

# Analisis Sentimen Persepsi Publik Terhadap Program MBG Pada Komentar YouTube Menggunakan Naïve Bayes dan Resampling

Lutfi Najib\*, Adzhal Arwani Mahfudh, Syaiful Bakhri

Fakultas Sains dan Teknologi, Program Studi Teknologi Informasi, Universitas Islam Negeri Walisongo, Semarang, Indonesia

Email: <sup>1</sup>\*2208096111@student.walisongo.ac.id, <sup>2</sup>adzhal@walisongo.ac.id, <sup>3</sup>syaifulbakhri@walisongo.ac.id

Email Penulis Korespondensi: 2208096111@student.walisongo.ac.id

Submitted: 12/02/2026; Accepted: 05/03/2026; Published: 06/03/2026

**Abstrak**—Program Makan Bergizi Gratis (MBG) yang diluncurkan pemerintah Indonesia pada tahun 2025 memunculkan berbagai respons publik di media sosial, khususnya pada platform YouTube sebagai ruang diskusi digital terbuka. Penelitian ini bertujuan menganalisis persepsi publik terhadap Program MBG melalui klasifikasi sentimen komentar YouTube menggunakan algoritma Multinomial Naïve Bayes dengan pembobotan Term Frequency–Inverse Document Frequency (TF-IDF). Dataset terdiri dari 1.082 komentar yang dikategorikan ke dalam tiga kelas sentimen: negatif, netral, dan positif. Distribusi data menunjukkan ketidakseimbangan kelas yang signifikan dengan dominasi sentimen negatif sebesar 70,61%. Model baseline menghasilkan akurasi 70,67% dengan F1-score macro 27,60%, yang mengindikasikan bias terhadap kelas mayoritas. Untuk mengatasi ketidakseimbangan tersebut, diterapkan teknik Random Oversampling (ROS) dan Synthetic Minority Over-sampling Technique (SMOTE). Hasil menunjukkan bahwa meskipun akurasi menurun hingga sekitar 51%, nilai F1-score macro meningkat menjadi 36,24% (SMOTE) dan 37,09% (ROS), yang menunjukkan peningkatan kemampuan model dalam mengenali kelas minoritas. Dalam konteks evaluasi kebijakan publik, peningkatan sensitivitas terhadap sentimen minoritas dinilai lebih representatif dibandingkan akurasi tinggi yang bias terhadap satu kelas. Temuan ini menegaskan pentingnya penanganan ketidakseimbangan kelas dalam analisis sentimen kebijakan publik berbasis media sosial.

**Kata Kunci:** Analisis Sentimen; Naïve Bayes; YouTube; TF-IDF; SMOTE; Random Oversampling; Pemetaan Opini

**Abstract**—The Free Nutritious Meal Program (MBG), launched by the Indonesian government in 2025, has generated diverse public responses on social media, particularly on YouTube as an open digital discussion space. This study aims to analyze public perception of the MBG program through sentiment classification of YouTube comments using the Multinomial Naïve Bayes algorithm combined with Term Frequency–Inverse Document Frequency (TF-IDF) weighting. The dataset consists of 1,082 comments categorized into three sentiment classes: negative, neutral, and positive. The data distribution reveals significant class imbalance, with negative sentiment dominating at 70.61%. The baseline model achieved an accuracy of 70.67% with a macro F1-score of 27.60%, indicating bias toward the majority class. To address this imbalance, Random Oversampling (ROS) and Synthetic Minority Over-sampling Technique (SMOTE) were applied. Although overall accuracy decreased to approximately 51% after resampling, the macro F1-score improved to 36.24% (SMOTE) and 37.09% (ROS), indicating enhanced performance in detecting minority classes. In the context of public policy evaluation, improved sensitivity to minority sentiment is considered more representative than high but biased accuracy. These findings highlight the importance of handling class imbalance in social media based sentiment analysis for public policy monitoring.

**Keywords:** Sentiment Analysis; Naïve Bayes; Youtube Comments; TF-IDF; SMOTE; Random Oversampling; Opinion Mapping

## 1. PENDAHULUAN

Indonesia masih menghadapi permasalahan gizi yang kompleks, terutama pada anak usia sekolah, yang berdampak pada kualitas sumber daya manusia di masa depan. Prevalensi stunting, keterbatasan akses terhadap pangan bergizi, serta ketimpangan distribusi bantuan sosial menjadi tantangan yang belum sepenuhnya teratasi. Data menunjukkan bahwa sekitar 4,5 juta anak Indonesia mengalami hambatan pertumbuhan fisik dan perkembangan kognitif akibat kekurangan gizi [1]. Kondisi ini mendorong pemerintah meluncurkan Program Makan Bergizi Gratis (MBG) pada tahun 2025 sebagai bagian dari strategi intervensi gizi nasional untuk meningkatkan kualitas nutrisi dan mendukung capaian pendidikan.

Program makan sekolah sebelumnya di Indonesia, seperti PROGAS, telah menunjukkan dampak positif terhadap peningkatan asupan gizi dan performa akademik siswa, meskipun menghadapi keterbatasan implementasi dan cakupan wilayah [2]. Penelitian lain menunjukkan bahwa program makan sekolah berkorelasi dengan peningkatan skor kognitif siswa, khususnya pada mata pelajaran matematika [3], serta berkontribusi terhadap perubahan perilaku konsumsi sarapan sehat [2]. Namun demikian, efektivitas kebijakan publik tidak hanya ditentukan oleh desain program, tetapi juga oleh penerimaan serta persepsi masyarakat terhadap pelaksanaannya. Persepsi publik dapat memengaruhi legitimasi kebijakan, tingkat partisipasi masyarakat, serta keberlanjutan program dalam jangka panjang.

Di era digital, media sosial tidak hanya berfungsi sebagai sarana komunikasi, tetapi juga sebagai ruang deliberatif dalam praktik demokrasi digital. Masyarakat dapat menyampaikan opini, kritik, maupun dukungan terhadap kebijakan pemerintah secara terbuka dan terdokumentasi. Dibandingkan platform seperti X (Twitter) yang membatasi panjang teks atau Instagram yang lebih berorientasi visual, YouTube sering menjadi medium distribusi konferensi pers pemerintah, siaran berita televisi, serta konten diskusi kebijakan publik. Kolom komentar pada YouTube memungkinkan pengguna menyampaikan opini yang lebih panjang, argumentatif, dan kontekstual, sehingga diskusi yang terjadi cenderung lebih mendalam. Karakteristik ini menjadikan YouTube relevan sebagai sumber data dalam menganalisis persepsi publik terhadap kebijakan nasional seperti Program Makan Bergizi Gratis (MBG) [4].

Dalam penelitian analisis sentimen, algoritma Multinomial Naïve Bayes sering digunakan karena sederhana dan efisien secara komputasi. Sejumlah penelitian menunjukkan bahwa kombinasi pembobotan TF-IDF dengan Naïve Bayes mampu menghasilkan performa klasifikasi yang stabil pada data teks berbahasa Indonesia [5] [6]. Dalam konteks pemantauan kebijakan publik seperti Program MBG, efisiensi komputasi menjadi aspek penting karena sistem analisis perlu memproses data dalam jumlah besar secara cepat dengan latensi rendah. Model yang ringan memungkinkan pembaruan model secara berkala dan pemantauan opini publik secara berkelanjutan (streaming data). Dibandingkan model berbasis deep learning seperti BERT yang memerlukan sumber daya komputasi tinggi dan waktu pelatihan lebih lama, Naïve Bayes lebih sesuai digunakan sebagai baseline monitoring system yang responsif terhadap dinamika opini masyarakat [7] [8].

Meskipun demikian, tantangan utama dalam analisis sentimen media sosial adalah ketidakseimbangan distribusi kelas (*class imbalance*), di mana satu kelas sentimen sering kali mendominasi jumlah data dibandingkan kelas lainnya [9]. Ketidakseimbangan ini dapat menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas, sehingga nilai akurasi terlihat tinggi tetapi kemampuan model dalam mengenali kelas minoritas rendah. Oleh karena itu, evaluasi performa tidak cukup hanya menggunakan akurasi, tetapi perlu mempertimbangkan metrik seperti *F1-score macro* yang lebih sensitif terhadap performa setiap kelas.

Beberapa penelitian telah mengusulkan teknik penanganan ketidakseimbangan kelas melalui metode *resampling*. Salah satu pendekatan yang banyak digunakan adalah *Random Oversampling* (ROS), yaitu teknik yang menyeimbangkan distribusi data dengan menggandakan sampel dari kelas minoritas. Penelitian pada deteksi serangan *Distributed Denial of Service* (DDoS) menunjukkan bahwa penerapan ROS mampu meningkatkan nilai *macro F1-score* serta memperbaiki performa klasifikasi pada dataset multi-kelas yang tidak seimbang [10]. Temuan tersebut mengindikasikan bahwa teknik oversampling efektif dalam meningkatkan keseimbangan performa model, meskipun dalam beberapa kasus dapat menyebabkan penurunan akurasi keseluruhan akibat redistribusi bobot prediksi. Selain ROS, teknik *Synthetic Minority Over-sampling Technique* (SMOTE) juga sering digunakan untuk menghasilkan sampel sintesis pada kelas minoritas guna memperkaya representasi data.

Namun demikian, sebagian besar penelitian terkait penanganan *class imbalance* masih berfokus pada domain kesehatan, keamanan jaringan, atau klasifikasi umum, dan belum banyak diterapkan secara spesifik pada analisis sentimen kebijakan publik berbasis komentar YouTube berbahasa Indonesia. Selain itu, beberapa studi analisis sentimen kebijakan publik masih menitikberatkan pada nilai akurasi tanpa mengevaluasi peningkatan keseimbangan performa antar kelas secara komprehensif melalui perbandingan *F1-score macro* sebelum dan sesudah *resampling*. Hal ini menunjukkan adanya kesenjangan penelitian dalam penerapan teknik *resampling* pada konteks analisis persepsi publik terhadap kebijakan nasional.

Berdasarkan kesenjangan tersebut, penelitian ini mengusulkan pendekatan analisis sentimen berbasis Multinomial Naïve Bayes dengan pembobotan TF-IDF untuk memetakan persepsi publik terhadap Program Makan Bergizi Gratis (MBG) melalui komentar YouTube. Penelitian ini tidak hanya membangun model baseline tanpa penanganan ketidakseimbangan kelas, tetapi juga menerapkan dua teknik *resampling*, yaitu *Random Oversampling* (ROS) dan *Synthetic Minority Over-sampling Technique* (SMOTE), guna mengevaluasi peningkatan keseimbangan performa model secara komprehensif. Dengan demikian, penelitian ini berfokus pada evaluasi keadilan performa antar kelas sentimen, bukan semata-mata pada peningkatan akurasi global.

Tujuan utama penelitian ini adalah (1) memetakan distribusi sentimen publik terhadap Program MBG secara terukur, (2) menganalisis pengaruh ketidakseimbangan kelas terhadap performa klasifikasi, serta (3) mengevaluasi efektivitas teknik *resampling* dalam meningkatkan kemampuan model mengenali kelas minoritas. Hasil penelitian ini diharapkan dapat memberikan kontribusi akademik dalam pengembangan analisis sentimen berbahasa Indonesia serta memberikan masukan berbasis data bagi pembuat kebijakan dalam mengevaluasi implementasi program gizi nasional [11].

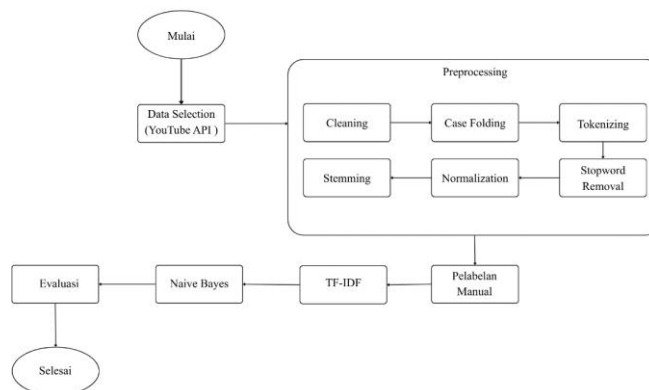
## 2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental berbasis *text mining* dan *machine learning* untuk mengklasifikasikan sentimen publik terhadap kebijakan Program Makan Bergizi Gratis (MBG). Pendekatan ini bertujuan memetakan persepsi masyarakat secara sistematis berdasarkan data komentar YouTube menggunakan algoritma *Multinomial Naïve Bayes* dengan pembobotan *Term Frequency–Inverse Document Frequency* (TF-IDF). Proses penelitian dilakukan secara terstruktur mulai dari pengumpulan data hingga evaluasi model.

### 2.1 Tahapan Penelitian

Tahapan penelitian merupakan bagian wajib dalam artikel ini dan menggambarkan alur kerja penelitian secara menyeluruh. Alur penelitian disusun berdasarkan pendekatan analisis sentimen berbasis *text mining* dan *machine learning* yang diterapkan pada komentar YouTube terkait Program Makan Bergizi Gratis (MBG). Secara umum, tahapan penelitian terdiri atas: (1) seleksi dan pengumpulan data menggunakan YouTube API, (2) preprocessing data teks, (3) pelabelan sentimen secara manual, (4) pembobotan fitur menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF), (5) klasifikasi sentimen menggunakan algoritma *Multinomial Naïve Bayes*, dan (6) evaluasi

kinerja model. Seluruh tahapan dilakukan secara berurutan untuk memastikan proses analisis yang sistematis, terstruktur, dan terukur. Alur tahapan penelitian ditunjukkan pada Gambar 1.



**Gambar 1.** Alur Kerja Penelitian

Tahap awal penelitian didasarkan pada identifikasi permasalahan dan kajian konseptual terkait implementasi kebijakan Program Makan Bergizi Gratis (MBG). Indonesia masih menghadapi tantangan dalam pemenuhan gizi anak usia sekolah, seperti tingginya prevalensi stunting, ketidakmerataan akses terhadap pangan bergizi, serta rendahnya literasi gizi masyarakat. Permasalahan tersebut berdampak langsung terhadap kualitas sumber daya manusia dan pencapaian *Sustainable Development Goals* (SDGs) di bidang kesehatan dan pendidikan [7]. Meskipun pemerintah telah meluncurkan Program MBG sebagai bentuk intervensi gizi nasional, keberhasilan implementasi kebijakan ini tidak hanya ditentukan oleh aspek teknis, tetapi juga sangat dipengaruhi oleh persepsi dan respons masyarakat. Oleh karena itu, pemetaan persepsi publik menjadi penting sebagai dasar evaluasi kebijakan berbasis data [12].

Secara metodologis, penelitian ini mengacu pada pipeline analisis sentimen yang umum digunakan dalam penelitian *machine learning*, yaitu pengumpulan data, preprocessing, ekstraksi fitur, klasifikasi, dan evaluasi. Metode TF-IDF digunakan pada tahap ekstraksi fitur karena mampu memberikan bobot lebih tinggi pada kata-kata yang informatif serta menurunkan bobot kata yang bersifat umum dalam korpus [13]. Kombinasi TF-IDF dengan algoritma *Multinomial Naïve Bayes* telah terbukti efisien dan efektif dalam mengklasifikasikan teks pendek dalam jumlah besar [6]. Dalam konteks bahasa Indonesia, tahapan preprocessing seperti cleaning, case folding, normalisasi, tokenisasi, stopword removal, dan stemming berperan penting dalam mengurangi noise serta meningkatkan stabilitas representasi fitur teks [9].

Selain itu, penelitian ini juga mempertimbangkan permasalahan ketidakseimbangan kelas (*class imbalance*) yang umum terjadi pada data sentimen media sosial. Ketidakseimbangan distribusi kelas dapat menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas dan mengabaikan kelas minoritas. Oleh karena itu, evaluasi kinerja model tidak hanya menggunakan metrik akurasi, tetapi juga mempertimbangkan *F1-score macro*, *F1-score weighted*, serta analisis confusion matrix agar performa model dapat dinilai secara lebih komprehensif [14].

## 2.2 Pengumpulan Data

Pengumpulan data dilakukan menggunakan pendekatan *text mining*, yaitu proses otomatis untuk mengekstraksi informasi bermakna dari data teks tidak terstruktur yang bersumber dari media sosial. Dalam penelitian ini, platform YouTube dipilih sebagai sumber data karena menyediakan ruang diskusi publik yang terbuka dan interaktif terkait kebijakan pemerintah, termasuk Program Makan Bergizi Gratis (MBG). Komentar pengguna pada video YouTube dinilai representatif dalam mencerminkan opini, dukungan, maupun kritik masyarakat terhadap implementasi kebijakan tersebut.

Data komentar diperoleh menggunakan YouTube Data API v3 dengan kata kunci “makan bergizi gratis”, “program makan bergizi gratis”, dan “program pemerintah 2025” untuk memastikan relevansi terhadap topik penelitian. Proses pengumpulan data terdiri atas dua tahap utama, yaitu *crawling* dan *scraping*. Tahap *crawling* dilakukan untuk menelusuri dan memilih sepuluh video YouTube yang relevan berdasarkan kesesuaian konten dan tingkat interaksi pengguna, seperti jumlah komentar dan penayangan. Pemilihan video dilakukan secara purposive untuk memastikan bahwa diskusi yang dianalisis benar-benar berkaitan dengan Program MBG.

Tahap *scraping* dilakukan untuk mengekstraksi komentar, waktu publikasi, serta metadata pendukung lainnya menggunakan bahasa pemrograman Python. Data yang berhasil dikumpulkan kemudian disimpan dalam format CSV (*Comma-Separated Values*) untuk memudahkan proses pengolahan lebih lanjut. Dari proses ini diperoleh 1.220 komentar mentah.

Selanjutnya dilakukan proses pembersihan awal berupa penghapusan duplikasi komentar dan penyaringan data yang tidak relevan. Komentar yang mengandung spam, ujaran kebencian, konten SARA, atau tidak berkaitan dengan topik penelitian dieliminasi untuk menjaga kualitas dataset. Setelah proses penyaringan, diperoleh 1.082 komentar unik yang digunakan sebagai data akhir penelitian. Seluruh identitas pengguna dianonimkan untuk menjaga etika penelitian digital dan privasi pengguna platform [15].

### 2.3 Preprocessing Data

Tahap preprocessing bertujuan untuk menyiapkan data teks mentah agar siap diolah secara komputasional serta meningkatkan kualitas representasi fitur yang akan digunakan dalam proses klasifikasi sentimen. Komentar YouTube umumnya masih mengandung berbagai elemen non-informatif seperti URL, emoji, simbol, karakter berulang, serta tanda baca berlebih yang dapat mengganggu proses analisis [16]. Oleh karena itu, dilakukan serangkaian tahapan pembersihan dan normalisasi teks secara sistematis.

Proses diawali dengan *cleaning*, yaitu penghapusan URL, emoji, simbol, angka yang tidak relevan, dan karakter khusus sehingga tersisa teks utama yang merepresentasikan opini pengguna. Selanjutnya dilakukan case folding dengan mengubah seluruh huruf menjadi huruf kecil untuk menyeragamkan token dan menghindari redundansi akibat perbedaan kapitalisasi [17]. Setelah itu, dilakukan proses normalisasi untuk mengonversi kata tidak baku, singkatan, atau bahasa gaul ke dalam bentuk baku bahasa Indonesia, seperti “gk” menjadi “tidak” dan “bgt” menjadi “banget” [18]. Tahap ini penting untuk mengurangi variasi kata yang memiliki makna serupa.

Teks yang telah dinormalisasi kemudian melalui proses tokenisasi untuk memecah kalimat menjadi unit kata (token) yang akan digunakan sebagai dasar pembentukan fitur [17]. Selanjutnya dilakukan stopword removal untuk menghapus kata-kata umum yang tidak memiliki kontribusi signifikan terhadap makna sentimen, seperti “yang”, “dan”, serta “di” [16]. Tahap terakhir adalah stemming, yaitu mengubah kata berimbuhan menjadi bentuk dasar menggunakan pustaka Sastrawi sehingga variasi kata turunan dapat disatukan dalam satu representasi fitur [19].

Rangkaian preprocessing ini bertujuan mengurangi noise, meningkatkan konsistensi struktur teks, serta memperbaiki kualitas matriks fitur yang akan dihasilkan pada tahap pembobotan TF-IDF. Dengan preprocessing yang terstruktur, model klasifikasi diharapkan mampu mengenali pola sentimen secara lebih stabil dan representatif terhadap opini publik yang dianalisis.

### 2.4 Pelabelan Sentimen

Pelabelan sentimen dilakukan secara manual oleh dua annotator independen dengan latar belakang akademik di bidang teknologi informasi dan linguistik. Setiap komentar dikelompokkan ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral. Proses pelabelan dilakukan dengan mempertimbangkan konteks kalimat, pilihan kata, serta kecenderungan opini yang disampaikan terhadap Program Makan Bergizi Gratis (MBG). Sentimen positif diberikan pada komentar yang menunjukkan dukungan, apresiasi, atau harapan positif terhadap program. Sentimen negatif diberikan pada komentar yang mengandung kritik, kekhawatiran, atau penilaian negatif terhadap implementasi kebijakan. Sementara itu, sentimen netral diberikan pada komentar yang bersifat informatif atau deskriptif tanpa kecenderungan evaluatif tertentu [8].

Untuk menjaga konsistensi pelabelan, kedua annotator menggunakan pedoman klasifikasi (annotation guideline) yang telah disusun sebelum proses anotasi. Pedoman tersebut mendefinisikan secara operasional indikator linguistik untuk masing-masing kategori sentimen, termasuk penggunaan kata evaluatif, ekspresi emosional, serta konteks implisit dalam kalimat. Setelah proses pelabelan selesai, tingkat kesepakatan antar penilai diukur menggunakan koefisien Cohen’s Kappa. Hasil perhitungan menunjukkan nilai Kappa sebesar 0,82, yang termasuk dalam kategori hampir sempurna (almost perfect agreement), sehingga menunjukkan tingkat konsistensi anotasi yang tinggi.

Pada kasus perbedaan label, dilakukan diskusi untuk mencapai konsensus akhir sebelum data digunakan dalam proses pelatihan model. Hasil akhir pelabelan menunjukkan distribusi kelas yang tidak seimbang, di mana sentimen negatif mendominasi dibandingkan sentimen netral dan positif. Kondisi ini menjadi pertimbangan penting dalam tahap penanganan ketidakseimbangan kelas.

### 2.5 Pembobotan TF-IDF

Metode Term Frequency–Inverse Document Frequency (TF-IDF) digunakan untuk mengubah data teks hasil *preprocessing* menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi. TF-IDF mengukur tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap keseluruhan dokumen dalam korpus [20]. Representasi TF-IDF menghasilkan matriks sparse yang efisien dan digunakan sebagai input utama dalam proses klasifikasi sentimen. Secara matematis, nilai TF dihitung menggunakan Persamaan (1), IDF menggunakan Persamaan (2), dan nilai TF-IDF diperoleh dari hasil perkalian TF dan IDF seperti ditunjukkan pada Persamaan (3). Pendekatan ini memungkinkan kata-kata yang sering muncul namun kurang informatif memperoleh bobot rendah, sementara kata-kata yang lebih spesifik dan relevan terhadap konteks pembahasan memperoleh bobot lebih tinggi. Hasil pembobotan TF-IDF membentuk matriks fitur berbentuk *sparse matrix* yang selanjutnya digunakan sebagai masukan pada proses klasifikasi sentimen menggunakan algoritma Multinomial Naïve Bayes.

$$TF(t, d) = \frac{f_{t,d}}{\sum_k f_{k,d}} \quad (1)$$

Dengan  $f_{t,d}$  adalah kemunculan kata  $t$  pada dokumen  $d$ ,  $\sum_k f_{k,d}$  adalah jumlah seluruh kata dalam dokumen  $d$ .

$$IDF(t) = \log\left(\frac{N}{df_t}\right) \quad (2)$$

Dengan  $N$  adalah jumlah total dokumen dalam korpus,  $df_t$  adalah jumlah dokumen yang mengandung kata  $t$ .

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (3)$$

Hasil pembobotan TF-IDF membentuk matriks fitur berbentuk *sparse matrix*, di mana setiap baris merepresentasikan satu komentar YouTube dan setiap kolom merepresentasikan satu kata unik hasil preprocessing. Matriks inilah yang digunakan sebagai masukan pada proses klasifikasi sentimen menggunakan algoritma *Multinomial Naïve Bayes*.

## 2.6 Penanganan Ketidakseimbangan Kelas

Distribusi data sentimen pada penelitian ini menunjukkan adanya ketidakseimbangan kelas (*class imbalance*), di mana kelas sentimen negatif mendominasi secara signifikan dibandingkan kelas netral dan positif. Ketidakseimbangan ini berpotensi menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas dan mengabaikan pola pada kelas minoritas. Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan dua teknik resampling, yaitu *Random Oversampling* (ROS) dan *Synthetic Minority Over-sampling Technique* (SMOTE).

*Random Oversampling* (ROS) dilakukan dengan menyalin secara acak data dari kelas minoritas hingga jumlahnya setara dengan kelas mayoritas. Teknik ini sederhana dan efektif dalam meningkatkan representasi kelas minoritas, namun berpotensi menyebabkan *overfitting* karena duplikasi data [10].

*Synthetic Minority Over-sampling Technique* (SMOTE) menghasilkan sampel sintesis baru pada kelas minoritas dengan cara melakukan interpolasi antar data yang berdekatan dalam ruang fitur. Berbeda dengan ROS, SMOTE tidak sekadar menduplikasi data, tetapi menciptakan variasi baru sehingga mampu memperkaya representasi kelas minoritas [21].

Proses resampling dilakukan setelah tahap pembobotan TF-IDF dan sebelum proses pelatihan model *Multinomial Naïve Bayes*. Evaluasi dilakukan dengan membandingkan performa model tanpa resampling, dengan SMOTE, dan dengan *Random Oversampling* menggunakan metrik *accuracy*, *precision*, *F1-score macro*, dan *F1-score weighted*.

## 2.7 Klasifikasi Multinomial Naïve Bayes

Klasifikasi sentimen pada penelitian ini menggunakan algoritma *Multinomial Naïve Bayes*, yaitu metode probabilistik yang menghitung peluang suatu dokumen termasuk ke dalam kelas sentimen tertentu berdasarkan distribusi kata yang muncul di dalamnya. Model Naïve Bayes didasarkan pada Teorema Bayes yang dinyatakan pada Persamaan (4), di mana probabilitas kelas  $c$  terhadap dokumen  $d$  dihitung berdasarkan probabilitas awal kelas dan probabilitas kemunculan kata-kata dalam dokumen tersebut. Untuk data teks yang direpresentasikan menggunakan bobot TF-IDF, probabilitas kelas dihitung menggunakan pendekatan Multinomial Naïve Bayes seperti ditunjukkan pada Persamaan (5) [22]. Kelas sentimen yang memiliki nilai probabilitas posterior terbesar akan dipilih sebagai hasil klasifikasi sebagaimana dinyatakan pada Persamaan (6). Pendekatan ini dipilih karena efisien secara komputasi dan efektif dalam menangani data teks berdimensi tinggi, sehingga sesuai digunakan sebagai model *baseline* dalam analisis sentimen komentar YouTube.

$$P(c|d) = \frac{P(c)P(d|c)}{P(d)} \quad (4)$$

$$P(d|c) = \prod_{i=1}^n P(w_i | c) \quad (5)$$

$$\hat{C} = \arg \max_{c \in C} P(c) \prod_{i=1}^n P(w_i | c) \quad (6)$$

Pendekatan ini dipilih karena efisien secara komputasi dan efektif dalam menangani data teks berdimensi tinggi, sehingga sesuai digunakan sebagai model *baseline* dalam analisis sentimen komentar YouTube.

## 2.8 Confusion Matrix

*Confusion matrix* digunakan sebagai alat evaluasi untuk menggambarkan perbandingan antara label aktual dan label hasil prediksi model pada setiap kelas sentimen. Matriks ini menyajikan jumlah prediksi benar (*true positive*) maupun kesalahan klasifikasi (*false positive* dan *false negative*) untuk masing-masing kategori sentimen, sehingga memungkinkan analisis yang lebih mendalam terhadap pola kesalahan model.

Pada model *baseline* tanpa penanganan ketidakseimbangan kelas, *confusion matrix* menunjukkan bahwa sebagian besar komentar berhasil diprediksi sebagai sentimen negatif, yang merupakan kelas mayoritas dalam dataset. Namun demikian, sejumlah komentar dengan label netral dan positif masih sering diprediksi sebagai negatif. Kondisi ini mengindikasikan adanya bias model terhadap kelas mayoritas, yang disebabkan oleh distribusi data yang tidak seimbang.

Analisis *confusion matrix* memperlihatkan bahwa meskipun nilai akurasi tergolong tinggi, kemampuan model dalam mengenali kelas minoritas masih terbatas. Hal ini sejalan dengan rendahnya nilai *F1-score macro* pada model *baseline*. Oleh karena itu, *confusion matrix* menjadi alat evaluasi yang penting untuk memahami secara rinci bagaimana ketidakseimbangan kelas memengaruhi kinerja model klasifikasi sentimen.

### 3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian dan pembahasan terkait analisis sentimen publik terhadap Program Makan Bergizi Gratis (MBG) berdasarkan komentar pengguna pada platform YouTube. Analisis dilakukan untuk mengelompokkan opini masyarakat ke dalam tiga kategori sentimen, yaitu negatif, netral, dan positif, menggunakan algoritma *Multinomial Naïve Bayes* dengan pembobotan TF-IDF. Hasil yang diperoleh tidak hanya menggambarkan kecenderungan persepsi publik terhadap implementasi program MBG, tetapi juga mengevaluasi kinerja metode klasifikasi yang digunakan, termasuk dampak ketidakseimbangan kelas terhadap performa model.

#### 3.1 Pengumpulan Data

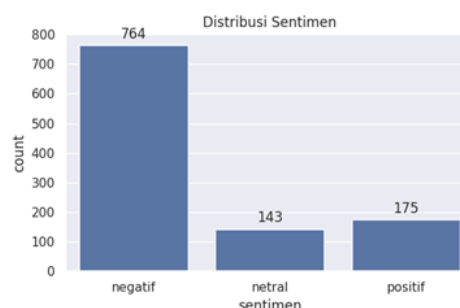
Data penelitian diperoleh melalui proses pengambilan komentar dari platform YouTube menggunakan YouTube Data API v3. Pengambilan data difokuskan pada sepuluh video yang membahas Program Makan Bergizi Gratis (MBG), yang dipilih berdasarkan relevansi topik serta tingkat interaksi pengguna, seperti jumlah komentar dan penayangan. Dari proses ini diperoleh sebanyak 1.220 komentar mentah.

Tahap selanjutnya adalah penghapusan duplikasi komentar untuk menghilangkan teks yang berulang akibat aktivitas *spam* atau pengulangan oleh akun yang sama. Proses ini penting untuk menjaga validitas distribusi data serta mencegah dominasi opini tertentu akibat komentar yang identik. Setelah dilakukan penyaringan dan penghapusan duplikasi, jumlah data yang digunakan dalam penelitian menjadi 1.082 komentar unik.

Berdasarkan hasil pelabelan manual terhadap 1.082 komentar tersebut, diperoleh distribusi sentimen sebagaimana ditunjukkan pada Tabel 1. Sebanyak 764 komentar (70,61%) dikategorikan sebagai sentimen negatif, 175 komentar (16,17%) sebagai sentimen positif, dan 143 komentar (13,22%) sebagai sentimen netral. Distribusi ini menunjukkan bahwa mayoritas komentar mengandung kritik atau kekhawatiran terhadap implementasi Program MBG.

**Tabel 1.** Distribusi Sentimen Komentar YouTube

No	Kategori Sentimen	Jumlah komentar	Persentase
1	Negatif	764	70,61%
2	Netral	143	13,22%
3	Positif	175	16,17%
	Total	1082	100%



**Gambar 2.** Diagram Batang Hasil Pengumpulan Data

Visualisasi distribusi sentimen ditunjukkan pada Gambar 2 dalam bentuk diagram batang. Terlihat bahwa kelas sentimen negatif mendominasi secara signifikan dibandingkan dua kelas lainnya. Ketimpangan distribusi ini mengindikasikan adanya class imbalance yang berpotensi memengaruhi performa model klasifikasi. Model pembelajaran mesin cenderung memprioritaskan kelas mayoritas selama proses pelatihan, sehingga berisiko menghasilkan prediksi yang bias terhadap sentimen negatif.

Dominasi sentimen negatif juga memberikan gambaran awal mengenai persepsi publik terhadap Program MBG. Tingginya proporsi komentar bernada kritik atau kekhawatiran menunjukkan bahwa implementasi program masih memunculkan perdebatan di ruang publik digital. Namun demikian, keberadaan sentimen positif dan netral mengindikasikan adanya variasi opini yang perlu dianalisis lebih lanjut melalui pendekatan klasifikasi otomatis.

Kondisi ketidakseimbangan kelas inilah yang menjadi dasar pertimbangan penerapan teknik resampling pada tahap selanjutnya, guna mengevaluasi apakah penanganan class imbalance dapat meningkatkan kemampuan model dalam mengenali sentimen minoritas secara lebih adil.

#### 3.2 Preprocessing

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam proses ekstraksi fitur dan klasifikasi sentimen. Proses ini bertujuan mengurangi *noise*, menyeragamkan struktur teks, serta meningkatkan stabilitas representasi fitur. Tahapan yang dilakukan meliputi *cleaning*, *case folding*, normalisasi, tokenisasi, *stopword removal*, dan *stemming*.

Pada tahap *cleaning*, elemen non-informatif seperti URL, emoji, simbol, angka yang tidak relevan, serta tanda baca berlebih dihapus dari teks. Proses ini menghasilkan komentar yang lebih fokus pada isi opini pengguna. Selanjutnya, *case folding* dilakukan dengan mengubah seluruh huruf menjadi huruf kecil untuk menghindari perbedaan token akibat variasi kapitalisasi.

Tahap normalisasi bertujuan mengonversi kata tidak baku atau singkatan menjadi bentuk baku bahasa Indonesia, sehingga variasi kata dengan makna serupa dapat diseragamkan. Setelah itu, tokenisasi memecah teks menjadi unit kata (token) yang akan digunakan sebagai dasar pembentukan fitur. Proses *stopword removal* kemudian menghilangkan kata-kata umum yang tidak memiliki kontribusi signifikan terhadap makna sentimen. Tahap terakhir, *stemming*, mengubah kata berimbuhan menjadi bentuk dasar sehingga variasi kata turunan dapat disatukan dalam satu representasi fitur.

Secara keseluruhan, tahapan *preprocessing* ini berkontribusi dalam meningkatkan konsistensi data serta mengurangi redundansi fitur. Dengan teks yang lebih terstruktur, proses pembobotan TF-IDF dapat menghasilkan representasi numerik yang lebih akurat dalam menangkap pola sentimen.

**Tabel 2.** Hasil *Preprocessing* Komentar YouTube

Tahapan	Kategori Sentimen
Teks Asli	Ini pelajaran masak,jangan asal asal
<i>Cleaning</i>	Ini pelajaran masak jangan asal asal
<i>Case Folding</i>	ini pelajaran masak jangan asal asal
Tokenisasi	['ini', 'pelajaran', 'masak', 'jangan', 'asal', 'asal']
<i>Stopword Removal</i>	['ini', 'pelajaran', 'masak', 'jangan', 'asal', 'asal']
<i>Stemming</i>	['ajar', 'masak']

### 3.3 Pelabelan data sentimen

Pelabelan data sentimen dilakukan secara manual oleh satu orang ahli bahasa dengan mengelompokkan setiap komentar YouTube ke dalam tiga kategori, yaitu negatif, netral, dan positif. Proses pelabelan mempertimbangkan konteks kalimat, pilihan kata, serta kecenderungan opini yang disampaikan pengguna terhadap Program Makan Bergizi Gratis (MBG). Pendekatan manual dipilih untuk memastikan kualitas anotasi data sebelum digunakan dalam proses pelatihan model klasifikasi.

Hasil pelabelan terhadap 1.082 komentar menunjukkan bahwa sentimen negatif mendominasi dengan jumlah 764 komentar (70,61%), diikuti oleh sentimen positif sebanyak 175 komentar (16,17%), dan sentimen netral sebanyak 143 komentar (13,22%). Dominasi sentimen negatif mengindikasikan bahwa sebagian besar komentar mengandung kritik, kekhawatiran, atau keraguan terhadap implementasi Program MBG. Sebaliknya, sentimen positif mencerminkan adanya dukungan dan apresiasi masyarakat terhadap tujuan program, sedangkan sentimen netral umumnya berisi tanggapan informatif atau pernyataan deskriptif tanpa kecenderungan emosional yang kuat.

Distribusi ini menunjukkan adanya ketidakseimbangan kelas (*class imbalance*) yang cukup signifikan, di mana proporsi sentimen negatif jauh lebih tinggi dibandingkan dua kelas lainnya. Kondisi ini berpotensi memengaruhi proses pelatihan model klasifikasi, karena algoritma cenderung memprioritaskan pola pada kelas mayoritas. Oleh karena itu, distribusi sentimen ini menjadi dasar pertimbangan dalam penerapan teknik penanganan ketidakseimbangan kelas pada tahap selanjutnya.

### 3.4 Hasil Pembobotan TF-IDF

Setelah melalui tahap *preprocessing*, seluruh komentar diubah menjadi representasi numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Proses ini menghasilkan matriks fitur berbentuk *sparse matrix*, di mana setiap baris merepresentasikan satu komentar dan setiap kolom merepresentasikan satu kata unik hasil *preprocessing*. Representasi numerik ini memungkinkan algoritma *Multinomial Naïve Bayes* memproses data teks secara efisien dalam ruang fitur berdimensi tinggi.

Berdasarkan hasil pembobotan, beberapa kata dengan nilai TF-IDF tertinggi antara lain “makan”, “mbg”, “program”, “gratis”, “anak”, dan “sekolah”, sebagaimana ditunjukkan pada Tabel 3. Kata-kata tersebut mencerminkan topik utama yang sering dibahas dalam komentar terkait Program Makan Bergizi Gratis. Tingginya bobot pada kata-kata tersebut menunjukkan bahwa diskusi publik berfokus pada aspek inti kebijakan, yaitu pemberian makanan gratis bagi anak sekolah.

Selain kata-kata yang bersifat deskriptif terhadap program, muncul pula kata-kata bernuansa kritis seperti “korupsi” dan “korban” dengan bobot yang relatif tinggi. Kemunculan kata-kata tersebut mengindikasikan adanya kekhawatiran publik terhadap transparansi, pengelolaan anggaran, atau potensi permasalahan dalam implementasi program. Pola ini sejalan dengan dominasi sentimen negatif yang telah diidentifikasi pada tahap pelabelan.

**Tabel 3.** Kata dengan Nilai TF-IDF Tertinggi

Ranking	Kata	Hasil TF-IDF
1	makan	0.126421
2	mbg	0.114286

Ranking	Kata	Hasil TF-IDF
3	program	0.090409
4	gratis	0.083387
5	anak	0.080107
6	sekolah	0.062013

Analisis terhadap hasil pembobotan TF-IDF menunjukkan bahwa fitur yang dihasilkan mampu merepresentasikan konteks kebijakan yang sedang dibahas secara relevan. Hal ini penting karena kualitas fitur sangat memengaruhi performa model klasifikasi. Representasi fitur yang baik memungkinkan model mengenali pola distribusi kata yang berkaitan dengan kategori sentimen tertentu.

Dengan demikian, hasil pembobotan TF-IDF tidak hanya berfungsi sebagai input numerik bagi algoritma klasifikasi, tetapi juga memberikan gambaran awal mengenai tema dominan dan kecenderungan opini dalam diskusi publik terkait Program MBG.

Meskipun kata seperti “makan”, “mbg”, dan “program” muncul dominan dalam Word Cloud pada seluruh kelas sentimen, kata-kata tersebut memiliki nilai IDF yang relatif rendah karena frekuensi kemunculannya tinggi di hampir seluruh dokumen. Dalam pembobotan TF-IDF, kata dengan kemunculan luas pada banyak dokumen cenderung memiliki daya diskriminatif yang rendah terhadap kategori sentimen tertentu. Oleh karena itu, meskipun kata-kata tersebut tampak dominan secara visual, proses klasifikasi Naïve Bayes lebih banyak dipengaruhi oleh kata-kata yang memiliki bobot lebih spesifik dan kontekstual, seperti “korupsi” pada sentimen negatif atau “bagus” dan “setuju” pada sentimen positif.

### 3.5 Klasifikasi Naïve Bayes

Proses klasifikasi sentimen dilakukan menggunakan algoritma *Multinomial Naïve Bayes* dengan beberapa skema pembagian data latih dan data uji (*train–test split*), yaitu 90%:10%, 80%:20%, 70%:30%, 60%:40%, dan 50%:50%. Pembagian data dilakukan secara *stratified* untuk menjaga proporsi distribusi kelas sentimen pada data latih maupun data uji, sehingga tidak terjadi distorsi distribusi selama proses pelatihan dan evaluasi model.

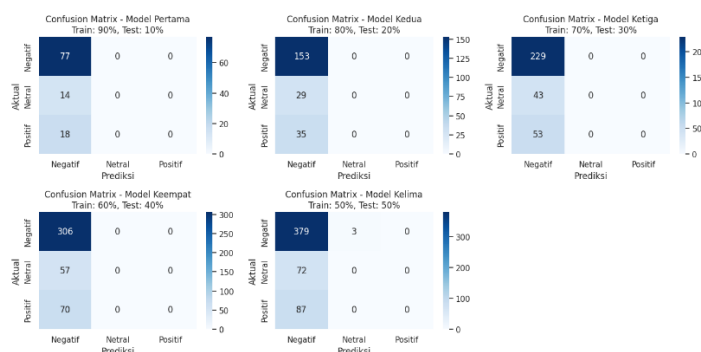
Hasil pengujian menunjukkan bahwa performa model relatif stabil pada seluruh skema pembagian data. Nilai akurasi berada pada kisaran 70%, dengan variasi yang tidak signifikan antar skema. Skema pembagian data 60%:40% menghasilkan performa terbaik dengan nilai akurasi sebesar 70,67% dan *F1-score weighted* sebesar 58,52%. Stabilitas ini menunjukkan bahwa model memiliki konsistensi dalam mempelajari pola distribusi kata pada data komentar YouTube.

Namun demikian, nilai *F1-score macro* yang diperoleh pada seluruh skema relatif rendah, dengan nilai tertinggi sebesar 27,60%. Perbedaan yang cukup jauh antara *F1-score weighted* dan *F1-score macro* menunjukkan adanya ketimpangan performa antar kelas. Nilai *weighted* yang lebih tinggi dipengaruhi oleh dominasi kelas negatif dalam dataset, sedangkan nilai *macro* yang rendah mengindikasikan bahwa model masih mengalami kesulitan dalam mengenali kelas minoritas, yaitu sentimen netral dan positif.

Kondisi ini mengindikasikan bahwa meskipun model mampu mencapai tingkat akurasi yang cukup baik secara keseluruhan, performa tersebut belum sepenuhnya merepresentasikan kemampuan model dalam mengklasifikasikan seluruh kategori sentimen secara seimbang. Bias terhadap kelas mayoritas menjadi faktor utama yang memengaruhi hasil evaluasi ini.

**Tabel 4.** Hasil Evaluasi Model *Multinomial Naïve Bayes*

Model	Train: test	Accuracy	F1-score (Macro)	F1-score (Weighted)
1	90:10	70.64%	27.10%	58.20%
2	80:20	70.51%	27.30%	58.10%
3	70:30	70.46%	27.40%	58.30%
4	60:40	70.67%	27.60%	58.52%
5	50:50	70.06%	26.90%	57.80%



**Gambar 3.** Confusion Matrix Model *Multinomial Naïve Bayes* Terbaik



### 3.8 Word Cloud Sentimen Netral

Word cloud sentimen netral ditunjukkan pada Gambar 6 dan merepresentasikan kata-kata yang sering muncul pada komentar yang bersifat informatif atau deskriptif tanpa kecenderungan emosional yang kuat. Kata-kata seperti “program”, “makan”, “gratis”, “anak”, dan “sekolah” mendominasi visualisasi ini.



Gambar 6. Word Cloud Sentimen Netral

Dominasi kata-kata tersebut menunjukkan bahwa komentar netral umumnya berisi informasi, tanggapan umum, atau diskusi terkait isi kebijakan tanpa disertai dukungan maupun penolakan yang eksplisit. Hal ini mengindikasikan bahwa sebagian masyarakat menggunakan kolom komentar YouTube sebagai sarana berbagi informasi atau menanggapi konten video secara objektif.

### 3.9 Analisis Penanganan Ketidakseimbangan Kelas (Baseline vs SMOTE vs ROS)

Untuk mengatasi permasalahan ketidakseimbangan kelas yang teridentifikasi pada model *baseline*, dilakukan penerapan dua teknik resampling, yaitu *Synthetic Minority Over-sampling Technique* (SMOTE) dan *Random Oversampling* (ROS). Evaluasi dilakukan dengan membandingkan performa model tanpa *resampling* (*baseline*), dengan SMOTE, dan dengan ROS menggunakan metrik *accuracy*, *F1-score macro*, dan *F1-score weighted*. Hasil perbandingan performa model ditunjukkan pada Tabel 5.

Tabel 5. Perbandingan Performa Model *Baseline*, SMOTE, dan ROS

Model	Accuracy	F1-score (Macro)	F1-score (Weighted)
Baseline (tanpa resampling)	70.67%	27.60%	58.52%
SMOTE	51.27%	36.24%	53.35%
Random Oversampling (ROS)	50.58%	37.09%	53.25%

Penurunan akurasi sebesar sekitar 20% setelah penerapan teknik resampling terlihat signifikan secara numerik. Namun demikian, interpretasi performa model pada dataset dengan ketidakseimbangan kelas tidak dapat hanya didasarkan pada akurasi global. Pada model *baseline*, akurasi tinggi (70,67%) terutama dipengaruhi oleh dominasi kelas negatif yang mencapai lebih dari 70% dari keseluruhan data. Dalam kondisi ini, model cenderung memprediksi sebagian besar komentar sebagai negatif, sehingga kemungkinan besar memiliki recall yang sangat rendah pada kelas netral dan positif. Dengan kata lain, model *baseline* berisiko gagal mendeteksi sentimen minoritas secara memadai meskipun menghasilkan akurasi tinggi.

Sebaliknya, model dengan resampling menunjukkan peningkatan signifikan pada *F1-score macro*, yang merepresentasikan rata-rata performa pada setiap kelas secara seimbang. Meskipun akurasi menurun menjadi sekitar 51%, peningkatan *F1-score macro* hingga 36–37% menunjukkan bahwa model menjadi lebih sensitif dalam mengenali kelas netral dan positif. Dalam konteks evaluasi kebijakan publik, kemampuan mendeteksi sinyal kritik maupun dukungan dari kelompok minoritas jauh lebih bernilai dibandingkan akurasi tinggi yang bias terhadap satu kategori saja. Oleh karena itu, model dengan resampling dinilai lebih representatif untuk memetakan persepsi publik secara menyeluruh.

### 3.10 Analisis Performa

Model *baseline* menghasilkan akurasi tertinggi sebesar 70,67%. Namun, nilai *F1-score macro* yang rendah (27,60%) menunjukkan bahwa model belum mampu mengenali kelas minoritas secara optimal. Tingginya akurasi pada *baseline* dipengaruhi oleh dominasi kelas negatif dalam dataset, sehingga model cenderung memprediksi komentar sebagai negatif.

Setelah penerapan SMOTE, nilai akurasi menurun menjadi 51,27%, tetapi *F1-score macro* meningkat menjadi 36,24%. Peningkatan ini menunjukkan bahwa model menjadi lebih seimbang dalam mengenali ketiga kelas sentimen. SMOTE membantu memperkaya representasi kelas minoritas dengan menghasilkan sampel sintetis, sehingga model tidak lagi terlalu bergantung pada pola kelas mayoritas. Penerapan *Random Oversampling* (ROS) menghasilkan akurasi sebesar 50,58% dengan *F1-score macro* tertinggi sebesar 37,09%. Meskipun akurasi sedikit lebih rendah

dibandingkan SMOTE, nilai *macro* yang lebih tinggi menunjukkan bahwa ROS memberikan keseimbangan performa yang lebih baik antar kelas. Hal ini mengindikasikan bahwa penduplikasian data minoritas secara langsung mampu meningkatkan sensitivitas model terhadap sentimen netral dan positif.

### 3.11 Trade-off Accuracy vs Class Fairness

Hasil ini menunjukkan adanya trade-off antara akurasi keseluruhan dan keseimbangan performa antar kelas. Model baseline memiliki akurasi tinggi namun menunjukkan bias signifikan terhadap kelas mayoritas, sehingga kurang efektif dalam mengenali opini netral dan positif. Sebaliknya, model dengan resampling memiliki akurasi lebih rendah tetapi menunjukkan peningkatan performa yang lebih merata antar kelas, sebagaimana tercermin dari kenaikan F1-score *macro*. Dalam konteks analisis sentimen kebijakan publik, model yang mampu menangkap variasi opini secara lebih adil dinilai lebih relevan dibandingkan model dengan akurasi tinggi namun mengabaikan sebagian kelompok opini masyarakat.

Dalam konteks analisis sentimen kebijakan publik, keseimbangan performa antar kelas menjadi aspek penting karena setiap kategori sentimen memiliki nilai informasi yang relevan untuk evaluasi kebijakan. Oleh karena itu, meskipun akurasi baseline lebih tinggi, model dengan teknik resampling dapat dianggap lebih representatif dalam memetakan persepsi publik secara menyeluruh.

### 3.12 Implikasi Metodologis

Temuan ini menegaskan bahwa ketidakseimbangan kelas berpengaruh signifikan terhadap performa model klasifikasi sentimen berbasis Naïve Bayes. Penerapan teknik resampling terbukti meningkatkan kemampuan model dalam mengenali kelas minoritas, meskipun berdampak pada penurunan akurasi global. Dengan demikian, pemilihan model terbaik perlu mempertimbangkan tujuan analisis, apakah lebih mengutamakan akurasi keseluruhan atau keseimbangan performa antar kategori sentimen.

Secara keseluruhan, *Random Oversampling* menghasilkan nilai *F1-score macro* tertinggi dan menunjukkan keseimbangan performa yang lebih baik dibandingkan SMOTE dan *baseline*, sehingga dapat dipertimbangkan sebagai pendekatan yang lebih optimal dalam konteks penelitian ini.

## 4. KESIMPULAN

Penelitian ini menganalisis persepsi publik terhadap Program Makan Bergizi Gratis (MBG) melalui klasifikasi sentimen komentar YouTube menggunakan algoritma Multinomial Naïve Bayes dengan pembobotan TF-IDF. Hasil pemetaan menunjukkan bahwa mayoritas komentar (70,61%) mengandung sentimen negatif, yang mengindikasikan tingginya tingkat kritik dan kekhawatiran masyarakat terhadap implementasi program. Model baseline menghasilkan akurasi 70,67%, namun nilai F1-score *macro* yang rendah (27,60%) menunjukkan adanya bias terhadap kelas mayoritas dan keterbatasan dalam mengenali sentimen minoritas. Penerapan teknik resampling (SMOTE dan Random Oversampling) meningkatkan F1-score *macro* hingga 36–37%, meskipun akurasi menurun menjadi sekitar 51%. Secara akademik, penurunan akurasi tersebut mencerminkan pergeseran fokus model dari dominasi kelas mayoritas menuju performa yang lebih seimbang antar kelas. Dalam konteks evaluasi kebijakan publik, kemampuan mendeteksi variasi opini, baik kritik, dukungan, maupun tanggapan netral—lebih bernilai dibandingkan akurasi tinggi yang hanya merepresentasikan satu sisi persepsi. Oleh karena itu, pendekatan dengan penanganan ketidakseimbangan kelas memberikan hasil pemetaan opini yang lebih representatif dan relevan sebagai bahan evaluasi implementasi Program MBG.

## REFERENCES

- [1] A. Wulandari, “Model Kyushoku Jepang dan Tantangan Implementasi Program Makan Bergizi Gratis di Indonesia,” *J. Pendidik. Tambusai*, vol. 9, no. 1, pp. 6948–6956, 2025, [Online]. Available: <https://jptam.org/index.php/jptam/issue/view/30>
- [2] R. Rimbawan, R. Nurdiani, P. H. Rachman, Y. Kawamata, and Y. Nozawa, “School Lunch Programs and Nutritional Education Improve Knowledge, Attitudes, and Practices and Reduce the Prevalence of Anemia: A Pre-Post Intervention Study in an Indonesian Islamic Boarding School,” *Nutrients*, vol. 15, no. 4, 2023, doi: 10.3390/nu15041055.
- [3] R. Nida and D. D. P. Sari, “School Meals Program and Its Impact Towards Student’s Cognitive Achievement,” *J. Econ. Res. Soc. Sci.*, vol. 7, no. 1, pp. 69–80, 2023, doi: <https://doi.org/10.18196/jerss.v7i1.17014>.
- [4] M. A. Muafa, C. Wahyudin, E. Salbiah, and O. Subagdja, “Efektivitas Program Pemberian Makanan Tambahan Pada Anak Stunting,” *Karimah Tauhid*, vol. 3, no. 4, pp. 4947–4953, 2024, doi: <https://doi.org/10.30997/karimahtauhid.v3i4.12978>.
- [5] A. Kusuma and A. Nugroho, “Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes,” *J. Ilm. Teknol. Inf. Asia*, vol. 15, no. 2, pp. 137–146, 2021, doi: 10.32815/jitika.v15i2.557.
- [6] C. A. Wicaksana, M. Fatkhurrohman, H. P. Pratama, R. Tryawan, Alimuddin, and R. Febriani, “Twitter Sentiment Analysis in Indonesian Language using Naive Bayes Classification Method,” *Proc. 2022 Int. Conf. Informatics, Electr. Electron.*, pp. 282–287, 2022, doi: 10.1109/ICIEE55596.2022.10010002.
- [7] I. Permana and K. D. Maani, “Publication Trend of Public Sentiment Towards Indonesia Government Policies,” *Sinkron*, vol. 8, no. 3, pp. 2061–2069, 2024, doi: 10.33395/sinkron.v8i3.13843.
- [8] A. Y. Pratama, G. A. Sanjaya, N. K. Lubis, and M. R. Aditya, “Analisis Sentimen Publik Terkait Danantara Menggunakan



- Algoritma IndoBERT pada Platform Media Sosial,” *METIK J.*, vol. 9, no. 1, 2025, doi: 10.47002/metik.v9i1.1055.
- [9] M. Zhikri and W. Istiono, “Handling Class Imbalance for Indonesian Twitter Sentiment Analysis A Comparative Study of Algorithms,” *J. Syst. Manag. Sci.*, vol. 14, no. 10, pp. 170–179, 2024, doi: 10.33168/jsms.2024.1010.
- [10] M. Al, G. Muttaqin, and G. A. Trisnopradi, “Optimasi Algoritma SVM dengan Teknik SMOTE dan Tuning Parameter pada Klasifikasi Balita Stunting,” *Build. Informatics, Technol. Sci.*, vol. 7, no. 3, pp. 1547–1556, 2025, doi: 10.47065/bits.v7i3.8330.
- [11] Bahrun and Wildan, “Stunting in Indonesian Children and Its Contributing Factors: Study through Bibliometric Analysis,” *JPUD - J. Pendidik. Usia Dini*, vol. 16, no. 2, pp. 271–293, 2022, doi: 10.21009/jpud.162.07.
- [12] A. Alrehaili, A. Alsaedi, and W. M. S. Yafooz, “Sentiment analysis of YouTube videos comments for children using machine learning and deep learning,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 40, no. 1, pp. 397–410, 2025, doi: 10.11591/ijeecs.v40.i1.pp397-410.
- [13] R. Rahmatulloh, M. I. Ibrahim, and M. R. Handayani, “Model Klasifikasi Naive Bayes untuk Pemetaan Persepsi Publik Secara Real-Time pada Media Sosial : Studi Kasus RUU TNI 2025,” *J. Pendidik. Teknol. Inf.*, vol. 5, no. 2, pp. 365–379, 2025, doi: <http://dx.doi.org/10.51454/decode.v5i2.1139>.
- [14] M. K. F. Mawar Hardiyanti, “Optimasi Analisis Sentimen Komentar Penonton Wayang Digital dengan SMOTE dan Algoritma Naive Bayes Optimizing Sentiment Analysis of Digital Wayang Viewer Comments using SMOTE and the Naive Bayes Algorithm,” *J. Sist. Inf.*, vol. 14, no. 3, pp. 1154–1164, 2025, doi: <https://doi.org/10.32520/stmsi.v14i3>.
- [15] Y. A. Mahmood and B. Mahmood, “A Web Scraper for Data Mining Purposes,” *Sistemasi*, vol. 13, no. 3, p. 1243, 2024, doi: 10.32520/stmsi.v13i3.4107.
- [16] I. R. I. A. Arrizqi Fauzy Aufar, Mochamad Alfian Rosid, Ade Eviyanti, “Optimizing Text Preprocessing for Accurate Sentiment Analysis on E-Wallet Reviews,” *J. Inf. Comput. Technol. Educ.*, vol. 7, no. 2, pp. 42–50, 2023, doi: 10.21070/jicte.v7i2.1650.
- [17] A. S. N. Handoko, A. Asrofiq, Junadhi, “Sentiment Analysis of Sirekap Tweets Using CNN Algorithm,” *J. Ilm. Penelit. dan Penerapan Teknol. Sist. Inf.*, vol. 8, no. 2, pp. 312–329, 2024, doi: <https://doi.org/10.29407/intensif.v8i2.23046>.
- [18] L. Ellyanti, Y. Ruldeviyani, L. E. Pradana, and A. Harjanto, “Sentiment Analysis of Twitter Users to the PeduliLindungi Using Naive Bayes Algorithm,” *Rekayasa Sist. dan Teknol. Inf.*, vol. 7, no. 2, pp. 414–421, 2023, doi: <https://doi.org/10.20473/jisebi.9.1.101-110>.
- [19] D. A. P. R. D. Ni Wayan Sumartini Saraswati, Christina Purnama Yanti, I Dewa Made Krihna Muku, “Evaluation Analysis of the Necessity of Stemming and Lemmatization in Text Classification,” vol. 24, no. 2, pp. 321–332, 2025, doi: 10.30812/matrik.v24i2.4833.
- [20] J. A. Putra, A. Dharmawan, and J. Gondohanindijo, “Sentimen Analisis Aplikasi Digital Mobile Menggunakan Naive Bayes dan Svm dengan Ekstraksi Fitur TF-IDF,” *J. Inf. Technol. Comput. Sci.*, vol. 7, no. 4, pp. 1139–1148, 2024, doi: <https://doi.org/10.31539/intecom.v7i4>.
- [21] C. S. Wildanil Khozi1, Jasim Nadheer Hussein, Ramadhan Rakhmat Sani, Fauzi Adi Rafrastara, Cinantya Paramita, “Mitigating Class Imbalance in DDoS Detection: The Impact of Random Over Sampling on Machine Learning Performance,” *ELKHA J. Tek. Elektro*, vol. 17, no. 2, pp. 109–117, 2025, doi: <https://doi.org/10.26418/elkha.v17i2.84652>.
- [22] W. B. Zulfikar, A. R. Atmadja, and S. F. Pratama, “Sentiment Analysis on Social Media Against Public Policy Using Multinomial Naive Bayes,” *Sci. J. Informatics*, vol. 10, no. 1, pp. 25–34, 2023, doi: 10.15294/sji.v10i1.39952.