Journal of Mathematics UNP
Vol. 11, No. 1, March 2026, pp. 18~28
ISSN: 2355-1658

Analisis Sentimen terhadap Program Makan Bergizi Gratis pada Media Sosial X Menggunakan Metode *Naive Bayes Classifier*

Vivi Delfitri Chandra, Meira Parma Dewi
Departemen Matematika, Universitas Negeri Padang

Article Info

Article history:

Received August, 19,2025
Revised November 18,2025
Accepted December 03,2025

Keywords:

Sentiment Analysis
X Social Media
Free Nutritious Meal Program
Naïve Bayes Classifier
SMOTE

Kata Kunci:

Analisis Sentimen
Media Sosial X
Program Makan Bergizi Gratis
Naïve Bayes Classifier
SMOTE

ABSTRACT

The rapid development of information and communication technology has positioned social media as a major platform for expressing public opinions on government policies, including the Free Nutritious Meal Program (MBG) launched in January 2025. This study aimed to analyze public sentiment toward the program using the Naïve Bayes Classifier. Secondary data were obtained from user posts on platform X through a data crawling technique and processed using cleaning, case folding, tokenizing, stopword removal, and stemming. Sentiment classification was conducted into positive, negative, and neutral categories, and the Synthetic Minority Over-sampling Technique (SMOTE) was applied to address data imbalance. Model performance was evaluated using accuracy, precision, recall, and F1-score. The results showed 407 positive comments, 383 negative comments, and 304 neutral comments. The best model with parameter α=0.3 achieved an accuracy of 83.76%. Overall, public sentiment toward the MBG Program tended to be positive, although continuous evaluation remained necessary.

ABSTRAK

Perkembangan teknologi informasi dan komunikasi sudah membuat media sosial berperan sebagai sarana utama khalayak untuk memberi opini terhadap kebijakan publik, termasuk Program Makan Bergizi Gratis (MBG) yang dilakukan peluncuran pada Januari 2025. Riset ini bertujuan guna menganalisis sentimen khalayak terhadap Program MBG memakai metode Naïve Bayes Classifier. Data sekunder berupa unggahan pengguna di platform X diperoleh melalui teknik *data crawling* dan pemrosesan lewat tahap *cleaning*, *case folding*, *tokenizing*, *stopword removal*, serta *stemming*. Selanjutnya, klasifikasi sentimen dilakukan ke dalam kategori positif, negatif, serta netral, melalui penerapan metode Synthetic Minority Over-sampling Technique (SMOTE) guna menyelesaikan permasalahan ketidakseimbangan data. Pengevaluasian model memakai metrik accuracy, precision, recall, serta F1-score. Hasil riset membuktikan 407 komentar positif, 383 komentar negatif, dan 304 komentar netral. Model terbaik dengan parameter α=0,3 menghasilkan akurasi 83,76%. Secara keseluruhan, sentimen masyarakat terhadap Program MBG cenderung positif, namun evaluasi berkelanjutan tetap diperlukan.

Penulis Korespondensi:

Vivi Delfitri Chandra

Departemen Matematika, Universitas Negeri Padang,
Email: vividelfitrichandra@gmail.com

1. PENDAHULUAN

Pada zaman sekarang khalayak dihadapkan adanya perkembangan teknologi informasi serta komunikasi yang cukup pesat, ditemukannya perihal tersebut memberi dorongan pada khalayak guna memakai sosial media dengan cara yang massif [1]. Perkembangan teknologi informasi dan komunikasi telah mendorong masyarakat untuk menggunakan media sosial secara masif sebagai sarana memberikan opini serta pandangan terhadap isu-isu publik. Media sosial seperti X (dahulu Twitter) memainkan peran penting dalam membentuk opini masyarakat secara real-time, terutama terkait kebijakan pemerintah yang bersifat kontroversial maupun strategis[1]. Salah satu kebijakan yang mencuat adalah Program Makan Bergizi Gratis (MBG) yang mulai diterapkan sejak Januari 2025 oleh pemerintah sebagai bagian dari visi Indonesia Emas 2045. MBG ditujukan untuk mendukung pertumbuhan generasi muda yang sehat dan cerdas, mencakup 82,9 juta sasaran penerima manfaat, seperti anak sekolah, balita, dan ibu hamil [2]. Kebijakan MBG harapannya tidak hanya sekedar mengurangi angka malnutrisi serta stunting anak-anak, namun juga sanggup memastikan kecukupan gizi untuk anak-anak selaras terhadap standar Angka Kecukupan Gizi [2].

Pemerintah mempunyai peranan yang krusial pada penyelenggaraan aturan MBG yang mana kebijakan strategis yang dilakukan pengambilan oleh pemerintah mulai dari perencanaan hingga penganggaran hendak menetapkan kesuksesan dari kebijakan MBG [3]. Dukungan komponen khalayak juga berperan sebagai bagian yang wajib dicermati serta berperan sebagai aspek yang menentukan dalam pelaksanaan MBG [4]. Meskipun demikian, pelaksanaan program ini menuai beragam respons publik yang tersebar di media sosial, mulai dari dukungan hingga kritik. Dengan demikian, pemetaan persepsi masyarakat terhadap MBG menjadi penting sebagai dasar evaluasi kebijakan publik. Pelaksanaan program MBG menimbulkan beragam tanggapan dari masyarakat.

Platform X (sebelumnya Twitter) berperan sebagai bagian dari media yang banyak dipakai guna menyampaikan sentimen tersebut, baik dalam bentuk dukungan maupun kritik. Twitter (X) memiliki keunggulan yakni jangkauan yang luas, mampu memperoleh jangkauan publik figur, media promosi lebih luas, banyak jaringan, serta lebih mudah dilakukan pengukuran terkait potensinya. Di bawah ini ialah fitur yang ada pada twitter yaitu Trending topic, Haastag, Retweet, dan Following[5]. Oleh karena itu, analisis terhadap opini masyarakat di media sosial menjadi penting sebagai bahan evaluasi terhadap implementasi kebijakan publik. Dalam konteks ini, analisis sentimen menjadi alat penting untuk mengkaji opini masyarakat melalui data teks. Analisis sentimen pada umumnya dipakai sebab peningkatan kebutuhan individu ataupun kelompok guna mengerti pendapat individu terhadap suatu hal. Analisis sentimen juga diberi pengaruh melalui data set yang dipakai hendak merasakan penanganan yang tidak sama [6]. Analisis sentimen juga diketahui sebagai opinion mining yang menjadi tahapan dalam menetapkan emosional pengguna melalui teknik analisis tulisan. Tak sekedar tokoh, analisis sentimen juga mampu menyelesaikan permasalahan terkait bisnis, program, produk, aplikasi, serta lainnya yang mampu diberi komentar oleh publik[7]. Hasil dilaksanakannya analisis sentimen juga mampu menjadi sebuah gambaran

bagi perusahaan, public figure, dan pemerintahan untuk menentukan langkah selanjutnya[8]. Sentimen dapat diklasifikasikan ke dalam kategori positif, negatif, dan netral dengan memanfaatkan algoritma pembelajaran mesin seperti Naïve Bayes Classifier[9]. Naïve Bayes termasuk supervised learning klasifikasi seperti contoh lainnya yaitu Support Vector Machine (SVM), K-Nearst Neighbor (KNN), Artificial Neural Network (ANN), Trees Gradient Boosted (TGB), dan Random Trees (RT) sedangkan regresi seperti Decision Tree, Logistic Regresion, dan Kernel Regresion[10]. Algoritma ini memiliki keunggulan dalam klasifikasi teks berbasis probabilistik, terutama pada data berbahasa Indonesia yang kompleks secara morfologis. Selain itu, ketidakseimbangan kelas dalam data sentimen yang dikumpulkan dari media sosial seringkali menjadi tantangan dalam membangun model klasifikasi yang akurat.

Untuk mengatasi hal tersebut, pendekatan Synthetic Minority Oversampling Technique (SMOTE) digunakan sebagai strategi augmentasi data sintetis untuk meningkatkan performa model klasifikasi[11]. Penelitian ini menggabungkan kedua pendekatan tersebut guna melakukan analisis sentimen khalayak pada Program MBG dalam media sosial X dengan tujuan memberikan informasi yang memiliki kerelevanan serta mampu digunakan untuk mengambil kebijakan. Selain itu pada penelitian sebelumnya penggunaan data yang dihimpun sebelum Program MBG beroperasi, sehingga tidak menangkap dinamika perubahan sentimen publik setelah implementasinya. Untuk mengatasi gap tersebut, penelitian ini menggunakan data pasca-implementasi sehingga mampu memberikan gambaran yang lebih komprehensif mengenai respons masyarakat yang sesungguhnya

2. METODE

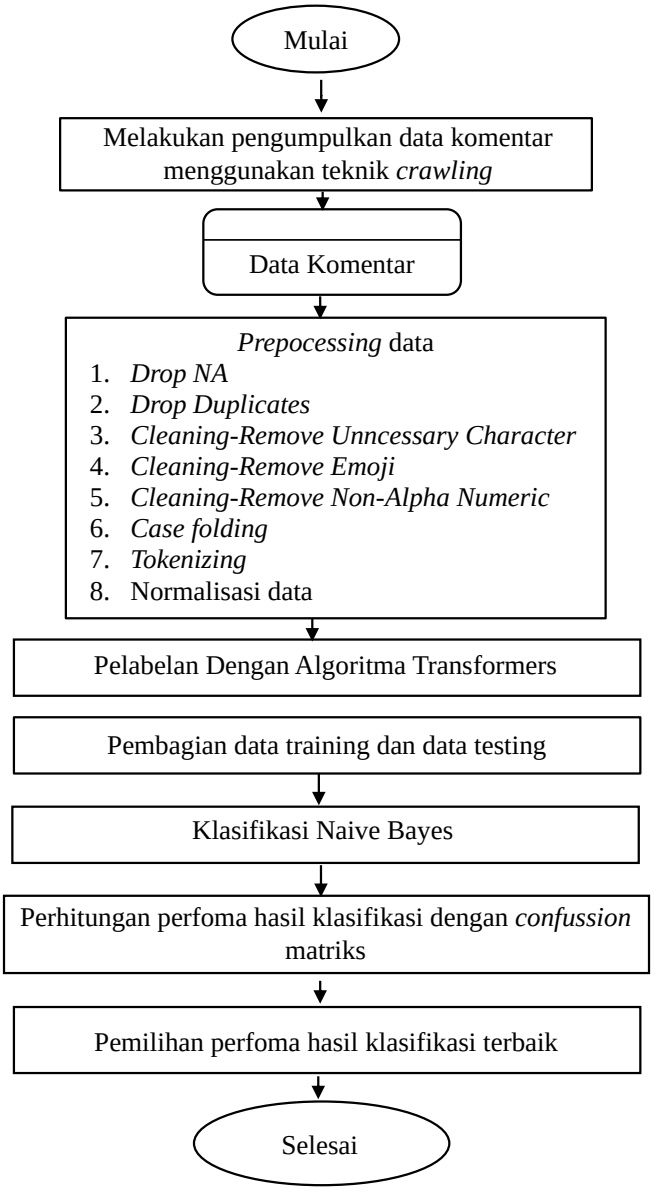
Jenis riset ini ialah riset terapan (*applied research*). Jenis data yang dipakai pada riset ini berupa data kualitatif. Sumber data menggunakan data dari respon pengguna media sosial X yang berisikan opini/tanggapan/komentar mereka terhadap program Makan Bergizi Gratis (MBG) pada tanggal 1 Januari 2025 sampai dengan 21 Juli 2025 sebanyak 1754 data berupa *tweet harvest*. Sebelum dilakukan analisis lebih lanjut, berikut disajikan contoh data yang sudah dilakukan klasifikasi ke dalam sentimen positif, netral, serta negatif. Rincian contoh tersebut ditampilkan pada Tabel. 1.

Tabel 1. Contoh Sentimen

Sentimen	Klasifikasi
@justlyaaaaa Makan Bergizi Gratis (MBG) yang dulu disebut Makan Siang Gratis merupakan bagian dari komitmen pemerintahan Presiden @prabowo dan Wakil Presiden @gibran_tweet dlm mendukung kesehatan masyarakat. Lanjutkan MBG Keren ï¸ •	Positif
@PraharaTemaram @tanyakanrl Bang program kuliah gratis udah ada sebetulnya kaya KIPK lPDP Bidikmisi dll . Tapi ko gak merata ? Dan banyak yang belum ngerasain Ya karena perlu anggaran banyak sama hal nya kaya program makan siang yang belum merata karena perlu anggaran banyak .	Netral
Eaaa penipu belain maling~~~ Gimana @Kepolisian_RI nggak diserang rakyat markupnya nggak kira-kira?! @prabowo emang nggak bisa pembukuan pantes makan siang gratis nggak keurus karena duitnya buat ginian? #buka_laporan_proses_pengadaan dong ah! Pake malu-malu. https://t.co/dkgtKHOFW7	Negatif

Berlandaskan data dalam Tabel 1, dapat dilihat bahwasanya setiap unggahan pengguna dilakukan pengklasifikasian ke dalam kategori sentimen yang tidak sama sesuai dengan konteks dan nada penyampaiannya. Misalnya, unggahan pertama menunjukkan apresiasi terhadap Program MBG sehingga dikategorikan sebagai sentimen positif. Sebaliknya, unggahan kedua berisi pandangan yang bersifat informatif dan tidak mengekspresikan dukungan maupun penolakan, sehingga diklasifikasikan sebagai sentimen netral. Adapun unggahan ketiga mengandung kritik secara langsung terhadap program dan aktor terkait, sehingga masuk ke dalam sentimen negatif. Klasifikasi ini mencerminkan kemampuan model dalam menangkap polaritas opini publik berdasarkan konten tekstual yang dianalisis.

Tahapan analisis data dalam kerangka riset ialah algoritma *Naive Bayes Classifier* mampu diamati dalam Gambar: 1.



Gambar 1. Flowchart Analisis Data

Berikut penjelasan flowchart analisis data dengan Algoritma *Naïve Bayes Classifier*:

1. Proses dimulai dengan pengumpulan data komentar menggunakan teknik *crawling* pada media sosial X dengan rentang waktu 1 januari 2025 sampai 21 Juli 2025.

2. Kemudian dilanjutkan dengan tahapan *preprocessing* yang mencakup penghapusan data kosong, duplikat, karakter tidak penting, emoji, serta normalisasi data.
3. Setelah data dibersihkan, dilakukan pelabelan sentimen secara otomatis menggunakan algoritma Transformers.
4. Data yang telah dilabel kemudian dibagi menjadi data latih dan data uji dengan perbandingan 80:20 untuk proses klasifikasi menggunakan algoritma *Naïve Bayes.*
5. Hasil klasifikasi dievaluasi menggunakan *confusion matrix* untuk memperoleh metrik performa seperti akurasi dan F1-score. Langkah terakhir adalah pemilihan model dengan hasil klasifikasi terbaik berdasarkan evaluasi tersebut.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data komentar (tweet) pada media sosial X tanggal 1 Januari 2025 sampai 21 Juli 2025 dengan kata kunci "Makan Siang Gratis" dan "Makan Bergizi Gratis" menggunakan teknik *crawling* pada google colab.

3.1. Deskripsi Data

Data yang digunakan pada penelitian ini merupakan data primer yang diperoleh secara langsung dari Twitter menggunakan metode Crawling Data dengan Python Programming di Google Colab.

3.1.1. *Crawling Data*

Gambar 2. *Crawling Data*

Pada Gambar 2. merupakan proses untuk mengambil data *tweet* dari X dimana pertama dilakukan penginstalan *library* yang dibutuhkan untuk proses *crawling* data *tweet*. Kemudian hasil *crawling* disimpan dalam bentuk tabel dengan format CSV.

3.1.2. *Prepocessing Data*

a. *Cleaning Data*

Bertujuan untuk pembersihan data *tweet* dari *username*, *links*, *hashtags*, *emoticon*, spesial karakter, angka dan kata yang hanya terdiri dari 1 huruf serta melakukan pengecekan ejaan atau kesalahan pengetikan menggunakan *tools google spreadsheets*.

Tabel 2. *Cleaning Data*

Sebelum Proses *Cleaning*	Setelah proses *Cleaning*
@prabowo Udah bener ada program sekolah gratis atau program satu rumah satu sarjana malah milih makan siang gratis. Diantara ketiga paslon emang ada plus minusnya tapi tolong lah masa milih yang gaada plusnya sama sekali.;;;;;	Udah bener ada program sekolah gratis atau program satu rumah satu sarjana malah milih makan siang gratis Diantara ketiga paslon emang ada plus minusnya tapi tolong lah masa milih yang gaada plusnya sama sekali.

b. *Case Folding*

Case Folding merupakan tahap dimana teks diproses oleh aplikası dengan merubah seluruh karakter huruf kecil, pada tahap ini, dimulai dari proses pembacaan dokumen CSV yang berisi kumpulan tweet, kemudian proses *case folding* dilakukan dengan memanfaatkan fungsi pada Pandas yaitu Series.str lower().

Tabel 3. *Case Folding*

Sebelum Proses *Case Folding*	Setelah Proses *Case Folding*
Senenarnya makan Siang gratis itu penting Buat anak Yatim Piatu seperti yayasan Pak Peter Bagi anak sekolah terutama di desa yang paling penting adalah infrastuktur jangan biarkan mereka pergi sekolah bergelayutan di tali sebagai jembatan darurat.	sebenarnya makan siang gratis itu penting buat anak yatim piatu seperti yayasan pak peter bagi anak sekolah terutama di desa yang paling penting adalah infrastuktur jangan biarkan mereka pergi sekolah bergelayutan di tali sebagai jembatan darurat.
Yang seperti ini GAK AKAN PAHAM GAK AKAN NGERTI Yg paham cuma makan siang gratis	yang seperti ini gak akan paham gak akan ngerti yg paham cuma makan siang gratis

c. *Tokenizing*

Tahap ini adalah proses yang dilakukan untuk memisahkan kalimat atau teks menjadi kata-kata atau token. Input darı proses ini berupa hasil *cleansing* sebelumnya dan diproses menggunakan *library* NLTK.

Tabel 4. *Tokenizing*

Sebelum Proses *Tokenizing*	Setelah Proses *Tokenizing*
pentingnya makan siang gratis	[‘pentingnya’,’makan’,’siang’,’gratis’]
makan siang gratis program gagal	[‘makan’,’siang’,’gratis’,’program’,’gagal’]

d. *Normalization*

Setelah dilakukannya tokenizing, proses selanjutnya ialah normalisasi yang bertujuan untuk mengubah bentuk tidak baku dari data teks menjadi kata bakunya termasuk singkatan-singkatan yang dibantu dengan kamus kumpulan kata baku berbentuk excel, seperti “sdh” menjadi “sudah”.

Tabel 5. *Normalization*

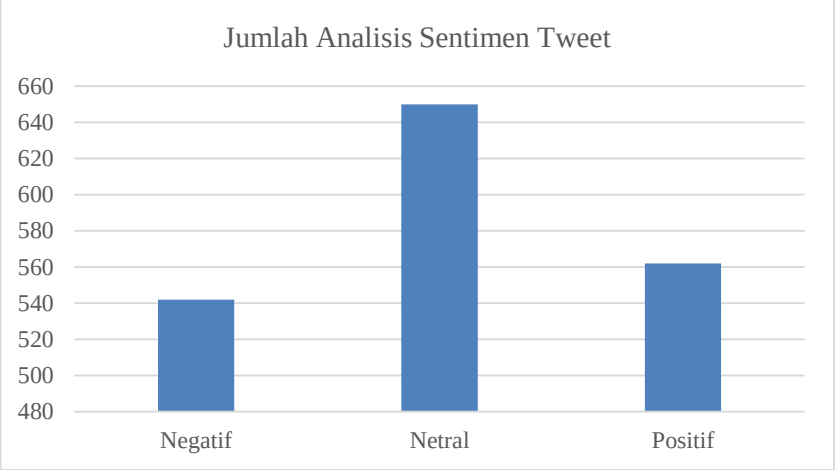
Sebelum Proses Normalisasi	Setelah Proses Normalisasi
[‘jd’,’sudah’,’makan’,’siang’,’bergizi’,’gratis’,’berapa’,’lama’,’pak’,’bs’,’numbuh’,’tinggi’,’segitu’,’nya’]	[‘jadi’,’sudah’,’makan’,’siang’,’bergizi’,’gratis’,’berapa’,’lama’,’pak’,’bisa’,’tumbuh’,’tinggi’,’segitu’,’nya’]

3.2. Pembahasan

3.1.1. Labeling Data

Pelabelan data dilakukan dengan menggunakan algoritma transformers. Dengan transformers ini memiliki mekanisme yang otomatis untuk melabelkan suatu data. Model transformator digunakan untuk mengekstraksi fitur untuk setiap kata, dan model *Natural Language Processing* (NLP) modern yang paling terkenal terdiri dari banyak model atau varian seperti itu. BERT (*Bidirectional Encoder Representations from Transformers*) adalah arsitektur transformer khusus encoder pra-terlatih dengan representasi bahasa dua arah yang mendalam. Itu dilatih sebelumnya menggunakan data tanpa label yang dikumpulkan dari BooksCorpus dengan 800 juta kata dan Wikipedia dengan 2,5 miliar kata[12].

Setelah dilakukan pelabelan sentimen untuk setiap data set, pada Gambar 3. ditunjukkan hasil dari pelabelan menggunakan algoritma transformers adalah untuk data tweet sentimen masyarakat terhadap program MBG memiliki 542 data Negatif, 650 data Netral, dan 562 data Positif.



Gambar 3. Jumlah Analisis Sentimen

3.1.2. Pemrosesan Label Sentimen

Langkah pertama adalah melakukan pemetaan label sentimen dari bentuk kategorikal ke bentuk numerik, yaitu: negatif menjadi 0, netral menjadi 1, dan positif menjadi 2. Proses ini penting karena sebagian besar algoritma pembelajaran mesin memerlukan label dalam bentuk numerik. Setelah itu, data yang memiliki nilai kosong pada kolom label dibersihkan, dan tipe data kolom sentimen diubah ke bentuk integer agar sesuai dengan format yang dibutuhkan. Penelitian ini memanfaatkan berbagai pustaka Python dalam proses analisis sentimen. Tahap awal mencakup instalasi dan pemanggilan pustaka seperti Sastrawi untuk *stemming* teks bahasa Indonesia, serta nltk untuk tokenisasi dan penghapusan *stopwords*. Pustaka pandas, numpy, matplotlib, dan seaborn digunakan untuk manipulasi data dan visualisasi. Ekstraksi fitur dilakukan dengan metode TF-IDF menggunakan TfidfVectorizer, sedangkan seleksi fitur dilakukan menggunakan metode *chi-square* melalui SelectKBest. Data kemudian dibagi menjadi data latih dan uji menggunakan train_test_split, lalu diklasifikasikan dengan algoritma *Complement Naïve Bayes* dari pustaka scikit-learn. Ketidakseimbangan kelas pada data ditangani menggunakan metode SMOTE. Sebelum pemodelan, dilakukan proses transformasi label sentimen menjadi bentuk numerik, di mana label 'negative', 'neutral', dan 'positive' masing-masing dikonversi menjadi 0, 1, dan 2. Proses ini bertujuan untuk mempermudah proses klasifikasi oleh model pembelajaran mesin. Evaluasi performa model dilakukan dengan metrik akurasi, confusion matrix, dan classification report untuk menilai ketepatan klasifikasi model terhadap data komentar masyarakat.

3.1.3. Pembersihan dan Seleksi Kolom

Pada tahap ini data yang memiliki nilai kosong pada kolom label dibersihkan, dan tipe data kolom sentiment diubah ke bentuk integer agar sesuai dengan format yang dibutuhkan. Sebelum proses pemodelan dilakukan, data perlu disesuaikan dan dibersihkan agar dapat diproses secara optimal oleh algoritma pembelajaran mesin. Pertama, dilakukan pemilihan dua kolom utama dari *dataframe*, yaitu kolom sentiment dan processed_text, yang masing-masing merepresentasikan label sentimen dan komentar yang telah diproses. Selanjutnya, nama kolom processed_text diganti menjadi komentar untuk memudahkan interpretasi. Langkah berikutnya adalah menghapus data yang tidak memiliki nilai pada kolom sentiment dengan menggunakan fungsi dropna(). Penghapusan ini bersifat penting untuk

menjaga kualitas data latih, karena nilai sentimen yang kosong akan mengganggu proses pembelajaran model. Setelah itu, dilakukan konversi tipe data pada kolom sentiment menjadi tipe data numerik integer. Konversi ini diperlukan agar label sentimen dapat dikenali dan diproses dengan baik oleh algoritma klasifikasi yang umumnya hanya menerima input numerik untuk label target. Tahapan ini menjadi bagian krusial dari proses praproses data karena menjamin bahwa hanya data yang lengkap dan terstruktur yang digunakan dalam tahap pelatihan model, sehingga kualitas prediksi dapat ditingkatkan dan kesalahan klasifikasi akibat data yang tidak konsisten dapat diminimalkan.

3.1.4. TF-IDF

Pada implementasinya, digunakan fungsi TfidfVectorizer() dari pustaka scikit-learn, dengan beberapa parameter yang disesuaikan untuk mengoptimalkan hasil ekstraksi fitur. Parameter ngram_range=(1,2) digunakan untuk mempertimbangkan *unigram* dan *bigram* secara bersamaan, sehingga model dapat menangkap konteks kata tunggal maupun pasangan kata. Tokenisasi dilakukan menggunakan RegexpTokenizer dari pustaka NLTK, yang telah disesuaikan sebelumnya. Selain itu, parameter max_features=7000 digunakan untuk membatasi jumlah fitur yang dihasilkan agar model tetap efisien, sementara sublinear_tf=True diterapkan agar frekuensi kata dalam dokumen ditransformasikan secara logaritmik, yang dapat meningkatkan performa klasifikasi. Parameter min_df=2 memastikan bahwa hanya kata-kata yang muncul setidaknya dua kali dalam korpus yang akan dipertimbangkan sebagai fitur, sehingga mengurangi kemungkinan memasukkan kata-kata yang kurang relevan atau terlalu jarang. Hasil dari proses ini berupa matriks fitur X yang merepresentasikan setiap komentar dalam bentuk vektor numerik berdasarkan bobot TF-IDF-nya, sedangkan y menyimpan label sentimen yang sesuai. Matriks ini kemudian akan digunakan sebagai input dalam tahap pelatihan dan pengujian model klasifikasi.

3.1.5. Seleksi Fitur Menggunakan Chi-Square

Setelah proses ekstraksi fitur dilakukan menggunakan TF-IDF, langkah selanjutnya adalah seleksi fitur untuk mengurangi dimensi data dan meningkatkan efisiensi serta performa model klasifikasi. Pada penelitian ini, digunakan metode *SelectKBest* dengan fungsi evaluasi *chi-square* (χ^2), yang umum digunakan untuk data kategorikal seperti data sentimen. Metode *chi-square* mengevaluasi sejauh mana hubungan antara fitur dan variabel target (label sentimen) bersifat signifikan secara statistik. Fitur-fitur dengan nilai chi-square tertinggi dianggap paling relevan dalam membedakan kelas sentimen, sehingga dipilih sebagai input utama untuk pelatihan model. Dalam implementasinya, hanya 3.000 fitur terbaik yang dipertahankan dari total 7.000 fitur awal hasil TF-IDF. Pemilihan jumlah ini bertujuan untuk menyeimbangkan antara kompleksitas model dan akurasi prediksi. Proses ini tidak hanya mempercepat waktu pelatihan, tetapi juga membantu mengurangi risiko *overfitting* dengan mengabaikan fitur-fitur yang kurang informatif. Dengan demikian, tahap seleksi fitur ini berfungsi sebagai filter awal untuk memastikan bahwa hanya informasi yang paling relevan yang disertakan dalam proses pembelajaran model klasifikasi.

3.1.6. Pembagian Data Training dan Data Testing

Setelah dilakukan seleksi fitur, langkah selanjutnya adalah memisahkan data menjadi dua bagian utama, yaitu data pelatihan (*training set*) dan data pengujian (*testing set*). Pemisahan ini dilakukan menggunakan fungsi train_test_split dari pustaka *scikit-learn*, dengan proporsi 80% data untuk pelatihan dan 20% data untuk pengujian, yang ditentukan oleh parameter test_size=0.2. Untuk menjaga proporsi distribusi label sentimen pada data pelatihan dan pengujian tetap seimbang, digunakan teknik *stratified sampling* melalui parameter stratify=y. Pendekatan ini penting dalam konteks klasifikasi, khususnya ketika data memiliki distribusi kelas yang tidak seimbang, karena memungkinkan model untuk

belajar dari representasi yang konsisten terhadap masing-masing kelas. Selain itu, parameter random_state=42 digunakan untuk menjamin reprodusibilitas hasil, sehingga pemisahan data dapat diulang dengan hasil yang sama. Proses ini menghasilkan empat buah variabel, yaitu X_train dan y_train yang digunakan untuk melatih model, serta X_test dan y_test yang digunakan untuk mengevaluasi performa model setelah pelatihan selesai.

3.1.7. Penerapan SMOTE

Pada tahap ini, diterapkan metode *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menangani permasalahan ketidakseimbangan kelas pada data latih (X_train, y_train). Ketidakseimbangan kelas sering kali menjadi hambatan dalam performa model klasifikasi karena model cenderung lebih mempelajari pola dari kelas mayoritas dan mengabaikan kelas minoritas. SMOTE bekerja dengan menghasilkan sampel sintetis dari kelas minoritas berdasarkan interpolasi antara titik data aktual dan tetangganya yang paling dekat dalam ruang fitur. Dengan demikian, SMOTE membantu menciptakan distribusi kelas yang lebih seimbang tanpa sekadar menggandakan data yang sudah ada. Pemanggilan SMOTE(random_state=42) digunakan untuk menjamin bahwa proses pembangkitan data sintetis dapat direproduksi. Kemudian, metode fit_resample(X_train, y_train) digunakan untuk menghasilkan data latih yang telah seimbang, yang selanjutnya akan digunakan dalam proses pelatihan model klasifikasi. Langkah ini bertujuan untuk meningkatkan kinerja klasifikasi, terutama dalam mengenali kelas minoritas, serta mengurangi bias model terhadap kelas mayoritas.

3.1.8. Fungsi Evaluasi Model

Setelah data siap, dilakukan proses pelatihan dan evaluasi model. Fungsi evaluate_model() digunakan untuk melatih model, melakukan prediksi terhadap data uji, serta menghitung dan menampilkan metrik performa model, seperti akurasi, confusion matrix, dan classification report (yang mencakup precision, recall, dan f1-score). Visualisasi hasil prediksi ditampilkan melalui heatmap confusion matrix untuk mempermudah interpretasi.

Tabel 6. Hasil *Confusion Matrix*

	Aktual Positif	Aktual Negatif
Prediksi Positif	203	19
Prediksi Negatif	38	91

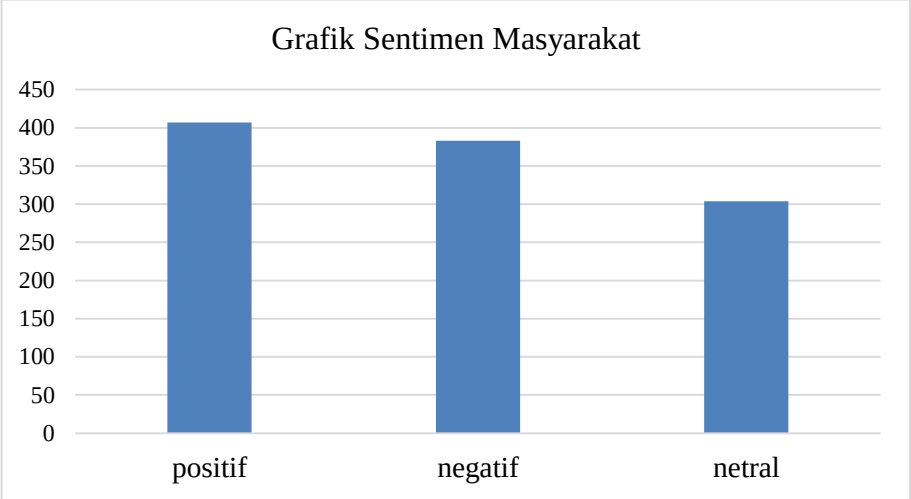
Tabel 6 menyajikan hasil *cofussion matrix* dari proses klasifikasi menggunakan metode *Naïve Bayes Classifier*. Berdasarkan tabel tersebut, Berdasarkan tabel tersebut, model berhasil mengidentifikasi 203 data sebagai positif secara benar (*True Positive*) dan 91 data sebagai negatif secara benar (*True Negative*). Sementara itu, terdapat 19 data yang salah diprediksi sebagai positif (*False Positive*) dan 38 data yang salah diprediksi sebagai negatif (*False Negative*). Temuan ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang cukup baik dalam membedakan kelas positif dan negatif pada dataset penelitian.

3.1.9. *Tuning Parameter Alpha* pada Complement Naïve Bayes

Tahap selanjutnya adalah melakukan tuning parameter alpha pada model Complement Naïve Bayes. Parameter alpha berperan sebagai smoothing parameter (additive smoothing), yang berfungsi untuk menghindari nilai probabilitas nol dan membantu generalisasi model, terutama pada fitur yang jarang muncul. Empat nilai alpha yang diuji adalah 0.1, 0.3, 0.5, dan 1.0. Hasil evaluasi dari masing-masing nilai alpha dibandingkan untuk melihat pengaruh smoothing terhadap kinerja klasifikasi.

3.1.10. Hasil Sentimen

Pada tahap ini akan ditampilkan hasil analisis sentimen masyarakat terhadap program MBG pada media sosial X menggunakan metode Naïve Bayes Classifier.



Gambar 4. Grafik Sentimen Masyarakat

Berdasarkan hasil analisis sentimen yang telah dilakukan terhadap data teks yang diperoleh sentimen yang telah diklasifikasikan selanjutnya divisualisasikan dalam bentuk diagram batang sebagaimana ditunjukkan pada Gambar 4. Distribusi sentimen sebanyak 407 data sentimen positif, 383 data sentimen negatif, dan 304 data sentimen netral. Distribusi ini menunjukkan bahwa secara umum, tanggapan yang diberikan oleh pengguna terhadap isu atau topik yang dianalisis cenderung bersifat positif, meskipun jumlah sentimen negatif dan netral juga cukup signifikan.

```
Complement Naive Bayes (alpha=0.3) Accuracy: 83.76%
Confusion Matrix:
[[ 91   6  11]
 [  8 101  21]
 [  7   4 102]]
Classification Report:
              precision    recall  f1-score   support

           0       0.86      0.84      0.85       108
           1       0.91      0.78      0.84       130
           2       0.76      0.90      0.83       113

    accuracy                           0.84       351
   macro avg       0.84      0.84      0.84       351
weighted avg       0.85      0.84      0.84       351
```

Gambar 5. Hasil Akurasi

Berdasarkan hasil pada Gambar 5, model klasifikasi yang digunakan, yaitu *Naïve Bayes Classifier*, telah dilatih dan diuji dengan pendekatan *preprocessing* teks, ekstraksi fitur menggunakan TF-IDF dengan n-gram, serta seleksi fitur menggunakan metode chi-square. Selain itu, penerapan teknik SMOTE untuk menyeimbangkan kelas pada data latih berhasil meningkatkan performa model dalam mengenali kelas minoritas. Dari proses evaluasi model, diperoleh akurasi sebesar 83,76%, dengan nilai alpha 0,3 yang tergolong dalam kategori "baik". Hal ini menunjukkan bahwa model yang dibangun mampu mengklasifikasikan sentimen dengan tingkat ketepatan yang baik dan dapat diandalkan dalam mengidentifikasi pola-pola sentimen dalam data teks.Secara keseluruhan, hasil penelitian ini mengindikasikan bahwa algoritma *Naïve Bayes Classifier*, apabila didukung

dengan tahapan preprocessing dan tuning parameter yang tepat, dapat digunakan secara efektif dalam klasifikasi sentimen pada data media sosial. Temuan ini memberikan kontribusi yang berarti dalam penerapan machine learning pada bidang analisis opini publik dan dapat menjadi dasar untuk penelitian lanjutan dengan cakupan data yang lebih luas atau pendekatan metode yang lebih kompleks.

4. KESIMPULAN

Hasil penelitian menunjukkan bahwa sentimen masyarakat terhadap Program MBG di platform X cenderung positif, dengan 407 komentar positif, 383 negatif, dan 304 netral. Model *Naïve Bayes Classifier* mampu memetakan polaritas opini publik secara proporsional dan mencapai akurasi 83,76%, yang termasuk kategori baik. Penelitian ini berbeda dari studi sebelumnya yang menggunakan data pra-implementasi, karena penelitian ini menganalisis data yang dikumpulkan setelah program berjalan sehingga memberikan gambaran sentimen yang lebih aktual dan relevan. Penelitian ini terbatas pada data periode Januari–Juli 2025 sehingga belum mampu merepresentasikan dinamika sentimen jangka panjang. Meski demikian, temuan ini dapat menjadi dasar bagi pemerintah untuk mengevaluasi pelaksanaan Program MBG, khususnya dengan memperhatikan sentimen negatif yang muncul sebagai bentuk kritik masyarakat. Selain itu, hasil penelitian ini membuka peluang pengembangan lebih lanjut, baik melalui penerapan metode klasifikasi yang lebih kompleks maupun perluasan sumber dan rentang waktu data untuk memperoleh pemahaman yang lebih komprehensif terhadap respons publik.

REFERENSI

[1] D. Wiryany, S. Natasha, dan R. Kurniawan, "Perkembangan Teknologi Informasi dan Komunikasi terhadap Perubahan Sistem Komunikasi Indonesia," *J. Nomosleca*, vol. 8, no. 2, hal. 242–252, 2022, doi: 10.26905/nomosleca.v8i2.8821.

[2] N. Desiani dan A. Syafiq, "Efektivitas Program Makan Gratis pada Status Gizi Siswa Sekolah Dasar: Tinjauan Sistematis," *Malahayati Nurs. J.*, vol. 7, no. 1, hal. 27–48, 2025.

[3] W. T. Aji, "Makan Bergizi Gratis di Era Prabowo-Gibran: Solusi untuk Rakyat atau Beban Baru?," *NAAFI J. Ilm. Mhs.*, vol. 1, no. 3, hal. 215–226, 2025.

[4] A. H. Pramudji dan J. Andiani, "PENGELOLAAN SAMPAH RUMAH TANGGA DI RUMAH BARU INDONESIA," *PUSTAKA ADITYA PERPINDAHAN IBU KOTA NEGARA DI MATA DIASPORA JEPANG*, hal. 93.

[5] I. Taufik, "Analisis Sentimen Terhadap Tokoh Publik Menggunakan Algoritma Support Vector Machine (SVM)," 2018, *Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta*.

[6] W. Widayat, "Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning," *J. Media Inform. Budidarma*, vol. 5, no. 3, hal. 1018, 2021.

[7] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, dan S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci., vol. 5, no. 2, p. 293, 2020, doi 10.24114/cess. v5i2. 18186*, 2020.

[8] A. P. Natasuwarna, "Seleksi Fitur Support Vector Machine pada Analisis Sentimen Keberlanjutan Pembelajaran Daring.," *Techno. Com*, vol. 19, no. 4, 2020.

[9] R. Yunita dan M. Kamayani, "Perbandingan Algoritma SVM Dan Naïve Bayes Pada Analisis Sentimen Penghapusan Kewajiban Skripsi," *Indones. J. Comput. Sci.*, vol. 12, no. 5, hal. 2879–2890, 2023, doi: 10.33022/ijcs.v12i5.3415.

[10] A. Roihan, P. A. Sunarya, dan A. S. Rafika, "Pemanfaatan machine learning dalam berbagai bidang," *J. Khatulistiwa Inform.*, vol. 5, no. 1, hal. 490845, 2020.

[11] R. Saputra dan F. N. Hasan, "Analisis Sentimen Terhadap Program Makan Siang & Susu Gratis Menggunakan Algoritma Naive Bayes," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 3, hal. 411–419, 2024, doi: 10.47233/jteksis.v6i3.1378.

[12] MP Firdaus, "Perbandingan Algoritma K-Nearest Neighbor (KNN) dan Naive Bayes Classifier (NBC) dengan pelabelan Transformers serta Ektraksi Fitur TF-IDF dan N-Gram untuk Analisis Sentimen Terhadap Penundaan Pemilu," *Perbandingan Algoritm. K-Nearest Neighbor dan Naive Bayes Classif. dengan pelabelan Transform. serta Ektraksi Fitur TF-IDF dan N-Gram untuk Anal. Sentimen Terhadap Penundaan Pemilu*, hal. 5–24, 2023, [Daring]. Tersedia pada: https://repository.uinjkt.ac.id/dspace/handle/123456789/72466