

Penerapan Multinomial Naïve Bayes dan Chi-Square pada Analisis Sentimen Makan Bergizi Gratis

Ni Putu Alya Noviyanti^{a1}, I Gusti Ngurah Anom Cahyadi Putra^{a2}

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
¹noviyanti.2308561064@student.unud.ac.id
²anom.cp@unud.ac.id

Abstract

The Free Nutritious Meal Program (MBG) is a policy initiated by the government in order to create a healthy superior generation. This policy has generated various responses from the public, both positive, negative, and neutral. This study aims to analyze public sentiment towards the program by utilizing 2.165 comments from YouTube. Data were analyzed using the Multinomial Naïve Bayes algorithm after going through the labeling, text preprocessing, and TF-IDF representation stages. Feature selection was carried out with 10 threshold values of the Chi-Square method to evaluate the effect of the number of features on model performance. Hyperparameter tuning with alpha and fit_prior was carried out through Grid Search with 5-fold cross validation. The results showed that the best performance was achieved when the threshold Chi-Square was 100% or when all features were used, which resulted in an accuracy of 83.48% and a macro F1 score of 83.42%. Visualization of the sentiment distribution shows the dominance of negative sentiment at 35.6%, followed by positive at 34.9%, and neutral at 29.5%, indicating public dissatisfaction with the policy.

Keywords: Free Nutritious Meal Program, Sentiment Analysis, Multinomial Naïve Bayes, Chi-Square, TF-IDF, Hyperparameter Tuning

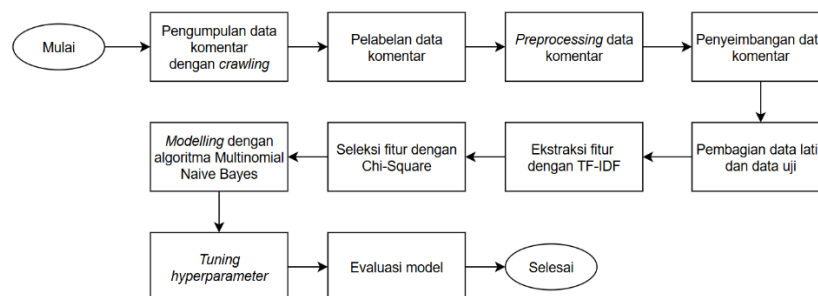
1. Pendahuluan

Peningkatan kualitas sumber daya manusia merupakan fondasi utama dalam mewujudkan visi Indonesia Emas 2045. Untuk mendukung hal tersebut, pemerintah Indonesia terus menggulirkan berbagai kebijakan strategis, salah satunya adalah Program Makan Bergizi Gratis (MBG) yang menyasar siswa PAUD hingga SMA, serta ibu hamil dan menyusui [1]. Program ini mulai dilaksanakan secara serentak pada 6 Januari 2025 dengan target penerima mencapai 82,9 juta orang. Implementasinya dimulai dari 26 provinsi dan kian akan menjangkau seluruh 38 provinsi di Indonesia. Dalam pelaksanaannya, MBG juga melibatkan pelaku UMKM lokal dalam proses pengadaan dan distribusi makanan, yang diharapkan dapat mendorong pertumbuhan ekonomi daerah serta memperkuat ketahanan pangan nasional. Secara strategis, MBG tidak hanya bertujuan untuk memenuhi kebutuhan gizi masyarakat, tetapi juga diharapkan mampu menurunkan angka kemiskinan, meningkatkan produktivitas generasi muda, dan membentuk sumber daya manusia yang sehat, cerdas, serta kompetitif di tingkat global [2]. Namun demikian, kebijakan ini tidak terlepas dari perhatian publik. Di era digital saat ini, masyarakat semakin aktif menyuarakan pendapat mereka, baik berupa dukungan, kritik, maupun sikap netral melalui media sosial seperti YouTube. Respons-respons ini penting untuk dianalisis guna memahami pandangan masyarakat terhadap pelaksanaan kebijakan MBG. Analisis sentimen merupakan salah satu bidang dalam ilmu komputer yang berfokus pada pengolahan data teks untuk mengevaluasi sikap atau opini masyarakat terhadap suatu hal. Penerapan analisis sentimen pada kasus MBG dalam penelitian sebelumnya telah dilakukan, salah satunya oleh peneliti yang menggunakan algoritma Naïve Bayes dengan data *crawling* dari akun YouTube Sekretariat Negara yang berjumlah 1.470 komentar, dan diimplementasikan melalui RapidMiner versi 9.10 dengan menghasilkan akurasi 84,69%, *recall* 53,73%, dan AUC sebesar 0,525 [3]. Penelitian kedua menggunakan algoritma serupa dengan ekstraksi fitur TF-IDF dan *dataset* dari Kaggle,

menghasilkan akurasi 65%, *precision* 71%, *recall* 65%, dan *F1-score* 62% [4]. Sementara itu, penelitian ketiga menggunakan data hasil *crawling* dari media sosial X dan memperoleh akurasi 69,3%, *precision* 72%, *recall* 57,05%, serta *F1-score* 54,5% [5]. Berbeda dari penelitian-penelitian terdahulu, penelitian ini juga mengangkat objek program Makan Bergizi Gratis (MBG), namun dengan pendekatan algoritma Multinomial Naïve Bayes dan seleksi fitur menggunakan metode Chi-Square. Selain itu, penelitian ini juga menggunakan data yang dikumpulkan dengan *crawling* komentar masyarakat di *platform* YouTube dengan jumlah data sebanyak 2.165. Tujuan dari penelitian ini adalah untuk membangun model analisis sentimen yang memiliki performa tinggi, menganalisis pengaruh jumlah fitur atau *threshold Chi-Square* terhadap performa model, serta memetakan sebaran sentimen masyarakat terhadap program MBG.

2. Metode Penelitian

Pada bagian metode penelitian, terdapat gambaran alur mengenai tahapan penelitian yang dilaksanakan oleh penulis. Data *input* yang digunakan adalah berupa komentar masyarakat yang diperoleh dari *platform* YouTube, khususnya dari kanal Seputar iNews RCTI, KOMPASTV, METRO TV, detikcom, dan CNN INDONESIA. *Input* tersebut kemudian dilabeli ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral untuk mengetahui respons masyarakat yang setuju, tidak setuju, bahkan netral terhadap pemberlakuan MBG. Adapun *output* dari penelitian ini adalah berupa model klasifikasi sentimen terbaik yang dihasilkan melalui implementasi algoritma Multinomial Naïve Bayes dengan evaluasi terhadap pengaruh seleksi fitur menggunakan *threshold* Chi-Square dan *tuning hyperparameter*. Selain itu, *output* juga mencakup distribusi sentimen untuk masing-masing kelas sentimen.



Gambar 1. Alur Metode Penelitian

Gambar 1 menyajikan alur metode penelitian dalam analisis sentimen terhadap Program Makan Bergizi Gratis (MBG). Penelitian dimulai dengan pengumpulan komentar masyarakat dari YouTube menggunakan teknik *crawling*, lalu dilakukan pelabelan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Data kemudian diproses melalui tahapan *preprocessing*, penyeimbangan data, dan pembagian data. Ekstraksi fitur dilakukan menggunakan TF-IDF, diikuti dengan seleksi fitur Chi-Square. Model dibangun menggunakan algoritma Multinomial Naïve Bayes dengan *tuning hyperparameter* dan selanjutnya dievaluasi untuk mengukur performa klasifikasi.

2.1. Pengumpulan Data

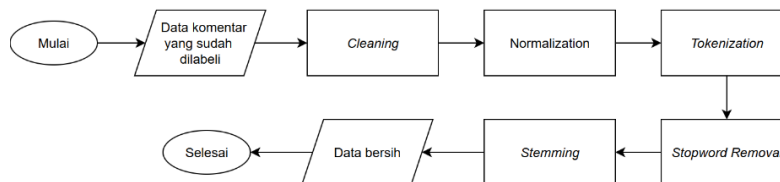
Data yang digunakan dalam penelitian ini adalah berupa komentar masyarakat yang diperoleh melalui proses *crawling* dari *platform* YouTube. Komentar diambil dari 11 video yang membahas Program Makan Bergizi Gratis (MBG) pada periode 2 Januari hingga 11 Januari 2025. Video tersebut berasal dari kanal-kanal berita nasional terverifikasi, yaitu seperti Seputar iNews RCTI, KOMPASTV, METRO TV, detikcom, dan CNN INDONESIA. Pemilihan kanal dilakukan berdasarkan kredibilitas sumber serta relevansi konten terhadap topik MBG. Hasil *crawling* memperoleh sebanyak 2.165 komentar, yang kemudian disimpan dalam format excel (.xlsx). Namun, untuk memastikan setiap komentar bersifat unik, dilakukan penghapusan data duplikat, sehingga diperoleh hanya sebanyak 2.116 data yang bisa digunakan untuk tahap selanjutnya.

2.2. Pelabelan Data

Data yang telah diperoleh pada tahap *crawling* selanjutnya diberi label dengan tiga kelas, yaitu label positif, negatif, dan netral untuk menggambarkan respons masyarakat terhadap pemberlakuan kebijakan Makan Bergizi Gratis (MBG). Proses pelabelan awal dilakukan secara otomatis menggunakan model IndoBERT untuk memberikan sentimen otomatis dalam bahasa Indonesia. Model IndoBERT digunakan untuk mempercepat proses pelabelan terhadap ribuan komentar. Namun, untuk menjaga akurasi dan validitas data, hasil pelabelan otomatis kemudian ditinjau dan divalidasi kembali secara manual. Tahapan validasi ini penting dilakukan guna mengoreksi ketidaksesuaian label yang diberikan oleh model IndoBERT, khususnya terhadap komentar yang bersifat ambigu. Hasil akhir dari proses pelabelan ini adalah bahwa data yang digunakan terdiri atas 978 komentar netral, 660 komentar negatif, dan 478 komentar positif.

2.3. Preprocessing Data

Preprocessing merupakan tahap awal yang dilakukan dalam analisis sentimen untuk menghasilkan data yang lebih baik, sehingga nantinya memberikan performa yang tinggi dalam membangun model klasifikasi [6]. Tahap *preprocessing* dalam penelitian ini mencakup beberapa langkah yang disajikan dalam gambar 2.



Gambar 2. Alur Tahap *Preprocessing*

Setiap tahap *preprocessing* secara detail dijabarkan sebagai berikut.

- a. **Cleaning**
Cleaning merupakan proses penghapusan karakter-karakter yang tidak diperlukan dalam data, yaitu seperti emoji, angka, simbol, URL, serta tanda baca untuk membuat data menjadi lebih bersih dan hanya memiliki fitur yang relevan dengan kelasnya [6].
- b. **Normalization**
Tahap *normalization* dilakukan dengan mengubah kata tidak baku atau slang menjadi bentuk baku yang sesuai dengan KBBI untuk menghasilkan representasi kata-kata yang baku, sehingga memudahkan dalam analisis dan pengolahan data lebih lanjut.
- c. **Tokenization**
Tokenization merupakan proses pemecahan teks menjadi unit-unit kata (token) sehingga dapat diproses lebih lanjut untuk menyederhanakan representasi teks dan fokus pada informasi yang sifatnya lebih penting.
- d. **Stopword Removal**
Stopword removal merupakan proses penghapusan kata-kata yang tidak penting dalam suatu teks. Keberadaan *stopword* dapat mengganggu model dalam menangkap informasi penting, sehingga perlu untuk dilakukan penghapusan guna meningkatkan kualitas fitur.
- e. **Stemming**
Stemming merupakan proses pengubahan kata menjadi bentuk kata dasarnya untuk mengurangi variasi fitur yang memiliki makna sama, namun karena terdapat imbuhan pada kata tersebut menyebabkan adanya perbedaan makna kata [6].

2.4. Penyeimbangan Data

Setelah tahap pelabelan data, diperoleh bahwa distribusi kelas antar sentimen tidak seimbang, yang mana jumlah komentar netral lebih mendominasi dibandingkan komentar positif dan negatif. Ketidakseimbangan data akan menyebabkan terjadinya kesalahan pengambilan keputusan saat klasifikasi sentimen, yaitu cenderung lebih memilih kelas mayoritas dibandingkan dengan kelas

minoritas [7]. Untuk mengatasi permasalahan tersebut, dilakukan tahap penyeimbangan data dengan *oversampling*.

2.5. Pembagian Data

Setelah data diseimbangkan, tahap selanjutnya adalah melakukan pembagian data (*split* data) untuk keperluan pelatihan dan pengujian model. Pada penelitian ini, data dibagi dengan rasio 80:20, yaitu 80% digunakan sebagai data latih dan 20% sebagai data uji.

2.6. Ekstraksi Fitur dengan *Term Frequency-Inverse Document Frequency* (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan metode yang digunakan untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma *machine learning*. TF-IDF menghitung frekuensi kata yang muncul dan seberapa banyak kata tersebut terdapat dalam dokumen [8]. Metode TF-IDF menggabungkan dua konsep, yaitu *Term Frequency* (TF) yang menyatakan jumlah kata t yang muncul dalam dokumen d dan *Inverse Document Frequency* (IDF) untuk menguji seberapa relevan kata tersebut [8]. Suatu kata t dalam dokumen d dengan jumlah total kata sebanyak n , memiliki nilai TF-IDF yang dapat dihitung menggunakan persamaan berikut.

$$tf(t, d) = \left(\frac{f(t, d)}{n} \right) \quad (1)$$

$$idf(t) = \log \left(\frac{n}{1 + df(t)} \right) \quad (2)$$

$$w_{t,d} = tf_{t,d} \times idf_t \quad (3)$$

2.7. Seleksi Fitur dengan Chi-Square

Seleksi Fitur dengan Chi-Square merupakan metode yang dapat dilakukan untuk mereduksi fitur-fitur yang memiliki ketergantungan rendah dalam sistem [9]. Metode ini bekerja dengan mengukur sejauh mana distribusi suatu fitur atau kata bergantung pada label kelas. Semakin tinggi nilai *threshold* Chi-Square, maka semakin besar pula pengaruhnya terhadap perbedaan antar kelas, sehingga fitur tersebut dianggap lebih relevan untuk digunakan dalam proses klasifikasi. Rumus perhitungan metode Chi-Square disajikan pada persamaan (4) dan (5) berikut.

$$x^2(t, c) = \left(\frac{N(AD - CB)^2}{(A+C)(B+D)(A+B)(C+D)} \right) \quad (4)$$

Dalam persamaan (4), x^2 merupakan nilai Chi-Square fitur t dalam kelas c , dan N merupakan jumlah keseluruhan dokumen. Nilai A merupakan jumlah dokumen pada kelas c yang mengandung fitur t , kemudian nilai B merupakan jumlah dokumen selain kelas c yang mengandung fitur t , selanjutnya nilai C merupakan jumlah dokumen pada kelas c yang tidak mengandung fitur t , serta nilai D merupakan jumlah dokumen selain dari kelas c yang tidak mengandung fitur t [9]. Persamaan (4) digunakan untuk memperoleh nilai bobot Chi-Square tiap kelas.

$$X^2(t) = \sum_{c=1}^k x^2(t, c) \quad (5)$$

Persamaan (5) digunakan untuk menghitung total skor Chi-Square fitur t dalam kasus multi-kelas. Nilainya diperoleh dengan menjumlahkan skor Chi-Square dari masing-masing kelas, yaitu dari $c = 1$ hingga k (jumlah seluruh kelas label) [9].

2.8. Multinomial Naïve Bayes

Multinomial Naïve Bayes merupakan algoritma *machine learning* yang berbasis probabilitas dan merupakan turunan dari Teorema Bayes, yang banyak digunakan dalam bidang *Natural Language Processing* (NLP). Algoritma ini bekerja berdasarkan konsep *term frequency*, yaitu

menghitung seberapa sering sebuah kata muncul dalam suatu dokumen. Multinomial Naïve Bayes tidak hanya mempertimbangkan keberadaan kata dalam dokumen, tetapi juga frekuensi kemunculannya, sehingga dapat memberikan bobot lebih pada kata-kata yang sering muncul [10]. Rumus perhitungan Multinomial Naïve Bayes disajikan pada persamaan (6), (7), dan (8) berikut.

$$P(C) = \frac{N_c}{N} \quad (6)$$

Berdasarkan persamaan (6), N_c merupakan jumlah kelas C pada seluruh dokumen dan N merupakan jumlah seluruh dokumen. Persamaan probabilitas dari kata ke- n ditentukan dengan persamaan (7) berikut.

$$P(X_n|C) = \left(\frac{N_{xn,c} + \alpha}{N(C) + V} \right) \quad (7)$$

Pada persamaan (7), $N_{xn,c}$ merupakan jumlah term X_n yang ditentukan dari keseluruhan data kelas C . $N(C)$ merupakan jumlah *term* dari keseluruhan data *training* pada kelas C , α adalah parameter *laplace smoothing*, dan V merupakan jumlah keseluruhan kata pada data *training* [11]. Rumus Multinomial Naïve Bayes yang digunakan dalam pembobotan TF-IDF disajikan sebagai berikut.

$$P(X_n|C) = \frac{\sum tf(X_n, d \in C) + \alpha}{\sum N_{d \in C} + V} \quad (8)$$

Dilihat dari persamaan (8), $\sum tf(X_n, d \in C)$ merupakan jumlah pembobotan term X_n dari seluruh data latih kelas C , sedangkan $\sum N_{d \in C}$ merupakan jumlah bobot dari seluruh *term* data latih kelas C [11].

2.9. Tuning Hyperparameter

Tuning Hyperparameter merupakan sebuah upaya yang dilakukan untuk mengoptimalkan model klasifikasi. Upaya dengan *tuning* ini akan melakukan pencarian secara menyeluruh terhadap seluruh kombinasi *hyperparameter* yang telah ditentukan [12]. Dalam penelitian ini, digunakan dua kombinasi *hyperparameter* yaitu *alpha* dan *fit_prior*. Untuk menentukan kombinasi *hyperparameter* terbaik, diterapkan *grid search* menggunakan *5-fold cross-validation*. *Grid search* mampu memberikan fleksibilitas tinggi dalam alokasi waktu ataupun sumber daya komputasi [12].

2.10. Evaluasi

Tahap evaluasi dilakukan untuk membandingkan antara hasil *testing* dari model yang telah dibangun. Metode evaluasi yang diterapkan adalah *confusion matrix* yang merupakan tabel penyajian hasil klasifikasi antara jumlah data uji yang benar dan yang salah untuk setiap kelas [12]. Berikut disajikan tabel *confusion matrix*.

Tabel 1. Confusion Matrix

		Kelas Aktual		
		Positif	Negatif	Netral
Kelas Prediksi	Positif	TP	FP	FP
	Negatif	FN	TN	TN
	Netral	FN	TN	TN

Berdasarkan *confusion matrix*, maka dapat diperoleh nilai *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* mengukur proporsi prediksi yang benar dari seluruh data yang diuji, *precision* menunjukkan ketepatan prediksi positif, *recall* menilai sejauh mana model mampu menemukan

semua data relevan, dan *F1-score* merepresentasikan keseimbangan antara *precision* dan *recall* [12]. Perhitungan seluruh nilai evaluasi tersebut disajikan pada persamaan sebagai berikut.

$$Accuracy = \frac{TP + TN}{(TP + FP + TN + FN)} \quad (9)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (10)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (11)$$

$$F1-Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (12)$$

Keterangan:

- TP (*True Positive*): Jumlah data yang diklasifikasikan dengan benar oleh model sebagai bagian dari suatu kelas.
- TN (*True Negative*): Jumlah data yang secara tepat tidak dimasukkan oleh model ke dalam kelas tertentu.
- FP (*False Positive*): Jumlah data yang secara keliru diprediksi oleh model sebagai termasuk dalam suatu kelas, padahal seharusnya tidak.
- FN (*False Negative*): Jumlah data yang seharusnya termasuk dalam suatu kelas, namun tidak berhasil dikenali oleh model dan justru diprediksi sebagai kelas lain.

3. Hasil dan Diskusi

Penelitian ini bertujuan untuk mengklasifikasikan komentar masyarakat terkait Program Makan Bergizi Gratis (MBG) ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral menggunakan algoritma Multinomial Naïve Bayes dan bahasa pemrograman Python. Selain itu, penelitian ini juga mengevaluasi pengaruh seleksi fitur menggunakan variasi 10 *threshold* Chi-Square, serta melakukan *tuning hyperparameter alpha* dan *fit_prior* untuk mengoptimalkan performa model. Kemudian untuk evaluasi dilakukan dengan menggunakan *5-fold cross-validation* pada setiap konfigurasi.

3.1. Preprocessing

Sebelum memasuki tahap pemodelan, data terlebih dahulu melalui proses *preprocessing* untuk membersihkan dan menormalkan teks, sehingga dapat meningkatkan kualitas dan kinerja model yang dibangun. Hasil dari tahap *preprocessing* ditampilkan pada Tabel 2 berikut.

Tabel 2. Hasil Tahap *Preprocessing*

No.	Tahapan	Hasil
1	Data awal	ALHAMDULILAH SYUKURI SAJA DARIPADA TDK SAMA SEKALI
2	<i>Cleaning</i>	alhamdulillah syukuri saja daripada tdk sama sekali
3	<i>Normalization</i>	alhamdulillah syukuri saja daripada tidak sama sekali
4	<i>Tokenization</i>	['alhamdulillah', 'syukuri', 'saja', 'daripada', 'tidak', 'sama', 'sekali']
5	<i>Stopword removal</i>	['alhamdulillah', 'syukuri', 'daripada', 'sama', 'sekali']
6	<i>Stemming</i>	['alhamdulillah', 'syukur', 'daripada', 'sama', 'sekali']

3.2. Penyeimbangan Data

Penyeimbangan data menggunakan teknik *oversampling* dilakukan dengan cara menduplikasi data dari kelas minoritas secara acak hingga jumlah data di seluruh kelas sentimennya menjadi

sama. Berdasarkan hasil dari tahap pelabelan bahwa data yang digunakan terdiri atas 978 komentar netral, 660 komentar negatif, dan 478 komentar positif. Oleh karena itu, pada tahap *oversampling* ini, jumlah data setiap kelas sentimen disamakan menjadi 978 komentar, sehingga diperoleh total datanya adalah berjumlah 2.934 komentar.

3.3. Pembagian Data

Proses pembagian dilakukan menggunakan metode *stratified splitting*, yang mempertahankan proporsi distribusi label pada masing-masing kelas sentimen agar tetap seimbang di kedua *subset* data. Berdasarkan hasil *oversampling*, total data berjumlah 2.934 komentar, yang kemudian dibagi menjadi 2.347 data latih dan 587 data uji. Rincian distribusi label pada data latih adalah 783 komentar negatif, 782 komentar positif, dan 782 komentar netral. Sementara itu, pada data uji terdapat 195 komentar negatif, 196 komentar positif, dan 196 komentar netral.

3.4. Ekstraksi Fitur dengan TF-IDF

Pada tahap ini, teks hasil *preprocessing* diubah menjadi representasi numerik yang dapat diproses oleh algoritma *machine learning*. Proses ini menghasilkan sebanyak 3.487 fitur unik yang merepresentasikan kata-kata dalam komentar masyarakat. Nilai TF-IDF yang dihasilkan mencerminkan tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap keseluruhan korpus yang kemudian digunakan sebagai input untuk tahap seleksi fitur menggunakan metode Chi-Square sebelum dilakukan proses klasifikasi dengan algoritma Multinomial Naïve Bayes.

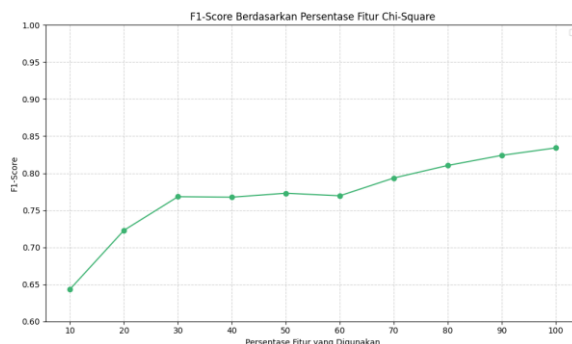
3.5. Implementasi Multinomial Naïve Bayes dengan *Tuning Hyperparameter* dan Seleksi Fitur Chi-Square

Algoritma Multinomial Naïve Bayes diimplementasikan menggunakan data yang telah dibagi pada tahap sebelumnya, yaitu 80% untuk data latih dan 20% untuk data uji [13]. Dalam proses ini, digunakan 10 variasi nilai *threshold* Chi-Square berdasarkan total 3.487 fitur hasil ekstraksi TF-IDF. Setiap variasi nilai *threshold* tersebut diuji secara terpisah untuk mengevaluasi pengaruh seleksi fitur terhadap performa model. Selanjutnya, dilakukan *tuning hyperparameter* untuk setiap skenario jumlah fitur dengan menggunakan kombinasi parameter *alpha* dan *fit_prior*, yang dievaluasi melalui *grid search* dan *5-fold cross-validation* untuk memperoleh konfigurasi parameter terbaik pada masing-masing tingkat fitur. Berdasarkan skenario tersebut, diperoleh hasil sebagai berikut.

Tabel 3. Hasil Pengujian Model dengan Variasi *Threshold* Chi-Square

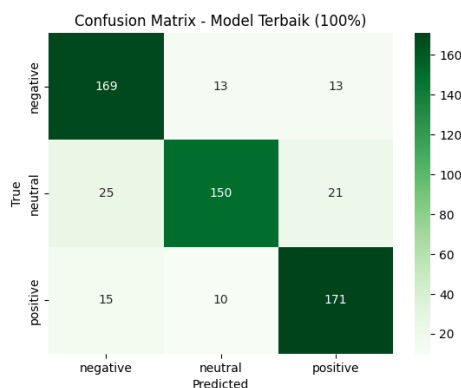
Chi-Square threshold	Best Params	Accuracy	Precision	Recall	F1-Score
10% (348)	'alpha': 0.1, 'fit_prior': True	0.6474	0.6603	0.6475	0.6435
20% (697)	'alpha': 0.1, 'fit_prior': True	0.7240	0.7275	0.7241	0.7230
30% (1046)	'alpha': 0.1, 'fit_prior': False	0.7700	0.7746	0.7701	0.7683
40% (1394)	'alpha': 0.1, 'fit_prior': True	0.7683	0.7701	0.7683	0.7677
50% (1743)	'alpha': 0.1, 'fit_prior': True	0.7734	0.7745	0.7734	0.7730
60% (2092)	'alpha': 0.1, 'fit_prior': True	0.7700	0.7708	0.7700	0.7695
70% (2440)	'alpha': 0.1, 'fit_prior': True	0.7939	0.7973	0.7939	0.7935
80% (2789)	'alpha': 0.1, 'fit_prior': True	0.8109	0.8119	0.8109	0.8105
90% (3138)	'alpha': 0.1, 'fit_prior': True	0.8245	0.8262	0.8246	0.8241
100% (3487)	'alpha': 0.1, 'fit_prior': True	0.8348	0.8366	0.8348	0.8342

Tabel 3 menunjukkan bahwa performa model meningkat seiring bertambahnya proporsi fitur yang dipilih dengan metode Chi-Square. Performa terbaik diperoleh ketika penggunaan 100% *threshold*. Meskipun sebagian besar *threshold* menunjukkan performa optimal dengan kombinasi *hyperparameter alpha: 0.1* dan *fit_prior: True*, namun pada *threshold* Chi-Square 30%, kombinasi terbaik justru diperoleh pada *alpha: 0.1* dan *fit_prior: False*.



Gambar 3. Grafik *F1-Score* dengan Kombinasi *Threshold* Chi-Square

Gambar 3 menunjukkan bahwa seiring meningkatnya jumlah fitur, nilai *F1-Score* juga mengalami peningkatan secara konsisten. Kenaikan ini menunjukkan bahwa semakin banyak fitur relevan yang digunakan, maka kemampuan model dalam menangkap pola sentimen makin baik. Hal ini memperkuat temuan bahwa model terbaik justru diperoleh saat tidak dilakukan reduksi fitur, yang mana hal tersebut menandakan bahwa informasi penting tersebar luas di seluruh fitur, dan tidak cukup hanya di sebagian kecil fitur. Grafik ini mempertegas bahwa pemilihan *threshold* Chi-Square sangat memengaruhi performa model, dan pemanfaatan seluruh fitur terbukti paling optimal untuk kasus MBG ini.



Gambar 4. Visualisasi *Confusion Matrix* Model Terbaik

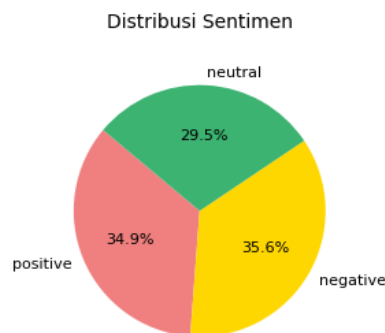
Pada Gambar 4, ditunjukkan bahwa untuk kelas negatif, sebanyak 169 data berhasil diklasifikasikan dengan benar, sementara 13 data salah diklasifikasikan sebagai netral dan 13 lainnya sebagai positif. Untuk kelas netral, terdapat 150 data yang diklasifikasikan dengan benar, 25 salah sebagai negatif, dan 21 sebagai positif. Sedangkan pada kelas positif, sebanyak 171 data terklasifikasi dengan benar, 15 data salah diklasifikasikan sebagai negatif, dan 10 lainnya sebagai netral. Gambaran ini selengkapny juga ditampilkan dalam laporan klasifikasi kinerja model pada Tabel 4 berikut.

Tabel 4. Hasil Pengujian Model dengan Laporan Klasifikasi

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
negatif	0.81	0.87	0.84	195

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
netral	0.87	0.77	0.81	196
positif	0.83	0.87	0.85	196
<i>accuracy</i>			0.83	587
<i>macro avg</i>	0.84	0.83	0.83	587
<i>weighted avg</i>	0.84	0.83	0.83	587

Tabel 4 menunjukkan bahwa model terbaik memiliki performa terbaik pada kelas positif dan negatif, namun sedikit lemah pada kelas netral. Akurasi keseluruhan mencapai 83%, dengan nilai rata-rata metrik yang konsisten antara *precision*, *recall*, dan *f1-score* yang berada di angka 83%-84%, sehingga menandakan bahwa performa model cukup seimbang di semua kelas.



Gambar 5. Hasil Prediksi Perbandingan Sentimen Masyarakat

Berdasarkan hasil pengujian model terhadap kasus program Makan Bergizi Gratis, diperoleh bahwa respons masyarakat lebih condong ke negatif dibandingkan dengan positif ataupun netral. Terlihat pada Gambar 5, persentase sentimen negatif adalah sebesar 35,6%, yang sedikit lebih tinggi dibandingkan sentimen positif yang hanya sebesar 34,9%, lalu disusul oleh sentimen netral yang jauh lebih rendah, yaitu sebesar 29,5%. Persentase tersebut menunjukkan bahwa sebagian besar masyarakat kurang menyambut baik program MBG tersebut.

Seluruh proses komputasi dalam penelitian ini dilakukan menggunakan bahasa pemrograman Python pada lingkungan Google Colab. Beberapa pustaka yang digunakan meliputi *pandas* dan *NumPy* untuk pengolahan data, *scikit-learn* untuk klasifikasi dan evaluasi model, serta *imbalanced-learn* untuk penyeimbangan data. Proses pelabelan awal dengan *IndoBERT* dilakukan menggunakan *Transformers Library* dari *Hugging Face*, serta untuk visualisasi digunakan *matplotlib* dan *seaborn*.

4. Kesimpulan

Berdasarkan hasil eksperimen yang telah dilakukan, dapat disimpulkan bahwa implementasi algoritma Multinomial Naïve Bayes dengan seleksi fitur Chi-Square dan *tuning hyperparameter* berhasil memberikan performa yang optimal dalam mengklasifikasikan sentimen masyarakat terhadap Program Makan Bergizi Gratis (MBG). Model terbaik diperoleh saat menggunakan 100% *threshold* hasil seleksi Chi-Square, dengan kombinasi *hyperparameter alpha* = 0.1 dan *fit_prior* = *True*, yang menghasilkan akurasi sebesar 83.48%, *precision* 83.66%, *recall* 83.48%, dan *F1-score* 83.42%. Temuan ini menunjukkan bahwa pemanfaatan seluruh fitur TF-IDF setelah *preprocessing* dan seleksi Chi-Square justru menghasilkan kinerja tertinggi, yang membantah asumsi bahwa reduksi fitur selalu meningkatkan akurasi. Hasil evaluasi menggunakan *confusion matrix* dan *classification report* juga menunjukkan bahwa model mampu mendeteksi ketiga kelas sentimen dengan seimbang. Proses validasi menggunakan *5-fold cross-validation* turut

memastikan bahwa model memiliki performa yang stabil dan tidak *overfitting*. Jika dilihat dari persentase terhadap sebaran sentimen, diperoleh bahwa komentar masyarakat cenderung bernada negatif sebesar 35,6%, diikuti oleh positif 34,9%, dan netral 29,5%, yang mengindikasikan bahwa masih adanya keraguan masyarakat terhadap kebijakan MBG tersebut. Untuk penelitian selanjutnya, disarankan untuk mengembangkan pendekatan dengan algoritma berbasis *deep learning* serta mempertimbangkan elemen bahasa non-formal seperti bahasa daerah ataupun sarkasme guna menangkap nuansa sentimen secara lebih komprehensif.

Daftar Pustaka

- [1] D. Ni Kadek Trisna Cintya, "Resmi Dimulai Hari Ini, Siapa Saja Penerima Manfaat Makan Bergizi Gratis? | tempo.co." Accessed: Jun. 27, 2025. [Online]. Available: <https://www.tempo.co/politik/resmi-dimulai-hari-ini-siapa-saja-penerima-manfaat-makan-bergizi-gratis--1190382>
- [2] C. Sugiarto, E., "Makan Bergizi Gratis dan SDM Unggul | Sekretariat Negara." Accessed: Jun. 26, 2025. [Online]. Available: https://www.setneg.go.id/baca/index/makan_bergizi_gratis_dan_sdm_unggul
- [3] B. Rahmatullah, S. A. Saputra, P. Budiono, and D. P. Wigandi, "Sentimen Analisis Makan Bergizi Gratis Menggunakan Algoritma Naive Bayes," vol. 05, no. 01, 2025, doi: <https://doi.org/10.46229/jifotech.v5i1.978>.
- [4] R. Amelia, B. Siregar, A. R. Manik, A. Syahri, M. B. Akbar, and F. Ramadhani, "Penerapan Naïve Bayes Untuk Analisis Sentimen Netizen Terhadap Program Makan Siang Gratis Pada Media Social X," vol. 9, no. 4, pp. 5621–5628, 2025, doi: <https://doi.org/10.36040/jati.v9i4.13870>.
- [5] Y. Z. Vebrian, T. Informatika, and U. N. Waluyo, "A Sentiment Analysis Of Free Meal Plans On Social Media Using Naïve Bayes Algorithms Analisis Sentimen Terhadap Rencana Makan Gratis Di Sosial Media X Menggunakan," vol. 10, no. 1, 2025, doi: <https://doi.org/10.35314/3m2fcz69>.
- [6] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: [10.30865/mib.v5i2.2835](https://doi.org/10.30865/mib.v5i2.2835).
- [7] Y. A. Sir and A. H. H. Soepranoto, "Pendekatan Resampling Data Untuk Menangani Masalah Ketidakseimbangan Kelas," *J. Komput. dan Inform.*, vol. 10, no. 1, pp. 31–38, 2022, doi: [10.35508/jicon.v10i1.6554](https://doi.org/10.35508/jicon.v10i1.6554).
- [8] I. Widaningrum, D. Mustikasari, R. Arifin, S. L. Tsaqila, and D. Fatmawati, "Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) dan K-Means Clustering Untuk Menentukan Kategori Dokumen," *Pros. Semin. Nas. Sist. Inf. dan Teknol.*, pp. 145–149, 2022.
- [9] A. P. Permana, T. Chamidy, and C. Crysdian, "Klasifikasi Ulasan Fasilitas Publik Menggunakan Metode Naïve Bayes dengan Seleksi Fitur Chi-Square," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 2, pp. 112–124, 2023, doi: [10.14421/jiska.2023.8.2.112-124](https://doi.org/10.14421/jiska.2023.8.2.112-124).
- [10] Yuyun, Nurul Hidayah, and Supriadi Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 820–826, 2021, doi: [10.29207/resti.v5i4.3146](https://doi.org/10.29207/resti.v5i4.3146).
- [11] A. Agung, I. Intan, P. Sari, and I. K. Gede, "Penggunaan Metode SVM dan Naive Bayes pada Analisis Sentiment Ulasan Aplikasi Edlink," vol. 3, pp. 575–584, 2025, doi: <https://doi.org/10.24843/JNATIA.2025.v03.i03.p11>.
- [12] Joshua Agung Nurcahyo and Theopilus Bayu Sasongko, "Hyperparameter Tuning Algoritma Supervised Learning untuk Klasifikasi Keluarga Penerima Bantuan Pangan Beras," *Indones. J. Comput. Sci.*, vol. 12, no. 3, pp. 1351–1365, 2023, doi: [10.33022/ijcs.v12i3.3254](https://doi.org/10.33022/ijcs.v12i3.3254).
- [13] I. G. Widnyana and A. Eka, "Analisis Sentimen pada Sengketa Pilpres 2024 dengan Multinomial Naïve Bayes dan Chi-Square," vol. 3, pp. 489–500, 2025, doi: <https://doi.org/10.24843/JNATIA.2025.v03.i03.p03>.