BINER
Jurnal Ilmiah Informatika dan Komputer
Vol. 5 | No. 2 | Juli 2026

p-ISSN : 2828-0210 | e-ISSN : 2828-0229
https://ojs.unsiq.ac.id/index.php/biner

PERBANDINGAN RANDOM FOREST, NAIVE BAYES, DAN HEURISTIC POUR UNTUK ANALISIS SENTIMEN MBG

Ita Hardianti [1)], Feby Charlos [2)], Kanim [3)]
[1,2,3)] Program Studi Sistem Informasi, Universitas Pamulang, Kota Serang, Indonesia
Email: dosen03355@unpam.ac.id, dosen03043@unpam.ac.id, dosen03351@unpam.ac.id

Dikirim: 31 Mei 2026; Disetujui: 10 Juli 2026; Dipublikasikan: 31 Juli 2026

ABSTRAK

Program Makan Bergizi Gratis (MBG) merupakan kebijakan intervensi gizi nasional yang memicu beragam respons masyarakat di media sosial, khususnya YouTube. Penelitian ini bertujuan menganalisis sentimen publik terhadap program MBG, membandingkan kinerja Random Forest, Naive Bayes, dan Heuristic Pour, serta mengevaluasi pengaruh Synthetic Minority Oversampling Technique (SMOTE) dalam menangani ketidakseimbangan kelas. Sebanyak 2.278 komentar YouTube periode Agustus 2024–Januari 2025 diproses melalui tahapan *cleaning*, *tokenizing*, stopword removal, stemming, dan pembobotan TF-IDF. Evaluasi model dilakukan menggunakan 10-fold cross-validation berdasarkan *confusion matrix*, *Precision*, *Recall*, F1-score, dan Area Under Curve (AUC). Hasil penelitian menunjukkan bahwa Random Forest memberikan performa terbaik dengan akurasi 0,88 dan AUC 0,91. Penerapan SMOTE meningkatkan kemampuan model dalam mendeteksi sentimen positif sebagai kelas minoritas. Distribusi sentimen didominasi sentimen negatif lebih dari 87,5%. Penelitian ini memberikan kontribusi empiris terkait *trade-off* antara akurasi dan interpretabilitas dalam analisis sentimen berbahasa Indonesia.. Kebaruan penelitian ini terletak pada perbandingan komprehensif pendekatan *machine learning* dan rule-based dipadukan dengan SMOTE pada analisis sentimen komentar YouTube berbahasa Indonesia terkait kebijakan MBG. Penelitian ini memberikan bukti empiris mengenai efektivitas penanganan ketidakseimbangan kelas dan menjadi referensi dalam pengembangan analisis sentimen untuk evaluasi kebijakan publik

Kata Kunci : Analisis Sentimen, Random Forest, Naive Bayes, Heuristic Pour, SMOTE.

ABSTRACT

The Free Nutritious Meal (MBG) program is a national nutrition intervention policy that has generated diverse public responses on social media, particularly YouTube. This study analyzes public sentiment toward the MBG program, compares the performance of Random Forest, Naive Bayes, and Heuristic Pour, and evaluates the effect of the Synthetic Minority Oversampling Technique (SMOTE) on class imbalance. A total of 2,278 YouTube comments collected from August 2024 to January 2025 were preprocessed through cleaning, tokenization, stopword removal, stemming, and TF-IDF weighting. Model performance was evaluated using 10-fold cross-validation based on the confusion matrix, Precision, Recall, F1-score, and Area Under the Curve (AUC). Random Forest achieved the best performance, with an accuracy of 0.88 and an AUC of 0.91, while SMOTE significantly improved the detection of positive sentiment as the minority class. Negative sentiment dominated the dataset. The novelty of this study lies in the comprehensive comparison of machine learning and rule-based approaches integrated with SMOTE for Indonesian YouTube sentiment analysis. The findings provide empirical evidence on effective class imbalance handling and support the development of analysis sentiment models for public policy evaluation.

Keywords *: Sentiment Analysis, Random Forest, Naive Bayes, Heuristic Pour, SMOTE.*

1. PENDAHULUAN

Stunting adalah masalah gizi jangka panjang yang berdampak besar pada produktivitas, perkembangan kognitif, dan kualitas sumber daya manusia. Survei Status Gizi Indonesia oleh WHO dan UNICEF menunjukkan bahwa tingkat stunting di Indonesia masih akan mencapai 21,6% pada tahun 2022. Angka menunjukkan Indonesia masih menghadapi banyak tantangan untuk mencapai target penurunan prevalensi stunting menjadi 14% sebagaimana ditetapkan dalam strategi nasional percepatan penurunan stunting. Pencapaian target tersebut memerlukan sinergi berbagai sektor melalui beberapa kebijakan, tidak hanya berorientasi pada pelayanan kesehatan, tetapi juga pada peningkatan kualitas konsumsi pangan, edukasi gizi, ketahanan pangan keluarga, serta pemerataan akses terhadap makanan bergizi. Upaya tersebut menjadi bagian penting dalam mendukung terwujudnya visi Indonesia Emas 2045 dengan meningkatkan kualitas sumber daya manusia sejak usia dini. [1]. Pemerintah Indonesia menetapkan tujuan untuk mengurangi tingkat stunting hingga 14% pada tahun 2024 sebagai bagian dari agenda Indonesia Emas 2045, yang mensyaratkan laju penurunan rata-rata 3,8% per tahun [2]. Sebagai respons terhadap urgensi tersebut, berbagai program intervensi gizi berbasis kebijakan terus diluncurkan. Program ini dirancang untuk memperbaiki status gizi masyarakat sekaligus mendukung peningkatan kualitas pendidikan melalui penyediaan makanan bergizi secara rutin. Selain berfungsi sebagai intervensi kesehatan, Program MBG juga diharapkan mampu memberikan dampak sosial dan ekonomi melalui peningkatan kualitas hidup masyarakat serta penguatan ketahanan pangan nasional. Dengan cakupan penerima manfaat yang sangat luas dan implementasi yang dilakukan secara bertahap di berbagai wilayah Indonesia, Program MBG menjadi salah satu kebijakan publik yang memperoleh perhatian tinggi dari masyarakat maupun berbagai pemangku kepentingan[3].

Sebagai kebijakan publik berskala nasional yang diimplementasikan secara masif pada tahun 2024, Program MBG memicu respons yang beragam dari masyarakat yang termanifestasi secara luas di platform media sosial. YouTube, sebagai platform dengan penetrasi pengguna terbesar di Indonesia, menjadi ruang diskursus publik yang kaya akan data opini berbentuk komentar tidak terstruktur. Studi internasional membuktikan bahwa data media sosial terkait kebijakan pangan dan gizi memiliki potensi besar sebagai sumber informasi real time untuk evaluasi kebijakan berbasis data [4]. Namun, kompleksitas linguistik data tersebut yang mencakup bahasa informal, slang, sarkasme, dan code-switching mengharuskan pendekatan analitis yang sistematis dan tervalidasi secara metodologis.

Pendekatan mengekstraksi dari data opini berbentuk teks ialah Analisis sentimen, sebagai metode dalam Natural Language Processing (NLP), telah berkembang menjadi instrumen utama untuk mengklasifikasikan opini publik tertentu, seperti sentimen positif, negatif, maupun netral, berdasarkan pola linguistik yang terkandung dalam suatu dokumen teks [5]. Dalam praktiknya, dua paradigma pendekatan dominan digunakan pendekatan *machine learning* yang bersifat *data-driven* dan adaptif, serta pendekatan rule-based yang bersifat transparan dan berbasis aturan linguistik. Algoritma Random Forest dan Naive Bayes merupakan representasi tipikal pendekatan *machine learning* dalam analisis sentimen, dengan keunggulan pada kemampuan menangani data berdimensi tinggi dan efisiensi komputasi. Sementara itu, pendekatan rule-based seperti Heuristic Pour menawarkan transparansi proses klasifikasi yang penting dalam konteks pengambilan kebijakan publik [5].

Secara umum, metode analisis sentimen dapat dikelompokkan ke dalam dua paradigma utama, yaitu pendekatan *machine learning* dan pendekatan rule-based. Pendekatan *machine learning* memanfaatkan proses pembelajaran dari data latih untuk membangun model klasifikasi sehingga mampu mengenali pola yang kompleks pada data teks. Algoritma seperti dua metode yang paling umum digunakan adalah Random Forest dan Naive Bayes. Ini karena keduanya memiliki tingkat akurasi yang relatif tinggi, efisiensi komputasi yang baik, dan kemampuan untuk menangani data berukuran besar seperti representasi Frekuensi Term Inverse Dokumen (TF–IDF).

Random Forest bekerja menggunakan konsep *ensemble learning* melalui pembentukan sejumlah pohon keputusan yang digabungkan untuk membuat prediksi yang lebih konsisten dan tahan terhadap *overfitting*, sedangkan Naive Bayes memanfaatkan probabilitas bersyarat berdasarkan Teorema Bayes dengan asumsi independensi antarfitur sehingga memiliki proses komputasi yang sederhana namun efektif pada klasifikasi dokumen teks. [5] Penelitian lain dilakukan oleh Anggriyani dan Fakhriza yang membandingkan Random Forest dan Naive Bayes dalam analisis sentimen Program Makan Bergizi Gratis pada media sosial X. Penelitian tersebut menunjukkan bahwa pendekatan *machine learning* mampu menghasilkan tingkat akurasi yang relatif tinggi dalam mengklasifikasikan opini publik. Meskipun demikian, fokus penelitian masih terbatas pada perbandingan antaralgoritma *machine learning* sehingga belum memberikan gambaran mengenai bagaimana performa metode tersebut apabila dibandingkan dengan pendekatan rule-based yang memiliki karakteristik berbeda, khususnya dalam aspek transparansi proses klasifikasi *model explainability*. Dengan demikian, penelitian tersebut lebih menitikberatkan pada peningkatan akurasi tanpa mengevaluasi keseimbangan antara akurasi dan interpretabilitas model. [6] Pendekatan berbeda dikembangkan oleh Mengevaluasi metode pelabelan berbasis leksikon menggunakan InSet dan VADER dengan algoritma Support Vector Machine (SVM) pada platform X. Penelitian tersebut menunjukkan bahwa kualitas pelabelan awal memiliki pengaruh terhadap performa klasifikasi yang dihasilkan. Akan tetapi, penelitian tersebut lebih berfokus pada strategi pelabelan dibandingkan evaluasi karakteristik algoritma klasifikasi. Selain itu, penelitian tersebut belum menginvestigasi pengaruh distribusi data tidak seimbang terhadap kemampuan model dalam mengenali kelas minoritas sehingga evaluasi performa masih berorientasi pada nilai akurasi secara keseluruhan [6]. Di tingkat internasional, penelitian mengenai analisis sentimen komentar YouTube juga menunjukkan perkembangan yang cukup signifikan. memanfaatkan algoritma *machine learning* untuk menganalisis komentar YouTube mengenai produk pangan, pendekatan *Natural Language Processing* untuk mengeksplorasi opini masyarakat mengenai ketahanan pangan melalui media sosial. Hasil penelitian tersebut menunjukkan bahwa komentar YouTube memiliki potensi sebagai sumber data yang kaya karena mampu menggambarkan opini masyarakat secara lebih panjang dan argumentatif dibandingkan platform media sosial lainnya. Namun demikian, sebagian besar penelitian internasional belum secara khusus mempelajari opini publik tentang kebijakan nasional seperti Program Makan Bergizi Gratis karena fokusnya pada masalah pangan umum, ulasan produk, atau masalah kesehatan masyarakat di Indonesia. Selain itu, sebagian besar penelitian tersebut masih menggunakan pendekatan *machine learning* tanpa melakukan evaluasi terhadap metode berbasis aturan *rule-based* maupun strategi penanganan ketidakseimbangan data secara sistematis.

klasifikasi, menggunakan platform Twitter/X sebagai sumber data, serta belum mengevaluasi pengaruh ketidakseimbangan kelas secara sistematis. Selain itu, penelitian sebelumnya belum membandingkan pendekatan *machine learning* dengan metode rule-based sehingga *trade-off* antara akurasi dan interpretabilitas belum dapat dijelaskan secara komprehensif. Kondisi tersebut menunjukkan masih adanya ruang penelitian yang perlu dikaji lebih lanjut.

Ditinjau secara keseluruhan, penelitian-penelitian terdahulu memberikan kontribusi penting terhadap perkembangan analisis sentimen, namun masih menyisakan beberapa keterbatasan metodologis. Pertama, sebagian besar penelitian hanya mengevaluasi algoritma dalam paradigma *machine learning*, sehingga belum tersedia bukti empiris mengenai perbandingan langsung dengan pendekatan rule-based yang memiliki tingkat interpretabilitas lebih tinggi. Akibatnya, keseimbangan antara akurasi dan explainability sebagai dua aspek penting dalam pengembangan sistem kecerdasan buatan belum banyak dibahas secara komprehensifi [7].

Kedua, penelitian sebelumnya masih didominasi penggunaan data dari platform

Twitter/X, sedangkan komentar YouTube belum banyak dimanfaatkan sebagai sumber utama analisis sentimen Program Makan Bergizi Gratis. Padahal, karakteristik komentar YouTube yang lebih panjang, argumentatif, dan kontekstual memungkinkan informasi yang diperoleh menjadi lebih kaya dibandingkan media sosial berbasis *microblogging*. Perbedaan karakteristik tersebut berpotensi menghasilkan pola linguistik yang berbeda sehingga performa suatu algoritma tidak dapat langsung digeneralisasi dari satu platform ke platform lainnya.

Ketiga, sebagian besar penelitian terdahulu belum memberikan perhatian yang memadai terhadap permasalahan *class imbalance*, padahal distribusi data sentimen pada media sosial umumnya didominasi oleh satu kelas tertentu. Ketidakseimbangan tersebut menyebabkan model cenderung menghasilkan nilai akurasi yang tinggi tetapi memiliki kemampuan yang rendah dalam mengenali kelas minoritas. Oleh karena itu, evaluasi model seharusnya tidak hanya didasarkan pada nilai akurasi, tetapi juga mempertimbangkan *Precision*, *Recall*, *F1-score, dan Area Under Curve (AUC)* agar kemampuan model dalam mengklasifikasikan seluruh kelas dapat diukur secara lebih objektif. Penggunaan teknik Synthetic Minority Oversampling Technique (SMOTE) salah satu pendekatan yang banyak direkomendasikan untuk mengurangi bias terhadap kelas mayoritas, namun implementasinya pada penelitian analisis sentimen Program MBG masih relatif terbatas.

Berdasarkan telaah kritis terhadap penelitian-penelitian sebelumnya, dapat diidentifikasi tiga research gap yang menjadi dasar dilaksanakannya penelitian ini. Pertama, belum terdapat penelitian yang secara bersamaan membandingkan pendekatan *machine learning*, yaitu Random Forest dan Naive Bayes, dengan pendekatan rule-based, yaitu Heuristic Pour, pada analisis sentimen Program Makan Bergizi Gratis. Padahal, perbandingan lintas paradigma diperlukan untuk memperoleh pemahaman mengenai keunggulan dan keterbatasan masing-masing metode, khususnya dalam aspek akurasi dan interpretabilitas.

Kedua, belum banyak penelitian yang memanfaatkan komentar YouTube sebagai korpus utama dalam analisis sentimen Program Makan Bergizi Gratis, meskipun platform tersebut merupakan salah satu media dengan tingkat interaksi publik yang tinggi serta menyediakan komentar yang lebih panjang dan kaya konteks dibandingkan platform media sosial lainnya.

Ketiga, masih terbatas penelitian yang secara eksplisit mengevaluasi efektivitas SMOTE dalam meningkatkan kemampuan model mendeteksi kelas minoritas pada data sentimen Program MBG yang memiliki distribusi tidak seimbang. Sebagian besar penelitian terdahulu hanya melaporkan nilai akurasi tanpa mengkaji perubahan performa model setelah penerapan teknik penyeimbangan data.

Berdasarkan gap penelitian bertujuan untuk menganalisis persepsi masyarakat terhadap Program Makan Bergizi Gratis dengan menggunakan data komentar YouTube, membandingkan performa Random Forest, Naive Bayes, dan Heuristic Pour, serta mengevaluasi pengaruh penerapan SMOTE terhadap peningkatan kemampuan model dalam mengklasifikasikan kelas minoritas. Evaluasi dilakukan menggunakan 10-fold cross-validation dengan metrik accuracy, *Precision*, *Recall*, F1-score, *confusion matrix*, dan Area Under Curve (AUC) sehingga menghasilkan penilaian performa yang lebih komprehensif.

Kebaruan penelitian ini terletak pada integrasi tiga aspek yang belum banyak dikaji secara bersamaan dalam penelitian sebelumnya. Pertama, penelitian ini menyajikan perbandingan langsung antara pendekatan *machine learning* dan rule-based pada analisis sentimen kebijakan publik menggunakan dataset yang sama sehingga mampu menunjukkan *trade-off* antara akurasi dan interpretabilitas model. Kedua, penelitian ini memanfaatkan komentar YouTube sebagai sumber data utama untuk merepresentasikan opini masyarakat terhadap Program Makan Bergizi Gratis, yang masih relatif jarang digunakan dibandingkan Twitter/X. Ketiga, penelitian ini mengintegrasikan penerapan SMOTE dalam proses klasifikasi untuk mengevaluasi pengaruh penanganan

ketidakseimbangan data terhadap peningkatan kemampuan model dalam mendeteksi kelas minoritas [8].

2. METODE

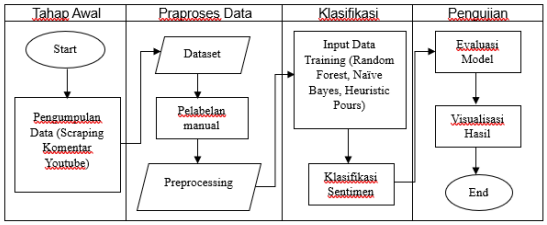
2.1. Desain Penelitian

Penelitian ini membandingkan kinerja tiga metode klasifikasi sentimen, yaitu Random Forest, Naive Bayes, dan Heuristic Pour, terhadap komentar YouTube yang berkaitan dengan Program Makan Bergizi Gratis (MBG). Penelitian disusun secara sistematis mulai dari pengumpulan pelabelan sentimen, *text preprocessing*, pembobotan fitur dengan menggunakan Frekwensi Waktu-Inverse Dokumen (TF–IDF), teknik penanganan ketidakseimbangan data menggunakan Teknik Penambahan Minoritas Sintetis (SMOTE), dan proses klasifikasi hingga evaluasi performa model.

Seluruh tahapan eksperimen dilakukan menggunakan perangkat lunak Python pada lingkungan Google Colaboratory. Proses pengolahan data memanfaatkan beberapa pustaka (*library*) seperti pandas untuk manipulasi data, Sastrawi untuk proses stemming bahasa Indonesia, scikit-learn untuk pembobotan TF–IDF, pembangunan model Random Forest dan Naive Bayes, serta evaluasi performa klasifikasi, sedangkan implementasi SMOTE menggunakan pustaka imbalanced-learn.

Data set dibagi menjadi sepuluh bagian dengan proporsi yang sama menggunakan metode 10-fold cross-validation. Sembilan bagian digunakan sebagai set pelatihan (training set) pada setiap iterasi, dan satu bagian digunakan sebagai set pengujian (testing set) pada setiap iterasi. Proses ini diulang sepuluh kali, sehingga setiap data memiliki kesempatan untuk digunakan, menjadi data uji tepat satu kali. Nilai performa akhir diperoleh dari rata-rata seluruh iterasi.

Alur penelitian secara keseluruhan ditampilkan pada Gambar 1.



Gambar 1. Alur Penelitian

2.2. Pengumpulan Data

Dataset yang digunakan terdiri dari 2.278 komentar YouTube yang diperoleh melalui proses web scraping menggunakan Google Colaboratory. Kriteria pengambilan data meliputi: komentar berasal dari periode Agustus 2024 hingga Januari 2025, Proses pengambilan data dilakukan berdasarkan beberapa kriteria, yaitu: (1) komentar menggunakan bahasa Indonesia, (2) komentar berkaitan dengan Program Makan Bergizi Gratis, (3) komentar bukan merupakan spam, iklan, maupun tautan eksternal, (4) komentar yang sama (*duplicate comments*) dihapus sehingga setiap komentar hanya muncul satu kali, dan (5) komentar yang kosong atau tidak memiliki makna setelah proses pembersihan data tidak disertakan dalam penelitian.

Setelah proses seleksi, diperoleh sebanyak 2.278 komentar, yang terdiri atas 1.994 komentar bersentimen negatif dan 284 komentar bersentimen positif. Distribusi tersebut menunjukkan ketidakseimbangan kelas (*class imbalance*) dengan rasio sekitar 7:1, sehingga diperlukan teknik penanganan data tidak seimbang sebelum proses klasifikasi

Tabel 1. Tampilan Dataset Hasil Crawling YouTube

Tanggal	Label	Text Display
05/02/2024	Negatif	Kong WOWO bikin mental orang jadi pengemis ancur ancuran
05/02/2024	Positif	Pak Ganjar figur yg berisi ilmu, pengetahuannya luas
05/02/2024	Negatif	kekenyangen, dia gak tahu stunting itu apa
05/02/2024	Positif	Capres cerdas dan berpengalaman
05/02/2024	Positif	GANJAR MAHFUD SAT SET 21

2.3. Pelabelan Data

Sebelum proses klasifikasi dilakukan, seluruh komentar diberikan label sentimen secara manual untuk membentuk *ground truth* yang digunakan sebagai acuan dalam proses pelatihan dan evaluasi model klasifikasi. Pelabelan dilakukan langsung oleh peneliti terhadap seluruh 2.278 komentar dengan mengacu pada makna keseluruhan kalimat serta konteks pembahasan mengenai Program Makan Bergizi Gratis.

Kategori sentimen dibedakan menjadi dua kelas, yaitu positif dan negatif. Komentar diberi label positif apabila mengandung dukungan, apresiasi, persetujuan, harapan, maupun penilaian yang menunjukkan persepsi positif terhadap Program MBG. Sebaliknya, komentar diberi label negatif apabila mengandung kritik, penolakan, ketidakpuasan, keraguan, sindiran, maupun penilaian yang menunjukkan persepsi negatif terhadap kebijakan tersebut.

Dalam proses pelabelan, penentuan kelas sentimen tidak hanya didasarkan pada keberadaan kata-kata yang bernilai positif atau negatif, tetapi juga mempertimbangkan konteks kalimat secara utuh. Pendekatan ini dilakukan karena komentar pada media sosial sering menggunakan bahasa informal, singkatan, ungkapan ironi, negasi, maupun kalimat yang maknanya bergantung pada konteks. Setiap komentar dibaca secara menyeluruh sebelum diberikan label sehingga makna yang diperoleh tidak hanya bergantung pada polaritas kata secara individual.

Untuk menjaga konsistensi pelabelan, peneliti terlebih dahulu menyusun pedoman pelabelan yang digunakan secara seragam selama proses anotasi berlangsung. Komentar yang memiliki makna ambigu atau mengandung lebih dari satu kecenderungan sentimen dianalisis kembali dengan mempertimbangkan tujuan utama komentar terhadap Program MBG. Apabila sebuah komentar mengandung unsur positif dan negatif secara bersamaan, label ditentukan berdasarkan sentimen yang paling dominan dalam keseluruhan isi komentar. Setelah seluruh proses pelabelan selesai, dilakukan pemeriksaan ulang (*review*) terhadap dataset untuk memastikan tidak terdapat inkonsistensi label maupun kesalahan anotasi sebelum data digunakan pada tahap *preprocessing* dan klasifikasi.

2.4. Penanganan Ketidakseimbangan Data

Distribusi data yang tidak seimbang menyebabkan bias pada model klasifikasi. Oleh karena itu, teknik oversampling kelas minoritas sintetis (SMOTE) digunakan untuk menghasilkan data sintetis untuk kelas minoritas. Metode ini menghasilkan sampel baru yang didasarkan pada kedekatan antar data dalam ruang fitur, yang meningkatkan kemampuan model untuk mengidentifikasi pola yang berkaitan dengan kelas minoritas. Tahapan pelabelan manual ini menghasilkan 1.994 komentar negatif dan 284 komentar positif, sehingga distribusi data menunjukkan kondisi tidak seimbang (imbalanced dataset). Kondisi tersebut menjadi salah satu pertimbangan utama dalam penelitian ini untuk menerapkan teknik Synthetic Minority Oversampling Technique (SMOTE) sebelum proses pembangunan model klasifikasi.

2.5. Metode Klasifikasi

a. Random Forest

Random Forest merupakan teknik berbasis ensemble yang menggunakan indeks Gini dan dapat menghasilkan klasifikasi yang lebih akurat melalui pembentukan atribut pada setiap node secara acak. Random Forest terdiri dari kumpulan pohon keputusan; data diklasifikasikan ke suatu kelas menggunakan kumpulan pohon keputusan tersebut [9]. Pemilihan fitur pada setiap titik dalam pohon keputusan didasarkan pada indeks Gini, yang dihitung dengan persamaan berikut:

$$\text{Gini}(Si) = 1 - \sum \rho i^2 \qquad (1)$$

dengan ρi merupakan frekuensi relatif kelas Ci di dalam himpunan, dan c adalah jumlah kelas. Sebagai contoh, dari data pada Tabel 1 dapat diperoleh distribusi label positif dan negatif sebelum pemisahan node akar sebagai berikut.

Data sebelum pemisahan node akar

Kelas	Jumlah
Positif	5
Negatif	6
Total Data	11

Proporsi Kelas

$\rho_{positif}= \frac{5}{11} = 0{,}455$

$\rho_{negatif}= \frac{6}{11} = 0{,}545$

Indeks Gini Sebelum

$Gini_{sebelum} = 1 - (\rho^2_{positif} + \rho^2_{negatif})$

$= 1 – (0{,}455^2 + 0{,}545^2)$

$= 1 – (0{,}207 + 0{,}297)$

$= 0{,}496$

Distribusi Data yang mengandung Kata Negatif Node 1 – mengandung kata negatif

$Gini_{Ya}= 1 – ((\frac{1}{6})^2 + (\frac{5}{6})^2)$

$= 1 – (0{,}028 + 0{,}694)$

$= 0, 278$

Perhitungan Node 2 tidak mengandung kata negatif (tidak)

Kelas	Jumlah
Positif	4
Negatif	1
Total	5

$Gini_{tidak}= 1 –(((\frac{4}{5}))^2 + (\frac{1}{5})^2)$

$= 1 – (0{,}64 + 0{,}04)$

$=0{,}32$

Perhitungan Indeks Gini Sesudah (Weighted Gini)

$Gini_{Sesudah} = (\frac{6}{11} \text{ x } 0{,}278) + (\frac{5}{11} \text{ x } 0{,}32)$

$= 0{,}152 + 0{,}145$

$= 0{,}297$

Keputusan pemilihan fitur

Tahap	Nilai Gini
Sebelum split	0,496
Sesudah split	0,297

Diperoleh nilai Gini sesudah lebih kecil fitur kemunculan kata bernada negatif sebagai atribut pemisah pada simpul internal pohon keputusan pada algoritma Random Forest. Node akar diperoleh nilai Indeks Gini sebesar 0,496. Setelah dilakukan pemisahan berdasarkan fitur kemunculan kata bernada negatif, diperoleh nilai Indeks Gini tertimbang sebesar 0,297.

b. Naive Bayes

Algoritma klasifikasi probabilistik Naive Bayes didasarkan pada Teorema Bayes, yang menganggap bahwa setiap fitur independen satu sama lain. Metode ini banyak digunakan dalam analisis sentimen karena memiliki keunggulan dalam kesederhanaan model, efisiensi komputasi, dan kemampuan menangani data teks. Sebagai berikut, teori Bayes dirumuskan sebagai berikut:

$$P(C|X) = \frac{P(X|C_i).P(X|C_i)}{P(X)} \quad (2)$$

Keterangan:

$P(C \mid X)$: probabilitas kelas terhadap data

$P(X \mid C)$: probabilitas data terhadap kelas

$P(C)$: probabilitas awal kelas

$P(X)$: probabilitas data

Naive Bayes menggunakan analisis sentimen untuk menentukan kemungkinan bahwa suatu dokumen akan termasuk dalam kelas positif atau negatif berdasarkan kemunculan kata-kata di dalamnya. Kelas dengan nilai kemungkinan tertinggi dipilih untuk diklasifikasikan. Tabel 1 menunjukkan jumlah data sentimen positif 5 dan negatif 6, jika data digunakan dengan cara yang sama.

Kelas	Jumlah
Positif	5
Negatif	6
Total Data	11

1. Menghitung Probabilitas Prior
   P(Positif)=5/11=0,455
   P(Negatif)=6/11=0,545
2. Menghitung Dokumen Uji dengan kalimat “makan gratis buruk” kata yang dianalisis:

Kata	Positif	Negatif
makan	3	2
gratis	4	3
buruk	1	5

3. Probabilitas Likelihood (Contoh Sederhana)

Kata	Positif	Negatif
makan	3	2
gratis	4	3

Total kata positif = 20, negatif = 25, gunakan *laplace smoothing*:

P(kata|kelas) = $\frac{f+1}{total+V}$

(V = jumlah kosakata, missal 10)

Kelas Positif

$$P(makan \mid positif) = (3+1)/(20+10) = 4/30 = 0{,}133$$
$$P(gratis \mid positif) = (4+1)/(20+10) = 5/30 = 0{,}167$$

Kelas Negatif

$$P(makan \mid negatif) = (2+1)/(25+10) = 3/35 = 0{,}086$$
$$P(gratis \mid negatif) = (3+1)/(25+10) = 4/35 = 0{,}114$$

Hitung Posterior

Positif

P(Positif|X) = 0,455×0,133×0,167

=0,455×0,022211=0,01010

4. Keputusan Klasifikasi

P(Negatif|X)>P(Positif|X), maka dokumen diklasifikasikan Negatif.perhitungan Naïve Bayes, probabilitas posterior kelas negatif lebih besar dibandingkan kelas positif. Hal ini menunjukkan bahwa kata "buruk" memiliki pengaruh signifikan dalam meningkatkan peluang dokumen termasuk ke dalam kelas negatif. Dengan demikian, metode Naïve Bayes mampu melakukan klasifikasi berdasarkan distribusi probabilitas kata dalam masing-masing kelas.

c. Heuristic Pour

Algoritma Heuristic Pour dikembangkan sebagai metode heuristik pada permasalahan flow shop scheduling untuk menentukan urutan pekerjaan (job sequencing) dan meminimalkan makespan. Metode ini menggunakan serangkaian aturan (heuristic rules) tanpa mengevaluasi seluruh kemungkinan solusi. Dalam penelitian ini, konsep Heuristic Pour diadaptasi ke dalam domain analisis sentimen. Objek evaluasi diubah dari urutan pekerjaan menjadi dokumen teks, sedangkan fungsi prioritas diubah menjadi penilaian polaritas berbasis aturan linguistik. Setiap komentar diperlakukan sebagai unit analisis dan memperoleh skor berdasarkan kombinasi indikator linguistik.

Berbeda dengan machine learning, metode ini tidak memerlukan proses pelatihan menggunakan data berlabel.

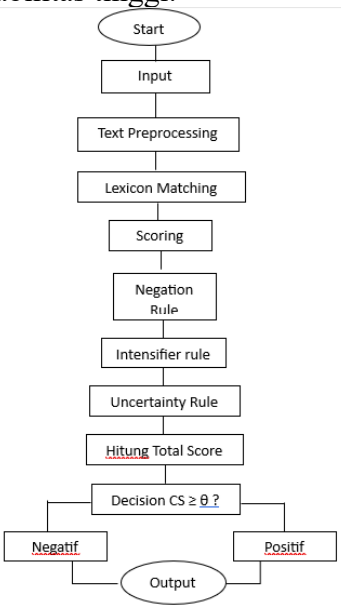
Klasifikasi dilakukan berdasarkan polaritas kata, negasi, kata penguat (intensifier), kata pelemah (downtoner), dan indikator ketidakpastian (uncertainty expressions).

Pendekatan ini memiliki tingkat interpretabilitas yang tinggi sehingga alasan klasifikasi sentimen positif atau negatif dapat dijelaskan secara eksplisit. Secara konseptual, proses adaptasi Heuristic Pour dalam penelitian ini terdiri atas lima tahapan utama, yaitu:

1. Identifikasi polaritas kata, yaitu menentukan nilai awal setiap kata berdasarkan kamus sentimen bahasa Indonesia.
2. Pemberian skor awal, yaitu memberikan bobot positif, negatif, atau netral terhadap setiap token hasil *preprocessing*.
3. Penerapan aturan heuristik, yaitu melakukan penyesuaian skor apabila ditemukan kata negasi, penguat, pelemah, ataupun ekspresi ketidakpastian yang memengaruhi polaritas kata.
4. Perhitungan skor sentimen dokumen, yaitu menjumlahkan seluruh skor kata setelah proses penyesuaian sehingga diperoleh skor akhir setiap komentar.

5. Penentuan kelas sentimen, yaitu mengklasifikasikan komentar ke dalam kelas positif atau negatif berdasarkan nilai skor akhir yang diperoleh.

Pemilihan Heuristic Pour didasarkan pada fleksibilitas konsep heuristiknya yang memungkinkan pengembangan aturan secara bertahap sesuai karakteristik bahasa Indonesia. Berbeda dengan metode leksikon konvensional yang umumnya hanya menjumlahkan skor polaritas kata, pendekatan yang diusulkan memberikan ruang untuk menerapkan aturan linguistik secara berurutan, seperti penanganan negasi, kata penguat, dan ekspresi ketidakpastian. Dengan demikian, metode ini tidak hanya menghasilkan klasifikasi sentimen, tetapi juga menyediakan proses pengambilan keputusan yang transparan dan mudah ditelusuri, sehingga sesuai untuk mendukung evaluasi kebijakan publik yang memerlukan tingkat interpretabilitas tinggi.



Gambar 2. Flowchart Heuristic Pour

Direpresentasikan dalam matriks skor Pij. Skor sentimen total dokumen dihitung sebagai:

$$CSi = \sum_{j=1}^{m} \text{Pij} \quad (3)$$

CSi = adalah skor sentimen total komentar ke-i;

Pij = adalah skor polaritas token ke-j setelah penerapan aturan heuristik;

$\sum_{j=1}^{m}$ = adalah jumlah token pada komentar ke-i.

Skor polaritas awal setiap token diperoleh melalui pencocokan dengan kamus sentimen bahasa Indonesia. Selanjutnya, skor tersebut disesuaikan berdasarkan keberadaan aturan linguistik berupa negasi, kata penguat (*intensifier*), kata pelemah (*downtoner*), dan ekspresi ketidakpastian (*uncertainty expression*). Seluruh skor token setelah penyesuaian dijumlahkan untuk memperoleh skor sentimen total setiap komentar. Penelitian ini menggunakan klasifikasi biner yang terdiri atas kelas positif dan negatif agar hasil metode *rule-based* dapat dibandingkan secara langsung dengan Random Forest dan Naive Bayes menggunakan *ground truth* yang sama. Nilai ambang klasifikasi ditetapkan sebesar $\theta = 0$. Komentar dengan skor sentimen total lebih besar dari nol diklasifikasikan sebagai sentimen positif, sedangkan komentar dengan skor kurang dari atau sama dengan nol diklasifikasikan sebagai sentimen negatif. Skor nol dimasukkan ke dalam kelas negatif karena penelitian tidak menggunakan kelas netral dan keputusan tersebut diterapkan secara konsisten pada seluruh data.

Rumus :

$$y^i = \begin{cases} Positif, CS_i > 0 \\ Negatif, CS_i < 0 \end{cases}$$

y^i = kelas sentimen hasil prediksi komentar ke-i;

CS_i= skor sentimen total komentar ke-i;

$\theta = 0$= nilai ambang klasifikasi.

Dok 1: " Kong WOWO bikin mental orang jadi pengemis hancur "

Penentuan mengandung Polaritas Kata *(Lexicon)*.

Kata	Polaritas	Nilai
Pengemis	Negatif	-1
Ancur	Negatif	-1

Tahap Sebelum (Skor Sentimen Awal / Raw Score) hanya penentuan polaritas kata.

Misalkan perhitungan

$CS_i = P_{(pengemis)} + P_{(hancur)}$

$= (-1) + (-1)$

$= -2$

Dokumen mengandung sentimen negatif kuat namun belum final.

Skor Sesudah

Interpretasi Skor Sentimen

nilai ambang: θ=0

Aturan klasifikasi:

Positif → $CS_i > 0$

Netral → $-0 \leq CS_i \leq 0$

Negatif → $CS_i < 0$

Karena :

CS1=−2<0, maka sentimen dokumen = Negatif

Dalam aturan heuristik akan berubah jika suatu dokumen terdapat negasi (tidak), intensifier dan ketidakpastian.

Contoh : ‘mental orang tidak hancur’

Kata	Polaritas	Skor
Hancur	Negatif	-1

$CS_{sebelum}$ = -1 (dokumen terindikasi negatif)

Perhitungan Sesudah Heuristic Pour

$CS_{sesudah}$=$CS_{sebelum}$ x (−1)

$CS_{sesudah}$= (-1) x (-1) = 1 (terindikasi positif).

Dokumen yang mengandung kata negasi mengalami perubahan skor sentimen setelah penerapan aturan Heuristic Pourada frasa “mental orang tidak hancur”, kata “hancur” memiliki polaritas awal negatif dengan skor −1. Kata negasi “tidak” yang muncul sebelum kata “hancur” membalik nilai polaritas melalui perkalian dengan faktor −1. Dengan demikian, skor kata berubah dari −1 menjadi +1 sehingga frasa tersebut memiliki kecenderungan sentimen positif.

2.6 Evaluasi Performa Model

Performa Random Forest, Naive Bayes, dan adaptasi Heuristic Pour dievaluasi dengan menggunakan *matrix confusion*, akurasi, ketepatan, pengecekan, skor F1, dan *Area Under the Curve (AUC)*, karena distribusi kelas dataset tidak seimbang, penggunaan beberapa metrik diperlukan. Oleh karena itu, satu-satunya nilai akurasi belum cukup untuk menunjukkan kemampuan model untuk mengidentifikasi kelas minoritas. Akurasi menunjukkan proporsi seluruh prediksi yang diklasifikasikan dengan benar. *Precision* menunjukkan tingkat *recall* menunjukkan kemampuan model untuk menemukan seluruh data yang sebenarnya termasuk dalam kelas, sedangkan prediksi tersebut berlaku untuk kelas tertentu. F1-nilai memberikan evaluasi yang lebih seimbang karena merupakan rata-rata harmonik antara Precision dan Recall. AUC digunakan untuk mengukur kemampuan model dalam membedakan kelas positif dan negatif pada berbagai nilai ambang klasifikasi.

3. HASIL DAN PEMBAHASAN

Studi ini melihat bagaimana tiga teknik klasifikasi sentimen—Random Forest, Naive Bayes, dan adaptasi Heuristic Pour—bekerja dengan 2.278 komentar YouTube tentang Program Makan Bergizi Gratis (MBG). 1.994 komentar negatif dan 284 komentar positif ditemukan dalam dataset, dengan rasio ketidakseimbangan sekitar 7:1. Sentimen negatif mendominasi distribusi sebesar 87,53%, sedangkan sentimen positif hanya 12,47%. Ketidakseimbangan kelas menjadi aspek penting dalam interpretasi hasil karena nilai akurasi yang tinggi belum tentu menunjukkan bahwa model yang baik dapat mengidentifikasi perasaan positif bagi kelompok minoritas. Untuk menilai kinerja,

matriks confusion, akurasi, ketepatan, pengecualian, skor F1, dan Area Under the Curve atau disingkat (AUC) digunakan. Untuk mendapatkan gambaran yang lebih luas tentang kinerja, berbagai metrik digunakan, terutama karena distribusi kelas yang tidak seimbang dapat menyebabkan model berkonsentrasi pada kelas mayoritas. Akibatnya, evaluasi tidak hanya mempertimbangkan akurasi keseluruhan, tetapi juga mempertimbangkan skor Recall dan F1 untuk setiap kelas.

3.1. Hasil Evaluasi Metode Random Forest

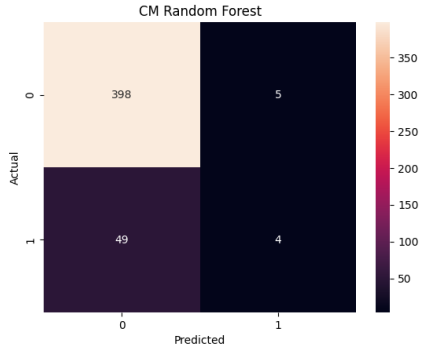
Hasil klasifikasi menggunakan model Random Forest ditampilkan pada Tabel 2.

Tabel 2. Hasil Akurasi, Presisi, *Recall*, dan F1-Score Metode Random Forest

Kelas	*Precision*	*Recall*	F1-Score
Negative	0,89	0,99	0,93
Positif	0,50	0,07	0,13
Accuracy	0,88	0,88	0,88

Random Forest mampu mencapai akurasi sebesar 0,88 dengan performa yang sangat tinggi pada kelas negatif (*Recall* = 0,99) dan nilai F1-score sebesar 0,93. Hasil menunjukkan kemampuan Random Forest untuk mengidentifikasi komentar negatif. Namun, performa pada kelas positif masih relatif rendah dengan *Precision* sebesar 0,50, *Recall* sebesar 0,07, dan F1-score sebesar 0,13.

Hasil *confusion matrix* Random Forest dapat dilihat dati gambar 3, gambar 3 menunjukkan bahwa sebanyak 398 komentar negatif berhasil diklasifikasikan dengan benar sebagai sentimen negatif (*true negative*), sedangkan lima komentar negatif salah diklasifikasikan sebagai sentimen positif (*false positive*). Pada kelas positif, hanya empat komentar yang berhasil diklasifikasikan dengan benar (*true positive*), sedangkan 49 komentar positif salah diprediksi sebagai sentimen negatif (*false negative*).



Gambar 3. *Confusion matrix* Random Forest

Karena sebagian besar data Random Forest berasal dari kelas negatif, tingkat akurasi yang tinggi harus diinterpretasikan dengan hati-hati. Dengan nilai recall kelas negatif 0,99, model menunjukkan kemampuan yang luar biasa untuk mempelajari pola dominan pada kelas mayoritas. Di sisi lain, dengan nilai recall kelas positif 0,07, sebagian besar komentar positif masih belum dapat dikenali oleh model. Akibatnya, akurasi sebesar 0,88 lebih dipengaruhi oleh keberhasilan model dalam mengklasifikasikan kelas mayoritas daripada hasil yang seimbang untuk kedua kelas.

Random Forest menggunakan variasi sampel dan karakteristik untuk membuat sejumlah pohon keputusan, dan kemudian menggunakan mekanisme suara mayoritas untuk menentukan hasil klasifikasi. Dengan fitur ini, model dapat mempelajari hubungan nonlinier antar fitur dan membuat prediksi yang lebih stabil daripada dengan menggunakan satu pohon keputusan. Namun, sebagian besar pohon memperoleh pola yang lebih tinggi dari kelas mayoritas pada dataset yang tidak seimbang, yang meningkatkan kemungkinan sentimen negatif dalam keputusan akhir.

Pola yang dipelajari oleh model juga dapat dipengaruhi oleh representasi TF-IDF. Untuk membuatnya lebih mudah dikenali, kata-kata negatif muncul secara eksplisit dan berulang memperoleh bobot yang relatif konsisten. Sebaliknya, komentar positif dapat menggunakan ekspresi yang lebih beragam, implisit, atau bergantung pada konteks. Di bawah keadaan ini, pola kelas positif memiliki representasi yang lebih sedikit dibandingkan dengan pola kelas negatif.

Hasil studi ini sejalan dengan penelitian Hidayat dan Ratnaningsih [8], yang

menunjukkan bahwa Random Forest mampu mengklasifikasika sentimen Program MBG dengan baik. Namun, penelitian ini menemukan bahwa kemampuan deteksi kelas minoritas yang baik tidak selalu diikuti oleh akurasi yang tinggi. Oleh karena itu, evaluasi model pada data tidak seimbang harus mempertimbangkan kinerja per kelas, bukan hanya akurasi sebagai indikator utama.

3.2. Hasil Evaluasi Metode Naive Bayes

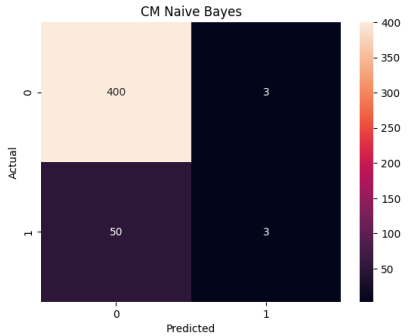
Hasil pengujian menggunakan pendekatan Naive Bayes ditampilkan pada Tabel 3.

Tabel 3. Hasil Akurasi, Presisi, *Recall*, dan F1-Score metode Naive Bayes

Kelas	*Precision*	*Recall*	F1-Score
Negative	0,88	0,99	0,93
Positif	0,50	0,03	0,07
Accuracy	0,88	0,88	0,88

Hasil pengujian Naive Bayes menunjukkan performa akurasi 0,88 atau sama dengan Random Forest. Pada kelas negatif, model memperoleh *Precision* sebesar 0,88, *Recall* sebesar 0,99, dan F1-score sebesar 0,93. Sebaliknya, pada kelas positif, *Precision* yang diperoleh sebesar 0,50, *Recall* sebesar 0,03, dan F1-score sebesar 0,07.

Hasil *confusion matrix* Naive Bayes ditampilkan pada Gambar 4.



Gambar 4. *Confusion matrix* Naive Bayes

Confusion matrix pada model Naive Bayes menunjukkan 400 data negatif berhasil diprediksi dengan benar (True Negative) dan hanya 3 kesalahan (False Positive), yang mencerminkan kemampuan model menangkap pola dominan pada kelas mayoritas. Namun, hanya tiga data dalam kelas positif yang diklasifikasikan dengan benar *(True Positive*), dan lima puluh data kelas positif yang salah diprediksi sebagai negatif *(False Negative)*. Hasil menunjukkan bahwa Naive Bayes lebih baik memprediksi kelas mayoritas daripada Random Forest.

Meskipun kedua model memiliki nilai akurasi yang sama, mereka tidak bekerja sama sepenuhnya. Random Forest mendapatkan skor Recall dan F1 kelas yang lebih tinggi dibandingkan Naive Bayes, yang menunjukkan bahwa evaluasi berdasarkan akurasi saja dapat menyembunyikan perbedaan kemampuan model untuk mengidentifikasi kelas minoritas.

Naive Bayes percaya bahwa jika class sudah diketahui, setiap kata akan muncul secara independen dari yang lain. Akibatnya, dia memiliki tingkat recall yang rendah untuk class yang positif. Asumsi tersebut menyederhanakan proses klasifikasi dan membuat model efisien pada data berdimensi tinggi. Namun, hubungan antarkata, urutan kata, negasi, dan konteks kalimat tidak dapat direpresentasikan secara optimal. Pada komentar media sosial, perubahan makna sering ditentukan oleh kombinasi beberapa kata. Sebagai contoh, frasa “tidak buruk” memiliki makna yang berbeda dengan kata “buruk” apabila dianalisis secara terpisah.

Naive Bayes tetap menunjukkan performa yang sangat baik pada kelas negatif karena jumlah data negatif yang lebih besar menghasilkan estimasi probabilitas kata yang lebih stabil. Kata-kata yang menunjukkan kritik, penolakan, atau ketidakpuasan juga cenderung muncul secara lebih eksplisit sehingga mudah dipelajari melalui distribusi frekuensi. Sebaliknya, keterbatasan jumlah komentar positif menyebabkan estimasi probabilitas pada kelas tersebut menjadi kurang representatif.

Temuan ini mendukung penelitian Sitanggang dkk. [10] [11] dan Saputra dan Hasan [12], yang menunjukkan bahwa Naive Bayes efektif digunakan pada analisis sentimen Program Makan Siang Gratis. Namun, penelitian ini menunjukkan bahwa efektivitas tersebut perlu ditinjau berdasarkan distribusi performa setiap kelas. Akurasi yang tinggi belum cukup untuk menyimpulkan bahwa model memiliki kemampuan klasifikasi yang seimbang apabila *Recall* kelas minoritas masih sangat rendah.

3.3. Hasil Evaluasi Metode Heuristic Pour

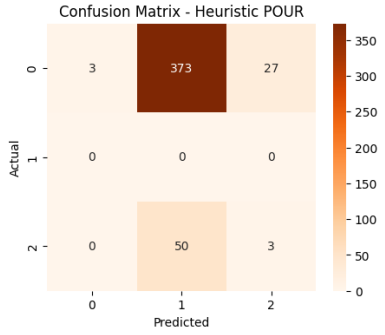
Hasil pengujian menggunakan metode Heuristic Pour ditampilkan pada Tabel 4.

Tabel 4. Hasil Akurasi, Presisi, *Recall*, dan F1-Score Metode Heuristic Pour

Index	*Precision*	*Recall*	F1-Score
Negative	1,00	0,009	0,019
Positif	0,125	0,057	0,078
Accuracy	0,015	0,015	0,015

Adaptasi Heuristic Pour menghasilkan akurasi sebesar 0,015 dan menunjukkan performa yang jauh lebih rendah dibandingkan Random Forest dan Naive Bayes. Pada kelas negatif, *Precision* mencapai 1,00, dan *Recall* hanya sebesar 0,009 dan F1-score sebesar 0,019. Nilai *Precision* yang tinggi tidak menunjukkan performa keseluruhan yang baik karena jumlah data yang berhasil diprediksi sebagai negatif sangat terbatas. Pada kelas positif, model menghasilkan *Precision* sebesar 0,125, *Recall* sebesar 0,057, dan F1-score sebesar 0,078.

Namun demikian, metode Heuristic Pour tetap memiliki nilai penting dalam konteks interpretabilitas model *(explainability)*. Berbeda dengan model *machine learning* yang bersifat black-box, pendekatan ini memungkinkan proses klasifikasi dijelaskan secara transparan melalui aturan yang digunakan. Oleh karena itu, meskipun performanya rendah, metode ini dapat berperan sebagai baseline interpretatif dalam pengembangan model hybrid di masa mendatang. Hasil *confusion matrix* Heuristic Pour ditampilkan pada Gambar 5.



Gambar 5. *Confusion matrix* Metode Heuristic Pour

Dari gambar 5 *confusion matrix* hanya empat komentar negatif dan tiga komentar positif yang berhasil diklasifikasikan dengan benar. Sebagian besar komentar tidak dapat dipetakan secara tepat oleh aturan linguistik yang digunakan. Hasil tersebut menunjukkan bahwa adaptasi Heuristic Pour dalam konfigurasi saat ini belum mampu menangkap keragaman pola bahasa pada komentar YouTube mengenai Program MBG.

Rendahnya performa Heuristic Pour dapat dipengaruhi oleh beberapa faktor. Pertama, metode berbasis aturan sangat bergantung pada cakupan dan kualitas kamus sentimen. Kata tidak baku, singkatan, kesalahan penulisan, istilah politik, bahasa daerah, serta kosakata baru yang tidak terdapat dalam kamus tidak memperoleh skor polaritas yang sesuai. Akibatnya, sebagian komentar menghasilkan skor yang tidak merepresentasikan makna sebenarnya. Kedua, aturan negasi, penguat, pelemah, dan ketidakpastian belum sepenuhnya mampu menangani struktur bahasa yang kompleks. Meskipun aturan negasi dapat membalik polaritas kata tertentu, makna komentar tidak selalu dapat ditentukan melalui perubahan skor secara langsung. Posisi kata, jarak antara negasi dan kata sentimen, serta konteks keseluruhan kalimat dapat menghasilkan interpretasi yang berbeda.

Ketiga, komentar media sosial sering mengandung ironi, sarkasme, metafora, dan makna implisit. Ungkapan tersebut sulit diklasifikasikan menggunakan aturan statis karena polaritas tekstual dapat berbeda dengan maksud sebenarnya. Sebagai contoh, komentar yang menggunakan kata positif dalam bentuk sindiran dapat memperoleh skor positif meskipun makna keseluruhannya bersifat negatif.

Keempat, berbeda dengan Random Forest dan Naive Bayes, adaptasi Heuristic Pour tidak melakukan proses pembelajaran berdasarkan pola empiris dalam dataset. Metode ini menerapkan aturan yang telah ditentukan sebelumnya sehingga tidak dapat memperbarui pola klasifikasi secara otomatis ketika menemukan variasi bahasa baru. Kondisi tersebut menjelaskan mengapa metode *machine learning* memiliki performa yang lebih tinggi pada data komentar YouTube yang bersifat dinamis dan heterogen.

Meskipun performanya rendah, adaptasi Heuristic Pour memiliki keunggulan pada aspek interpretabilitas. Setiap hasil klasifikasi dapat ditelusuri melalui kata yang teridentifikasi, skor polaritas awal, aturan linguistik yang diterapkan, dan skor akhir dokumen. Transparansi tersebut berbeda dengan Random Forest yang memiliki struktur keputusan lebih kompleks. Oleh karena itu, Heuristic Pour dalam penelitian ini lebih tepat diposisikan sebagai baseline rule-based yang interpretatif, bukan sebagai pengganti langsung model *machine learning*.

3.4. Dampak Ketidakseimbangan Data dan Peran SMOTE

Distribusi data yang tidak seimbang (1.994 negatif dan 284 positif) menjadi faktor utama yang memengaruhi performa seluruh model. Ketimpangan ini mengakibatkan kecenderungan model untuk mengoptimalkan prediksi untuk kelas mayoritas, mengabaikan kelas minoritas. Penerapan *SMOTE* dalam penelitian ini terbukti mampu meningkatkan *Recall* pada kelas positif dari 0,16 menjadi 0,62 pada skenario tertentu. Hal ini menunjukkan bahwa penambahan data sintetis dapat membantu model mengenali pola kelas minoritas dengan lebih baik.

Temuan tersebut sejalan dengan Obiedat dkk [7] , yang menjelaskan bahwa distribusi kelas tidak seimbang dapat menyebabkan nilai akurasi global memberikan gambaran performa yang kurang representatif. Hasil penelitian ini memperkuat temuan tersebut pada konteks komentar YouTube berbahasa Indonesia mengenai kebijakan publik. Dengan demikian, kontribusi penerapan SMOTE tidak hanya terletak pada peningkatan jumlah data kelas minoritas, tetapi juga pada peningkatan sensitivitas model terhadap pola sentimen yang sebelumnya kurang terwakili.

Namun, peningkatan tersebut belum sepenuhnya optimal. Hal ini kemungkinan disebabkan oleh distribusi fitur yang masih overlap antar kelas, representasi TF-IDF yang belum mampu menangkap konteks semantik, serta variasi bahasa pada data yang tinggi. Dengan demikian, meskipun SMOTE efektif mengatasi ketidakseimbangan data secara kuantitatif, kualitas representasi fitur tetap menjadi faktor penentu utama dalam meningkatkan performa klasifikasi.

3.5. Visualisasi Data

Visualisasi kata yang paling sering muncul dalam komentar publik digambarkan dengan visualisasi cloud kata terdapat pada Gambar 6.

Gambar 6. Word Cloud

Word cloud menunjukkan bahwa pembahasan masyarakat paling banyak berfokus pada program makan bergizi gratis, dengan kata dominan seperti "gratis", "siang", "makan", "gibran", "anak", dan "program". Kemunculan kata "mahfud", "rakyat", dan "ganjar" menandakan adanya respons publik yang aktif berupa dukungan, pertanyaan, maupun keraguan, sedangkan kata "stunting", "gizi", dan "sekolah" menunjukkan keterkaitan program dengan isu kesehatan dan pendidikan anak.

3.6. Perbandingan dan Implikasi Metode

Studi ini membandingkan tiga metode untuk analisis sentimen random forest, naive bayes dan heuristic pour :

Tabel 5. Perbandingan Kinerja Metode Random Forest, Naive Bayes, dan Heuristic Pour

Metode	Accuracy	*Precision*	*Recall*	F1-Score
Random Forest	0,880	0,850	0,880	0,840
Naive Bayes	0,880	0,840	0,880	0,830
Heuristic Pour	0,015	0,890	0,015	0,020

Metode berbasis *machine learning* (Random Forest dan Naive Bayes) dan metode berbasis aturan (Heuristic Pour). Random Forest dan Naive Bayes menunjukkan kinerja yang cukup baik dengan nilai akurasi yang sama (0,88), tetapi kesamaan akurasi ini tidak segera menunjukkan kesetaraan kinerja secara keseluruhan. Nilai *Recall* yang sangat tinggi pada kelas negatif (mendekati 1) menunjukkan

bahwa model mengenali pola dominan dengan baik namun, tidak dapat menggeneralisasi pola pada kelas positif, yang menyebabkan mayoritas kelas bias. Dengan menggunakan metode kelompok, keunggulan Random Forest yakni mengidentifikasi hubungan *non-linear* antar elemen, sehingga lebih adaptif terhadap variasi pola dalam data. Sementara itu, Naive Bayes tetap memiliki keunggulan berupa struktur model yang sederhana, waktu komputasi yang efisien, serta kemampuan menangani data berdimensi tinggi. Namun, asumsi independensi antarf fitur membatasi kemampuan model dalam memahami hubungan kontekstual antarkata. Oleh karena itu, metode ini sesuai digunakan sebagai model dasar yang efisien, tetapi memerlukan pengembangan representasi fitur untuk meningkatkan kemampuan klasifikasi pada pola bahasa yang kompleks. Adaptasi Heuristic Pour menghasilkan performa terendah, tetapi menyediakan proses klasifikasi yang lebih transparan. Hasil tersebut memperlihatkan adanya *trade-off* antara performa prediktif dan interpretabilitas. Model *machine learning* mampu mempelajari pola data secara lebih adaptif, sedangkan metode rule-based memungkinkan setiap keputusan ditelusuri berdasarkan aturan yang digunakan. Namun, interpretabilitas yang tinggi tidak secara otomatis menghasilkan akurasi yang tinggi apabila aturan dan sumber daya linguistik belum mampu mencakup variasi bahasa dalam dataset.

Perbandingan ketiga metode menunjukkan adanya *trade-off* antara akurasi dan interpretabilitas. Model *machine learning* memberikan performa klasifikasi yang tinggi namun dengan interpretasi terbatas *(black-box),* sedangkan Heuristic Pour menawarkan transparansi tinggi namun performa rendah. Temuan ini mengindikasikan tidak ada satu metode yang sepenuhnya unggul, sehingga penelitian ini mengusulkan arah pengembangan melalui pendekatan hybrid yang mengintegrasikan kekuatan *machine learning* dalam menangkap pola kompleks dengan keunggulan metode heuristik dalam memberikan interpretasi yang transparan.

Penelitian ini memiliki beberapa keterbatasan. Pertama, data hanya berasal dari YouTube sehingga hasil penelitian belum dapat digeneralisasi terhadap seluruh pengguna media sosial di Indonesia. Kedua, proses pelabelan dilakukan oleh satu peneliti sehingga konsistensi label belum dievaluasi menggunakan kesepakatan antarpelabel. Ketiga, representasi TF-IDF belum mampu menangkap konteks, urutan kata, ironi, dan sarkasme secara mendalam. Keempat, aturan pada adaptasi Heuristic Pour masih bergantung pada cakupan kamus sentimen dan aturan linguistik yang ditentukan. Keterbatasan tersebut dapat menjadi dasar pengembangan penelitian berikutnya melalui penggunaan data lintas platform, validasi oleh lebih dari satu anotator, representasi berbasis *word embedding* atau model transformer, serta pengembangan metode hibrida yang menggabungkan kemampuan prediktif *machine learning* dengan transparansi pendekatan rule-based.

4. PENUTUP

4.1. Kesimpulan

Penelitian ini menunjukkan bahwa pendekatan *machine learning* memiliki kemampuan klasifikasi yang lebih baik dibandingkan adaptasi metode *rule-based* dalam menganalisis sentimen masyarakat terhadap Program Makan Bergizi Gratis berdasarkan komentar YouTube. Random Forest dan Naive Bayes mampu mengenali pola sentimen negatif sebagai kelas mayoritas dengan baik. Meskipun kedua metode menghasilkan tingkat akurasi yang sama, Random Forest menunjukkan performa keseluruhan yang relatif lebih baik berdasarkan nilai F1-score dan kemampuan diskriminasi model. Namun, rendahnya kemampuan kedua model dalam mengenali sentimen positif menunjukkan bahwa nilai akurasi tidak dapat digunakan sebagai satu-satunya indikator performa pada dataset dengan distribusi kelas yang tidak seimbang.

Penerapan *Synthetic Minority Oversampling Technique (SMOTE)* memberikan pengaruh positif terhadap kemampuan model dalam mengenali sentimen positif sebagai kelas minoritas. Temuan tersebut menunjukkan bahwa penanganan ketidakseimbangan kelas merupakan tahapan penting dalam pengembangan model analisis sentimen karena dapat mengurangi kecenderungan model untuk memprioritaskan kelas mayoritas. Meskipun

demikian, peningkatan jumlah data sintetis belum sepenuhnya mengatasi keterbatasan representasi *TF-IDF* dalam memahami konteks, hubungan antarkata, ironi, dan sarkasme pada komentar media sosial.

Adaptasi Heuristic Pour sebagai pendekatan *rule-based* menghasilkan performa klasifikasi yang lebih rendah dibandingkan Random Forest dan Naive Bayes. Keterbatasan tersebut dipengaruhi oleh ketergantungan metode terhadap cakupan kamus sentimen dan aturan linguistik yang belum mampu merepresentasikan keragaman bahasa informal pada komentar YouTube. Namun, pendekatan tersebut memberikan tingkat interpretabilitas yang lebih tinggi karena setiap keputusan klasifikasi dapat ditelusuri melalui skor polaritas dan aturan yang diterapkan. Dengan demikian, adaptasi Heuristic Pour memiliki potensi untuk dikembangkan sebagai komponen interpretatif dalam pendekatan hibrida, bukan sebagai pengganti langsung model *machine learning*. pada evaluasi komparatif antara pendekatan *machine learning* dan adaptasi pendekatan *rule-based* pada data komentar YouTube berbahasa Indonesia, serta pengujian pengaruh SMOTE terhadap kemampuan deteksi kelas minoritas. Hasil penelitian memberikan bukti empiris mengenai adanya *trade-off* antara performa prediktif dan interpretabilitas dalam analisis sentimen kebijakan publik. Secara praktis, model yang dikembangkan dapat menjadi referensi awal dalam pemanfaatan opini masyarakat di media sosial sebagai informasi pendukung untuk mengevaluasi komunikasi dan implementasi Program Makan Bergizi Gratis.

4.2. Keterbatasan dan Saran

Penelitian ini memiliki beberapa keterbatasan. Dataset hanya berasal dari YouTube sehingga hasil analisis belum dapat merepresentasikan opini masyarakat pada seluruh platform media sosial. Pelabelan data dilakukan oleh satu peneliti sehingga tingkat kesepakatan antarpelabel belum dapat diukur. Selain itu, penggunaan TF-IDF masih terbatas dalam merepresentasikan konteks semantik, urutan kata, ironi, dan sarkasme. Adaptasi Heuristic Pour juga masih menggunakan cakupan kamus sentimen dan aturan linguistik yang terbatas sehingga belum mampu menangani seluruh variasi bahasa informal pada komentar media sosial.

Penelitian selanjutnya disarankan menggunakan data dalam jumlah besar dan berasal dari berbagai *platform*, seperti YouTube, Instagram, TikTok, dan X, agar diperoleh representasi opini publik yang lebih luas. Proses anotasi dapat melibatkan lebih dari satu pelabel dan dievaluasi menggunakan metode kesepakatan antar label untuk meningkatkan reliabilitas *ground truth*. Pengembangan model dapat memanfaatkan representasi teks kontekstual dan metode *deep learning*, seperti *Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM)*, maupun model berbasis transformer seperti *IndoBERT.*

Pengembangan adaptasi Heuristic Pour dapat dilakukan melalui perluasan sentimen, normalisasi bahasa informal, pengaturan cakupan negasi, optimasi bobot aturan, serta penambahan mekanisme untuk menangani ironi dan sarkasme. Integrasi aturan Heuristic Pour dengan model *machine learning* juga dapat dikembangkan sebagai pendekatan hibrida untuk memperoleh keseimbangan yang lebih baik antara akurasi klasifikasi dan interpretabilitas hasil.

5. DAFTAR PUSTAKA

[1] S. A. Sb, R. Hipni, and E. Kristiana, “HUBUNGAN PENGETAHUAN DAN PARITAS DENGAN KEJADIAN STUNTING PADA BALITA UMUR 1-3 TAHUN DI WILAYAH KERJA PUSKESMAS BATULICIN TAHUN 2024,” vol. 2, no. 1, pp. 265–277, 2025.

[2] S. Amara, M. Khadapi, S. Informasi, and S. Kaputama, “Analisis Sentimen Masyarakat terhadap Program Makan Siang Gratis di Indonesia Tahun 2024 Menggunakan Long Short-Term Memory (LSTM) kesulitan dalam pemenuhan kebutuhan primer , salah satunya ialah pangan . Produksi pangan Untuk dapat menghadapi masalah i,” *Merkurius J. Ris. Sist. Inf. dan Tek. Inform.*, vol. 3, no. 4, pp. 150–160, 2025.

[3] Team of Reference, *Indonesia Sehat*. 2016. [Online]. Available: www.eagleinstitute.id/submission

[4] N. Kamalawati, N. R. Febrina, N. Wijayanti, and Y. Hanoselina, "Ekspansi Program Makan Bergizi Gratis (MBG) dalam Penanggulangan Stunting dan Peningkatan Kualitas Sumber Daya Manusia Ekspansi Program Makan Bergizi Gratis (MBG) dalam Penanggulangan Stunting dan Peningkatan Kualitas Sumber Daya Manusia," *J. Intelek Insa. Cendikia*, vol. 2 No: 12, pp. 19033–19045, 2025.

[5] M. Wankhade, A. C. S. Rao, and C. Kulkarni, *A survey on sentiment analysis methods, applications, and challenges*, vol. 55, no. 7. Springer Netherlands, 2022. doi: 10.1007/s10462-022-10144-1.

[6] W. Anggriyani and M. Fakhriza, "Analisis Sentimen Program Makan Gratis Pada Media Sosial X Menggunakan Metode NLP," *J. Comput. Syst. Informatics*, vol. 5, no. 4, pp. 1033–1042, 2024, doi: 10.47065/josyc.v5i4.5826.

[7] R. Obiedat *et al.*, "Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution," *IEEE Access*, vol. 10, pp. 22260–22273, 2022, doi: 10.1109/ACCESS.2022.3149482.

[8] R. Hidayat and D. J. Ratnaningsih, "Analisis Sentimen Program Makanan Bergizi Gratis Menggunakan Algoritma Random Forest dan Naive Bayes," vol. 5, no. 1, pp. 395–400, 2025, doi: 10.47065/comforch.v5i1.2355.

[9] W. A. Rayadhani and M. Rahardi, "Comparative Analysis of Random Forest , SVM , and Naive Bayes for Cardiovascular Disease Prediction," vol. 9, no. 6, pp. 3234–3243, 2025.

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

[11] A. Sitanggang, Y. Umaidah, Y. Umaidah, R. I. Adam, and R. I. Adam, "Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naïve Bayes," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4902.

[12] D. T. Attaulah and D. Soyusiawaty, "Analisis Sentimen Program Makan Siang Gratis di Twitter/X menggunakan Metode BI-LSTM," *Edumatic J. Pendidik. Inform.*, vol. 9, no. 1, pp. 294–303, 2025, doi: 10.29408/edumatic.v9i1.29725.

[13] A. Molenaar, D. Lukose, L. Brennan, E. L. Jenkins, and T. A. McCaffrey, "Using Natural Language Processing to Explore Social Media Opinions on Food Security: Sentiment Analysis and Topic Modeling Study," *J. Med. Internet Res.*, vol. 26, no. 1, 2024, doi: 10.2196/47826.

[14] M. Tsiourlini, K. Tzafilkou, D. Karapiperis, and C. Tjortjis, "Text Analytics on YouTube Comments for Food Products," *Inf.*, vol. 15, no. 10, 2024, doi: 10.3390/info15100599.

[15] A. Cahya Kamilla, N. Priyani, R. Priskila, and V. Handrianus Pranatawijaya, "Analisis Sentimen Film Agak Laen Dengan Kecerdasan Buatan: Text Mining Metode Naïve Bayes Classifier," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 2923–2928, 2024, doi: 10.36040/jati.v8i3.9587.

[16] L. P. Mudarakola, R. K. Gatla, A. S. N. Raju, A. Y. Jaffar, A. Alzahrani, and A. A. Dessalegn, "Multi stage sentiment analysis for product reviews on Twitter using optimized machine learning algorithm," *Sci. Rep.*, vol. 15, no. 1, pp. 1–18, 2025, doi: 10.1038/s41598-025-23451-8.

[17] S. Tahun, D. Wilayah, K. Pkm, and K. Palembang, "Faktor Faktor Yang Berhubungan Dengan Kejadian Stunting Pada Balita Usia 2," vol. 5, no. 2, pp. 11396–11406, 2025.

[18] A. A. Sy, Y. Franciska, and A. Noviyanti, "The Influence Of Feeding Habbits On The Incidence Of Stunting,"

vol. 3, 2023, doi: 10.36086/maternalandchild.v3i2.2074.

[19] N. Technologies, "Sentiment analysis of Twitter data by making use of SVM , Random Forest and Decision Tree algorithm," pp. 193–198, 2021, doi: 10.1109/CSNT.2021.35.

[20] Hastuti, H. F. Maulana, and E. Satria, "Sentiment Analysis and Topic Modeling of Twitter Conversations in Indonesia's 2024 Presidential Election," *J. Pekommas*, vol. 10, no. 1, pp. 17–28, 2025, doi: 10.56873/jpkm.v9i1.5545.

[21] A. Hermawan, I. Jowensen, J. Junaedi, and Edy, "Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine," *JST (Jurnal Sains dan Teknol.*, vol. 12, no. 1, pp. 129–137, 2023, doi: 10.23887/jstundiksha.v12i1.52358.

[22] R. F. Pramesti, A. A. Firdaus, K. Yulita, and M. Thoyyibah, "Analisis Efisiensi Apbn Era Prabowo: Kajian Ekonomi Dan Analisis Sentimen Publik," *Jesya*, vol. 8, no. 2, pp. 1147–1161, 2025, doi: 10.36778/jesya.v8i2.2054.

[23] N. N. Azzat and M. C. Zulfa, "Optimasi Penjadwalan Produksi Dengan Algoritma Heuristik Pour Untuk Reduksi Makespan Pada CV CJ Furniture," *Integr. J. Ilm. Tek. Ind.*, vol. 8, no. 1, pp. 14–22, 2023, doi: 10.32502/js.v8i1.5977.

[24] S. Sutarmi, W. Warijan, T. Indrayana, D. P. P. B, and I. Gunawan, "Machine Learning Model For Stunting Prediction," *J. Heal. Sains*, vol. 4, no. 9, pp. 10–23, 2023, doi: 10.46799/jhs.v4i9.1073.

[25] R. N. Handayani, I. Najiyah, and D. A. Wisnuwardana, "Classification of Bulughul Maraam Categories: Prohibitions, Recommendations, and Information Using Extreme Learning Machine and Fasttext," *J. Online Inform.*, vol. 8, no. 2, pp. 242–251, 2023, doi: 10.15575/join.v8i2.1205.

[26] M. C. Buzea, S. Trausan-Matu, and T. Rebedea, "A Three Word-Level Approach Used in Machine Learning for Romanian Sentiment Analysis," *Proc. - RoEduNet IEEE Int. Conf.*, vol. 2019-Octob, no. July, 2019, doi: 10.1109/ROEDUNET.2019.8909458.

[27] H. Aji and F. D. Astuti, "K-Means Approach to Identifying High-Risk Stunting Areas in Indonesia," *bit-Tech*, vol. 8, no. 2, pp. 1680–1688, 2025, doi: 10.32877/bt.v8i2.3042.

[28] Aiswarya A S and H. Rajeev, "Youtube Comment Sentimental Analysis," *Indian J. Data Min.*, vol. 4, no. 1, pp. 5–8, 2024, doi: 10.54105/ijdm.a1633.04010524.