

# Analisis Sentimen Kebijakan Makan Bergizi Gratis Menggunakan IndoBERT dan Machine Learning

Erik Setiadi<sup>1</sup>, Danang Arbian Sulistyio<sup>1</sup>

<sup>1</sup>Teknik Informatika, Fakultas Teknologi dan Desain, Institut Teknologi dan Bisnis Asia Malang

<sup>1</sup>[eriksetiadi2k24@gmail.com](mailto:eriksetiadi2k24@gmail.com), <sup>2</sup>[danangarbian@gmail.com](mailto:danangarbian@gmail.com) \*

## Abstract

*Social media has increasingly become a primary channel for the public to express opinions and reactions toward government policies, including Indonesia's Makan Bergizi Gratis (MBG) program. The large volume of online discussions and the diversity of public perspectives highlight the importance of sentiment analysis to assess public acceptance and identify potential challenges in policy implementation. This study aims to analyze the distribution of public sentiment toward the MBG policy and to evaluate the performance of machine learning models for sentiment classification. The dataset consists of 12,389 Indonesian-language tweets collected from platform X. Sentiment labeling was performed automatically using a hybrid labeling approach that combines a domain-specific lexicon-based method with the IndoBERT deep learning model to improve contextual understanding. Sentiment classification was conducted using hybrid features, including TF-IDF trigrams, IndoBERT embeddings, and lexicon-based features. Three classification models Random Forest, XGBoost, and an Ensemble model were trained and evaluated on a SMOTE-balanced dataset to address class imbalance. The results reveal that public sentiment toward the MBG policy is predominantly negative (68.6%), followed by positive (19.5%) and neutral (11.9%) sentiments. Among the evaluated models, Random Forest achieved the best performance, obtaining an F1-score of 0.9383 using K-Fold cross-validation and 0.9363 on the final test set. This study concludes that the proposed hybrid approach is effective and reliable for classifying public sentiment toward government policies in the Indonesian language.*

*Keywords: sentiment analysis, makan bergizi gratis, hybrid method, random forest, IndoBERT*

## Abstrak

Media sosial telah berkembang menjadi ruang utama bagi masyarakat dalam mengekspresikan opini terhadap berbagai kebijakan publik, termasuk kebijakan Makan Bergizi Gratis (MBG) yang diluncurkan pemerintah Indonesia. Tingginya intensitas diskusi dan beragamnya pandangan publik terhadap kebijakan ini menjadikan analisis sentimen penting untuk memahami tingkat penerimaan masyarakat serta potensi permasalahan dalam implementasinya. Penelitian ini bertujuan untuk menganalisis distribusi sentimen publik terhadap kebijakan MBG serta mengevaluasi kinerja beberapa model *machine learning* dalam melakukan klasifikasi sentimen. Data yang digunakan berupa 12.389 tweet berbahasa Indonesia yang dikumpulkan dari platform X. Pelabelan sentimen dilakukan secara otomatis menggunakan metode *hybrid labeling* yang menggabungkan pendekatan leksikon berbasis domain dengan model *deep learning* IndoBERT guna meningkatkan akurasi pelabelan. Selanjutnya, proses klasifikasi dilakukan menggunakan fitur hibrida yang mencakup TF-IDF trigram, *embedding* IndoBERT, dan fitur leksikon, dengan tiga model yang dibandingkan yaitu Random Forest, XGBoost, dan Ensemble. Untuk mengatasi ketidakseimbangan data, teknik SMOTE diterapkan pada data latih. Hasil penelitian menunjukkan bahwa sentimen publik didominasi oleh sentimen negatif sebesar 68,6%, diikuti oleh sentimen positif 19,5% dan netral 11,9%. Model Random Forest menunjukkan kinerja terbaik dengan nilai F1-Score sebesar 0,9383 pada *K-Fold Cross-Validation* dan 0,9363 pada pengujian akhir. Penelitian ini menyimpulkan bahwa pendekatan hibrida yang diusulkan efektif dan andal dalam mengklasifikasikan sentimen publik terhadap kebijakan pemerintah berbahasa Indonesia.

Kata kunci: analisis sentimen, makan bergizi gratis, metode hibrida, random forest, indoBERT

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

## 1. Pendahuluan

Era digital telah mengubah media sosial, khususnya X (sebelumnya Twitter), menjadi sebuah forum publik global tempat masyarakat secara bebas mengekspresikan opini, sentimen, dan respons terhadap berbagai isu [1]. Opini publik yang terekam secara masif ini telah menjadi sumber data yang sangat berharga untuk memahami respons masyarakat terhadap kebijakan pemerintah [2]. Analisis sentimen terhadap kebijakan publik, seperti kebijakan PPKM [3], kenaikan harga bahan bakar [4], hingga kenaikan PPN [5], telah banyak dilakukan untuk memberikan masukan kepada pemangku kepentingan.

Secara khusus, platform X menjadi sumber data yang populer untuk analisis sentimen karena sifatnya yang *real-time* dan ringkas, memungkinkan peneliti untuk menangkap reaksi instan publik [1, 2]. Namun, analisis sentimen berbahasa Indonesia memiliki tantangan unik. Bahasa Indonesia bersifat sangat dinamis, dipenuhi dengan bahasa *slang*, kata tidak baku, dan variasi ejaan yang terus berkembang. Selain itu, masyarakat sering menggunakan sarkasme atau kalimat berkonotasi negatif untuk mengekspresikan sentimen yang sebenarnya positif (atau sebaliknya), yang tidak dapat ditangkap oleh metode berbasis leksikon sederhana [7]. Oleh karena itu, diperlukan sebuah metode yang tidak hanya mengandalkan pencocokan

kata kunci, tetapi juga mampu memahami konteks kalimat.

Kebijakan "Makan Bergizi Gratis" (MBG) merupakan salah satu program baru yang memicu diskusi publik secara luas di Indonesia. Sebagai program yang memiliki implikasi sosial dan anggaran yang besar, program ini menuai beragam respons, mulai dari dukungan hingga kritik [1]. Memahami polaritas (positif, negatif, atau netral) dari sentimen publik ini menjadi krusial untuk mengevaluasi penerimaan masyarakat dan mengidentifikasi potensi masalah dalam implementasinya.

Untuk menganalisis data tekstual dalam jumlah besar tersebut, metode komputasi seperti *Natural Language Processing* (NLP) dan *machine learning* telah terbukti sangat efektif [6, 7]. Berbagai penelitian sebelumnya telah berhasil menerapkan beragam algoritma untuk klasifikasi sentimen di Indonesia. Model *machine learning* klasik seperti Random Forest [2, 8, 10, 19] dan XGBoost [3, 11] dikenal memiliki performa yang kuat dan stabil. Di sisi lain, model *deep learning* berbasis *transformer* seperti IndoBERT telah menunjukkan keunggulan dalam memahami konteks kalimat berbahasa Indonesia [5, 9, 21]. Beberapa penelitian juga menunjukkan bahwa pendekatan *hybrid* yang menggabungkan kekuatan leksikon berbasis aturan dengan *machine learning* mampu meningkatkan akurasi, terutama dalam menangani sarkasme dan konteks spesifik [6, 7, 4].

Meskipun telah banyak penelitian dilakukan, masih terdapat celah penelitian (*research gap*). Pertama, analisis sentimen terhadap kebijakan MBG yang spesifik ini masih terbatas, terutama yang menggunakan dataset skala besar pasca-pengumuman resmi [1]. Kedua, terdapat kebutuhan untuk membandingkan secara langsung model *machine learning* klasik (Random Forest, XGBoost) dengan pendekatan *hybrid* yang memanfaatkan fitur *embedding* dari IndoBERT [5, 9, 12]. Belum banyak penelitian yang mengadu performa model-model ini secara komprehensif pada domain kebijakan publik yang sama di Indonesia.

Oleh karena itu, penelitian ini dilakukan untuk mengisi celah tersebut. Tujuan utama dari penelitian ini adalah (1) untuk mengidentifikasi dan menganalisis distribusi sentimen publik (positif, negatif, dan netral) masyarakat Indonesia terhadap kebijakan "Makan Bergizi Gratis" berdasarkan data dari media sosial X; dan (2) untuk merancang, mengimplementasikan, dan mengevaluasi arsitektur *hybrid* guna menemukan model klasifikasi terbaik antara Random Forest, XGBoost, dan Ensemble dengan menggabungkan fitur TF-IDF, fitur leksikon [6, 7], dan fitur *embedding* IndoBERT.

## 2. Metode Penelitian

Metode penelitian yang digunakan adalah analisis sentimen kuantitatif menggunakan teknik *hybrid*

*machine learning*. Penelitian ini dibagi menjadi beberapa tahapan utama, yang diuraikan secara rinci dalam sub-bab berikut.

### 2.1. Penelitian Terkait

Untuk memvalidasi pemilihan metode, telah dilakukan tinjauan terhadap 25 studi sebelumnya. Tinjauan ini menunjukkan bahwa topik analisis sentimen terhadap kebijakan pemerintah dan program publik merupakan bidang penelitian yang relevan di Indonesia. Beberapa penelitian telah menganalisis sentimen terhadap program makan siang gratis [1], kebijakan vaksinasi COVID-19 [2], kebijakan PPKM [3], dan kenaikan PPN [5]. Pendekatan *hybrid* yang menggabungkan leksikon dengan *machine learning* juga telah terbukti efektif dalam berbagai studi [4, 6, 7, 8, 15]. Penggunaan model individual seperti Random Forest [9, 12, 16, 17], XGBoost [11, 18, 19], dan IndoBERT [10, 13, 14, 20, 21] secara spesifik untuk analisis sentimen di Indonesia juga telah divalidasi. Penelitian lain juga telah menerapkan metode ini pada berbagai kebijakan pemerintah [22-24] serta pengembangan model bahasa untuk konteks regional Asia [25].

### 2.2. Pengumpulan Data

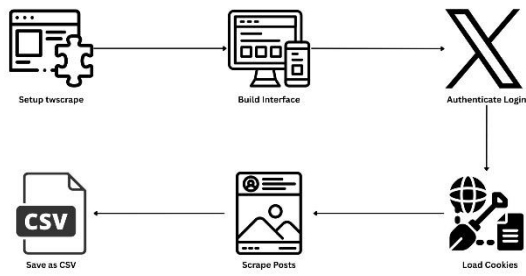
Pengumpulan data, atau *data scraping*, dilakukan dari platform media sosial X (sebelumnya Twitter). Proses ini menargetkan tweet-tweet publik berbahasa Indonesia yang relevan dengan topik "kebijakan makan bergizi gratis", mirip dengan pendekatan yang dilakukan pada penelitian terkait [1]. Total data mentah yang berhasil dikumpulkan adalah 12.389 tweet. Dataset bersih (*clean dataset*) yang telah melalui *preprocessing* dan pelabelan dalam penelitian ini dapat diakses secara publik melalui repositori GitHub peneliti (tersedia untuk sementara di: <https://github.com/rikk-se/Dataset-Program-Makan-Bergizi-Gratis>). Beberapa contoh data mentah disajikan pada Tabel 1.

Tabel 1. Contoh Data Mentah

Username	Teks Mentah
Sa***21	Sementara ada nya masalah keracunan ini BGN mencari cara menangani masalah ini, dan BGN akhirnya melakukan evaluasi SOP program ini agar tidak terjadi lagi Lanjutkan MBG
Hy***aa	Hadehhhhh, udah deh mending kasih uang ke ortunya aja, suruh masak sendiri. Ga semua masakan itu cocok ya, kadang ada yg alergi atau intoleran. Tolong!!! Masa gini aja ga paham paham sih ðŸ™f ngabisin dana, tapi belum merata, kan bisa tuh prioritasin daerah yg memang butuh dulu
Mi***74	Meremehkan nyawa manusia dgn memberikan makanan yg tidak higienis atau mungkin basi sehingga membuat anak bukan tambah sehat malah keracunan . Program apa ini ?

Pengambilan data menggunakan *tool* Twscrape yang memanfaatkan otentikasi akun untuk mengakses *postingan* dan menyimpannya dalam format CSV. Alur

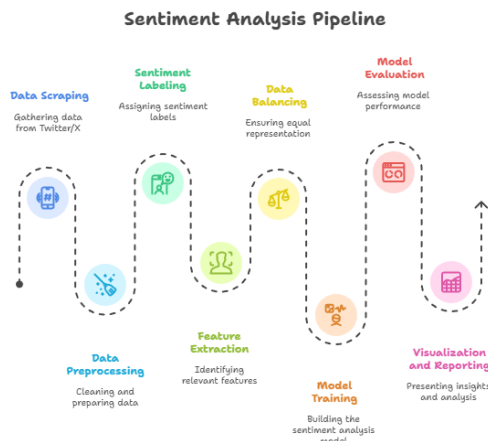
proses *scraping* data yang terdiri dari setup *tool*, otentikasi, pemuatan *cookies*, dan penyimpanan ke CSV diilustrasikan secara rinci pada Gambar 1.



Gambar 1. Alur Proses Scraping Data

### 2.3. Alur Penelitian

Secara keseluruhan, penelitian ini mengikuti *pipeline* analisis sentimen yang sistematis, seperti yang ditunjukkan pada Gambar 2. Alur ini merangkum seluruh tahapan inti penelitian, mulai dari pengumpulan data mentah (dijelaskan di 2.2), dilanjutkan dengan pemrosesan data dan pelatihan model, hingga penyajian hasil dan laporan akhir. Detail dari setiap tahapan dalam alur ini akan dijelaskan pada sub-bab berikutnya.



Gambar 2. Alur Pipeline Analisis Sentimen

### 2.4. Data Preprocessing

Data mentah pada Tabel 1 kemudian melalui tahap *preprocessing* atau pembersihan teks untuk menghilangkan noise dan menyeragamkan data [2, 3]. Proses ini sangat krusial untuk meningkatkan kinerja model. Tahapan ini dibagi menjadi empat proses utama yang berurutan.

Tahap pertama adalah *case folding*, di mana seluruh teks diubah menjadi format huruf kecil. Hasil dari tahap ini ditunjukkan pada Tabel 2.

Tabel 2. Contoh Hasil Case Folding

Teks Asli (Dari Tabel 1)	Hasil Case Folding
@UserB Setuju bgt, program MBG ini bisa jadi solusi gizi, tapi... dananya dari mana??	@userb setuju bgt, program mbg ini bisa jadi solusi gizi, tapi... dananya dari mana??

Jgn korupsi lagi ya #MBG 🙄 <a href="https://t.co/xyz">https://t.co/xyz</a>	jgn korupsi lagi ya #mbg 🙄 <a href="https://t.co/xyz">https://t.co/xyz</a>
---	---

Tahap kedua adalah *cleansing* (pembersihan karakter). Proses ini bertujuan untuk menghapus semua karakter dan simbol yang tidak relevan, yang mencakup penghapusan tautan URL, mention (@username), hashtag (#), emoji, dan karakter numerik serta tanda baca. Hasil dari tahap ini ditunjukkan pada Tabel 3.

Tabel 3. Contoh Hasil Cleansing

Teks Sebelumnya (Case Folding)	Hasil Cleansing
@userb setuju bgt, program mbg ini bisa jadi solusi gizi, tapi... dananya dari mana?? jgn korupsi lagi ya #mbg 🙄 <a href="https://t.co/xyz">https://t.co/xyz</a>	setuju bgt program mbg ini bisa jadi solusi gizi tapi dananya dari mana jgn korupsi lagi ya mbg

Tahap ketiga adalah *slangword normalization* (normalisasi slang). Tahap ini memanfaatkan kamus *slang* yang telah disusun untuk mengubah kata-kata tidak baku menjadi kata baku (contoh: 'gak' menjadi 'tidak', 'bgt' menjadi 'banget', 'dgn' menjadi 'dengan'). Hasil dari tahap ini ditunjukkan pada Tabel 4.

Tabel 4. Contoh Hasil Slangword Normalization

Teks Sebelumnya (Cleaned)	Hasil Normalisasi
setuju bgt program mbg ini bisa jadi solusi gizi tapi dananya dari mana jgn korupsi lagi ya mbg	setuju banget program mbg ini bisa jadi solusi gizi tapi dananya dari mana jangan korupsi lagi ya mbg

Tahap terakhir adalah *stopword removal*. Proses ini menggunakan daftar *stopword* Bahasa Indonesia yang dimodifikasi untuk mempertahankan kata-kata kunci yang memengaruhi sentimen, seperti kata-kata negasi (contoh: 'tidak', 'belum') dan kata-kata intensitas (contoh: 'sangat', 'kurang'). Hasil akhir preprocessing ditunjukkan pada Tabel 5.

Tabel 5. Contoh Hasil Akhir Stopword Removal

Teks Sebelumnya (Normalized)	Hasil Akhir (Stopword Removal)
setuju banget program mbg ini bisa jadi solusi gizi tapi dananya dari mana jangan korupsi lagi ya mbg	setuju banget program mbg jadi solusi gizi tapi dananya mana jangan korupsi mbg

### 2.5. Hybrid Sentiment Labeling

Tahap ini adalah inti dari metode penelitian untuk menentukan label sentimen (positif, negatif, atau netral) pada 12.389 data bersih. Alih-alih pelabelan manual, penelitian ini menerapkan metode *hybrid labeling* yang menggabungkan dua pendekatan untuk akurasi yang lebih kontekstual [4, 6, 8, 15].

Pendekatan pertama adalah *lexicon-based*. Metode ini menggunakan kamus sentimen Bahasa Indonesia yang telah dikustomisasi dan dipertajam dengan *domain-specific terms* (seperti 'korupsi', 'bergizi', 'higienis') serta penanganan negasi dan *intensifiers*.

Pendekatan kedua adalah model *deep learning* IndoBERT. Model yang digunakan adalah *w11wo/indonesian-roberta-base-sentiment-classifier*.

Penggunaan model berbasis BERT dan IndoBERT telah terbukti efektif dalam penelitian sebelumnya untuk klasifikasi sentimen di Indonesia [2, 5, 9, 12].

Skor dari kedua pendekatan ini kemudian digabungkan melalui logika *weighted voting* dan prioritas untuk menghasilkan satu label sentimen akhir. Contoh hasil dari proses pelabelan ini ditunjukkan pada Tabel 6.

Tabel 6. Contoh Hasil Proses Hybrid Labeling

Teks Bersih	Prediksi Lexicon	Prediksi IndoBERT	Label Final (Metode)
setuju banget program mbg jadi solusi gizi tapi dananya mana jangan korupsi mbg	negatif	negatif	negatif (agreement)
programnya bagus tapi awas banyak celah korupsi	positif	negatif	negatif (bert_prioty)
keren semoga program makan bergizi gratis pak presiden berjalan lancar...	positif	positif	positif (agreement)

## 2.6. Feature Engineering dan Klasifikasi

Setelah data dilabeli, dilakukan *feature engineering* untuk mengubah teks menjadi vektor numerik. Tiga jenis fitur diekstraksi dan digabungkan (*concatenated*) menjadi satu vektor fitur, seperti yang diilustrasikan pada Gambar 3.

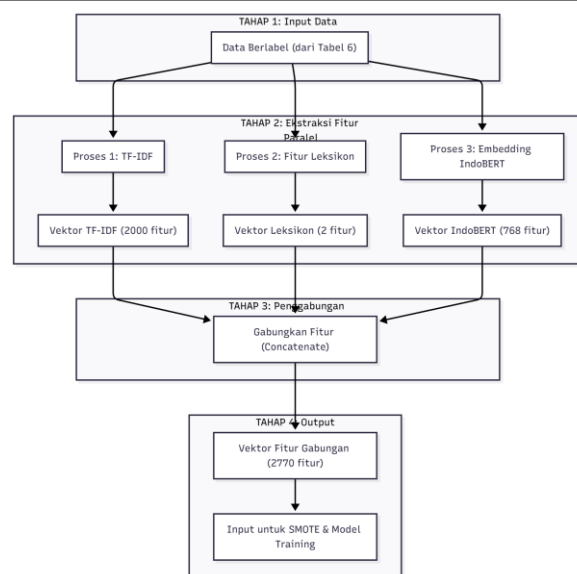
Fitur pertama adalah TF-IDF. TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah metode statistik untuk mengukur seberapa penting suatu kata (*term*) dalam sebuah dokumen terhadap keseluruhan korpus data. Bobot akhir (*W*) dari sebuah kata dihitung menggunakan Rumus 1 [8].

$$W_{t,d} = tf_{t,d} \times \log \frac{N}{df_t} \quad (1)$$

dengan  $W_{t,d}$  adalah bobot TF-IDF untuk kata (*term*)  $t$  pada dokumen  $d$ ;  $tf_{t,d}$  adalah frekuensi kemunculan kata  $t$  pada dokumen  $d$ ;  $N$  adalah jumlah total dokumen dalam korpus; dan  $df_t$  adalah jumlah dokumen yang mengandung kata  $t$ . Fitur TF-IDF ini diatur untuk menangkap *n-gram* (1,3) dan dibatasi hingga 2000 fitur teratas.

Fitur kedua adalah fitur berbasis leksikon. Fitur ini mencakup skor numerik sentimen dan skor kepercayaan yang dihasilkan oleh proses *hybrid labeling* pada tahap sebelumnya.

Fitur ketiga adalah *embedding* kalimat. Fitur ini berupa vektor 768 dimensi yang diekstraksi dari model IndoBERT (*indolem/indobert-base-uncased*).



Gambar 3. Ilustrasi Proses Feature Engineering (Penggabungan Vektor Fitur)

Dataset yang dihasilkan bersifat tidak seimbang (*imbalanced*). Untuk mengatasi hal ini, teknik oversampling SMOTE (*Synthetic Minority Over-sampling Technique*) diterapkan hanya pada data latih.

Tiga model *machine learning* kemudian dilatih di atas data yang seimbang ini, yaitu: (1) Random Forest, yang telah terbukti andal [2, 8, 10, 19], (2) XGBoost, yang dikenal memiliki performa tinggi [3, 11], dan (3) Ensemble (*Voting Classifier*).

## 2.7. Evaluasi Model

Untuk mengevaluasi kinerja dan stabilitas ketiga model, digunakan metode *Stratified K-Fold Cross-Validation* dengan  $K=5$ . Kinerja model diukur menggunakan empat metrik evaluasi standar yang didasarkan pada *Confusion Matrix* [1-4]. *Confusion Matrix* adalah tabel yang memvisualisasikan kinerja model dengan membandingkan prediksi model dengan kelas aktual. Komponen utamanya adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

Metrik pertama adalah Akurasi (*Accuracy*), yang mengukur seberapa banyak data yang diklasifikasikan dengan benar oleh model dari keseluruhan data. Akurasi dihitung menggunakan Rumus 2.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

dengan  $TP$  adalah data positif yang diprediksi benar,  $TN$  adalah data negatif yang diprediksi benar,  $FP$  adalah data negatif yang salah diprediksi positif, dan  $FN$  adalah data positif yang salah diprediksi negatif. Meskipun umum, akurasi bisa menyesatkan pada data *imbalanced*.

Metrik kedua adalah Presisi (*Precision*), yang mengukur tingkat ketepatan model dari prediksi positif yang dilakukannya. Presisi penting untuk mengetahui

seberapa banyak prediksi positif yang benar-benar positif. Presisi dihitung menggunakan Rumus 3.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Metrik ketiga adalah Perolehan (*Recall*), yang mengukur seberapa baik model dapat menemukan kembali (mendeteksi) semua data yang seharusnya positif. *Recall* sangat penting jika kerugian akibat *False Negative* tinggi. *Recall* dihitung menggunakan Rumus 4.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

Metrik keempat adalah F1-Score. Karena Presisi dan *Recall* seringkali *trade-off* (satu naik, yang lain turun), F1-Score digunakan sebagai rata-rata harmonik dari keduanya. Metrik ini sangat ideal untuk mengukur kinerja pada dataset yang tidak seimbang (seperti pada penelitian ini) karena F1-Score hanya akan tinggi jika kedua Presisi dan *Recall* tinggi. F1-Score dihitung menggunakan Rumus 5. Model dengan F1-Score rata-rata tertinggi dari hasil *cross-validation* dipilih sebagai model terbaik, yang hasilnya disajikan pada Bagian 3.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

### 2.8. Desain Algoritma

Ringkasan dari keseluruhan *pipeline* metodologi penelitian (*preprocessing*, *labeling*, *feature engineering*, *balancing*, dan *training*) yang diimplementasikan dari kode program disajikan dalam bentuk desain algoritma di bawah ini.

```

Program Jurnal
Mulai
    Load semua CSV → gabungkan → hapus duplikat
    FOR setiap teks:
        preprocess(teks)
        label_lex = lexicon_predict(teks)
        label_bert = bert_predict(teks)
        final_label =
            hybrid_decision(label_lex, label_bert)
    Extract fitur:
        bert_embed = BERT(teks)
        tfidf_feat = TFIDF(teks)
        lex_feat = lexicon_score(teks)
    Gabungkan fitur
    Balance dataset dengan SMOTE
    Lakukan K-Fold:
        Train RF, XGB, Ensemble
        Simpan metrik
    Train final model pada seluruh data
    Buat visualisasi & laporan
Selesai
    
```

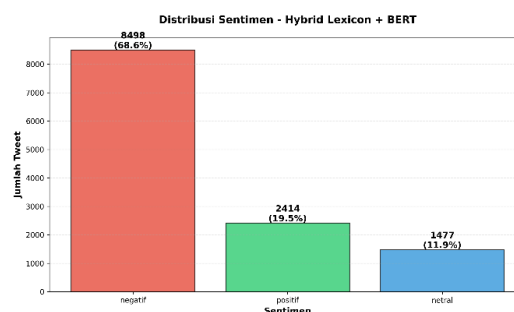
### 3. Hasil dan Pembahasan

Bagian ini menyajikan temuan dari analisis sentimen dan evaluasi model klasifikasi. Sub-bab pertama memaparkan hasil pengumpulan data dan kinerja

model, diikuti dengan pembahasan mendalam mengenai temuan tersebut.

#### 3.1. Hasil Penelitian

Penelitian ini berhasil menganalisis total 12.389 tweet terkait kebijakan makan bergizi gratis. Setelah melalui tahap *hybrid labeling* yang mengkombinasikan metode *lexicon-based* dan IndoBERT, data terdistribusi ke dalam tiga kategori sentimen. Hasil visualisasi pada Gambar 4 menunjukkan bahwa sentimen negatif merupakan yang paling dominan, mencakup 8.498 tweet (68,6%). Diikuti oleh sentimen positif sebanyak 2.414 tweet (19,5%) dan sentimen netral sebanyak 1.477 tweet (11,9%).



Gambar 4. Distribusi Sentimen Publik (Hybrid Labeling)

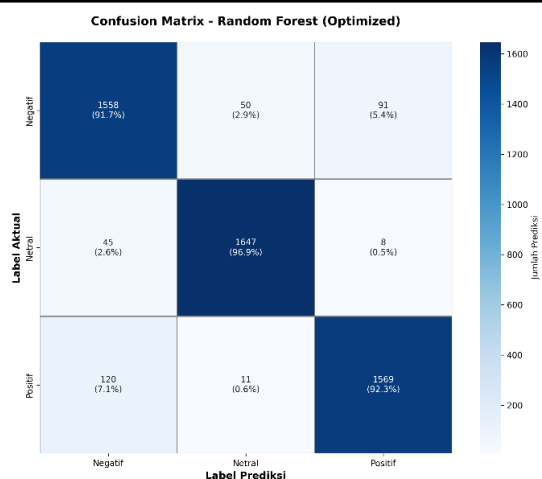
Selanjutnya, dilakukan perbandingan tiga model *machine learning* untuk menentukan arsitektur klasifikasi terbaik. Model yang diuji adalah Random Forest, XGBoost, dan Ensemble (*Voting Classifier*). Evaluasi dilakukan menggunakan metode *Stratified K-Fold Cross-Validation* dengan K=5 untuk memastikan stabilitas dan keandalan model. Hasil komparasi rata-rata (*mean*) dan standar deviasi (*std*) dari kelima fold tersebut dirangkum secara detail pada Tabel 7.

Tabel 7. Perbandingan Kinerja Model (K-Fold=5)

Model	Accuracy (Mean)	Precision (Mean)	Recall (Mean)	F1-Score (Mean)
Random Forest	0.9382	0.9386	0.9382	0.9383
XGBoost	0.9174	0.9178	0.9174	0.9167
Ensemble	0.9250	0.9251	0.9250	0.9244

Berdasarkan data pada Tabel 7, model Random Forest menunjukkan kinerja tertinggi di semua metrik evaluasi, dengan F1-Score rata-rata mencapai 0.9383. Model ini terbukti paling unggul dibandingkan XGBoost (F1: 0.9167) dan Ensemble (F1: 0.9244). Oleh karena itu, model Random Forest (Optimized) dipilih sebagai model final untuk pengujian lebih lanjut.

Kinerja detail dari model Random Forest terbaik kemudian dievaluasi pada *test set* terpisah. Hasil evaluasi disajikan dalam bentuk *confusion matrix* pada Gambar 5. *Confusion matrix* menunjukkan kemampuan model dalam memprediksi setiap kelas secara akurat.



Gambar 5. Confusion Matrix Model Random Forest (Optimized)

Secara keseluruhan, model Random Forest (Optimized) pada *test set* final berhasil mencapai akurasi 0.9363, presisi 0.9364, recall 0.9363, dan F1-Score 0.9363. Nilai F1-Score yang sangat tinggi ini mengindikasikan bahwa model memiliki keseimbangan yang sangat baik antara presisi dan recall dan mampu menangani data *imbalanced* dengan efektif.

### 3.2. Pembahasan

Temuan utama dari penelitian ini adalah sentimen masyarakat terhadap kebijakan makan bergizi gratis yang didominasi oleh respons negatif (68,6%). Hal ini mengindikasikan adanya kekhawatiran, kritik, atau penolakan yang signifikan dari publik terkait implementasi kebijakan tersebut, seperti yang tercermin dalam rekomendasi laporan analisis. Sebaliknya, proporsi sentimen positif (19,5%) dan netral (11,9%) berada pada level yang jauh lebih rendah. Temuan ini menjawab pertanyaan penelitian utama mengenai bagaimana respons publik terhadap kebijakan tersebut.

Dari sisi teknis, kinerja model yang sangat tinggi (F1-Score 0.9383) bukanlah kebetulan. Kinerja ini dicapai melalui serangkaian optimasi metodologis yang telah diuraikan pada Bagian 2. Faktor kunci keberhasilan ini mencakup penggunaan *hybrid labeling* yang menggabungkan kekuatan lexicon dan IndoBERT, *preprocessing* teks yang ditingkatkan dengan normalisasi *slang* dan penanganan negasi, serta pengayaan lexikon dengan *domain-specific terms* seperti 'korupsi', 'bergizi', dan 'higienis'.

Kinerja F1-Score 0.9383 yang dicapai oleh model Random Forest dalam penelitian ini menunjukkan keunggulan yang signifikan. Sebagai perbandingan, penelitian [2] yang juga menggunakan Random Forest (dikombinasikan dengan BERT) untuk sentimen vaksinasi COVID-19, topik yang juga kompleks, melaporkan akurasi 93%. Sementara itu, penelitian [3] yang menggunakan XGBoost untuk sentimen PPKM melaporkan akurasi 90,5%. Keunggulan kinerja dalam penelitian kami dapat diatribusikan pada dua faktor

utama: (1) penggunaan *feature engineering hybrid* yang menggabungkan tiga sumber (TF-IDF, Leksikon, dan *Embedding* IndoBERT) menjadi satu vektor fitur yang kaya, dan (2) efektivitas SMOTE dalam menangani dataset yang sangat tidak seimbang (68,6% negatif).

Analisis lebih dalam pada *confusion matrix* (Gambar 5) menunjukkan bahwa model Random Forest sangat efektif dalam mengklasifikasikan ketiga kelas. Kelas 'Netral' memiliki akurasi prediksi tertinggi (96,9%). Namun, kesalahan klasifikasi terbesar terjadi ketika model memprediksi 120 tweet yang seharusnya 'Positif' sebagai 'Negatif', yang merupakan 7,1% dari total data positif. Analisis kualitatif pada *error* ini menunjukkan bahwa kesalahan dominan terjadi pada kalimat sarkastik atau ironis. Sebagai contoh, tweet seperti "Bagus sekali programnya, pasti anggarannya tidak akan dikorupsi..." memiliki kata-kata positif ('bagus'), tetapi konteksnya sangat negatif. Model saat ini, meskipun *hybrid*, masih kesulitan membedakan sarkasme yang kompleks ini, yang menjadi salah satu keterbatasan utama penelitian.

Implikasi dari temuan sentimen ini sangat signifikan. Dominasi sentimen negatif (68,6%) bukanlah sekadar angka statistik, melainkan representasi dari kekhawatiran publik yang nyata. Analisis kualitatif terhadap sampel tweet negatif menunjukkan bahwa kekhawatiran utama masyarakat terbagi menjadi tiga kluster: (1) potensi korupsi dan kebocoran anggaran, (2) keraguan terhadap kualitas gizi dan higienitas makanan di lapangan, dan (3) transparansi sumber pendanaan program. Temuan ini sejalan dengan penelitian kebijakan publik lainnya [4, 5, 20] yang juga menemukan skeptisisme publik terhadap program berskala besar. Oleh karena itu, temuan ini dapat menjadi masukan penting bagi pemangku kepentingan untuk meningkatkan transparansi dan komunikasi publik terkait implementasi kebijakan MBG.

### 4. Kesimpulan

Penelitian ini berhasil menganalisis sentimen masyarakat terhadap kebijakan Makan Bergizi Gratis menggunakan 12.389 data tweet dari media sosial X. Hasil analisis menunjukkan bahwa sentimen publik didominasi oleh respons negatif sebesar 68,6%, sementara sentimen positif dan netral masing-masing hanya mencapai 19,5% dan 11,9%. Temuan ini mengindikasikan adanya kekhawatiran dan kritik yang signifikan dari masyarakat terhadap implementasi kebijakan tersebut.

Dari sisi metodologi, penelitian ini membuktikan bahwa arsitektur *hybrid* yang menggabungkan pendekatan *lexicon-based* dan *machine learning* menggunakan Random Forest mampu menghasilkan kinerja klasifikasi yang sangat baik, dengan nilai F1-Score sebesar 0,9363 pada pengujian akhir. Meskipun demikian, penelitian ini memiliki keterbatasan karena hanya menggunakan satu platform media sosial dan

masih menghadapi tantangan dalam mendeteksi kalimat sarkasme yang kompleks. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan data dari berbagai platform dalam rentang waktu yang lebih panjang serta mengembangkan metode khusus untuk meningkatkan kemampuan deteksi sarkasme.

### Daftar Rujukan

- [1] Sitanggang, A., Umaidah, Y., & Adam, R. I., 2024. Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naïve Bayes. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(3). doi: <http://dx.doi.org/10.23960/jitet.v12i3.4902>.
- [2] Elhan, A., Hardhienata, M. K. D., Herdiyeni, Y., Wijaya, S. H., & Adisantoso, J., 2022. Analisis Sentimen Pengguna Twitter terhadap Vaksinasi COVID-19 di Indonesia menggunakan Algoritma Random Forest dan BERT. *Jurnal Ilmu Komputer dan Agri-Informatika*, 9(2), pp. 199–211. doi: <http://dx.doi.org/10.29244/jika.9.2.199-211>.
- [3] Widiarta, I. P. A. P., Dwiyanaputra, R., & Aranta, A., 2023. Analisis Sentimen Masyarakat Terhadap Kebijakan Penerapan PpkM Di Media Sosial Twitter Dengan Menggunakan Metode Xgboost. *Jurnal Teknologi Informasi Komputer dan Aplikasi (JTika)*, 5(2), pp. 154–163. doi: <http://dx.doi.org/10.29303/jtika.v5i2.342>.
- [4] 'Adilah, H. S., & Alit, R., 2023. Analisis Sentimen Masyarakat Twitter Terhadap Kebijakan Pemerintah Dalam Menaikkan Harga Bahan Bakar Minyak Dengan Menggunakan Metode Support Vector Machine. *Jurnal Informatika Computer Science (JINACS)*, 5(02), pp. 201–215. doi: <http://dx.doi.org/10.26740/jinacs.v5n02.p201-215>.
- [5] Manoppo, M. R., dkk., 2025. Analisis Sentimen Publik Di Media Sosial Terhadap Kenaikan Ppn 12% Di Indonesia Menggunakan Indobert. *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, 4(2), pp. 152–163. doi: <http://dx.doi.org/10.69916/jkbt.v4i2.322>.
- [6] Syahrul, E., & Fatharani, D., 2025. Hybrid Sentiment Analysis Of Maxim App Users Using Support Vector Machine And Lexicon-Based Approach. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(3S1). doi: <http://dx.doi.org/10.23960/jitet.v13i3S1.8148>.
- [7] Najib, A. C., Irsyad, A., Qandi, G. A., & Rakhmawati, N. A., 2019. Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter. *Fountain of Informatics Journal*, 4(2), pp. 41–48. doi: <http://dx.doi.org/10.21111/fij.v4i2.3573>.
- [8] Baskoro, B. B., Susanto, I., & Khomsah, S., 2021. Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR). [Daring]. Tersedia di: <https://journal.itelkompwt.ac.id/index.php/inista/article/download/218/98>.
- [9] Al-Kadzim, M. G., Rasim, R., & Herbert, H., 2024. Analisis Perubahan Sentimen Publik di Media Sosial X terhadap Konflik Palestina-Israel Menggunakan Model IndoBERT. *Digital Transformation Technology*, 4(2), pp. 1167–1174. doi: <http://dx.doi.org/10.47709/digitech.v4i2.5312>.
- [10] Indrayanto, C. G., Ratnawati, D. E., & Rahayudi, B. Analisis Sentimen Data Ulasan Pengguna Aplikasi MyPertamina di Indonesia pada Google Play Store menggunakan Metode Random Forest.
- [11] Yulistiani, Y., & Styawati, S., 2024. Analisis Sentimen Terhadap Calon Presiden Indonesia 2024 dengan Metode Extreme Gradient Boosting (XGBOOST). *Jurnal Informatika: Jurnal Pengembangan IT*, 9(3), pp. 322–328. doi: <http://dx.doi.org/10.30591/jpit.v9i3.6127>.
- [12] Akhdaan, D. A., Sutanto, T. E., & Liebenlito, M., 2024. Confident Learning pada IndoBERT: Peningkatan Kinerja Klasifikasi Sentimen. *Indonesian Journal of Computer Science*, 13(5). doi: <http://dx.doi.org/10.33022/ijcs.v13i5.4401>.
- [13] Daulay, I., Firdaus, M., Ardiansyah, B., Hutagaol, R., & Rahmaddeni, R., 2023. Analisis Sentimen Opini Publik Terhadap Penerimaan Bantuan Subsidi Upah (BSU) Menggunakan Algoritma Naive Bayes. *SENTIMAS Seminar Nasional Penelitian dan Pengabdian Masyarakat*, pp. 155–162.
- [14] Ahda, F. A., & Zainuddin, M., 2019. Prediksi Kepuasan Pelayanan Perpustakaan Menggunakan Algoritma Decision Tree (C4.5). *Jurnal Teknologi Informasi*, 10(2), pp. 143–150. doi: <http://dx.doi.org/10.36382/jti-tki.v10i2.368>.
- [15] Anggraini, W. P., & Utami, M. S., 2021. Klasifikasi Sentimen Masyarakat Terhadap Kebijakan Kartu Pekerja Di Indonesia. *Faktor Exacta*, 13(4), pp. 255. doi: <http://dx.doi.org/10.30998/faktorexacta.v13i4.7964>.
- [16] Nurhasananda, A., & Akbar, M., 2025. Analisis Sentimen Masyarakat Terhadap Kebijakan Undang-Undang Tentara Nasional Indonesia (UU TNI) Menggunakan Support Vector Machine. *Jurnal Komputer Informasi dan Teknologi*, 5(1), pp. 14–14. doi: <http://dx.doi.org/10.53697/jkomitek.v5i1.2603>.
- [17] Romadhona, W., & Isnain, A. R., 2024. Analisis Sentimen Pengguna Media Sosial Terhadap Kebijakan Kenaikan Pajak Hiburan Menggunakan Metode Svm (Support Vector Machine). *JUPI Jurnal Ilmiah Penelitian dan Pembelajaran Informatika*, 9(4), pp. 2185–2195. doi: <http://dx.doi.org/10.29100/jupi.v9i4.5603>.
- [18] Lestari, S., Mupaat, M., & Erfina, A., 2022. Analisis Sentimen Masyarakat Indonesia terhadap Pemindahan Ibu Kota Negara Indonesia pada Twitter. *JUSIFO Jurnal Sistem Informasi*, 8(1), pp. 13–22. doi: <http://dx.doi.org/10.19109/jusifo.v8i1.12116>.
- [19] Adhan, S. N., Wibawa, G. N. A., Arisona, D. C., Yahya, I., & Ruslan, R., 2024. Analisis Sentimen Ulasan Aplikasi Wattpad Di Google Play Store Dengan Metode Random Forest. *Anoatik Jurnal Teknologi Informasi dan Komputer*, 2(1), pp. 6–15. doi: <http://dx.doi.org/10.33772/anoatik.v2i1.32>.
- [20] Sulisty, D. A., Wibawa, A. P., Prasetya, D. D., & Ahda, F. A., 2023. LSTM-Based Machine Translation for Madurese-Indonesian. *Journal of Applied Data Sciences*, 4(3), pp. 189–199. doi: <http://dx.doi.org/10.47738/jads.v4i3.113>.
- [21] Kalalo, A. N., Juhriah, E., & Tantowi, A. A., 2025. Analisa Sentimen Indonesia Mengenai Krl Commuter Line Menggunakan Metode Naïve Bayes Dan Support Vector Machine. *String Satuan Tulisan Riset dan Inovasi Teknologi*, 9(3), pp. 345–353. doi: <http://dx.doi.org/10.30998/string.v9i3.28706>.
- [22] Ahda, F. A., Wibawa, A. P., Prasetya, D. D., & Sulisty, D. A., 2024. Comparison of Adam Optimization and RMS prop in Minangkabau-Indonesian Bidirectional Translation with Neural Machine Translation. *JOIV International Journal on Informatics Visualization*, 8(1), pp. 231–238. doi: <http://dx.doi.org/10.62527/joiv.8.1.1818>.
- [23] Rahman, R. A., Saputra, R. F., & Yaqin, M. A., 2024. Implementasi Metode Agile pada Pengembangan Sistem Informasi Manajemen Masjid Berbasis Website. *SISFO*, 11(2).
- [24] Sulisty, D., Ahda, F., & Fitria, V. A., 2021. Epistemologi dalam Natural Language Processing. *Jurnal Inovasi Teknologi dan Edukasi Teknik*, 1(9), pp. 652–664. doi: <http://dx.doi.org/10.17977/um068v1i92021p652-664>.
- [25] Naraswati, N. P. G., Nooraeni, R., Rosmilda, D. C., Desinta, D., Khairi, F., & Damaiyanti, R., 2021. Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification. *SISTEMASI*, 10(1), pp. 222. doi: <http://dx.doi.org/10.32520/stmsi.v10i1.1179>.