Journal of Information Technology and Computer Science (INTECOMS)
Volume 9 Nomor 4, Tahun 2026
e-ISSN : 2614-1574
p-ISSN : 2621-3249

KOMPARASI SVM DAN INDOBERT DALAM ANALISIS SENTIMEN KOMENTAR YOUTUBE TERHADAP KEBIJAKAN MAKAN BERGIZI GRATIS

COMPARISON OF SVM AND INDOBERT IN SENTIMENT ANALYSIS OF YOUTUBE COMMENTS ON THE FREE NUTRITIOUS MEAL POLICY

M. Fikri Haikal Hidasalim[1], Jeffri Alfa Razaq [2]
Fakultas Teknologi Informasi Dan Industri, Program Studi Teknik Informatika, Universitas StikubankSemarang[1,2]
mfikrihaikalhidasalim@gmail.com[1]

ABSTRACT

This study compares the performance of Support Vector Machine (SVM) and IndoBERT for sentiment analysis of YouTube comments regarding the Free Nutritious Meal (MBG) program. It utilizes 3,177 comments (January 2025–March 2026) classified into positive, negative, and neutral sentiments, evaluated with an 80:20 split. SVM applies full preprocessing and TF-IDF feature extraction, while IndoBERT uses minimal preprocessing and fine-tuning. Results show IndoBERT significantly outperforms SVM, achieving 96.52% accuracy and a 95% F1-Score, compared to SVM's 86.87% accuracy and 85% F1-Score. IndoBERT's bidirectional capability effectively captures informal contexts, making it highly recommended for government policy public opinion monitoring.

Keywords: *Sentiment analysis, Support Vector Machine, IndoBERT, TF-IDF, Natural Language Processing*

ABSTRAK

Penelitian ini membandingkan kinerja Support Vector Machine (SVM) dan IndoBERT untuk analisis sentimen komentar YouTube terkait Program Makan Bergizi Gratis (MBG). Menggunakan 3.177 data komentar (Januari 2025–Maret 2026) yang diklasifikasikan menjadi sentimen positif, negatif, dan netral. Pengujian dilakukan menggunakan pembagian data 80:20. SVM menerapkan praproses penuh dan ekstraksi fitur TF-IDF, sedangkan IndoBERT menggunakan praproses minimal dan *fine-tuning*. Hasil menunjukkan IndoBERT secara signifikan lebih unggul dengan akurasi 96,52% dan F1-Score 95%, dibandingkan SVM (akurasi 86,87%, F1-Score 85%). Keunggulan ini berkat kemampuan bidireksional IndoBERT dalam memahami konteks bahasa informal. IndoBERT sangat direkomendasikan untuk sistem pemantauan opini publik pada ranah kebijakan pemerintah.

Kata kunci : Analisis sentimen, *Support Vector Machine*, IndoBERT, TF-IDF, *Natural Language Processing*

PENDAHULUAN

Transisi kepemimpinan nasional di Indonesia membawa regulasi strategis berupa Program Makan Bergizi Gratis (MBG) yang memicu perbincangan masif melintasi berbagai ekosistem digital [1]. Gelombang opini masyarakat ini terekam dengan jelas dalam ruang interaksi virtual yang mencerminkan secara langsung tingginya atensi publik terhadap regulasi pemerintah tersebut [1]. Membludaknya tumpukan respons digital ini menjadikan upaya pemetaan sentimen audiens sebagai instrumen evaluasi empiris yang sangat krusial guna mengukur tingkat penerimaan masyarakat secara objektif [2].

Artikulasi pandangan warga negara terbukti semakin menguat pada beragam jejaring sosial lintas format [3]. Untuk memahami fenomena tersebut, analisis sentimen hadir sebagai metode komputasi yang mengelompokkan polaritas emosional teks ke dalam dimensi positif, negatif, dan netral, sehingga keberadaannya menjadi sebuah kebutuhan primer [4]. Meskipun platform media sosial sangat beragam, YouTube memiliki karakteristik linguistik yang jauh lebih menantang dan menjadi arena utama bagi masyarakat luas untuk menyuarakan opini asinkron yang kompleks terkait anggaran dan distribusi kebijakan [5]. Fenomena ribuan komentar pada video bertema MBG di YouTube ini menciptakan tumpukan data berskala besar yang merupakan representasi reaksi nyata masyarakat sekaligus aset informasi berharga bagi pengambil kebijakan [6].

Mengingat ekstraksi informasi secara manual dari ribuan komentar tersebut rentan terhadap bias manusia, dibutuhkan intervensi teknologi *Natural Language Processing* (NLP) untuk pengklasifikasian otomatis [6]. Dalam ranah ini, algoritma *machine learning* konvensional seperti *Support Vector Machine* (SVM) telah lama diakui ketangguhannya dalam mengonstruksi batas pemisah pada data teks berdimensi tinggi [7]. SVM juga menawarkan efisiensi komputasi yang sangkil [8]. Sebaliknya, inovasi *Deep Learning* menghadirkan model *Transformer* seperti IndoBERT yang didesain khusus dengan korpus bahasa Indonesia [9]. IndoBERT memiliki keunggulan dalam memahami makna kalimat secara dua arah yang sangat penting untuk memproses teks media sosial [9].

Studi komparatif sebelumnya menunjukkan bahwa model IndoBERT kerap melampaui kemampuan *machine learning* tradisional dalam menangani kompleksitas konteks bahasa Indonesia, baik pada isu tata kota [10] maupun evaluasi spesifik kebijakan MBG [11]. Meski begitu, SVM tetap mempertahankan relevansinya berkat tingkat kestabilan dan kesederhanaannya [9]. Kesenjangan riset saat ini terletak pada masih minimnya studi yang membenturkan performa SVM dan IndoBERT secara komparatif khusus pada data komentar YouTube mengenai MBG, yang teksnya jauh lebih berantakan dan kompleks dibandingkan platform lain [11]. Oleh karena itu, riset ini bertujuan mengomparasi akurasi dan efektivitas kedua model tersebut untuk menemukan pendekatan paling optimal dalam sistem pemantauan opini publik [11].

TINJAUAN PUSTAKA

2.1. Kajian Penelitian Terdahulu

Eksplorasi literatur sebelumnya menjadi pijakan komparatif yang esensial untuk mengidentifikasi keunggulan serta batasan dari berbagai metode komputasi. Sebuah studi yang diinisiasi oleh Rayyana Muntaza [1] membuktikan bahwa algoritma *Support Vector Machine* (SVM) berbasis *kernel linear* memiliki stabilitas tinggi dalam mengklasifikasikan respons publik terkait program Makan Bergizi Gratis (MBG) di YouTube, asalkan data mentah tersebut telah melewati fase pembersihan teks yang ketat. Dalam ranah perbandingan algoritma, riset Sidauruk [2] mengenai sentimen pembangunan Ibu Kota Negara memperlihatkan superioritas model IndoBERT dibandingkan kombinasi TF-IDF dan SVM. Keunggulan ini didasari oleh kapabilitas *contextual embedding* pada IndoBERT yang sangat piawai dalam menerjemahkan struktur bahasa informal beserta makna tersiratnya.

Pada pengujian lain, Anas Qolbu dan Fitriyati [3] menakar efektivitas *Naive Bayes*, SVM, dan IndoBERT dalam skenario penanganan data yang tidak seimbang. Hasil eksperimen tersebut mengonfirmasi bahwa IndoBERT memimpin pada perolehan metrik *F1 Score*, meskipun SVM tetap menawarkan nilai tambah berupa efisiensi waktu komputasi dan konsumsi sumber daya perangkat keras yang lebih ringan. Khusus pada analisis kebijakan MBG, Setiadi dan Sulistyo [4] menyoroti bahwa tahapan penyesuaian akhir (*fine tuning*) pada model pralatih IndoBERT mampu secara presisi mendeteksi nuansa kritik dan dukungan warga internet yang melampaui kapasitas algoritma pembelajaran mesin tradisional.

Berdasarkan sintesis dari berbagai literatur tersebut, ditarik sebuah kesenjangan riset yang signifikan. Hingga saat ini, kajian yang menghadapkan secara langsung performa SVM klasik dengan model *deep learning* IndoBERT pada domain spesifik komentar YouTube mengenai kebijakan MBG masih sangat terbatas. Penelitian ini hadir untuk mengisi celah tersebut dengan menguji model mana yang paling adaptif dalam memproses bahasa publik yang heterogen.

2.2. Konsep Analisis Sentimen

Analisis sentimen atau penambangan opini merupakan salah satu cabang dari

disiplin *Natural Language Processing* (NLP) yang dirancang untuk mengomputasi afeksi emosional teks ke dalam polaritas positif, negatif, maupun netral [5]. Pada implementasi kebijakan publik seperti MBG, metode ini bertindak sebagai instrumen evaluasi digital untuk memetakan persepsi masyarakat secara instan. Fenomena interaksi di media sosial melahirkan volume data berskala raksasa yang tidak mungkin diolah secara manual, sehingga adopsi kecerdasan buatan menjadi sebuah urgensi. Namun, pengolahan bahasa Indonesia di platform seperti YouTube menghadirkan kerumitan tersendiri akibat maraknya penggunaan bahasa gaul, sarkasme, dan singkatan tidak baku. Hal ini menuntut penggunaan algoritma yang tidak sekadar berhitung frekuensi kata, melainkan mampu menalar ambiguitas konteks secara mendalam demi mencapai reliabilitas pengujian yang tinggi.

2.3. Dinamika Data Platform YouTube

YouTube kini tidak hanya berfungsi sebagai medium hiburan, melainkan telah bertransformasi menjadi ruang diskursus publik untuk merespons kebijakan strategis pemerintah. Platform ini menyimpan karakteristik teks yang amat rumit, mencakup transkripsi bahasa lisan, campuran dialek daerah, hingga simbol visual pengganti emosi. Syafia *et al.* [6] menegaskan bahwa anomali pola interaksi ini menuntut penanganan ekstraksi fitur yang spesifik. Tingginya tingkat ambiguitas dan kesalahan tik pada komentar YouTube membuat proses analisisnya membutuhkan algoritma yang sangat fleksibel [1]. Firdaus dan Trisnawarman [7] juga membuktikan perlunya model bahasa tingkat lanjut untuk menaklukkan istilah nonformal, sementara Huwaida *et al.* [8] menekankan bahwa besarnya volume opini ini merupakan representasi paling murni dari reaksi riil masyarakat.

2.4. Integrasi Machine Learning dan Text Mining

Pembelajaran mesin (*machine learning*) menawarkan kerangka sistematis untuk menstrukturkan data teks yang semula acak menjadi informasi klasifikasi yang berharga. Kesuksesan penambangan teks ini sangat bergantung pada tahapan prapemrosesan. Merujuk pada pandangan Amanda Salsabila *et al.* [9], prapemrosesan harus mencakup penyeragaman karakter huruf, pembersihan simbol dan tautan, normalisasi kosakata tidak baku, hingga pemecahan token kata. Setelah data bersih, tahapan ekstraksi fitur mengambil alih peran utama. Model konvensional umumnya menerapkan pembobotan *Term Frequency Inverse Document Frequency* (TF IDF) [2], sedangkan arsitektur mutakhir seperti IndoBERT mengandalkan mekanisme *contextual embedding* untuk menangkap relasi semantik antarkata secara utuh [3].

2.5. Support Vector Machine (SVM)

SVM adalah algoritma pembelajaran terarah yang sangat populer untuk tugas klasifikasi teks. Prinsip utama algoritma ini adalah mencari bidang pemisah ruang (*hyperplane*) paling optimal yang sanggup membelah sebaran data antarkelas dengan jarak margin paling maksimal [9]. Karena SVM tidak dapat memproses teks secara mentah, metode TF IDF digunakan untuk menerjemahkan nilai linguistik menjadi vektor numerik. Keistimewaan SVM terletak pada daya tahannya saat mengeksekusi data dengan dimensi fitur yang membengkak tanpa terjebak pada persoalan *overfitting*. Algoritma ini sering kali didapuk sebagai model tolok ukur (*benchmark*) yang solid pada berbagai riset penambangan opini.

2.6. Arsitektur Model IndoBERT

IndoBERT merupakan inovasi representasi bahasa berbasis arsitektur *Transformer* yang dilatih secara khusus menggunakan miliaran leksikon bahasa Indonesia. Berbeda dengan model klasik yang memproses kata secara terpisah, IndoBERT mengadopsi mekanisme atensi

mandiri (*self attention*) dua arah [10]. Sistem ini memampukan model untuk mengenali konteks sebuah kata dengan melihat kaitan kata sebelum dan sesudahnya secara bersamaan. Kekuatan pemahaman semantik inilah yang membuat IndoBERT sangat tangguh dalam membongkar makna tersirat pada frasa media sosial. Melalui proses *fine tuning* yang tepat, model ini dapat menghasilkan klasifikasi yang cerdas. Terlebih lagi, integrasi skor kepercayaan pada IndoBERT berfungsi sebagai parameter pelengkap untuk menjamin konsistensi hasil prediksi pada data yang bersifat dinamis [11].

METODE PENELITIAN

3.1. Desain dan Pendekatan Penelitian

Penelitian ini mengadopsi pendekatan kuantitatif melalui desain eksperimen komparatif guna mengevaluasi secara objektif dan empiris kinerja dua algoritma klasifikasi pada analisis sentimen teks berbahasa Indonesia. Komparasi difokuskan pada *Support Vector Machine* (SVM) yang merepresentasikan algoritma *machine learning* konvensional, dan model IndoBERT yang mewakili pendekatan *deep learning* berarsitektur *Transformer*. Objek kajian dipusatkan pada korpus komentar pengguna YouTube terkait program Makan Bergizi Gratis (MBG), yang merekam spektrum opini publik mulai dari dukungan, kritik, hingga pandangan netral. Rangkaian tahapan penelitian menstrukturkan prosedur mulai dari pengadaan data, anotasi label, prapemrosesan linguistik, ekstraksi fitur, pemodelan, hingga evaluasi komprehensif untuk memastikan validitas hasil.

3.2. Alur Operasional Penelitian

Arsitektur penelitian dirancang secara sistematis. Fase inisiasi bermula dari pengumpulan data komentar YouTube melalui teknik *web scraping*, dilanjutkan dengan pelabelan sentimen secara manual untuk menciptakan *ground truth*. Data mentah tersebut kemudian dinormalisasi melalui tahapan prapemrosesan teks guna mereduksi derau linguistik. Pada fase transformasi fitur, aliran komputasi dibagi menjadi dua cabang: ekstraksi berbasis bobot frekuensi (TF-IDF) untuk *pipeline* SVM, dan representasi *contextual embedding* untuk *pipeline* IndoBERT. Kedua model dilatih secara terpisah menggunakan partisi data latih, dan diakhiri dengan pengujian kinerja prediktif menggunakan matriks evaluasi standar (*Accuracy, Precision, Recall*, dan *F1-Score*).

3.3. Pengumpulan Data

Pengadaan data primer bertumpu pada ekstraksi teks dari platform YouTube, yang direpresentasikan sebagai lokus sentral diskursus asinkron dengan tingkat interaktivitas tinggi. Melalui teknik penambangan *scraping*, diekstraksi sebanyak 3.177 komentar unik dari berbagai video kanal berita dan diskusi yang secara spesifik mengulas implementasi kebijakan MBG. Data mentah ini diarsipkan dalam format *Comma Separated Values* (CSV) guna memfasilitasi kelancaran proses komputasi pada tahapan lanjutan.

3.4. Anotasi dan Pelabelan Data

Pelabelan data (*data annotation*) dieksekusi secara manual oleh peneliti untuk menjamin akurasi interpretasi makna. Pendekatan manual ini menjadi krusial untuk menavigasi kompleksitas bahasa media sosial yang sarat akan sarkasme, singkatan tak baku, dan ambiguitas konteks. Setiap entri diklasifikasikan ke dalam tiga kelas polaritas diskrit:

Positif: Mengindikasikan dukungan, optimisme, atau afirmasi terhadap program MBG.

Negatif: Mengandung elemen kritik, penolakan, sinisme, atau kekhawatiran masyarakat.

Netral: Bersifat informatif murni atau ketiadaan tendensi emosional yang condong ke salah satu kutub.

3.5. Prapemrosesan Data (*Text Preprocessing*)

Prapemrosesan beroperasi sebagai fase vital untuk mentransformasi teks nir-struktur yang dipenuhi *noise* menjadi format siap komputasi. Terdapat diferensiasi penanganan antar-algoritma: SVM menuntut prapemrosesan radikal untuk mereduksi dimensi fitur, sementara IndoBERT dikondisikan dengan prapemrosesan moderat guna menjaga keutuhan rantai konteks semantik.

1. *Text Cleaning*: Eliminasi karakter non-alfabetis yang minim nilai sentimen, seperti tautan (URL), angka, simbol, dan tanda baca.
2. *Case Folding*: Penyeragaman seluruh kapitalisasi huruf menjadi huruf kecil (*lowercase*) untuk menekan redundansi fitur leksikal.
3. *Tokenization*: Segmentasi untaian kalimat panjang menjadi token-token kata tunggal secara terpisah.
4. *Stopword Removal*: Reduksi kata hubung fungsional (seperti "dan", "yang", "di") yang memiliki frekuensi tinggi namun bobot maknanya rendah. Pada IndoBERT, tahap ini diaplikasikan sangat terbatas agar struktur tata bahasa tetap utuh.
5. *Stemming*: Transformasi morfologis untuk memangkas imbuhan sehingga menyisakan kata dasar (*root word*). Tahap ini diimplementasikan secara eksklusif untuk SVM. IndoBERT tidak memerlukan *stemming* karena telah dilengkapi mekanisme pemecah kata internal (*WordPiece tokenizer*).

3.6. Partisi Dataset

Korpus data yang telah disanitasi dipartisi menggunakan metode pembagian acak (*random splitting*) dengan rasio 80:20. Proporsi 80% dialokasikan sebagai data latih (*training data*) bagi mesin untuk mengenali pola linguistik, sedangkan sisa 20% diisolasi sebagai data uji (*testing data*) guna memvalidasi kemampuan generalisasi model terhadap data baru. Rasio ini menjamin ekuilibrium optimal antara ketersediaan data untuk pembelajaran dan objektivitas pengujian kinerja.

3.7. Ekstraksi Fitur Berbasis TF-IDF

Sebelum dikonsumsi oleh algoritma SVM, teks kualitatif ditransformasikan menjadi vektor matematis melalui formulasi *Term Frequency–Inverse Document Frequency* (TF-IDF). Metode ini menyeimbangkan urgensi sebuah kata berdasarkan densitas kemunculannya di satu dokumen berbanding terbalik dengan frekuensinya di seluruh korpus.

1. Term Frequency (TF): Menghitung frekuensi kemunculan sebuah kata *t* di dalam dokumen tunggal *d*.
$$TF(t,d) = f(t,d)$$
2. Inverse Document Frequency (IDF): Menakar seberapa spesifik/informatif sebuah kata, menggunakan total dokumen *N* dan jumlah dokumen yang memuat kata tersebut $df(t)$.
$$IDF(t) = \log\left(\frac{N}{df(t)}\right)$$
3. Nilai TF-IDF: Bobot representasi akhir diperoleh dari perkalian kedua komponen di atas.
$$TF\text{-}IDF(t,d) = TF(t,d) \times IDF(t)$$

3.8. Implementasi Support Vector Machine (SVM)

SVM diposisikan sebagai model *supervised learning* yang diunggulkan dalam menangani matriks fitur berdimensi masif hasil kalkulasi TF-IDF. Prinsip fundamental algoritma ini adalah mencari *hyperplane* (bidang batas) terbaik yang membelah sebaran vektor data antar-kelas sentimen di ruang multidimensi. Titik-titik data terluar yang membatasi bidang ini disebut sebagai *support vectors*.

Konsep Hyperplane: Batas linear pemisah kelas dirumuskan dalam persamaan matematika:

$$w \cdot x + b = 0$$

(Keterangan: *w* = vektor bobot, *x* = vektor fitur input, *b* = bias).

Optimalisasi Margin: SVM bekerja dengan menargetkan margin terlebar antarkelas

untuk menekan risiko *overfitting*. Kalkulasi pelebaran margin didefinisikan melalui:

$$\text{Margin} = \frac{2}{|w|}$$

3.9. Implementasi Model IndoBERT

Sebagai kerangka perbandingan, IndoBERT merepresentasikan *language model* berbasis *Transformer* yang telah menelan prapelatihan korpus raksasa berbahasa Indonesia. Berbeda dari TF-IDF, mekanisme atensi mandiri (*self-attention*) dua arah pada IndoBERT memampukan mesin untuk menafsirkan arti kata secara dinamis bergantung pada struktur kalimat yang mengapitnya.

Mekanisme Arsitektur: Setelah ditokenisasi, mesin menyematkan penanda [CLS] di awal kalimat sebagai representasi makna komprehensif, dan [SEP] sebagai demarkasi akhir. Rangkaian token ini dikonversi menjadi *embedding* numerik dan didorong melewati lapisan-lapisan *encoder transformer* untuk merekam korelasi kontekstual. Lapisan [CLS] terakhir kemudian diekstraksi untuk memprediksi kelas target.

Prosedur Pelatihan (*Fine-Tuning*): Varian model *indobenchmark/indobert-base-p2* diselaraskan ulang menggunakan data latih penelitian. Memanfaatkan integrasi pustaka PyTorch, proses iterasi (*epoch*) dioptimasi menggunakan algoritma AdamW guna memperbarui parameter bobot hingga batas *loss* minimum tercapai. Probabilitas puncak dari distribusi tiga kelas sentimen (Positif, Negatif, Netral) akan ditetapkan sebagai output akhir.

3.10. Metrik Evaluasi Kinerja

Kapasitas presisi kedua model ditakar menggunakan kerangka matriks kebingungan (*Confusion Matrix*), yang menyilangkan label aktual dengan prediksi mesin ke dalam empat kuadran: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dari matriks tersebut, diekstraksi empat indikator evaluasi komprehensif:

Accuracy: Merepresentasikan rasio prediksi yang secara absolut benar dibandingkan keseluruhan populasi sampel uji.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: Memproyeksikan kualitas tebakan positif mesin (rasio ketepatan prediksi positif terhadap total seluruh prediksi positif).

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall: Mengukur tingkat kepekaan (sensitivitas) model dalam merengkuh seluruh data positif yang benar-benar eksis.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-Score: Menyajikan ekuilibrium antara *Precision* dan *Recall* melalui perhitungan rerata harmonik. Metrik ini sangat ideal dan tepercaya untuk studi teks klasifikasi yang kerap menghadapi distribusi data asimetris.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

HASIL DAN PEMBAHASAN

4.1. Pengumpulan dan Karakteristik Dataset

Dataset yang dianalisis dalam penelitian ini bersumber dari opini publik di platform YouTube terkait implementasi kebijakan "Makan Bergizi Gratis" (MBG). Ekstraksi data dilakukan menggunakan YouTube Data API v3 pada video dari kanal berita nasional dan instansi pemerintah dengan interaksi tertinggi selama periode Maret 2025 hingga Maret 2026. Pemilihan rentang waktu ini merepresentasikan fase krusial dari wacana hingga uji coba kebijakan.

Setelah proses penyaringan komentar *spam* dan *bot*, terkumpul 3.177 data valid yang kemudian dianotasi secara manual ke dalam tiga kelas sentimen: Positif, Netral, dan Negatif. Rincian distribusi data dapat dilihat pada Tabel 1.

Tabel 1. Distribusi Sentimen Opini Kebijakan MBG

Kelas Sentimen	Jumlah Komentar	Persentase
Negatif	1.734	54,58%
Positif	754	23,73%

Netral	689	21,69%
Total	3.177	100%

Dominasi sentimen negatif (54,58%) mengindikasikan tingginya polarisasi dan kekhawatiran publik terhadap aspek teknis maupun penganggaran program MBG. Sampel variasi gaya bahasa (*natural language*) warganet yang mencakup bahasa gaul, singkatan, dan majas disajikan pada Tabel 2.

Tabel 2. Sampel Komentar Berdasarkan Label Sentimen

No	Cuplikan Komentar Asli	Label
1	58% pasukan Konoha, hanya bisa lapar aja, mau racunkah di makan juga? Mungkin otaknya udah di cuci bahkan di racuni, salam waras	Negatif
2	Pola pemerintahan dari dulu sama saja. Keras teriak anti korupsi keluar lingkaran kekuasaan. Rakyat senang dan bertepuk tangan	Negatif
3	Marah, sedih, kecewa, gk ngerti lagi gw	Negatif
4	Semoga program MBG BS dpt mningkatkn ank2 penerus bangsa	Netral
5	Enak e, pak Prabowo haturnuhun anak saya kopen	Positif

4.2. Prapemrosesan Data

Praproses teks dilaksanakan melalui lima tahapan berurutan yang saling terhubung. Setiap tahapan menghasilkan transformasi spesifik pada data tekstual mentah. Kelima sampel komentar dari Tabel 4 digunakan secara konsisten untuk mengilustrasikan output pada setiap tahap, sehingga perubahan kumulatif antarstep dapat diamati secara sistematis.

4.2.1 *Text Cleaning*

Text cleaning bertujuan mengeliminasi elemen-elemen yang tidak memiliki kontribusi semantik dalam penentuan sentimen. Komponen yang dihapus meliputi angka, tanda baca, simbol matematika, URL, karakter khusus, dan emoji. Setelah proses ini, teks hanya menyisakan kata-kata yang secara langsung merepresentasikan opini pengguna. Tabel 5 menampilkan perbandingan teks sebelum dan sesudah *cleaning*.

Tabel 3. Hasil Tahap Text Cleaning

No	Data Sebelum Cleaning	Data Sesudah Cleaning
1	58% pasukan Konoha, hanya bisa lapar aja, mau racunkah di makan juga? Mungkin otaknya udah di cuci bahkan di racuni, salam waras	pasukan Konoha hanya bisa lapar aja mau racunkah di makan juga Mungkin otaknya udah di cuci bahkan di racuni salam waras
2	Pola pemerintahan dari dulu sama saja. Keras teriak anti korupsi keluar lingkaran kekuasaan. Rakyat senang dan bertepuk tangan	Pola pemerintahan dari dulu sama saja Keras teriak anti korupsi keluar lingkaran kekuasaan Rakyat senang dan bertepuk tangan
3	Marah, sedih, kecewa, gk ngerti lagi gw	Marah sedih kecewa gk ngerti lagi gw
4	Semoga program MBG BS dpt mningkatkn ank2 penerus bangsa ID🔥	Semoga program MBG BS dpt mningkatkn ank penerus bangsa
5	Enak e, pak Prabowo haturnuhun anak saya kopen	Enak e pak Prabowo haturnuhun anak saya kopen

Berdasarkan Tabel 3, tanda baca seperti koma dan tanda tanya berhasil dieliminasi. Pada komentar nomor 4, karakter angka "2" dalam "ank2" dan emoji bendera Indonesia serta simbol api turut dihapus. Teks yang dihasilkan lebih bersih

namun masih mempertahankan keberagaman kapitalisasi yang akan diseragamkan pada tahap berikutnya.

4.2.2 *Case Folding*

Penyeragaman abjad atau *case folding* merupakan proses mengubah seluruh karakter huruf menjadi huruf kecil (*lowercase*). Tahapan ini dilakukan guna mencegah terjadinya redundansi fitur. Tanpa proses ini, sistem komputasi secara bawaan akan memperlakukan kata dengan kapitalisasi berbeda, misalnya "Makan", "MAKAN", dan "makan", sebagai tiga entitas terpisah meskipun ketiganya memiliki makna identik. Tabel 6 menyajikan hasil transformasi *case folding* pada kelima sampel komentar.

Tabel 4. Hasil Tahap *Case Folding*

No	Sebelum Case Folding	Sesudah Case Folding
1	pasukan Konoha hanya bisa lapar aja mau racunkah di makan juga Mungkin otaknya udah di cuci bahkan di racuni salam waras	pasukan konoha hanya bisa lapar aja mau racunkah di makan juga mungkin otaknya udah di cuci bahkan di racuni salam waras
2	Pola pemerintahan dari dulu sama saja Keras teriak anti korupsi keluar lingkaran kekuasaan Rakyat senang dan bertepuk tangan	pola pemerintahan dari dulu sama saja keras teriak anti korupsi keluar lingkaran kekuasaan rakyat senang dan bertepuk tangan
3	Marah sedih kecewa gk ngerti lagi gw	marah sedih kecewa gk ngerti lagi gw
4	Semoga program MBG BS dpt mningkatkn ank penerus bangsa	semoga program mbg bs dpt mningkatkn ank penerus bangsa
5	Enak e pak Prabowo haturnuhun anak saya kopen	enak e pak prabowo haturnuhun anak saya kopen

Berdasarkan Tabel 4, seluruh huruf kapital telah dikonversi menjadi huruf kecil secara konsisten. Kata "Konoha" menjadi "konoha", "MBG" menjadi "mbg", dan "Prabowo" menjadi "prabowo". Transformasi ini memastikan bahwa setiap kata diperlakukan sebagai satu entitas tunggal, terlepas dari cara penulisannya dalam komentar asli.

4.2.3 *Tokenisasi*

Tokenisasi merupakan proses pemecahan rangkaian kalimat menjadi satuan kata tunggal yang disebut token. Melalui tokenisasi, setiap komentar yang semula berbentuk string kalimat diubah menjadi sebuah daftar (list) token yang terpisah. Representasi ini memungkinkan penghitungan frekuensi tiap kata pada algoritma SVM maupun pemrosesan urutan sekuensial pada model IndoBERT. Tabel 5 menampilkan hasil tokenisasi pada kelima sampel.

Tabel 5. Hasil Tahap Tokenisasi

No	Teks Input	Hasil Tokenisasi
1	pasukan konoha hanya bisa lapar aja...	['pasukan','konoha','hanya','bisa','lapar','aja','mau','racunkah','di','makan','juga','mungkin','otaknya','udah','di','cuci','bahkan','di','racuni','salam','waras']
2	pola pemerintahan dari dulu sama...	['pola','pemerintahan','dari','dulu','sama','saja','keras','teriak','anti','korupsi','keluar','lingkaran','kekuasaan','rakyat','senang','dan','bertepuk','tangan']
3	marah sedih kecewa gk ngerti lagi gw	['marah','sedih','kecewa','gk','ngerti','lagi','gw']

4	semoga program mbg bs dpt mningkatkn ank penerus bangsa	['semoga','program','mbg','bs','dpt','mningkatkn','ank','penerus','bangsa']
5	enak e pak prabowo haturnuhun anak saya kopen	['enak','e','pak','prabowo','haturnuhun','anak','saya','kopen']

Berdasarkan Tabel 5, setiap kalimat berhasil dipecah menjadi daftar token individual. Komentar nomor 1 menghasilkan 21 token, komentar nomor 2 menghasilkan 18 token, sementara komentar nomor 3 yang relatif pendek hanya menghasilkan 7 token. Hasil tokenisasi ini menjadi input bagi tahapan stopword removal berikutnya untuk mereduksi token yang tidak bermakna.

4.2.4 *Stopword Removal*

Stopword removal bertujuan mengeliminasi kata-kata fungsional yang memiliki intensitas kemunculan tinggi tetapi tidak menyumbang nilai informatif dalam penentuan polaritas sentimen, seperti kata hubung, kata ganti, dan partikel. Proses ini diterapkan secara penuh pada jalur SVM untuk mempertajam bobot fitur TF-IDF, sedangkan pada jalur IndoBERT diterapkan secara terbatas guna menjaga integritas konteks kalimat yang diperlukan oleh arsitektur Transformer. Tabel 6 menyajikan perubahan daftar token sebelum dan sesudah stopword removal.

Tabel 6. Hasil Tahap Stopword Removal

No	Sebelum Stopword Removal	Sesudah Stopword Removal
1	['pasukan','konoha','hanya','bisa','lapar','aja','mau','racunkah','di','makan','juga','mungkin','otaknya','udah','di','cuci','bahkan','di','racuni','salam','waras']	['pasukan','konoha','lapar','racunkah','makan','otaknya','cuci','racuni','salam','waras']
2	['pola','pemerintahan','dari','dulu','sama','saja','keras','teriak','anti','korupsi','keluar','lingkaran','asaan','rakyat','senang','dan','bertepuk','tangan']	['pola','pemerintahan','keras','teriak','anti','korupsi','keluar','lingkaran','kekuasaan','rakyat','senang','bertepuk','tangan']
3	['marah','sedih','kecewa','gk','ngerti','lagi','gw']	['marah','sedih','kecewa','ngerti']
4	['semoga','program','mbg','bs','dpt','mningkatkn','ank','penerus','bangsa']	['program','mbg','mningkatkn','ank','penerus','bangsa']
5	['enak','e','pak','prabowo','haturnuhun','anak','saya','kopen']	['enak','prabowo','haturnuhun','anak','kopen']

Berdasarkan Tabel 6, tampak reduksi token yang signifikan pada setiap komentar. Komentar nomor 1 mengalami penurunan dari 21 token menjadi 10 token, dengan kata-kata seperti "hanya", "bisa", "aja", "mau", "di", "juga", "mungkin", "udah", dan "bahkan" berhasil dieliminasi. Pada komentar nomor 3, daftar token terpangkas dari 7 menjadi 4 token, menyisakan kata-kata bermuatan sentimen negatif yang kuat seperti "marah", "sedih", "kecewa", dan "ngerti". Reduksi dimensi fitur ini secara langsung meningkatkan efisiensi komputasi dan kualitas bobot TF-IDF pada model SVM.

4.2.5 Stemming

Stemming merupakan tahap akhir praproses pada jalur SVM, yaitu proses mereduksi kata berimbuhan menjadi bentuk kata dasar (root word) melalui penghapusan awalan (prefix), akhiran (suffix), maupun

sisipan (infix) menggunakan library PySastrawi. Tujuannya adalah menyeragamkan variasi leksikal agar kata-kata yang memiliki akar makna serupa dikonsolidasikan ke dalam satu representasi tunggal. Tahapan ini tidak diterapkan pada jalur IndoBERT karena tokenizer internal WordPiece pada model tersebut telah mampu menangkap informasi morfologis secara inheren. Tabel 7 menyajikan hasil stemming pada kelima sampel komentar.

Tabel 7. Hasil Tahap Stemming

No	Sebelum Stemming	Sesudah Stemming
1	['pasukan','konoha','lapar','racunkah','makan','otaknya','cuci','racuni','salam','waras']	['pasuk','konoha','lapar','racun','makan','otak','cuci','racun','salam','waras']
2	['pola','pemerintahan','keras','teriak','anti','korupsi','keluar','lingkaran','kekuasaan','rakyat','senang','bertepuk','tangan']	['pola','perintah','keras','teriak','anti','korupsi','keluar','lingkar','kuasa','rakyat','senang','tepuk','tangan']
3	['marah','sedih','kecewa','ngerti']	['marah','sedih','kecewa','ngerti']
4	['program','mbg','mningkatkn','ank','penerus','bangsa']	['program','mbg','tingkat','ank','terus','bangsa']
5	['enak','prabowo','haturnuhun','anak','kopen']	['enak','prabowo','haturnuhun','anak','kopen']

Merujuk pada Tabel 7, tahapan *stemming* terbukti mampu menekan variasi morfologi pada beberapa kosakata. Sebagai contoh teks "pemerintahan" bertransformasi menjadi "perintah", "kekuasaan" menjadi "kuasa", "lingkaran" menjadi "lingkar", "bertepuk" menjadi "tepuk", serta "penerus" menyusut menjadi "terus". Khusus pada sampel opini ketiga tidak ditemukan adanya modifikasi mengingat seluruh token tersisa telah berbentuk kata dasar. Adapun istilah serapan dari dialek daerah semacam "haturnuhun" dan "kopen" maupun singkatan nonbaku "mbg" tidak mengalami perubahan wujud karena perbendaharaan kata tersebut belum terdaftar dalam kamus *stemmer* bahasa Indonesia. Hasil pamungkas dari prapemrosesan ini kemudian difungsikan sebagai masukan untuk ekstraksi fitur TF IDF pada fase pelatihan algoritma SVM.

Secara komprehensif kelima fase prapemrosesan tersebut sukses memproduksi korpus data yang jauh lebih terstandar serta bersih dari derau atau *noise*. Penyusutan jumlah token secara drastis dari sebuah opini dengan puluhan teks mentah menjadi sekumpulan istilah bermuatan semantik membuktikan tingginya efektivitas alur standardisasi ini. Dataset yang telah dibersihkan tersebut pada akhirnya siap diolah lebih lanjut untuk proses klasifikasi sentimen menggunakan kedua model pengujian.

4.3. Pembagian Dataset

Dari total 3.177 komentar, pembagian data dengan rasio 80:20 menghasilkan 2.545 data latih dan 632 data uji. Distribusi per kelas sentimen pada setiap partisi disajikan pada Tabel 8.

Tabel 8. Distribusi Pembagian Dataset Training dan Testing

Kategori Sentimen	Data Training	Data Testing	Total
Negatif	1.398	336	1.734
Netral	546	143	689
Positif	601	153	754
Total	2.545	632	3.177

4.4. Implementasi dan Evaluasi Model SVM

Setelah praproses selesai, fitur teks diubah menjadi representasi numerik menggunakan TF-IDF. Visualisasi bobot TF-IDF tertinggi menunjukkan bahwa kata-

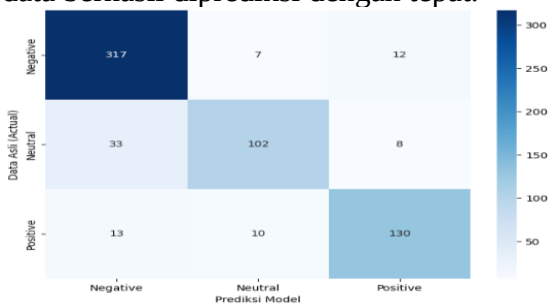
kata seperti "mbg", "racun", "makan", "gratis", "program", "anak", "gizi", dan "sekolah" memiliki bobot tertinggi dalam dataset. Hal ini mengindikasikan relevansi kuat antara topik kebijakan pangan dan pola kata yang dominan dalam komentar.

Model SVM kemudian dilatih menggunakan data latih dan dievaluasi pada data uji. Hasil evaluasi disajikan pada Tabel 9. Model SVM mencapai akurasi keseluruhan sebesar 86,87%. Performa terbaik ditunjukkan pada kelas negatif dengan recall 0,94, mengindikasikan kemampuan model yang baik dalam mendeteksi mayoritas data bertanda negatif. Namun, kelas netral menunjukkan nilai recall lebih rendah (0,71), yang mencerminkan kesulitan model dalam membedakan komentar netral dari komentar negatif akibat kemiripan pola leksikal.

Tabel 9. Classification Report Model SVM

Kelas	Precision	Recall	F1-Score	Support
Negatif	0,87	0,94	0,91	336
Netral	0,86	0,71	0,78	143
Positif	0,87	0,85	0,86	153
Accuracy	0,87	-	-	632
Macro Average	0,87	0,83	0,85	632
Weighted Average	0,87	0,87	0,87	632

Gambar 1 menampilkan Confusion Matrix model SVM. Dari visualisasi tersebut, sebanyak 317 dari 336 data negatif berhasil diprediksi benar. Pada kelas netral, terdapat 33 data yang salah diklasifikasikan sebagai negatif, yang menunjukkan adanya tumpang tindih pola kata antara kedua kelas tersebut. Pada kelas positif, 130 dari 153 data berhasil diprediksi dengan tepat.



Gambar 1. Confusion Matrix Model SVM

4.5. Implementasi dan Evaluasi Model IndoBERT

Pemodelan IndoBERT dilatih melalui tahapan penyesuaian akhir atau *fine tuning* selama tiga putaran *epoch* dengan memanfaatkan pengoptimal AdamW. Dinamika metrik kesalahan atau *loss* pada tiap putaran disajikan secara terperinci dalam Tabel 8. Nilai *loss* memperlihatkan penyusutan yang sangat konsisten, berawal dari angka 0,0510 di *epoch* perdana hingga menyentuh 0,0051 pada *epoch* ketiga. Tren penurunan ini menjadi indikator kuat bahwa alur pembelajaran model berlangsung secara konvergen serta sangat efektif. Adapun durasi eksekusi komputasi untuk satu putaran penuh menelan waktu kurang lebih 40 menit. Fakta waktu komputasi ini sekaligus merefleksikan tingginya tingkat kompleksitas arsitektur *deep learning* apabila dikomparasikan dengan efisiensi algoritma SVM.

Tabel 8. Proses Pelatihan Model IndoBERT

Epoch	Nilai Loss / Keterangan
1	Model mulai mempelajari pola dasar dari data teks; nilai loss: 0,0510
2	Penyesuaian parameter lebih lanjut; nilai loss: 0,1496
3	Peningkatan performa terlihat signifikan; nilai loss: 0,0051

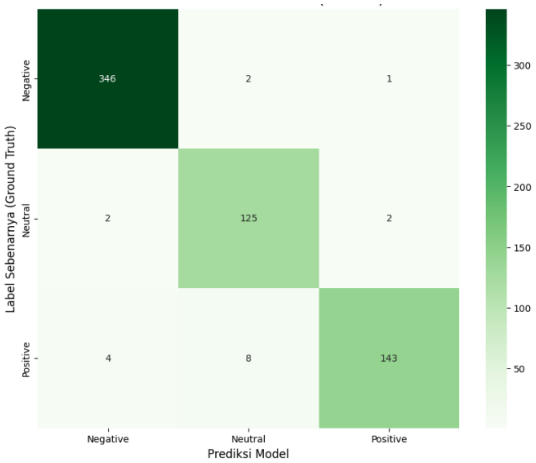
Hasil evaluasi model IndoBERT pada data uji disajikan pada Tabel 9. Model mencapai akurasi 96,52%, jauh melampaui performa SVM. Pada kelas negatif, model mencapai precision dan recall masing-masing 0,99, menunjukkan kemampuan hampir sempurna dalam mengenali data bertanda negatif. Kelas netral memperoleh recall 0,97, sedangkan kelas positif mendapatkan F1-Score sebesar 0,95.

Tabel 9. Classification Report Model IndoBERT

Kelas	Precision	Recall	F1-Score	Support
Negatif	0,99	0,99	0,99	349
Netral	0,89	0,97	0,93	129
Positif	0,98	0,92	0,95	155
Accuracy	0,97	-	-	633

Macro Average	0,95	0,96	0,95	633
Weighted Average	0,97	0,97	0,97	633

Gambar 2 menampilkan Confusion Matrix model IndoBERT. Sebanyak 346 dari 349 data negatif berhasil diklasifikasikan dengan benar. Pada kelas netral, 125 dari 129 data terprediksi tepat, sedangkan sisanya terdistribusi ke kelas negatif dan positif. Pada kelas positif, 143 dari 155 data berhasil diklasifikasikan dengan benar. Kesalahan prediksi yang tersisa sangat minimal dibandingkan hasil model SVM.



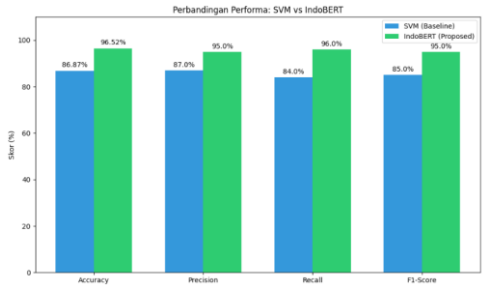
Gambar 2. Confusion Matrix Model IndoBERT

4.6. Analisis Komparasi Kinerja Algoritma

Tabel 10 memaparkan rekapitulasi komparasi performa dari kedua algoritma berdasarkan seluruh parameter pengujian. IndoBERT tampak mendominasi pada seluruh parameter ukur dengan marjin keunggulan yang sangat substansial.

Tabel 10. Rekapitulasi Komparasi Performa SVM dan IndoBERT

Algoritma	Akurasi	Presisi	Recall	F1 Score
SVM (TF IDF)	86,87%	87%	83%	85%
IndoBERT (*fine tuning*)	96,52%	95%	96%	95%
Selisih Margin	+9,65%	+8%	+13%	+10%



Gambar 3. Visualisasi Komparasi Kinerja Model SVM melawan IndoBERT

Merujuk pada pemaparan Tabel 14 beserta Gambar 3, terlihat jelas dominasi IndoBERT pada keseluruhan aspek evaluasi. Terdapat rentang akurasi sebesar 9,65 poin persentase di antara IndoBERT yakni 96,52 persen berbanding SVM pada angka 86,87 persen. Kesenjangan ini merefleksikan adanya disparitas mendasar perihal bagaimana kedua algoritma tersebut mengenali serta merepresentasikan korpus teks. Pendekatan SVM lewat ekstraksi TF IDF sebatas menghitung frekuensi kemunculan teks tanpa menelaah relasi semantik antarkata di dalam satu untaian kalimat. Implikasinya algoritma ini kerap menemui kendala saat harus mencerna makna gramatikal yang sangat bergantung pada struktur maupun susunan konteks. Kondisi linguistik semacam ini justru sangat lumrah ditemukan pada gaya bahasa informal di platform media sosial.

Sebaliknya IndoBERT dengan dukungan mekanisme *self attention* pada arsitektur *Transformer* sukses menjaring hubungan semantik antarkata lewat pendekatan dua arah secara simultan. Kapasitas komputasi ini memberi ruang bagi model untuk membedakan intensi dari sebuah leksikon identik namun berada pada konteks kalimat yang berlainan. Sebagai contoh terma "enak" berpotensi memuat muatan afeksi positif berupa apresiasi atau justru bermakna sarkasme sebagai bentuk kritik terselubung bergantung pada situasi kalimatnya. Tingkat pemahaman analitik inilah yang menyumbang kontribusi masif bagi IndoBERT dalam memilah opini bernuansa rumit dengan tingkat ketepatan yang luar biasa [11].

Indikator paling krusial dari supremasi IndoBERT tampak teramat jelas pada performa klasifikasi sentimen netral di mana metrik *recall* menyentuh angka 0,97 berbanding SVM yang hanya berada di level 0,71. Opini bersentimen netral pada dasarnya memiliki irisan karakter leksikal yang sering tumpang tindih dengan kelas polaritas lainnya. Contoh nyatanya adalah untaian kalimat informatif yang memuat kosakata serupa dengan opini berlabel negatif. IndoBERT terbukti sanggup mengisolasi kedua kelas tersebut lewat penalaran konteks secara holistik, sementara SVM sangat rentan terkecoh oleh sekadar kemiripan wujud teks tunggal.

Walaupun kalah telak dari segi akurasi, metode SVM tetap menawarkan keunggulan pragmatis yang pantas diperhitungkan. Durasi pembelajaran SVM terbukti jauh lebih ringkas serta terbebas dari keharusan menggunakan unit pemrosesan grafis berspesifikasi tinggi. Kondisi ini menempatkan SVM sebagai alternatif solusi paling logis apabila peneliti dihadapkan pada keterbatasan infrastruktur perangkat keras komputasi. Konklusi empiris ini sangat sejalan dengan penelitian Anas Qolbu dan Fitriyati [4] yang turut mengafirmasi relevansi SVM sebagai opsi pemodelan efisien pada berbagai skenario penyusunan dataset teks.

KESIMPULAN

Hasil studi menyimpulkan bahwa IndoBERT (akurasi 96,52%) terbukti lebih efektif daripada SVM (86,87%) dalam mengklasifikasikan sentimen komentar YouTube mengenai program Makan Bergizi Gratis. Arsitektur *transformer* pada IndoBERT memungkinkannya menangkap konteks bahasa secara mendalam, sehingga sanggup mengatasi kelemahan SVM saat mengenali kelas sentimen netral, meskipun SVM memiliki keunggulan dari segi efisiensi komputasi. Untuk penelitian selanjutnya, disarankan adanya penambahan volume dataset, optimasi *hyperparameter*, serta komparasi dengan model lain seperti *Random Forest* dan Regresi Logistik. Ke depannya, arsitektur ini sangat berpeluang diimplementasikan sebagai sistem pemonitor opini publik otomatis guna membantu pemerintah mengevaluasi kebijakan secara presisi.

DAFTAR PUSTAKA

[1] D. A.-T. Mukarom, G. P. Pangestu, M. F. Margadipraja, and H. M. Winata, "Analisis Sentimen Berbasis Aspek terhadap Program Makan Bergizi Gratis (MBG) di Twitter/X dengan Pendekatan Efektivitas Duncan," *Prosiding Seminar Nasional Informatika*, 2026.

[2] W. Rayyana Muntaza, "Analisis Sentimen Tanggapan Masyarakat terhadap Program Makan Bergizi Gratis pada Platform YouTube Menggunakan SVM," *Jurnal Informatika dan Sistem Informasi*, vol. 5, 2026.

[3] E. Triningsih, M. Afdal, I. Permana, and N. E. Rozanda, "Analisis Sentimen terhadap Program Makan Bergizi Gratis Menggunakan Algoritma Machine Learning pada Sosial Media X," *Technology and Science (BITS)*, vol. 6, no. 4, pp. 2240–2250, 2025, doi: 10.47065/bits.v6i4.6534.

[4] A. Anas Qolbu and N. Fitriyati, "Performa Naïve Bayes, SVM, dan IndoBERT pada Analisis Sentimen Twitter IndiHome dengan Strategi Penanganan Data Tidak Seimbang," *Fourier*, vol. 14, no. 1, pp. 29–44, 2025, doi: 10.14421/fourier.2025.141.29-44.

[5] E. Setiadi and D. Arbian Sulistyo, "Analisis Sentimen Kebijakan Makan Bergizi Gratis Menggunakan IndoBERT dan Machine Learning," *Jurnal Komputasi dan Teknologi Informasi*, 2025. [Online]. Available: https://github.com/rikk-se/Dataset_Program-Makan.

[6] M. F. Kono, I. N. Fajri, and Y. Pristyanto, "Public Sentiment Analysis on Corruption Issues in

Indonesia Using IndoBERT Fine-Tuning, Logistic Regression, and Linear SVM," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 5, 2025. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC.

[7] G. P. Pangestu and A. Maulana, "Implementasi Confidence Score pada Klasifikasi Sentimen Berbasis IndoBERT untuk Analisis Kebijakan Publik," *Jurnal Ilmu Komputer dan Informasi*, 2026.

[8] Wardianto, M. J. Putra, and M. Rohman, "Analisis Sentimen Publik Program Makan Bergizi Gratis Platform Instagram dengan Algoritma SVM," *SMARTICS Journal*, vol. 11, no. 1, pp. 14–20, 2025, doi: 10.21067/smartics.v11i1.11852.

[9] S. Amanda Salsabila, B. Priyatna, and A. Hananto, "Komparasi Kinerja Model Naive Bayes, SVM, dan BERT dalam Klasifikasi Teks," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 2, pp. 42–47, 2025, doi: 10.55123.

[10] M. P. Firdaus and D. Trisnawarman, "Analisis Sentimen Publik terhadap Program Tabungan Perumahan Rakyat Menggunakan Model IndoBERT Lite pada Komentar YouTube," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 1, pp. 359–368, 2025, doi: 10.57152/malcom.v5i1.1744.

[11] V. E. Sidauruk and Herowati, "Indobert-Based Sentiment Analysis of Political Discourse on Platform X: The Case of Prabowo-Gibran Administration," *Journal of Applied Informatics and Computing (JAIC)*, vol. 10, no. 1, 2026. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC.

[12] V. E. Sidauruk, "Indobert-Based Sentiment Analysis of Political Discourse on Platform X," *Journal of Applied Informatics and Computing (JAIC)*, vol. 10, no. 1, 2026.

[13] S. F. Huwaida, R. Kusumawati, and B. Isnaini, "Analisis Sentimen Komentar YouTube terhadap Pemindahan Ibu Kota Negara Menggunakan Metode Naïve Bayes," *Jambura Journal of Informatics*, vol. 6, no. 1, pp. 26–39, 2024, doi: 10.37905/jji.v6i1.24718.

[14] C. Setiawan, B. Rahayudi, and D. E. Ratnawati, "Analisis Sentimen Pengguna X (Twitter) terhadap Program Makan Bergizi Gratis Menggunakan IndoBERTweet dan BERTopic," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK)*, vol. 10, no. 4, 2026. [Online]. Available: http://j-ptiik.ub.ac.id.

[15] A. N. Syafia, M. F. Hidayattullah, and W. Suteddy, "Studi Komparasi Algoritma SVM dan Random Forest pada Analisis Sentimen Komentar YouTube BTS," *Jurnal Informatika*, vol. 8, no. 3, 2023.