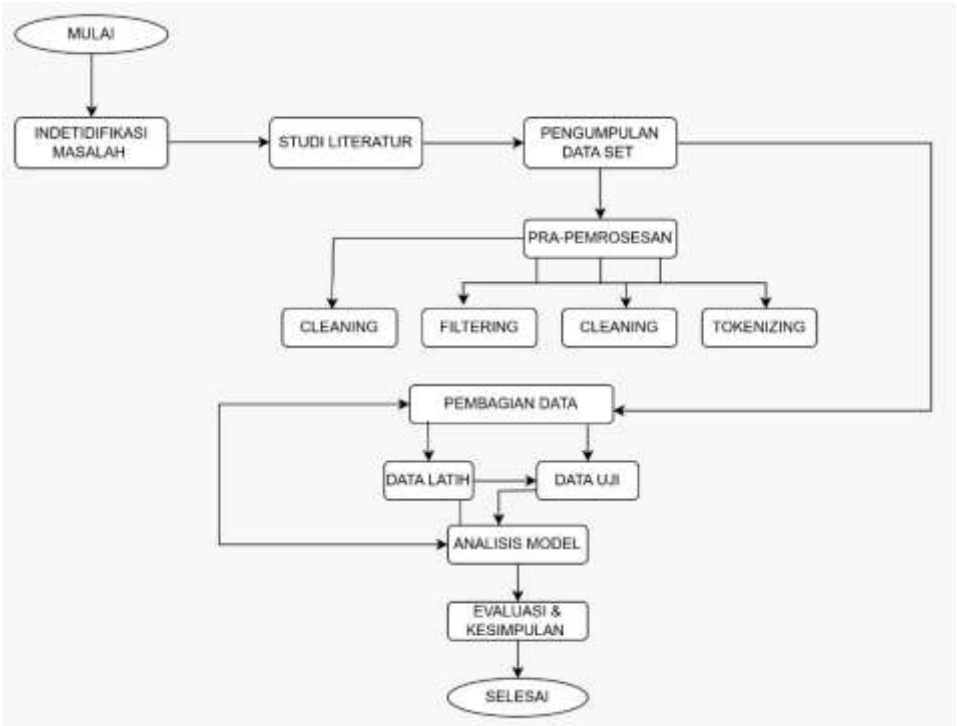
P(X): probabilitas total data

Asumsi independensi antar fitur dirumuskan sebagai:

$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i)$

Model ini memiliki keunggulan dalam kesederhanaan komputasi dan hasil yang kompetitif meskipun menggunakan pendekatan probabilistik dasar.

3. METODE PENELITIAN



Gambar 1. Tahapan penelitian.

Bagian ini memuat rancangan penelitian meliputi disain penelitian, populasi/ sampel penelitian, teknik dan instrumen pengumpulan data, alat analisis data, dan model penelitian yang digunakan. Metode yang sudah umum tidak perlu dituliskan secara rinci, tetapi cukup merujuk ke referensi acuan (misalnya: rumus uji-F, uji-t, dll). Pengujian validitas dan reliabilitas instrumen penelitian tidak perlu dituliskan secara rinci, tetapi cukup dengan mengungkapkan hasil pengujian dan interpretasinya. Keterangan simbol pada model dituliskan dalam kalimat.

Berikut ini disusun urutan prosedur yang digunakan dalam penelitian, mulai dari identifikasi masalah hingga tahap prediksi:

Identifikasi Masalah

Penelitian ini diawali dengan identifikasi masalah utama terkait persepsi dan sentimen masyarakat terhadap isu tertentu yang tersebar di media sosial. Fokus diarahkan pada pemahaman bagaimana opini publik terbentuk dan berubah dalam platform X (Twitter), serta bagaimana data teks tersebut dapat dimanfaatkan untuk analisis sentimen secara efektif. Identifikasi masalah ini menjadi dasar perumusan tujuan penelitian dan penentuan metode pengolahan data.

Studi Pustaka

Sebagai landasan teoritis, dilakukan kajian literatur mendalam mengenai penggunaan media sosial sebagai sumber data, teknik text mining dalam analisis teks, serta algoritma klasifikasi yang relevan seperti Naive Bayes. Studi pustaka bertujuan untuk memahami perkembangan metode yang telah ada, mengidentifikasi gap penelitian, dan menyusun kerangka kerja yang sesuai untuk mendukung proses ekstraksi informasi dari data teks tidak terstruktur pada media sosial.

Pengumpulan Data

Data dikumpulkan dari platform X menggunakan Application Programming Interface (API) resmi, yang memungkinkan pengambilan tweet secara otomatis sesuai dengan kriteria kata kunci dan rentang waktu tertentu. Data yang diperoleh berbentuk teks singkat yang merepresentasikan opini pengguna, yang kemudian disimpan dalam format terstruktur untuk diproses lebih lanjut. Pengambilan data dilakukan secara bertahap dan disesuaikan dengan kebutuhan analisis.

Pra-pemrosesan Data

Data mentah yang diperoleh mengalami serangkaian tahapan pra-pemrosesan untuk memastikan kualitas dan relevansi informasi. Tahapan ini meliputi normalisasi teks (case folding), pemisahan kata (tokenisasi), penghilangan kata umum yang tidak bermakna

(stopword removal), dan reduksi kata ke bentuk dasarnya (stemming). Proses ini penting untuk meminimalisir noise dan meningkatkan akurasi ekstraksi fitur dari teks.

Pembagian Dataset

Setelah pra-pemrosesan, dataset dibagi menjadi dua subset utama, yakni data pelatihan dan data pengujian, dengan rasio pembagian yang telah ditentukan (misalnya 80:20 atau 70:30). Pembagian ini dilakukan secara acak untuk menjaga representasi yang seimbang dan menghindari bias selama proses pelatihan model klasifikasi serta evaluasi performanya.

Analisis Data

Tahap analisis melibatkan ekstraksi fitur menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) yang mampu mengidentifikasi bobot kata penting dalam kumpulan dokumen. Selanjutnya, data yang telah diolah ini digunakan untuk membangun model klasifikasi dengan algoritma Naive Bayes. Evaluasi model dilakukan dengan mengukur metrik kinerja seperti akurasi, presisi, recall, dan F1-score, guna memastikan kemampuan prediksi yang memadai.

Prediksi

Model yang telah dilatih digunakan untuk melakukan prediksi sentimen pada data uji. Dengan menghitung probabilitas posterior berdasarkan fitur yang ada, model menentukan kelas sentimen (positif, negatif, atau netral) untuk setiap input teks. Hasil prediksi dianalisis untuk memberikan gambaran pola sentimen masyarakat dan potensi dinamika opini di media sosial yang relevan dengan isu penelitian.

4. HASIL DAN PEMBAHASAN

Pengumpulan Data

Proses akuisisi data dalam penelitian ini dilakukan melalui teknik web scraping yang diterapkan pada platform media sosial X (sebelumnya dikenal sebagai Twitter). Penarikan data dilakukan dengan memanfaatkan bahasa pemrograman Python dan ekosistem Node.js, menggunakan pustaka tweet-harvest yang dioperasikan di lingkungan Google Colaboratory (Google Colab).

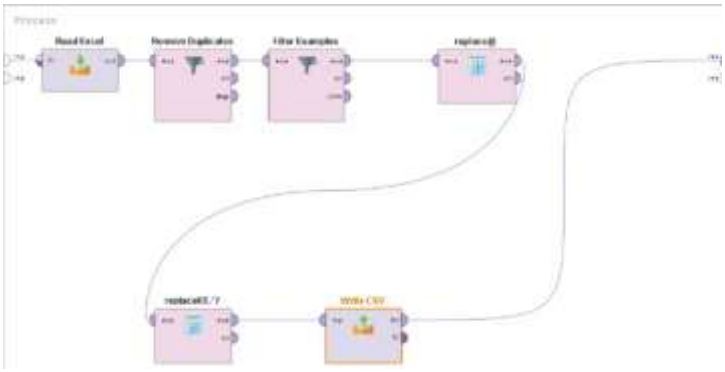
Adapun parameter pencarian yang digunakan meliputi kata kunci "program magang pemerintah" tanpa pembatasan tanggal spesifik. Pendekatan ini memungkinkan pengambilan data secara luas untuk menangkap opini publik yang lebih representatif terhadap topik tersebut. Dalam implementasinya, skrip yang digunakan ditulis dengan sintaks sebagai berikut:

```
# Crawl Data

filename = 'magangpemerintah.csv'
search_keyword = 'program magang pemerintah lang:id'
limit = 1000

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_k
```

Skrip ini berfungsi untuk mengekstraksi maksimal 1000 tweet yang mengandung frasa "program magang pemerintah" dalam bahasa Indonesia. Token API digunakan untuk mengautentikasi permintaan data melalui Twitter API v2. Proses ini menghasilkan dataset mentah dalam format CSV yang selanjutnya diproses untuk tahap pra-pemrosesan dan analisis sentimen.

Preprocesing Data

Sebelum data dianalisis lebih lanjut, langkah awal yang dilakukan adalah proses preprocessing, yaitu tahap pembersihan dan penyaringan data agar lebih layak untuk dianalisis. Prosedur ini sangat penting dalam analisis berbasis teks, karena data mentah dari media sosial umumnya mengandung banyak elemen yang tidak relevan, seperti duplikasi, simbol, atau entri yang tidak memiliki nilai analitis. Seluruh proses dilakukan menggunakan platform RapidMiner, dengan rancangan sebagai berikut:



Gambar 2. Desain Proses Preprocessing pada RapidMiner.

Proses dimulai dengan mengimpor data mentah melalui operator Read Excel, yang berfungsi untuk membaca file data dari sumber eksternal. Selanjutnya, operator Remove Duplicates digunakan untuk menghapus data yang berulang agar tidak mengganggu hasil analisis. Tahap ini penting untuk menjamin bahwa setiap opini atau komentar dianalisis hanya sekali.

Kemudian, data difilter menggunakan Filter Examples untuk menyeleksi hanya entri-entri yang memenuhi kriteria tertentu, seperti keberadaan teks dan label yang valid. Setelah itu, dilakukan proses standarisasi teks dengan operator Replace, yang bertujuan mengganti atau menghapus karakter-karakter tertentu, seperti tanda baca atau simbol yang tidak diperlukan.

Operator Replace (RTL?.?) melanjutkan proses pembersihan lanjutan terhadap entri yang masih mengandung karakter khusus atau pola tertentu yang dapat mengganggu proses pembobotan kata nantinya. Setelah seluruh tahap pembersihan selesai, data hasil preprocessing disimpan menggunakan Write CSV untuk keperluan analisis selanjutnya.

Dari total 1.000 data awal yang dikumpulkan dari media sosial X, hanya 944 data yang lolos tahap preprocessing. Hal ini menunjukkan bahwa sebanyak 56 data diidentifikasi sebagai duplikat atau tidak memenuhi kriteria dan karenanya dihapus dalam proses pembersihan. Tahap ini memastikan bahwa data yang dianalisis memiliki kualitas dan relevansi tinggi terhadap topik yang diteliti.

Pembagian Data

Setelah melalui tahapan pembersihan data, seperti penghapusan kata yang tidak relevan, penghilangan entri kosong, serta eliminasi data ganda, langkah berikutnya adalah memisahkan data menjadi dua bagian utama. Dalam studi ini, data yang telah bersih dibagi menjadi dua subset, yaitu 80% dialokasikan sebagai data latih dan 20% sebagai data uji. Pembagian ini dilakukan untuk memastikan bahwa proses pembelajaran dan pengujian dapat berlangsung secara seimbang dengan variasi data yang cukup.

Pada data latih yang telah diperoleh, peneliti melakukan pemberian label secara manual berdasarkan interpretasi terhadap isi setiap tweet. Setiap entri diklasifikasikan ke dalam dua kelompok sentimen, yaitu positif dan negatif, sesuai dengan makna yang terkandung dalam teks. Pemberian label ini sangat penting karena digunakan sebagai dasar pembelajaran algoritma berbasis supervised learning. Sementara itu, data uji dibiarkan tanpa label dan berfungsi sebagai sampel yang terpisah, untuk menguji kemampuan model dalam mengenali pola sentimen dari data yang tidak dikenali sebelumnya, apabila proses evaluasi dilakukan.

Implementasi Naive Bayes untuk Klasifikasi Sentimen

Data yang telah terkumpul dari media sosial kemudian dipisahkan menjadi dua kelompok utama. Sebanyak 756 data digunakan sebagai dasar untuk membentuk model pembelajaran, sedangkan 188 data lainnya berperan sebagai bahan evaluasi terhadap kemampuan prediksi dari model yang telah dibangun. Proses pemodelan ini dilakukan menggunakan perangkat lunak RapidMiner, yang memungkinkan perancangan alur analitik secara visual dan sistematis.

Langkah pertama diawali dengan mengimpor berkas data berformat CSV ke dalam lingkungan RapidMiner. Setelah itu, data yang sudah diberi label sentimen difilter melalui komponen "Filter Examples" agar hanya data relevan yang diteruskan ke tahap berikutnya. Atribut yang bersifat nominal kemudian dikonversi ke bentuk teks melalui proses transformasi "Nominal to Text", agar kompatibel dengan metode pengolahan dokumen.

Tahapan selanjutnya adalah proses pembentukan representasi numerik dari data teks menggunakan metode TF-IDF. Ini dilakukan melalui modul "Process Document from Data", yang secara otomatis mengolah dokumen teks mentah menjadi vektor angka berdasarkan frekuensi dan distribusi kata. Proses ini juga mencakup normalisasi teks, penghilangan kata umum (stopword removal), dan tokenisasi kata.

Rangkaian tahapan ini divisualisasikan melalui diagram desain proses di RapidMiner, seperti diperlihatkan pada gambar berikut:

Gambar 3. Process Documents from Data.

Hasil dari proses transformasi tersebut digunakan untuk melatih model klasifikasi menggunakan algoritma Naïve Bayes. Model ini secara otomatis mempelajari pola distribusi kata dalam setiap kategori sentimen, lalu membentuk suatu representasi matematis untuk keperluan prediksi. Baik model prediksi maupun dataset hasil pelatihan kemudian disimpan sebagai komponen yang siap digunakan untuk proses klasifikasi data uji berikutnya.

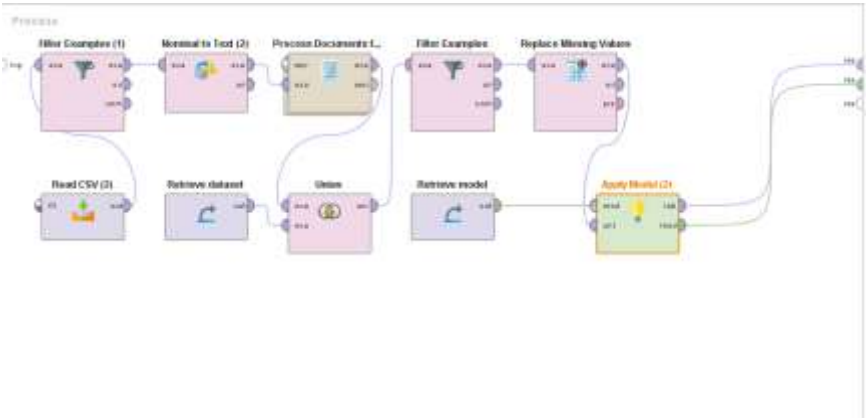
Rangkaian tahapan ini divisualisasikan melalui diagram desain proses di RapidMiner, seperti diperlihatkan pada gambar berikut:

Gambar 4. Rangkaian Proses Klasifikasi Sentimen Menggunakan Naïve Bayes di RapidMiner.

Keseluruhan proses ini bertujuan untuk mengubah data teks tidak terstruktur menjadi informasi prediktif yang bermanfaat, dalam konteks ini adalah opini publik terhadap program magang pemerintah. Dengan dukungan antarmuka visual yang disediakan oleh RapidMiner, analisis sentimen dapat dilakukan secara efisien dan akurat tanpa perlu menulis kode secara eksplisit.

Prediksi Sentimen terhadap Program Magang Pemerintah

Setelah proses pelatihan dilakukan menggunakan algoritma Naïve Bayes, langkah berikutnya adalah menerapkan model yang telah dibangun tersebut untuk mengklasifikasikan data uji yang belum memiliki label sentimen. Proses ini dilakukan menggunakan RapidMiner dengan alur desain sistem yang terlihat pada Gambar 3.



Gambar 5. Proses Prediksi Sentimen Menggunakan Algoritma Naive Bayes.

Model klasifikasi dibentuk berdasarkan 756 data pelatihan, dan digunakan untuk mengevaluasi 188 data uji yang diperoleh dari platform media sosial X (sebelumnya dikenal sebagai Twitter). Tujuan utama dari tahap ini adalah untuk mengidentifikasi kecenderungan opini publik terhadap pelaksanaan program magang yang diselenggarakan oleh pemerintah.

Hasil dari proses klasifikasi menunjukkan bahwa sebesar 52,3% dari total unggahan yang dianalisis memiliki kecenderungan sentimen positif terhadap program tersebut. Sementara itu, 47,7% lainnya diklasifikasikan sebagai sentimen negatif.

Tabel 1. Hasil Klasifikasi Sentimen Program Magang Pemerintah.

Sentimen	Jumlah Komentar	Persentase
Positif	108	52,3%
Negatif	80	47,7%

Temuan ini mengindikasikan bahwa tanggapan masyarakat dalam ruang digital terhadap kebijakan magang pemerintah cenderung terbagi hampir merata, dengan sentimen positif sedikit lebih dominan. Hal ini mencerminkan bahwa program tersebut mendapatkan respons yang beragam dari publik, baik dalam bentuk dukungan maupun kritik.

Pembahasan

Pengumpulan data dilakukan dari platform media sosial X (sebelumnya dikenal sebagai Twitter), dengan total 1.000 komentar yang berhasil diperoleh melalui teknik web scraping. Data yang terkumpul ini kemudian melalui proses preprocessing untuk menghilangkan unsur-unsur yang tidak relevan, seperti tanda baca, simbol, tautan, serta kata-kata kosong (stopwords) yang tidak memiliki nilai signifikan dalam analisis sentimen. Tahap ini juga mencakup normalisasi teks dan tokenisasi, yakni pemisahan kalimat menjadi satuan kata yang dapat diolah lebih lanjut.

Setelah proses pembersihan dilakukan, jumlah data berkurang menjadi 944 komentar yang layak untuk dianalisis. Selanjutnya, data ini dibagi ke dalam dua kelompok, yakni data latih dan data uji. Sebanyak 756 komentar digunakan sebagai data latih yang telah dilabeli berdasarkan kategori sentimen positif dan negatif. Data latih ini menjadi dasar dalam membangun model klasifikasi menggunakan algoritma Naïve Bayes.

Model yang telah dilatih kemudian diterapkan pada 188 komentar sisanya yang berperan sebagai data uji. Tujuan dari pengujian ini adalah untuk mengetahui efektivitas model dalam mengklasifikasikan sentimen dari komentar yang belum diberi label sebelumnya.

Hasil prediksi menunjukkan bahwa dari 188 komentar, sebanyak 108 di antaranya dikategorikan sebagai sentimen positif, yang mencerminkan tanggapan masyarakat yang cenderung mendukung atau menyambut baik program magang pemerintah. Sementara itu, 80 komentar lainnya mengandung sentimen negatif, menunjukkan adanya kritik atau ketidakpuasan terhadap kebijakan tersebut. Secara proporsional, komentar positif mencakup 52,3% dari total data uji, sementara komentar negatif berjumlah 47,7%.

Distribusi ini mengindikasikan bahwa persepsi publik terhadap program magang pemerintah terbilang cukup berimbang, meskipun terdapat kecenderungan dominan pada sisi positif. Dengan demikian, model Naïve Bayes mampu memberikan klasifikasi awal yang berguna dalam memahami dinamika opini publik di media sosial terhadap kebijakan yang tengah berjalan.

5. KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini mengkaji persepsi publik terhadap program magang pemerintah melalui analisis sentimen di media sosial X dengan menerapkan algoritma Naïve Bayes. Dari 1.000 data yang dikumpulkan, sebanyak 944 data valid diperoleh setelah melalui tahap preprocessing. Data tersebut kemudian dibagi menjadi data latih sebanyak 756 entri dan data uji sebanyak 188 entri. Proses klasifikasi sentimen pada data uji menunjukkan bahwa 52,3% komentar bersifat positif dan 47,7% bersentimen negatif. Hasil ini menunjukkan bahwa secara umum respons masyarakat terhadap program magang pemerintah cenderung positif, meskipun masih terdapat proporsi yang cukup besar dari opini negatif. Temuan ini mencerminkan adanya ketertarikan sekaligus kekhawatiran publik terhadap kebijakan tersebut, yang perlu diperhatikan oleh pemangku kebijakan untuk evaluasi dan perbaikan program ke depannya. Model Naïve Bayes terbukti dapat digunakan secara efektif dalam melakukan klasifikasi sentimen terhadap teks berbahasa Indonesia dalam konteks kebijakan publik.

Saran

Berdasarkan hasil yang diperoleh, disarankan agar instansi pemerintah memperhatikan opini negatif yang muncul sebagai bahan evaluasi program magang, khususnya terkait aspek pelaksanaan dan manfaat nyata bagi peserta. Penelitian selanjutnya dapat memperluas cakupan data, baik dari segi jumlah maupun platform media sosial lain untuk memperoleh representasi opini publik yang lebih komprehensif.

Selain itu, disarankan pula untuk melakukan perbandingan performa antara beberapa algoritma klasifikasi, seperti Support Vector Machine, Random Forest, atau Logistic Regression, guna meningkatkan akurasi dan reliabilitas dalam analisis sentimen. Pendekatan berbasis deep learning juga dapat menjadi alternatif yang menjanjikan dalam pengolahan data besar dan kompleks pada masa mendatang.

DAFTAR REFERENSI

Aldy, M. D., & Nasution, M. I. P. (2023). Implementasi big data di media sosial sebagai strategi komunikasi krisis pemerintah. Jurnal Sistem Informasi dan Ilmu Komputer, 1(3), 73–87. https://doi.org/10.59581/jusiik-widyakarya.v1i3.907

Anugerah, S. M., Wijaya, R., & Bijaksana, M. A. (2024). Sentiment analysis social media for disaster using Naïve Bayes and IndoBERT. INTEK: Jurnal Penelitian, 11(1), 51–58. https://doi.org/10.31963/intek.v11i1.4771

Artanto, F. A. (2024). Analisis sentimen opini publik terhadap fenomena bunuh diri mahasiswa di Twitter menggunakan algoritma Naive Bayes. SATESI: Jurnal Sains Teknologi dan Sistem Informasi, 4(1), 70–77. https://doi.org/10.54259/satesi.v4i1.2908

Fadli, M., Triayudi, A., & Handayani, E. T. E. (2024). Sentiment analysis on Twitter social media application on fuel oil price hike using Naïve Bayes and decision tree algorithms. SaNa: Journal of Blockchain, NFTs and Metaverse Technology, 2(2), 114–122. https://doi.org/10.58905/sana.v2i2.271

Hariguna, T., & Rachmawati, V. (2019). Community opinion sentiment analysis on social media using Naive Bayes algorithm methods. International Journal of Informatics and Information System, 2(1), 33–38. https://doi.org/10.47738/ijiis.v2i1.11

Harun, A., & Ananda, D. P. (2021). Analisis sentimen opini publik tentang vaksinasi Covid-19 di Indonesia menggunakan Naïve Bayes dan decision tree. MALCOM, 1(1), 58–64. https://doi.org/10.57152/malcom.v1i1.63

Kiu, E. R. T., Sembiring, O. B., & Ngafidin, K. N. M. (2023). Twitter sentiment analysis using the Naive Bayes algorithm in the case of the Indosurya savings and loans cooperative. CESS (Journal of Computer Engineering, System and Science), 8(2), 294–306. https://doi.org/10.24114/cess.v8i2.45527

Kristiyanti, D. A., & Hardani, S. (2023). Sentiment analysis of public acceptance of COVID-19 vaccine types in Indonesia using Naïve Bayes, support vector machine, and long short-term memory (LSTM). Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 7(3), 722–732. https://doi.org/10.29207/resti.v7i3.4737

Likhith, S. R., Ahuja, P., Prathibha, B. N., & Shankari, B. U. (2024). Sentiment analysis of COVID-19 Twitter data using machine learning. In N. R. Shetty, N. H. Prasad, & N. Nalini (Eds.), Advances in computing and information (ERCICA 2023) (Lecture Notes in Electrical Engineering, Vol. 1104). Springer. https://doi.org/10.1007/978-981-99-7622-5_13

Lisdiantini, N., Azis, A., Syafitri, E. M., & Thousani, H. F. (2022). Analisis efektivitas program magang untuk sinkronisasi link and match perguruan tinggi dengan dunia industri. Ecobis, 9(2). https://doi.org/10.36987/ecobi.v9i2.2491

Muzaki, A., & Witanti, A. (2021). Sentiment analysis of the community on Twitter toward the 2020 election during the COVID-19 pandemic using Naïve Bayes classifier. Jurnal Teknik Informatika (JUTIF), 2(2), 101–107. https://doi.org/10.20884/1.jutif.2021.2.2.51

Naraswati, N. P. G. (2021). Analisis sentimen publik dari Twitter tentang kebijakan penanganan COVID-19 di Indonesia dengan Naive Bayes classification. Sistemasi: Jurnal Sistem Informasi, 10(1), 222–238. https://doi.org/10.32520/stmsi.v10i1.1179

Novianti, E. W., & Wibowo, W. (2022). Analisis sentimen pengguna Twitter terhadap program Kartu Prakerja di tengah pandemi COVID-19 menggunakan metode Naïve Bayes classifier. Jurnal Sains dan Seni ITS, 11(1). https://doi.org/10.12962/j23373520.v11i1.63552

Puspita Sari, W., & Soegiarto, A. (2021). Indonesian government public relations in using social media. ICHELSS, 1(1), 495–508.

Ramadhani, S. H., & Wahyudin, M. I. (2022). Analisis sentimen terhadap vaksinasi AstraZeneca pada Twitter menggunakan metode Naïve Bayes dan K-NN. Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi), 6(4), 526–534. https://doi.org/10.35870/jtik.v6i4.530

Sari, Y. K., Rozi, F., Muhyiddin, S., & Sukmana, F. (2024). Sentiment analysis of public opinion on application X (Twitter) in Indonesia against ChatGPT using Naïve Bayes algorithm. JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), 9(4), 2473–2484. https://doi.org/10.29100/jipi.v9i4.7052

Subekti, I. H., Habibi, M., Murdiyanto, A. W., & Jannah, A. R. (2023). Analisis sentimen di media sosial Twitter dengan studi kasus Kartu Prakerja. Teknomatika, 14(2), 49–58. https://doi.org/10.30989/teknomatika.v14i2.1101

Susanto, A., Maula, M. A., Utomo, I., Mulyono, W., & Sarker, K. (2021). Sentiment analysis on Indonesia Twitter data using Naïve Bayes and K-Means method. Jurnal Aplikasi Ilmu Sosial, 6(1). https://doi.org/10.33633/jais.v6i1.4465

Syakir, A., & Hasan, F. N. (2023). Analisis sentimen masyarakat terhadap perilaku korupsi pejabat pemerintah berdasarkan tweet menggunakan Naive Bayes classifier. Jurnal Media Informatika Budidarma, 7(4), 1796. https://doi.org/10.30865/mib.v7i4.6648

Zulfikar, W. B., Atmadja, A. R., & Pratama, S. F. (2023). Sentiment analysis on social media against public policy using multinomial Naive Bayes. Scientific Journal of Informatics, 10(1), 25–34. https://doi.org/10.15294/sji.v10i1.39952