

Klasifikasi Opini Tidak Informatif Pada Program Makan Bergizi Gratis (MBG) Menggunakan *Random Forest*

Laila Syal Syabilla¹, Mei Intan Natasyah², Fathoni³, Jeremiah Alwin Siahaan⁴

Universitas Sriwijaya^{1,2,3,4}

09031382328125@student.unsri.ac.id¹, 09031282328047@student.unsri.ac.id², fathoni@unsri.ac.id³, 09031182328007@student.unsri.ac.id⁴

Diterima (18-03-2026)	Direvisi (30-03-2026)	Disetujui (15-04-2026)
--------------------------	--------------------------	---------------------------

Abstrak – Pemerintahan Prabowo-Gibran meluncurkan kebijakan strategis Program Makan Bergizi Gratis (MBG) untuk menjamin pemenuhan hak dasar anak atas pangan yang aman, sehat, dan bergizi. Urgensi program ini didasarkan pada angka stunting di Indonesia tahun 2024 yang mencapai 14%, sehingga peluncuran kebijakan Program Makan Bergizi Gratis (MBG) memicu diskusi publik yang masif di platform media sosial X dengan jumlah pengguna mencapai 24,7 juta orang. Namun, volume data yang besar tersebut menghadirkan masalah "Data Sampah" (*Noise*) berupa spam, promosi jualan, hingga akun bot yang berpotensi menyebabkan bias pada analisis opini publik terhadap program Program Makan Bergizi Gratis (MBG). Penelitian ini bertujuan membangun model klasifikasi opini tidak informatif dengan mengimplementasikan algoritma *Random Forest* berbasis *Knowledge Discovery in Database (KDD)* sebagai tahap pra-pemrosesan sebelum analisis sentimen lanjutan. Data yang digunakan berjumlah 10.000 *tweet* bersumber dari Kaggle, diproses melalui lima tahapan KDD meliputi *Data Selection*, *Data Preprocessing*, *Data Transformation*, *Data Mining*, dan *Data Evaluation* dengan menggunakan *RapidMiner*. Representasi fitur dilakukan dengan pembobotan TF-IDF dan validasi model menggunakan *k-fold Cross Validation* dengan $k=10$. Hasil evaluasi menunjukkan model mencapai akurasi 82,03%, *precision* 93,78%, *recall* 68,61%, dan *F-Measure* 79,24%. Hasil ini membuktikan bahwa pendekatan KDD berbasis *Random Forest* efektif digunakan sebagai *pipeline filter noise* yang terstruktur untuk teks media sosial berbahasa Indonesia, khususnya pada domain opini kebijakan pemerintah.

Kata Kunci: Opini Tidak Informatif, Program Makan Bergizi Gratis, *Random Forest*, *Knowledge Discovery in Database*, *Text Mining*.

Abstract - The Prabowo-Gibran government launched the Free Nutritious Meals (MBG) strategic policy to ensure the fulfillment of children's basic rights to safe, healthy, and nutritious food. The urgency of this program is based on Indonesia's stunting rate in 2024 which reached 14%, so the launch of the Free Nutritious Meals (MBG) policy triggered massive public discourse on the social media platform X with a user base reaching 24.7 million people. However, the large volume of data presents a "Data Noise" problem in the form of spam, product promotions, and bot accounts that could potentially cause bias in the analysis of public opinion regarding the Free Nutritious Meals (MBG) program. This study aims to build an uninformative opinion classification model by implementing the *Random Forest* algorithm based on the *Knowledge Discovery in Database (KDD)* framework as a pre-processing stage prior to further sentiment analysis. The data used consisted of 10,000 tweets sourced from Kaggle, analyzed using five sequential KDD phases *Data Selection*, *Data Preprocessing*, *Data Transformation*, *Data Mining*, and *Data Evaluation* implemented within the *Rapidminer* environment. Feature representation was performed using TF-IDF weighting and model validation was conducted using *k-fold Cross Validation* with $k=10$. Evaluation results show that the model achieved an accuracy of 82.03%, precision of 93.78%, recall of 68.61%, and *F-Measure* of 79.24%. These results demonstrate that the KDD-based *Random Forest* approach is effective as a structured noise filter pipeline for Indonesian social media text, particularly in the domain of government policy opinion.

Keywords: Uninformative Opinion, Free Nutritious Meals Program, *Random Forest*, *Knowledge Discovery in Database*, *Text Mining*.

I. PENDAHULUAN

Program Makan Bergizi Gratis (MBG) merupakan salah satu program yang dibentuk dengan tujuan menangani permasalahan kekurangan gizi dan meningkatkan kualitas Sumber Daya Manusia (SDM) Indonesia di era pemerintahan Prabowo-Gibran (Kiftiyah et al., 2025). Karena mengatasi permasalahan kekurangan gizi adalah salah satu dari tiga faktor utama untuk mencapai visi Indonesia Emas 2045, anak-anak usia sekolah menjadi sasaran utama program ini, dengan melalui penyediaan makanan bergizi secara gratis oleh pemerintah (Albaburrahim et al., 2025). Kebijakan strategis ini bertujuan untuk memenuhi hak dasar anak atas makanan yang aman, sehat, dan bergizi (Candra et al., 2025), sekaligus meningkatkan produktivitas jangka panjang individu demi kualitas SDM yang lebih baik (Habibah et al., 2025).

Urgensi program ini didasarkan pada angka stunting di Indonesia tahun 2024 yang mencapai 14% (Kiftiyah et al., 2025). Dengan sasaran masif sebanyak 82,9 juta penerima manfaat, kebijakan ini memicu perdebatan sosial yang luas di masyarakat (KS et al., 2025). Diskusi publik tersebut banyak terjadi di media sosial X yang kini menjadi ruang diskusi strategis dengan jumlah pengguna mencapai 24,7 juta orang pada awal 2024 (Mardiana & Abidin, 2025; Rahmatullah & Annisa, 2025). Persepsi Masyarakat terhadap program pemerintah sangat dipengaruhi oleh perkembangan diskusi di media sosial (Hermansyah & Hasibuan, 2025), namun volume data yang besar di platform X ini menghadirkan masalah "Data Sampah" (*Noise*).

Banyak ditemukan tweet berupa spam, promosi jualan, hingga akun bot yang tidak memiliki substansi informasi. Hal ini menjadi hambatan serius karena jika opini tidak informatif tersebut tidak dibuang, maka hasil analisis sentimen mengenai respons masyarakat terhadap program MBG akan menjadi tidak akurat atau bias. Oleh karena itu, diperlukan algoritma yang mampu menangani dimensi data besar dan menghasilkan prediksi akurat seperti *Random Forest* (Anam et al., 2026; Hapsari & Indriyanti, 2023). Penggunaan *framework* ini dinilai tepat karena terbukti mampu mengatasi data berdimensi tinggi dengan relasi yang kompleks (Ramadhan & Alfat, 2026).

Kajian opini publik seputar program MBG kini telah menjadi tren riset yang signifikan dengan berbagai pendekatan komputasi. Sebagai contoh, analisis sentimen terkait program ini telah dilakukan menggunakan algoritma

Random Forest yang dioptimasi dengan teknik SMOTE, sehingga mampu mengklasifikasikan polaritas sentimen positif dan negatif di media X dengan akurasi yang maksimal (Azhari & Parjito, 2024). Selain itu, evaluasi komparatif terhadap sepuluh algoritma pembelajaran mesin juga menunjukkan bahwa *Random Forest* memiliki performa yang stabil (Utama et al., 2025), dan bahkan mampu mencapai akurasi hingga 97% pada skala dataset yang lebih besar (Napiah et al., 2026). Pendekatan lain seperti integrasi *IndoBERT* (Sulistyo & Setiadi, 2025) maupun arsitektur *Capsule Network* (Raditya et al., 2026) turut membuktikan keandalan ekstraksi fitur dalam klasifikasi teks.

Namun, dibalik tingginya akurasi berbagai model analisis sentimen tersebut, terdapat tantangan mendasar berupa tingginya volume data tidak relevan atau *noise* pada platform X yang belum ditangani secara sistematis (Saleh & Imanda, 2025). Sebagian besar penelitian terdahulu langsung melakukan kategorisasi sentimen (positif, negatif, atau netral) tanpa memprioritaskan pembersihan opini yang tidak informatif seperti komentar spam, promosi, atau ulasan tanpa substansi informasi. Kemampuan metode klasifikasi yang terbatas dalam mengolah teks informal serta opini yang tidak relevan ini berisiko menyebabkan bias pada hasil analisis persepsi publik yang sebenarnya. Meskipun terdapat rekomendasi untuk menggunakan algoritma yang lebih stabil terhadap *noise* seperti *Random Forest* (Fathoni et al., 2025), penelitian yang secara khusus menerapkan tahapan KDD untuk memfilter opini tidak informatif pada isu MBG belum banyak dilakukan.

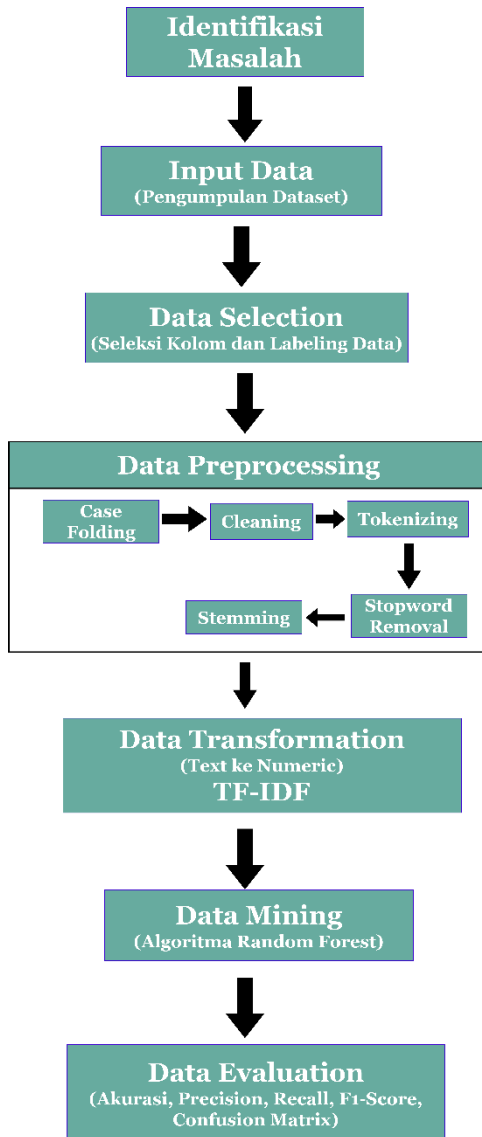
Berdasarkan tantangan tersebut, penelitian ini difokuskan pada pemodelan klasifikasi kualitas informasi ulasan menggunakan algoritma *Random Forest*. Dengan menerapkan pendekatan berbasis KDD, penelitian ini tidak hanya melakukan klasifikasi opini informatif dan tidak informatif, tetapi diharapkan mampu menyediakan data yang lebih bersih dan akurat bagi para pengambil kebijakan dalam memahami dinamika opini masyarakat terhadap kebijakan pemerintah.

Secara keseluruhan penelitian ini memiliki tujuan untuk mengidentifikasi karakter opini tidak informatif pada isu MBG dan menghasilkan model klasifikasi yang dapat memfilter data sebelum dianalisis lebih lanjut. Kontribusi penelitian ini adalah dengan memberikan referensi dalam pengembangan *text mining* berbahasa Indonesia, serta membantu pengambilan kebijakan dalam memahami

dinamika opini masyarakat yang lebih informatif melalui data media sosial.

II. METODOLOGI PENELITIAN

Proses klasifikasi dalam penelitian ini dilaksanakan dengan menerapkan kerangka berbasis KDD menggunakan algoritma *Random Forest*.



Sumber: Peneliti (2026)

Gambar 1. Diagram Alur Proses Klasifikasi

Pada gambar 1 mengilustrasikan alur proses klasifikasi yang terdiri dari 5 tahap utama yang diterapkan, yakni *Data Selection*, *Data Preprocessing*, *Data Transformation*, *Data Mining*, dan *Data Evaluation*(Athallah et al.,2025). Berikut penjabaran setiap tahap tersebut:

1. Metode Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle dan dapat diakses melalui tautan berikut: <https://www.kaggle.com/datasets/alzain03/program-makan-bergizi-gratis-10k-data>. *Dataset* ini mencakup sekitar 10.000 baris data yang dikumpulkan dari platform X terkait Program MBG. Selain 15 nama kolom yang tersedia pada *dataset* tersebut, ditambahkan satu kolom yaitu label sebagai variabel target melalui proses anotasi berbasis aturan (*rule-based labeling*). Kolom ini dikategorikan menjadi dua kelas, yakni nilai 1 untuk opini informatif dan 0 untuk opini tidak informatif, sebagai dasar pembelajaran model pada tahap *Data Mining*. *Dataset* ini akan diproses lebih lanjut untuk klasifikasi opini tidak informatif, seperti yang ditampilkan pada Tabel 1.

Tabel 1. *Dataset* Opini Tidak Informatif Program MBG

Nama Kolom	Deskripsi
conversation_id_str	ID unik percakapan yang menghubungkan berbagai cuitan dalam satu utas diskusi.
created_at	Waktu dan tanggal saat opini diunggah (digunakan untuk analisis aktualitas).
favorite_count	Jumlah apresiasi (like) yang diterima oleh cuitan.
full_text	Berisi teks opini masyarakat yang akan diklasifikasikan.
id_str	Identitas unik cuitan dalam format string sebagai referensi database.
image_url	Tautan gambar yang disertakan dalam cuitan (jika ada).
in_reply_to_screen_name	Nama pengguna yang dituju dalam sebuah balasan cuitan.
lang	Kode bahasa; digunakan untuk memvalidasi penggunaan Bahasa Indonesia (in).
location	Informasi geografis atau asal wilayah pengguna.
quote_count	Jumlah kutipan (<i>quotes</i>) yang diterima oleh cuitan.
reply_count	Jumlah balasan (<i>replies</i>) yang diterima oleh cuitan.
retweet_count	Jumlah penyebaran ulang (<i>retweets</i>) dari cuitan tersebut.

search_query	Kata kunci pencarian yang digunakan untuk menarik dataset ini dari platform X.
user_id_str	Identitas unik pengguna yang memposting cuitan.
username	Nama tampilan akun pengguna Twitter/X.

Sumber: Kaggle (2025)

2. Tahapan Knowledge Discovery in Database (KDD)

a. Data Selection

Dalam konteks penelitian ini, pemilihan data difokuskan pada karakteristik teks opini masyarakat terhadap Program Makan Bergizi Gratis yang memberikan kontribusi signifikan terhadap akurasi klasifikasi (Athallah et al., 2025; Rois et al., 2025). Atribut-atribut yang dipilih merupakan parameter penting yang secara teknis dianggap paling menentukan dalam prediksi (Rois et al., 2025).

b. Data Preprocessing

Pengolahan data dilakukan untuk memastikan dataset opini yang digunakan bersih, terorganisir, dan relevan (Nuru Aini et al., 2024; Sultan et al., 2025), dengan melalui langkah-langkah sistematis sebagai berikut:

- 1) *Cleaning*: Melakukan pembersihan data dengan mengecek *missing value*, menghapus data duplikat, serta menghilangkan gangguan (*noise*), seperti angka, tanda baca, simbol, kurung kotak, tautan (link), tag HTML, dan *new line* (\n) (Firdaus et al., 2024; Mahing et al., 2024; Nuru Aini et al., 2024).
- 2) *Case Folding (Lower Casting)*: Seluruh karakter huruf kapital dikonversi menjadi huruf kecil guna menyeragamkan format teks (Firdaus et al., 2024; Mahing et al., 2024).
- 3) *Stopword Removal*: Tahap ini mengeliminasi kata-kata umum yang sering muncul seperti kata 'the', 'and', dan 'a' (Mahing et al., 2024).
- 4) *Tokenizing*: Memecah kalimat menjadi satuan kata tunggal atau token untuk diolah ke dalam fitur *bag of words* (Firdaus et al., 2024).
- 5) *Stemming*: Mengembalikan kata ke bentuk dasarnya (Mahing et al., 2024) guna meminimalisir redundansi kata dan menyederhanakan dokumen.

c. Data Transformation

Tahap *Data Transformation* dilakukan untuk menyesuaikan format dari atribut dengan kebutuhan dari algoritma yang dipilih

(Athallah et al., 2025). Tujuannya adalah menyelaraskan karakteristik setiap atribut agar sesuai dengan persyaratan algoritma yang digunakan (Rois et al., 2025), sehingga proses penambangan data (*data mining*) dapat berjalan lebih efektif dan optimal. Dalam implementasi *Random Forest*, ulasan teks perlu diubah menjadi angka atau fitur baru yang lebih representatif agar model dapat mengenali pola informasi dengan lebih akurat.

d. Data Mining

Tahap *Data mining* dilakukan pembangunan model klasifikasi melalui penggunaan algoritma *Random Forest* (Athallah et al., 2025). Proses ini dilakukan untuk menemukan informasi berharga dari kumpulan data dengan memanfaatkan metode tertentu guna menghasilkan data *training* yang berkualitas. Algoritma ini dipilih karena kemampuannya dalam menangani data berjumlah besar, memiliki banyak atribut, serta kemampuannya untuk mengurangi *overfitting* melalui agregasi dari banyak *decision tree* (Athallah et al., 2025).

e. Data Evaluation

Kesimpulan penelitian diperoleh dengan mengevaluasi hasil klasifikasi guna menentukan tingkat akurasi terbaik dari masing-masing kelas (Athallah et al., 2025). Penilaian ini dilakukan menggunakan *Accuracy*, *Confusion Matrix*, *Precision*, *Recall*, dan *F1-score* (Idris et al., 2025). Ketentuan penilaian merujuk pada format *Confusion Matrix* yang disajikan pada Tabel 2 guna menggambarkan kelompok prediksi yang benar dan salah yang dibuat oleh pengklasifikasi (Saputra et al., 2024).

Tabel 2. *Confusion Matrix*

Predicted	Actual	
	True	False
True	TP (<i>True Positive</i>)	FP (<i>False Positive</i>)
False	TN (<i>True Negative</i>)	FN (<i>False Negative</i>)

Sumber: (Athallah et al., 2025)

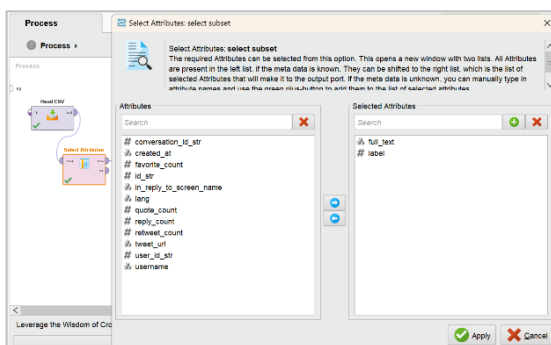
Pada tahap evaluasi ini dilakukan pengujian terhadap algoritma *Random Forest* untuk menghasilkan performa model yang optimal (Mawarni et al., 2023).

III. HASIL DAN PEMBAHASAN

Pada tahap ini menyajikan hasil implementasi setiap tahapan KDD dalam membangun model klasifikasi dengan algoritma *Random Forest*. Setiap tahapan dijabarkan secara sistematis mulai dari *Data Selection* hingga *Data Evaluation*, disertai visualisasi proses pada *RapidMiner* dan analisis hasil yang diperoleh.

1. Data Selection

Tahapan *Data Selection*, proses penyaringan dilakukan dengan operator *select attributes* dengan tipe *subset*. Pada 15 atribut yang dimiliki *dataset*, yang dipilih untuk dipertahankan adalah *full_text* dan label. Hal ini bertujuan untuk membuang dimensi data yang tidak relevan sehingga algoritma berfokus murni pada teks opini, yang membuat penjelasan klasifikasi nantinya menjadi lebih detail dan mudah dipahami.



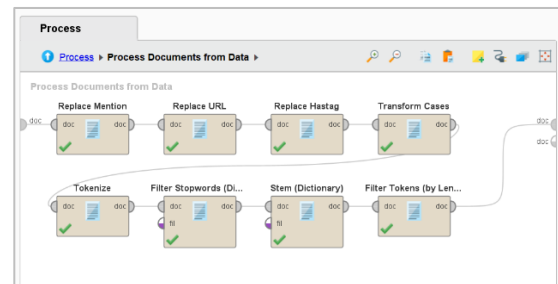
Sumber: Peneliti (2026)

Gambar 2. Tahap awal *Data Selection* pada *Rapidminer*

2. Data Preprocessing

Pada tahap *data preprocessing* dilaksanakan untuk membersihkan teks ulasan mentah agar siap diolah oleh algoritma melalui serangkaian operator secara berurutan. Langkah pertama adalah proses *cleaning* untuk membuang elemen pengganggu khas media sosial yang masih berbentuk kalimat utuh menggunakan operator *replace mention*, *replace URL*, dan *replace hashtag*. Teks yang sudah bersih dari atribut tersebut kemudian diseragamkan menjadi huruf kecil seluruhnya melalui *transform cases*. Setelah format seragam, kalimat dipecah menjadi satuan kata tunggal menggunakan operator *Tokenize*. Kata-kata tunggal ini kemudian diseleksi lebih lanjut dengan *Filter Stopwords (Dictionary)* untuk menghapus kata hubung atau kata umum yang tidak relevan. Setelah itu, diterapkan proses *Stemming*

menggunakan operator *Stem (Dictionary)* yang memetakan kata yang berimbuhan kembali ke kata dasarnya berdasarkan kamus bahasa indonesia guna meminimalisir redundansi makna. Rangkaian ini diakhiri dengan *Filter Tokens (by Length)* untuk membuang sisa-sisa token yang jumlah karakternya terlalu pendek atau terlalu panjang, sehingga hanya menyisakan kata-kata yang esensial untuk pemodelan.

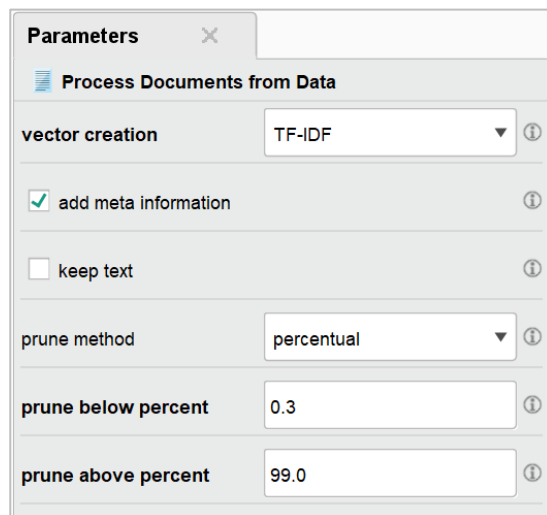


Sumber: Peneliti (2026)

Gambar 3. Proses *preprocessing* pada *Rapidminer*

3. Data Transformation

Setelah data teks dibersihkan, tahapan *data transformation* dilakukan dengan mengonversi teks menjadi representasi numerik (*text to numeric*) sehingga dapat diproses oleh model klasifikasi. Transformasi ini dilakukan dengan metode pembobotan *Term Frequency-Inverse Document Frequency (TF-IDF)* guna mengukur tingkat kepentingan sebuah kata dalam korpus teks *relative* terhadap keseluruhan data. Untuk mengoptimalkan fitur numerik yang dihasilkan, diterapkan teknik pemangkasan (*pruning*) dengan metode persentase (*percentual*). Parameter diatur dengan *prune below percent* sebesar 0.3 dan *prune above percent* sebesar 99.0. Langkah ini berfungsi untuk membuang kata-kata yang kemunculannya sangat langka (di bawah 0,3%) atau terlalu mendominasi (di atas 99,0%), sehingga algoritma *Random Forest* nantinya hanya memproses fitur numerik yang paling esensial dan berkualitas.

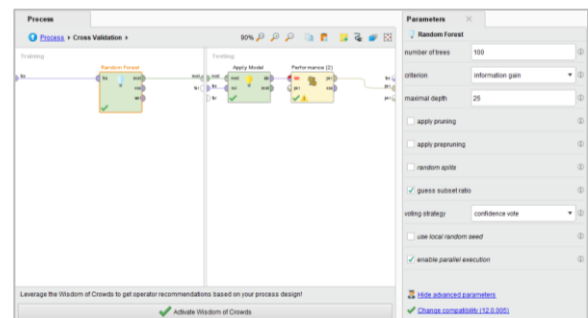


Sumber: Peneliti (2026)

Gambar 4. Proses *Data Transformation* dengan pembobotan TF-IDF pada *Rapidminer*

4. Data Mining

Pada tahap *data mining*, proses klasifikasi dilakukan dengan menggunakan algoritma *Random Forest* yang diuji dengan metode *k-fold Cross Validation* dengan nilai $k=10$. Konfigurasi *hyperparameter Random Forest* yang digunakan seperti *number of trees* sebesar 100, *criterion* dengan *information gain*, dan *maximal depth* sebesar 25, dan *voting strategy* berbasis *confidence vote*. Parameter *apply pruning* dan *apply prepruning* tidak diaktifkan karena algoritma *Random Forest* secara *inheren* sudah mengatasi *overfitting* melalui mekanisme agregasi *ensemble* dari 100 *decision tree* yang dibangun secara independen, sehingga *pruning* tambahan tidak diperlukan. Penerapan *k-fold Cross Validation* dengan nilai $k=10$ guna menjamin estimasi performa model yang lebih stabil dan objektif dibandingkan metode *split data* tunggal, khususnya pada *dataset* yang variabilitas linguistik tinggi. Dalam prosedur ini, *dataset* secara otomatis dibagi menjadi sepuluh *subset* yang sama rata, di mana pada setiap iterasi, sembilan *subset* dialokasikan sebagai *data training* untuk membangun struktur *ensemble tree* pada *Random Forest*, sementara satu *subset* sisanya berfungsi sebagai *data testing* untuk menguji performa model. Pendekatan iteratif ini memastikan bahwa setiap sampel data dalam korpus pernah berperan sebagai data uji, sehingga metrik evaluasi yang dihasilkan merupakan representasi yang kuat (*robust*) terhadap kemampuan generalisasi model dalam mengklasifikasikan opini tidak informatif Program MBG.



Sumber: Peneliti (2026)

Gambar 5. Konfigurasi metode *K-fold Cross Validation* ($k=10$) dan *Hyperparameter Random Forest* pada *RapidMiner*

5. Data Evaluation

Tahap evaluasi model dilaksanakan guna mengevaluasi kemampuan algoritma *Random Forest* dalam mengklasifikasikan opini tidak informatif pada Program MBG melalui empat parameter utama, yaitu *accuracy*, *precision*, *recall*, dan *f-measure*. Proses evaluasi dilakukan melalui aplikasi *RapidMiner*, dan *output dari confusion matrix* ditampilkan pada Gambar 6.

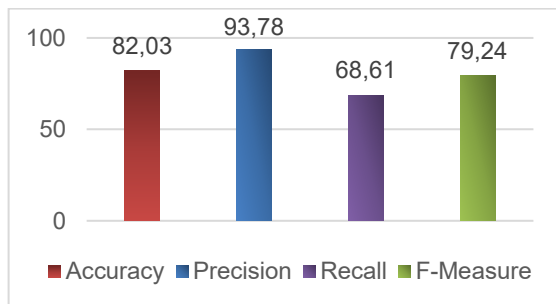
Criterion	Value
accuracy	82.03% +/- 0.77% (micro average: 82.03%)
precision	
recall	
f measure	

	true 1	true 0	class precision
pred. 1	7944	2580	75.25%
pred. 0	374	5638	93.78%
class recall	95.45%	68.61%	

Sumber: Peneliti (2026)

Gambar 6. Hasil *Confusion Matrix Model Random Forest*

Data pada Gambar 6 menunjukkan hasil dari total sampel, model dapat memprediksi 7.844 data sebagai *True Positive* (TP) dan 5.638 data sebagai *True Negative* (TN). Meskipun demikian, terdapat sejumlah kesalahan klasifikasi berupa 374 data *False Positive* (FP) dan 2.580 data *False Negative* (FN). Perbandingan keempat metrik evaluasi secara visual ditampilkan pada Gambar 7.



Sumber: Peneliti (2026)

Gambar 7. Perbandingan Hasil Evaluasi Model *Random Forest*

Gambar 7 menyajikan keempat perbandingan hasil matrik evaluasi dengan nilai *Precision* sebagai performa tertinggi sebesar (93,78%), selanjutnya nilai *Accuracy* (82,03%), *F-Measure* (79,24%) dan *Recall* sebagai nilai terendah sebesar (68,61%). Rekapitulasi hasil evaluasi selengkapnya ditampilkan pada Tabel 3.

Tabel 3. Rekapitulasi Hasil Evaluasi

Metrik Evaluasi	Hasil (%)
<i>Accuracy</i>	82,03%
<i>Precision</i>	93,78%
<i>Recall</i>	68,61%
<i>F-Measure</i>	79,24%

Sumber: Peneliti (2026)

Nilai *Precision* sebesar 93,78% menunjukkan bahwa model sangat akurat dalam meminimalisir kesalahan pelabelan opini yang sebenarnya bukan target klasifikasi. Sedangkan, nilai *Recall* sebesar 68,61% menggambarkan proporsi data tidak informatif yang berhasil dikenali dari keseluruhan data aktual kelas tidak informatif. Integrasi keduanya menghasilkan *F-Measure* sebesar 79,24%, yang membuktikan bahwa performa Algoritma *Random Forest* memiliki tingkat keselarasan yang baik dalam klasifikasi opini tidak informatif pada Program MBG. Nilai *Recall* yang lebih rendah dibandingkan *Precision* menandakan adanya *False Negative* yang signifikan, yang konsisten dengan temuan Saleh & Imanda (2025) bahwa variabilitas linguistik tinggi pada teks Twitter berbahasa Indonesia mencakup penggunaan *slang*, singkatan, dan sarkasme menjadi faktor pembatas utama dalam klasifikasi teks informal. Jika dibandingkan dengan penelitian Utama et al. (2025) yang mencapai akurasi 77% tanpa tahapan filter *noise*, model dalam penelitian ini menunjukkan peningkatan akurasi sebesar 5,03 poin *persentase*, yang menunjukkan bahwa integrasi antara *pipeline* KDD berkontribusi

positif terhadap kualitas klasifikasi.

IV. KESIMPULAN

Hasil dari penelitian ini berhasil membangun model klasifikasi opini tidak informatif pada Program MBG dengan mengimplementasikan algoritma *Random Forest* dalam *framework* KDD sebagai tahap pra-pemrosesan sebelum analisis sentimen lanjutan. Hasil evaluasi menunjukkan model mencapai akurasi 82,03%, *precision* 93,78%, *recall* 68,61%, dan *F-Measure* 79,24%, yang membuktikan bahwa pendekatan KDD berbasis *Random Forest* efektif digunakan sebagai *pipeline filter noise* yang terstruktur untuk teks media sosial berbahasa Indonesia pada domain opini kebijakan pemerintah.

Nilai *recall* yang lebih rendah dibandingkan *precision* menandakan adanya keterbatasan model dalam mengenali opini tidak informatif yang mengandung variabilitas linguistik tinggi seperti *slang* dan sarkasme. Untuk penelitian selanjutnya, disarankan mengeksplorasi teknik augmentasi data seperti SMOTE untuk menangani ketidakseimbangan kelas, serta mengintegrasikan *pre-trained language model* seperti IndoBERT guna meningkatkan kemampuan pemahaman konteks bahasa informal.

V. REFERENSI

- Albaburrahim, A., Putikadyanto, A. P. A., Efendi, A. N., Alatas, M. A., Romadhon, S., & Wachidah, L. R. (2025). Program makan bergizi gratis: Analisis kritis transformasi pendidikan Indonesia menuju generasi emas 2045. *Entita: Jurnal Pendidikan Ilmu Pengetahuan Sosial Dan Ilmu-Ilmu Sosial*, 767–780.
- Anam, M. H. K., Kurnianingtyas, D., & Soebroto, A. A. (2026). Implementasi Algoritma *Random Forest* Untuk Prediksi Churn Pada Pelanggan Retail Online. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 10(4).
- Athallah, D., Fathoni, F., & Rachmad, M. I. F. (2025). Klasifikasi Risiko Strok Menggunakan Algoritma *Random Forest* dengan Teknik Knowledge Discovery in Database. *Indonesian Journal Computer Science*, 4(1), 38–44.
- Azhari, M., & Parjito, P. (2024). Analisis sentimen opini publik program makan siang gratis dengan *random forest* pada media. *Building of Informatics, Technology and Science (BITS)*, 6(3), 1932–1942.
- Candra, A. N., Kamila, S. I., Victoria, M. Z. A.,

- Alviya, B., & Bintang, F. N. C. (2025). Failure of Governance and State Responsibility in the Free Nutritious Meals (FNM) Program. *Journal of Social, Policy and Development Studies*, 2(2), 27–38.
- Fathoni, F., Maretta, A. P., Kusuma, A. N., Sasmita, R. M., Rizkyllah, A. F., & Ibrahim, A. (2025). PERBANDINGAN METODE NAÏVE BAYES DAN K-NEAREST NEIGHBOR TERHADAP ANALISIS SENTIMEN ULASAN PROGRAM MAKAN SIANG GRATIS DI INDONESIA. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(4), 6385–6390.
- Firdaus, A. A., Hadiana, A. I., & Ningsih, A. K. (2024). Klasifikasi Sentimen pada Aplikasi Shopee Menggunakan Fitur Bag of Word dan Algoritma Random Forest. *Ranah Research: Journal of Multidisciplinary Research and Development*, 6(5), 1678–1683.
- Habibah, N., Aradytia, M. T., Izzah, K. N., Hawarie, K. A., & Maheswara, M. (2025). Reviewing the Effectiveness of the Free Nutritious Meals (MBG) Policy in Promoting Nutritional Equity Among School Children. *Journal of Social, Policy and Development Studies*, 2(2), 18–26.
- Hapsari, N. A., & Indriyanti, A. D. (2023). Analisis sentimen pada aplikasi dompet digital menggunakan algoritma Random Forest. *Journal of Emerging Information Systems and Business Intelligence*, 4(3), 186–192.
- Hermansyah, I., & Hasibuan, M. S. (2025). Sentiment Analysis of Free Nutritious Meal Programme on Social Media X using Linear Regression and Random Forest Algorithms. *PIKSEL: Penelitian Ilmu Komputer Sistem Embedded and Logic*, 13(1), 55–68.
- Idris, M., Rifai, A., & Tania, K. D. (2025). Sentiment analysis of Tokopedia app reviews using machine learning and word embeddings. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, 9(1), 210–219.
- Kiftiyah, A., Palestina, F. A., Abshar, F. U., & Rofiah, K. (2025). Program makan bergizi gratis (MBG) dalam perspektif keadilan sosial dan dinamika sosial-politik. *Pancasila: Jurnal Keindonesiaan*, 5(1), 101–112.
- KS, F. A. D., Jamaluddin, J., Sari, W. I. R., & Rachmaliana, L. (2025). Government Social Policy: Implementation of The Free Nutritious Meal Program (MBG). *FORUM EKONOMI: Jurnal Ekonomi, Manajemen Dan Akuntansi*, 27(Special Issue), 134–142.
- Mahing, N. F., Gunawan, A. L., Zen, A. F. A., Bachtiar, F. A., & Wicaksono, S. A. (2024). Klasifikasi Tingkat Stress dari Data Berbentuk Teks dengan Menggunakan Algoritma Support Vector Machine (SVM) dan Random Forest. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(5), 1067–1076.
- Mardiana, A., & Abidin, Z. (2025). The influence of user sentiment analysis on the free nutritious meal program using k-nearest neighbors and logistic regression methods. *Research Horizon*, 5(4), 1655–1668.
- Mawarni, A. C., Rusdah, R., Hin, L. L., & Anubhakti, D. (2023). Deteksi Dini Gejala Awal Penyakit Diabetes Menggunakan Algoritma Random Forest. *IDEALIS: InDonEsiA Journal Information System*, 6(2), 165–171.
- Napiah, M., Heristian, S., Raharjo, M., & Purnama, R. A. (2026). Analyzing Public Sentiment Toward Makanan Bergizi Gratis Program Using Machine Learning. *Computer Science (CO-SCIENCE)*, 6(1), 30–38.
- Nuru Aini, M. A., Agustin, I. T., & Toyibah, Z. B. (2024). Implementasi Algoritma Random Forest untuk Klasifikasi Bidang MSIB di Prodi Pendidikan Informatika. *Jurnal Informatika*, 11(1).
- Raditya, V. A., Saragih, T. H., Faisal, M. R., Abadi, F., & Muliadi, M. (2026). Analysis of Static and Contextual Word Embeddings in Capsule Network for Sentiment Analysis of The Free Nutritious Meal Program on Twitter. *Jurnal Teknik Informatika (Jutif)*, 7(1), 633–646.
- Rahmatullah, A., & Annisa, Q. (2025). Application Of TF-IDF And Word2vec For Feature Extraction In Sentiment Analysis Of Free Nutritious Food Policies. *Journal of Computer, Electronic, and Telecommunication (COMPLETE)*, 6(2).
- Ramadhan, N. F., & Alfat, L. (2026). APLIKASI PROFILING KEBUTUHAN PELAJARAN TAMBAHAN SISWA SMA MENGGUNAKAN ALGORITMA RANDOM FOREST. *Jurnal Sistem Informasi Dan Informatika*, 9(1), 46–54.
- Rois, M. I. A., Dwilestari, G., & Suarna, N. (2025). PREDIKSI PERSETUJUAN PINJAMAN MENGGUNAKAN DATASET LOAN APPROVAL MENGGUNAKAN ALGORITMA KLASIFIKASI. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(1), 1342–1347.
- Saleh, M. F., & Imanda, R. (2025). Public Sentiment Analysis of the Free Meal Program: A Comparison of Naive Bayes and Support Vector Machine Methods on the Twitter (X) Social Media Platform. *Journal of Applied Informatics and Computing*, 9(1),

- 131–139.
- Saputra, D. B., Atina, V., & Nastiti, F. E. (2024). Penerapan model CRISP-DM pada prediksi nasabah kredit menggunakan algoritma Random Forest. *IDEALIS: InDonEsiA Journal Information System*, 7(2), 240–247.
- Sulistyo, D. A., & Setiadi, E. (2025). Analisis Sentimen Kebijakan Makan Bergizi Gratis Menggunakan IndoBERT dan Machine Learning. *JURNAL FASILKOM*, 15(3), 507–513.
- Sultan, M., Darnila, E., & Afrillia, Y. (2025). PREDIKSI POTENSI WISATA MENGGUNAKAN ALGORITMA RANDOM FOREST DI KABUPATEN GAYO LUES: TOURISM POTENTIAL PREDICTION USING RANDOM FOREST ALGORITHM IN GAYO LUES DISTRICT. *RABIT: Jurnal Teknologi Dan Sistem Informasi Univrab*, 10(2), 741–751.
- Utama, D., Fahmi, R., & Pratama, M. R. K. (2025). Performance Comparison of 10 Machine Learning Algorithms in Sentiment Classification on Platform X Regarding the Government's Priority Program: Makanan Bergizi Gratis (MBG). *Teknika*, 14(3), 501–508.