



Topic Modeling and Sentiment Analysis of TikTok Comments on the Free Nutritious Meal Program Using BERTopic and IndoBERT Model

Pemodelan Topik dan Analisis Sentimen Komentar TikTok pada Program Makanan Bergizi Gratis Menggunakan BERTopic dan IndoBERT

Owen Thomas Kayadu^{1*}, Josua Josen A Limbong², Andreas Leonardo Sumendap³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Papua, Indonesia

E-Mail: ¹owenkaydu178@gmail.com, ²jja.limbong@unipa.ac.id, ³al.sumendap@unipa.ac.id

Received May 13th 2026; Revised Jun 08th 2026; Accepted Jul 17th 2026; Available Online Aug 01st 2026

Corresponding Author: Owen Thomas Kayadu

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

TikTok has become one of the most widely used social media platforms for expressing public opinions on various government policies, including the Free Nutritious Meal (MBG) Program. This study aims to analyze discussion topics and public sentiment toward the MBG Program based on TikTok comments using BERTopic and IndoBERT. The research process includes data collection, preprocessing, topic modeling, and sentiment classification. The results show that BERTopic successfully identified eight main topics related to nutrition, education, employment, economic conditions, budget transparency, and public support for the government. Topic 5 was the most dominant topic, containing 5,422 comments. Sentiment analysis using IndoBERT revealed that neutral sentiment was the most dominant category with 3,614 comments (43.29%), followed by positive sentiment with 2,611 comments (31.27%) and negative sentiment with 2,124 comments (25.44%). In addition, Topic 4 appeared most frequently across all sentiment categories. The IndoBERT model achieved an accuracy of 82.68%, indicating good performance in understanding the context of Indonesian-language comments. These findings demonstrate that the combination of BERTopic and IndoBERT can provide a comprehensive understanding of public discussions and sentiment regarding the MBG Program on TikTok.

Keywords: BERTopic, IndoBERT, Sentiment Analysis, Topic Modeling, TikTok Comments

Abstrak

TikTok menjadi salah satu media sosial yang banyak digunakan masyarakat untuk menyampaikan opini terhadap berbagai kebijakan publik, termasuk Program Makan Bergizi Gratis (MBG). Penelitian ini bertujuan untuk menganalisis topik pembahasan dan sentimen masyarakat terhadap Program MBG berdasarkan komentar TikTok menggunakan metode BERTopic dan IndoBERT. Tahapan penelitian meliputi pengumpulan data, preprocessing, pemodelan topik, dan klasifikasi sentimen. Hasil penelitian menunjukkan bahwa BERTopic berhasil mengidentifikasi delapan topik utama yang berkaitan dengan gizi, pendidikan, lapangan kerja, kondisi ekonomi, transparansi anggaran, dan dukungan terhadap pemerintah. Topik 5 menjadi topik yang paling dominan dengan jumlah 5.422 komentar. Analisis sentimen menggunakan IndoBERT menunjukkan bahwa sentimen netral mendominasi dengan 3.614 komentar (43,29%), diikuti oleh sentimen positif sebanyak 2.611 komentar (31,27%) dan sentimen negatif sebanyak 2.124 komentar (25,44%). Selain itu, Topik 4 menjadi topik yang paling banyak muncul pada ketiga kategori sentimen. Model IndoBERT memperoleh akurasi sebesar 82,68%, yang menunjukkan performa yang baik dalam memahami konteks komentar berbahasa Indonesia. Hasil penelitian menunjukkan bahwa kombinasi BERTopic dan IndoBERT mampu memberikan analisis yang komprehensif terhadap topik pembahasan dan opini masyarakat mengenai Program MBG di TikTok.

Kata Kunci: Analisis Sentimen, BERTopic, IndoBERT, Pemodelan Topik, TikTok

1. PENDAHULUAN

Perkembangan teknologi digital dan media sosial telah mengubah cara masyarakat menyampaikan dan membentuk opini publik. Salah satu platform yang banyak digunakan adalah TikTok, yang tidak hanya berfungsi sebagai media hiburan tetapi juga sebagai ruang diskusi berbagai isu sosial dan kebijakan publik



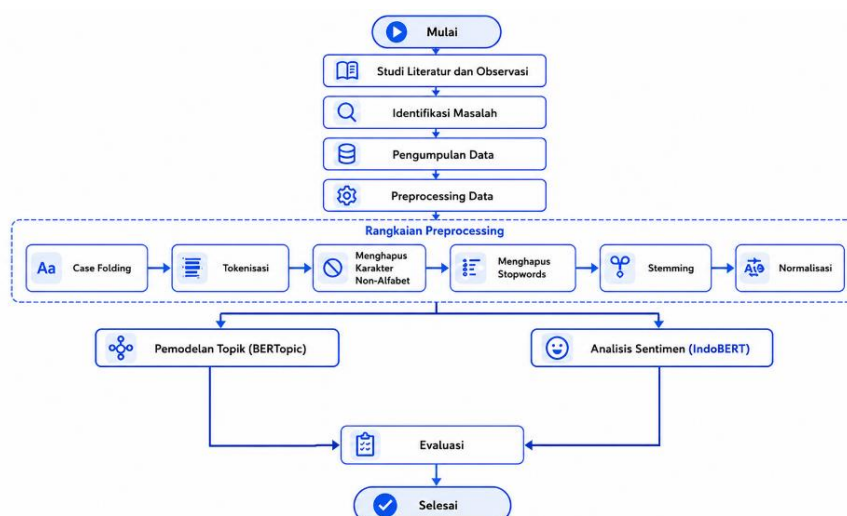
melalui fitur komentar. Berdasarkan laporan DataReportal tahun 2025, Indonesia memiliki sekitar 108 juta pengguna TikTok berusia 18 tahun ke atas, sehingga platform ini menjadi sumber data yang potensial untuk mengkaji opini masyarakat terhadap kebijakan pemerintah [1]. Salah satu kebijakan yang banyak diperbincangkan di media sosial adalah Program Makan Bergizi Gratis (MBG), yang bertujuan meningkatkan kualitas gizi dan sumber daya manusia Indonesia. Program ini memperoleh beragam respons dari masyarakat, mulai dari dukungan terhadap manfaatnya hingga kritik terkait anggaran dan implementasinya. Berbagai tanggapan tersebut banyak ditemukan di kolom komentar TikTok, yang merepresentasikan persepsi publik terhadap program tersebut [2]. Analisis komentar media sosial menghadapi berbagai tantangan karena data yang dihasilkan bersifat masif, singkat, tidak baku, dan tidak terstruktur. Karakteristik tersebut menyulitkan proses analisis secara manual, terutama dalam mengidentifikasi pola pembahasan dan kecenderungan opini masyarakat. Oleh karena itu, pendekatan *Text Mining* dan *Natural Language Processing* (NLP) diperlukan untuk mengekstraksi informasi secara efektif dari data komentar di media sosial [3].

Penelitian sebelumnya telah memanfaatkan metode seperti *Latent Dirichlet Allocation* (LDA) untuk pemodelan topik serta Naïve Bayes dan Support Vector Machine (SVM) untuk analisis sentimen [4]. Meskipun mampu memberikan gambaran umum mengenai opini publik, metode tersebut masih memiliki keterbatasan dalam memahami konteks semantik yang kompleks. Selain itu, penelitian terkait MBG masih didominasi oleh analisis sentimen tanpa pemetaan topik yang komprehensif dan lebih banyak menggunakan data dari platform X (Twitter), sehingga kajian berbasis komentar TikTok masih terbatas [5].

Berdasarkan kondisi tersebut, penelitian ini mengusulkan penggunaan BERTopic untuk pemodelan topik dan IndoBERT untuk analisis sentimen pada komentar TikTok terkait Program MBG. Kombinasi kedua metode ini diharapkan mampu menghasilkan identifikasi topik dan klasifikasi sentimen yang lebih kontekstual dibandingkan dengan pendekatan konvensional. Penelitian ini bertujuan mengungkap topik utama serta kecenderungan sentimen masyarakat terhadap Program MBG, sekaligus memberikan kontribusi metodologis dalam pemanfaatan NLP untuk analisis opini publik di media sosial [6].

2. METODE PENELITIAN

Melalui teknik *text mining*, studi ini menganalisis diskursus publik dan polaritas sentimen terhadap Program Makan Bergizi Gratis (MBG) di platform TikTok. Integrasi model BERTopic digunakan untuk memetakan kelompok topik utama, sementara IndoBERT diimplementasikan guna mendukung representasi teks serta analisis sentimen [7]. Pengumpulan data dilaksanakan pada 30 April 2026 dengan menyasar komentar pada video TikTok bertema MBG. Pengambilan sampel data dilakukan secara otomatis menggunakan layanan *TikTok Scraping API* dari API.com untuk menjamin efisiensi dan akurasi data yang diperoleh. Video yang digunakan berjumlah tiga dengan judul atau keterangan “Tanggapan Masyarakat Terhadap Program Makan Gratis Bergizi atau MBG”. Pemilihan video dilakukan karena isinya sesuai dengan topik penelitian dan memiliki tingkat interaksi yang cukup tinggi di TikTok. Dari proses pengumpulan data tersebut, diperoleh 8.197 komentar yang kemudian disimpan dalam bentuk dataset untuk dianalisis lebih lanjut [8]. Sebelum masuk ke tahap analisis, data komentar terlebih dahulu diproses. Proses ini diarahkan untuk membersihkan data teks guna menjamin kerapian serta kemudahan integrasi dengan model. Transformasi data dilakukan melalui serangkaian teknik, yaitu penghapusan karakter non-alfabet, *case folding*, konversi kata tidak baku, tokenisasi, eliminasi stopwords, serta *stemming*. Tahapan ini diperlukan karena komentar di TikTok biasanya ditulis dengan bahasa yang tidak baku, singkatan, simbol, serta bentuk penulisan yang beragam [9]. Secara umum metodologi penelitian ditunjukkan pada Gambar 1.

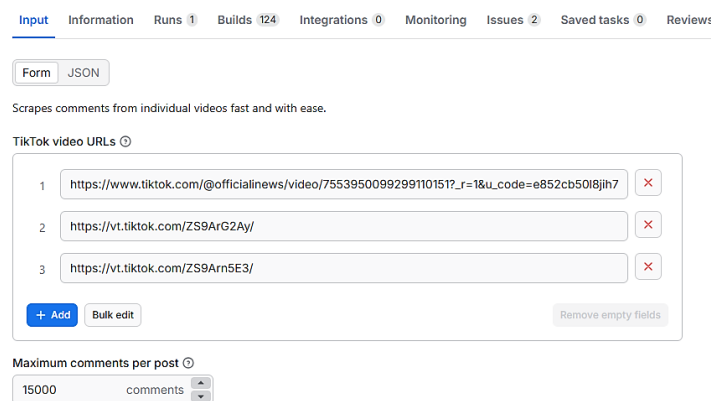


Gambar 1. Alur Penelitian

Setelah data selesai dibersihkan, komentar diubah ke dalam bentuk representasi teks menggunakan model BERT agar makna dalam kalimat dapat dipahami berdasarkan konteksnya. Hasil representasi tersebut kemudian digunakan dalam metode BERTopic untuk mengelompokkan komentar ke dalam beberapa topik berdasarkan kemiripan makna. Metodologi ini menggunakan BERTopic yang bekerja dengan *embedding* BERT, reduksi dimensi menggunakan UMAP, dan clustering menggunakan HDBSCAN. Topik yang teridentifikasi disarikan menjadi kata kunci utama untuk memudahkan analisis. Sejalan dengan itu, analisis sentimen dijalankan untuk mengkategorikan opini pengguna menjadi positif, netral, atau negatif terhadap Program MBG. Data hasil olahan kemudian dipetakan ke dalam bentuk visual seperti tabel, grafik, *word cloud*, hingga *confusion matrix*. Evaluasi terhadap kinerja model indoBERT menjadi tahap krusial untuk memastikan bahwa klasifikasi sentimen yang dihasilkan mampu mencerminkan realitas opini masyarakat secara akurat [10].

2.1. Pengumpulan Data

Proses akuisisi data dilakukan dengan menggunakan instrumen *TikTok Comments Scraper*. Data yang dihimpun terdiri dari sekumpulan komentar yang diekstraksi dari sejumlah unggahan video TikTok yang memiliki relevansi dengan fokus penelitian ini. Data diambil dengan cara memasukkan URL video TikTok ke dalam sistem scraper. Dalam gambar terlihat ada dua URL video yang dipakai sebagai sumber data. Di samping itu, jumlah komentar maksimum yang diambil dari setiap video dibatasi hingga 1500 komentar untuk menjamin data yang cukup representative [11].



Gambar 2. Pengumpulan Data

Pengumpulan data yang dilakukan pada 30 april 2026 dan menghasilkan dataset sebanyak 8.197 komentar. Data yang terkumpul meliputi teks komentar asli dari pengguna serta elemen metadata dan informasi interaksi terkait, yang secara keseluruhan menjadi basis data bagi tahapan analisis selanjutnya. Data itu selanjutnya dipakai sebagai bahan utama untuk analisis topik dan sentimen pada tahap berikutnya

2.2. Hapus Data Duplikat

Penghapusan data berulang bertujuan untuk memastikan bahwa setiap unit analisis dalam dataset bersifat unik. Dengan mengeliminasi duplikasi, peneliti dapat menghindari dominasi data tertentu yang dapat mendistorsi hasil analisis sentimen, sehingga validitas dan reliabilitas kesimpulan penelitian tetap terjaga [12]. Berdasarkan hasil `df.info()`, dataset memiliki 8.471 komentar pada kolom text tanpa nilai kosong setelah di bersihkan data menjadi 8447. Tahap ini penting dalam analisis sentimen agar komentar yang dianalisis tidak berulang, sehingga hasil klasifikasi sentimen menjadi lebih bersih dan tidak bias.

2.3. Cleaning Data

Tahap *cleaning* bertujuan untuk mentransformasi data mentah menjadi informasi yang lebih bersih dan konsisten. Hal ini dicapai dengan menghapus *noise* serta elemen teks yang tidak esensial, sehingga data memiliki format yang seragam dan siap digunakan dalam proses pemodelan topik maupun klasifikasi sentimen.

Tabel 1. Cleaning Data

Text	Cleaning
program yang bikin negri bangkrut,1009,20,2025	program yang bikin negri bangkrut
Terima kasih pak Prabowo, kami UMKM sdh merasakan manfaatnya	Terima kasih pak Prabowo kami UMKM sdh merasakan manfaatnya
klu sgitu byk ngapa sppi ngk langsung diangkat	klu sgitu byk ngapa sppi ngk langsung diangkat
Baguss pak dana triliunan begitu langsung di rampas	Baguss pak dana triliunan begitu langsung di rampas
Dimata pembenci program ini pasti ngk ada bagusny	Dimata pembenci program ini pasti ngk ada bagusny

Berdasarkan Tabel 1, terlihat bahwa kolom *text* berisi data komentar asli yang masih mengandung penulisan tidak baku, singkatan, serta tanda baca yang kurang terstruktur. Setelah dilakukan proses *cleaning*, pada kolom *cleaning* data telah diperbaiki menjadi lebih rapi, seperti penghilangan karakter yang tidak perlu, penyederhanaan teks, serta penyesuaian penulisan agar lebih konsisten. Proses ini bertujuan untuk mempermudah tahap analisis selanjutnya, seperti tokenisasi dan pemodelan topik, sehingga hasil yang diperoleh menjadi lebih akurat.

2.4. Case Folding

Tujuan utama dari *case folding* adalah untuk menstandarisasi format huruf menjadi *lowercase* guna menghindari ambiguitas makna yang disebabkan oleh perbedaan huruf besar dan kecil. Dengan menyeragamkan format teks, data menjadi lebih konsisten, yang pada akhirnya memudahkan model dalam mengenali pola kata secara efektif [13].

Tabel 2. Case Folding

<i>Text</i>	<i>Cleaning</i>	<i>Case Folding</i>
program yang bikin negri bangkrut,1009,20,2025	program yang bikin negri bangkrut	program yang bikin negri bangkrut
Terima kasih pak Prabowo, kami UMKM sdh merasakan manfaatnya	Terima kasih pak Prabowo kami UMKM sdh merasakan manfaatnya	terima kasih pak prabowo kami umkm sdh merasakan manfaatnya
klu sgitu byk ngapa sppi ngk langsung diangkat	klu sgitu byk ngapa sppi ngk langsung diangkat	klu sgitu byk ngapa sppi ngk langsung diangkat
Baguss pak dana triliunan begitu langsung di rampas	Baguss pak dana triliunan begitu langsung di rampas	baguss pak dana triliunan begitu langsung di rampas
Dimata pembenci program ini pasti ngk ada bagusnya	Dimata pembenci program ini pasti ngk ada bagusnya	dimata pembenci program ini pasti ngk ada bagusnya

Tabel 2 menunjukkan tahapan awal pengolahan data, yaitu *text*, *cleaning*, *case folding*, dan normalisasi. Setelah melalui tahap *cleaning*, dilakukan *case folding* untuk mengubah seluruh teks menjadi huruf kecil, seperti konversi dari “Terima kasih Pak Prabowo” menjadi “terima kasih pak prabowo”. Langkah berikutnya adalah normalisasi bahasa, di mana kata-kata tidak baku atau singkatan diperbaiki menjadi bentuk standarnya demi mendukung efektivitas pemrosesan model.

2.5. Normalisasi Kata

Normalisasi diarahkan pada perbaikan istilah-istilah tidak baku yang sering ditemukan dalam media sosial guna mencapai format teks yang homogen. Melalui pengubahan bentuk informal menjadi kata standar, proses ini memastikan bahwa informasi yang terkandung dalam teks lebih mudah diinterpretasikan oleh algoritma tanpa kehilangan konteks aslinya.

Tabel 3. Normalisasi Kata

<i>Text</i>	<i>Cleaning</i>	<i>Case Folding</i>	Normalisasi kata
program yang bikin negri bangkrut,1009,20,2025	program yang bikin negri bangkrut	program yang bikin negri bangkrut	program yang bikin negeri bangkrut
Terima kasih pak Prabowo, kami UMKM sdh merasakan manfaatnya	Terima kasih pak Prabowo kami UMKM sdh merasakan manfaatnya	terima kasih pak prabowo kami umkm sdh merasakan manfaatnya	terima kasih pak prabowo kami umkm sudah merasakan manfaatnya
klu sgitu byk ngapa sppi ngk langsung diangkat	klu sgitu byk ngapa sppi ngk langsung diangkat	klu sgitu byk ngapa sppi ngk langsung diangkat	kalu sgitu banyak ngapa sppi tidak langsung diangkat
Baguss pak dana triliunan begitu langsung di rampas	Baguss pak dana triliunan begitu langsung di rampas	baguss pak dana triliunan begitu langsung di rampas	bagus pak dana triliunan begitu langsung di rampas
Dimata pembenci program ini pasti ngk ada bagusnya	Dimata pembenci program ini pasti ngk ada bagusnya	dimata pembenci program ini pasti ngk ada bagusnya	dimata pembenci program ini pasti tidak ada bagusnya

Pada gambar 3, tahap normalisasi dilakukan setelah *case folding* untuk memperbaiki kata singkatan atau tidak baku menjadi bentuk yang lebih jelas. Contohnya, “sdh” menjadi “sudah”, “bnyk” menjadi “banyak”, dan “ngk” menjadi “tidak”. Proses ini membuat teks lebih rapi dan siap digunakan untuk analisis sentimen maupun pemodelan topik.

2.6. Tokenisasi

Tokenisasi bertujuan untuk menyederhanakan struktur teks dengan membaginya menjadi potongan-potongan kata agar lebih mudah diolah oleh model klasifikasi. Melalui unit-unit token tersebut, sistem dapat

melakukan kalkulasi terhadap kecenderungan emosi yang disampaikan pengguna TikTok mengenai isu penelitian, yang kemudian dikategorikan ke dalam tiga kelompok sentimen utama.

Tabel 4. Tokenisasi

<i>Text</i>	<i>Cleaning</i>	<i>Case Folding</i>	<i>Normalisasi Kata</i>	<i>Tokenisasi</i>
program yang bikin negri bangkrut,1009,20,2025	program yang bikin negri bangkrut	program yang bikin negri bangkrut	program yang bikin negri bangkrut	[program, yang, bikin, negeri, bangkrut]
Terima kasih pak Prabowo, kami UMKM sdh merasakan manfaatnya	Terima kasih pak Prabowo kami UMKM sdh merasakan manfaatnya	terima kasih pak prabowo kami umkm sdh merasakan manfaatnya	terima kasih pak prabowo kami umkm sudah merasakan manfaatnya	[terima, kasih, pak, prabowo, kami, umkm, sudah, merasakan, manfaatnya]
klu sgitu byk ngapa sppi ngk langsung diangkat	klu sgitu byk ngapa sppi ngk langsung diangkat	klu sgitu byk ngapa sppi ngk langsung diangkat	kalau sgitu banyak ngapa sppi tidak langsung diangkat	[kalau, sgitu, banyak, ngapa, sppi, tidak, langsung, diangkat]
Baguss pak dana triliunan begitu langsung di rampas	Baguss pak dana triliunan begitu langsung di rampas	baguss pak dana triliunan begitu langsung di rampas	bagus pak dana triliunan begitu langsung di rampas	[bagus, pak, dana, triliunan, begitu, langsung, di, rampas]
Dimata pembenci program ini pasti ngk ada bagusnya	Dimata pembenci program ini pasti ngk ada bagusnya	dimata pembenci program ini pasti ngk ada bagusnya	dimata pembenci program ini pasti tidak ada bagusnya	[dimata, pembenci, program, ini, pasti, tidak, ada, bagusnya]

Tabel 4 menunjukkan bahwa tahap tokenisasi bertujuan untuk memecah struktur kalimat ke dalam bentuk kata-kata individual yang terpisah satu sama lain. Contohnya, kalimat “program yang bikin negri bangkrut” diubah menjadi “[program, yang, bikin, negeri, bangkrut]”. Tahap ini membantu sistem mengenali setiap kata dalam komentar. Karena banyak komentar menggunakan bahasa tidak baku dan singkatan, proses normalisasi dan tokenisasi penting untuk meningkatkan hasil analisis sentimen.

2.7. Stopword Removal

Stopword removal merupakan tahapan eliminasi kata-kata umum yang dinilai tidak memiliki signifikansi semantik dalam analisis teks, seperti konjungsi atau preposisi (misalnya: "yang", "di", "dan"). Prosedur ini bertujuan untuk mengarahkan fokus sistem pada kata-kata kunci yang mengandung informasi esensial guna menentukan polaritas sentimen. Dalam konteks analisis sentimen, reduksi kata-kata non-informatif ini sangat krusial untuk meningkatkan efektivitas dan akurasi model klasifikasi [14].

Tabel 5. Stop Removal

<i>Text</i>	<i>Cleaning</i>	<i>Case Folding</i>	<i>Normalisasi Kata</i>	<i>Tokenisasi</i>	<i>Stop Removal</i>
program yang bikin negri bangkrut,1009,20,2025	program yang bikin negri bangkrut	program yang bikin negri bangkrut	program yang bikin negri bangkrut	[program, yang, bikin, negeri, bangkrut]	[program, yang, bikin, negeri, bangkrut]
Terima kasih pak Prabowo, kami UMKM sdh merasakan manfaatnya	Terima kasih pak Prabowo kami UMKM sdh merasakan manfaatnya	terima kasih pak prabowo kami umkm sdh merasakan manfaatnya	terima kasih pak prabowo kami umkm sudah merasakan manfaatnya	[terima, kasih, pak, prabowo, kami, umkm, sudah, merasakan, manfaatnya]	[terima, kasih, pak, prabowo, kami, umkm, sudah, merasakan, manfaatnya]
klu sgitu byk ngapa sppi ngk langsung diangkat	klu sgitu byk ngapa sppi ngk langsung diangkat	klu sgitu byk ngapa sppi ngk langsung diangkat	kalau sgitu banyak ngapa sppi tidak langsung diangkat	[kalau, sgitu, banyak, ngapa, sppi, tidak, langsung, diangkat]	[kalau, sgitu, banyak, ngapa, sppi, tidak, langsung, diangkat]
Baguss pak dana triliunan begitu langsung di rampas	Baguss pak dana triliunan begitu langsung di rampas	baguss pak dana triliunan begitu langsung di rampas	bagus pak dana triliunan begitu langsung di rampas	[bagus, pak, dana, triliunan, begitu, langsung, di, rampas]	[bagus, pak, dana, triliunan, begitu, langsung, di, rampas]

Text	Cleaning	Case Folding	Normalisasi Kata	Tokenisasi	Stop Removal
Dimata pembenci program ini pasti ngk ada bagusnya	Dimata pembenci program ini pasti ngk ada bagusnya	dimata pembenci program ini pasti ngk ada bagusnya	dimata pembenci program ini pasti tidak ada bagusnya	[dimata, pembenci, program, ini, pasti, tidak, ada, bagusnya]	[dimata, pembenci, program, ini, pasti, tidak, ada, bagusnya]

Tabel 5 menunjukkan alur prapemrosesan data, mulai dari pembersihan teks hingga *stopword removal*. Tujuannya adalah untuk menghilangkan *noise* dan menyederhanakan istilah. Contohnya, kalimat "program yang bikin negeri bangkrut" berubah menjadi [program, bikin, negeri, bangkrut]. Dengan membuang kata-kata yang tidak informatif, model dapat lebih fokus dalam menentukan sentimen setiap komentar.

2.8. Stemming

Stemming adalah proses mengubah kata berimbuhan menjadi kata dasar. Hal ini dilakukan untuk menyederhanakan variasi kata agar sistem dapat menganalisis teks dengan lebih mudah dan konsisten. Tujuan stemming adalah menyederhanakan kata sehingga kata yang memiliki makna sama dianggap seragam [15].

Tabel 6. Stemming Data

Text	Cleaning	Case Folding	Normalisasi Kata	Tokenisasi	Stopword Removal	Stemming Data
program yang bikin negeri bangkrut,1009,20,2025	program yang bikin negeri bangkrut	program yang bikin negeri bangkrut	program yang bikin negeri bangkrut	[program, yang, bikin, negeri, bangkrut]	[program, yang, bikin, negeri, bangkrut]	program yang bikin negeri bangkrut
Terima kasih pak Prabowo, kami UMKM sdh merasakan manfaatnya	Terima kasih pak Prabowo kami UMKM sdh merasakan manfaatnya	terima kasih pak prabowo kami umkm sdh merasakan manfaatnya	terima kasih pak prabowo kami umkm sudah merasakan manfaatnya	[terima, kasih, pak, prabowo, kami, umkm, sudah, merasakan, manfaatnya]	[terima, kasih, pak, prabowo, kami, umkm, sudah, merasakan, manfaatnya]	terima kasih pak prabowo kami umkm sudah merasakan manfaatnya
klu sgitu byk ngapa sppi ngk langsung diangkat	klu sgitu byk ngapa sppi ngk langsung diangkat	klu sgitu byk ngapa sppi ngk langsung diangkat	kalau sgitu banyak ngapa sppi tidak langsung diangkat	[kalau, sgitu, banyak, ngapa, sppi, tidak, langsung, diangkat]	[kalau, sgitu, banyak, ngapa, sppi, tidak, langsung, diangkat]	kalau sgitu banyak ngapa sppi tidak langsung diangkat
Baguss pak dana triliunan begitu langsung di rampas	Baguss pak dana triliunan begitu langsung di rampas	baguss pak dana triliunan begitu langsung di rampas	bagus pak dana triliunan begitu langsung di rampas	[bagus, pak, dana, triliunan, begitu, langsung, di, rampas]	[bagus, pak, dana, triliunan, begitu, langsung, di, rampas]	bagus pak dana triliunan begitu langsung di rampas
Dimata pembenci program ini pasti ngk ada bagusnya	Dimata pembenci program ini pasti ngk ada bagusnya	dimata pembenci program ini pasti ngk ada bagusnya	dimata pembenci program ini pasti tidak ada bagusnya	[dimata, pembenci, program, ini, pasti, tidak, ada, bagusnya]	[dimata, pembenci, program, ini, pasti, tidak, ada, bagusnya]	dimata pembenci program ini pasti tidak ada bagusnya

Tabel 6 menunjukkan alur *preprocessing* lengkap, termasuk tahap *stemming* yang mengubah kata ke bentuk dasar. Contohnya, kata "pembenci" diringkas menjadi "benci" dan "merasakannya" menjadi "rasa". Dengan menyederhanakan struktur kata, sistem dapat lebih mudah mengenali inti informasi untuk mengklasifikasikan sentimen pada setiap komentar.

2.9 Pelabelan Data

Pelabelan sentimen pada penelitian ini dilakukan secara otomatis menggunakan model IndoBERT. Setiap komentar diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral, berdasarkan konteks kalimat pada komentar. Pemilihan IndoBERT dilakukan karena model ini telah dilatih menggunakan korpus bahasa Indonesia dalam jumlah besar sehingga mampu memahami struktur bahasa, penggunaan kata tidak baku, singkatan, serta konteks kalimat yang umum ditemukan pada komentar media sosial seperti TikTok.

Dibandingkan dengan metode *lexicon-based* seperti InSet yang mengandalkan daftar kata positif dan negatif, IndoBERT memiliki keunggulan dalam memahami makna berdasarkan konteks kalimat secara keseluruhan. Dengan demikian, komentar yang mengandung kata ambigu atau memiliki makna berbeda dalam konteks tertentu dapat diklasifikasikan dengan lebih akurat. Oleh karena itu, IndoBERT dipilih sebagai metode pelabelan sentimen karena mampu memberikan hasil klasifikasi yang lebih baik pada data komentar berbahasa Indonesia yang bersifat dinamis dan tidak terstruktur [10].

Tabel 7. Pelabelan Data

<i>Stemming Data</i>	<i>Score</i>	<i>Sentiment</i>
program yang bikin negeri bangkrut	0	Netral
terima kasih pak prabowo kami umkm sudah merasakan manfaatnya	0	Netral
kalau sgitu banyak ngapa sppi tidak langsung diangkat	2	Positif
bagus pak dana triliunan begitu langsung di rampas	0	Netral
dimata pembenci program ini pasti tidak ada bagusny	-1	Negatif

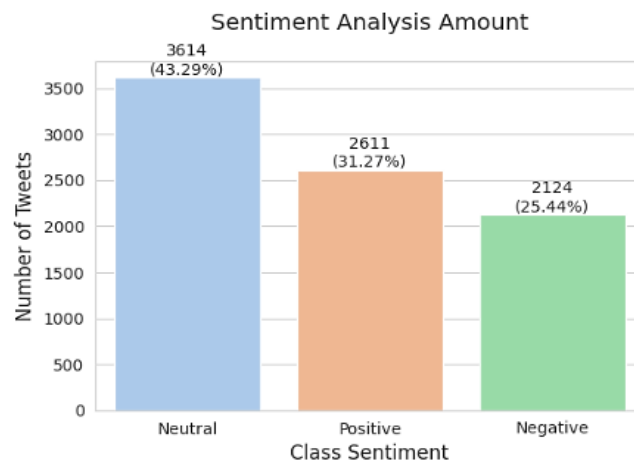
Berdasarkan tabel 7, komentar yang telah melalui *preprocessing* diberi label sentimen menggunakan IndoBERT, yaitu Negative untuk kritik, Positive untuk dukungan, dan Neutral untuk komentar yang tidak menunjukkan kecenderungan tertentu. Pelabelan ini membantu menganalisis opini masyarakat, sementara kombinasi IndoBERT dan BERTopic digunakan untuk memberikan pemahaman yang lebih komprehensif terhadap opini publik pada komentar TikTok.

3. HASIL DAN PEMBAHASAN

Melalui metode BERTopic, data komentar TikTok dianalisis hingga membentuk topik-topik utama dengan pemetaan sentimen positif, negatif, dan netral. Penyajian data dilakukan melalui tabel dan visualisasi untuk memperjelas temuan tanpa pengulangan informasi, sekaligus membuktikan efektivitas model dalam mengelompokkan data penelitian [16], [17].

3.1. Class Sentiment

Gambar 3 menunjukkan bahwa komentar netral menjadi kategori yang paling banyak muncul. Dari seluruh data yang dianalisis, terdapat 3.614 komentar netral atau 43,29%. Komentar positif berjumlah 2.611 komentar atau 31,27%, sedangkan komentar negatif berjumlah 2.124 komentar atau 25,44%. Hasil ini menunjukkan bahwa respons masyarakat terhadap Program MBG cenderung netral, meskipun komentar yang berisi dukungan dan kritik juga cukup banyak ditemukan [4].



Gambar 3. Class Sentiment

3.2. Wordcloud

Visualisasi *wordcloud* digunakan untuk menonjolkan kata-kata yang memiliki intensitas kemunculan tinggi dalam data teks. Semakin besar ukuran teks dalam *wordcloud*, semakin sering kata tersebut digunakan dalam komentar. Metode ini mempermudah peneliti untuk menangkap inti dari opini publik secara cepat dan melihat kecenderungan sentimen melalui kata-kata yang mendominasi data [18].

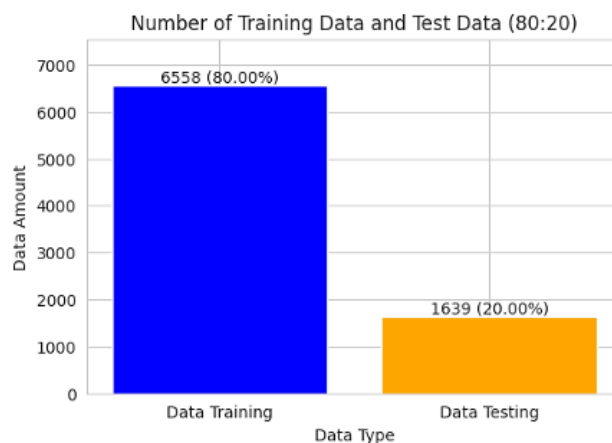
Gambar 4 merupakan beberapa hasil *wordcloud* pada setiap kategori sentimen dari komentar TikTok mengenai program MBG di Indonesia [19].



Gambar 4. Wordcloud Sentiment

3.3. Training Data

Dalam analisis sentimen, *data training* berperan dalam proses pembelajaran pola oleh model, sedangkan *data testing* berfungsi untuk menguji validitas hasil pelatihan. Perbandingan kedua data ini memungkinkan peneliti untuk mengetahui tingkat akurasi serta reliabilitas performa model [20].

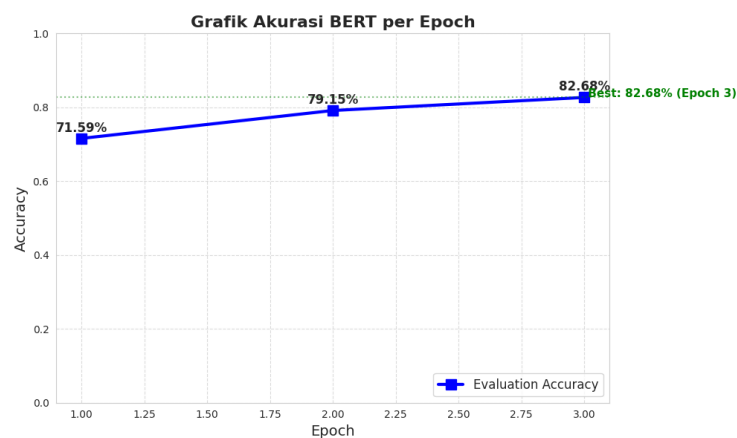


Gambar 5. Training Data

Berdasarkan gambar 5, data dibagi menjadi *data training* sebanyak 6.558 data (80%) dan *data testing* sebanyak 1.639 data (20%). Mayoritas data dialokasikan untuk pelatihan agar model dapat mempelajari pola secara optimal, sementara sisanya digunakan untuk mengevaluasi akurasi prediksi pada analisis sentimen [21].

3.4. Akurasi Model

Pembagian data dilakukan dengan proporsi 80% data latih (6.558) dan 20% data uji (1.639) untuk memastikan model dapat mempelajari pola komentar secara efektif sebelum dilakukan pengujian [22]. Kinerja model terbukti meningkat di setiap iterasi, dengan perolehan akurasi sebesar 71,59% pada *epoch* awal, 79,15% pada *epoch* kedua, hingga mencapai akurasi tertinggi sebesar 82,68% pada *epoch* terakhir [23].



Gambar 6. Akurasi Model

3.5. Classification Report

Model *indoBERT* menunjukkan performa yang solid dengan akurasi 83%. Detail *classification report* mencatat *f1-score* sebesar 0,81 untuk sentimen negatif dan 0,82 untuk netral, sementara sentimen positif

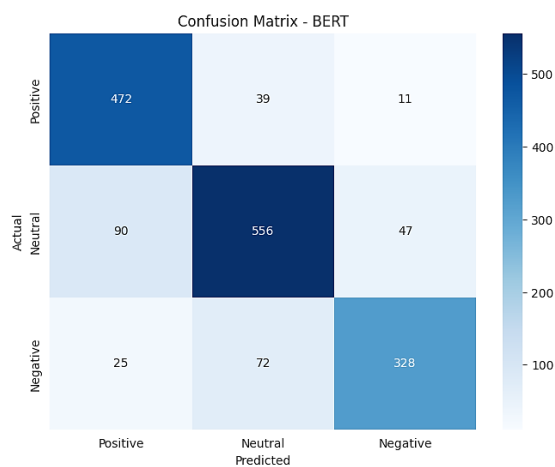
mencapai hasil tertinggi yakni 0,85. Angka-angka tersebut membuktikan kemampuan model dalam mengenali pola sentimen pada komentar TikTok secara akurat, khususnya pada opini-opini yang bersifat positif [24].

Tabel 8. Clasification Report

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
Negatif	0.85	0.77	0.81	425
Netral	0.83	0.80	0.82	693
Positif	0.80	0.90	0.85	522
<i>Accuracy</i>	-	-	0.83	1640
Macro avg	0.83	0.83	0.83	1640
Weighted avg	0.83	0.83	0.83	1640

3.6. Confusion Matrix

Confusion matrix berfungsi mengevaluasi efektivitas model dengan membandingkan nilai prediksi dan nilai aktual. Dengan metode ini, distribusi klasifikasi yang benar maupun salah pada setiap kelas sentimen dapat dipetakan secara jelas untuk menilai performa model secara menyeluruh [25].



Gambar 7. Confusion Matrix

Pengujian model *indoBERT* menghasilkan akurasi 83% dengan *f1-score* tertinggi pada kelas positif (0,85). Berdasarkan data *confusion matrix*, model berhasil memprediksi dengan benar 472 komentar positif, 556 netral, dan 328 negatif. Meskipun terdapat sejumlah *misclassification*, hasil ini membuktikan reliabilitas model *indoBERT* dalam mengelompokkan opini publik ke dalam kategori sentimen yang sesuai.

3.7. Distribusi Kata dan Distribusi Persentimen

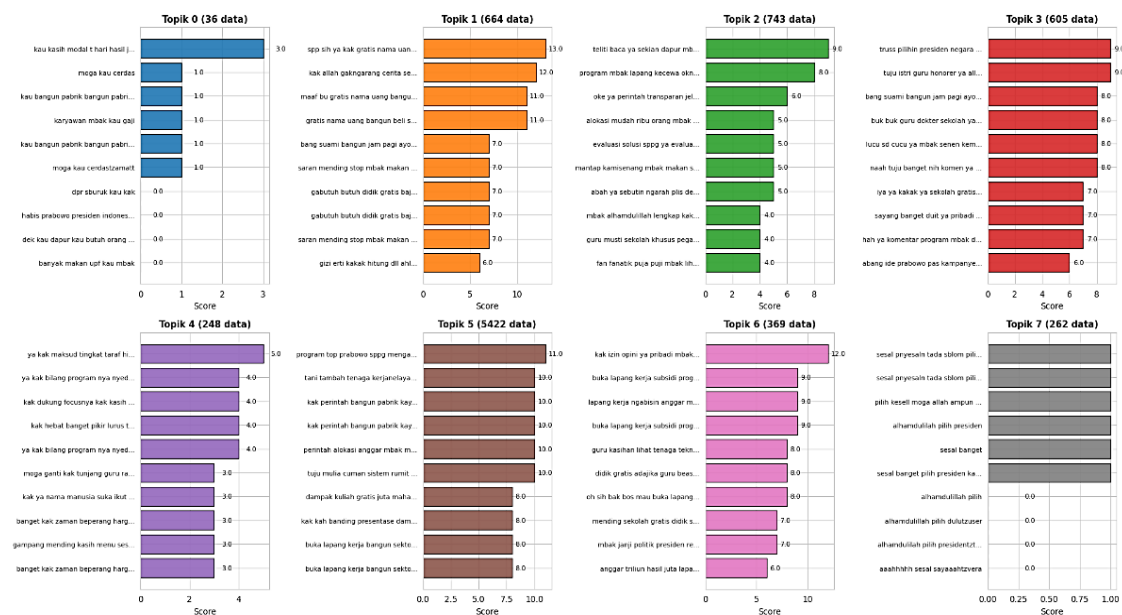
Berdasarkan hasil pemodelan topik menggunakan *BERTopic*, diperoleh delapan topik utama terkait program Makan Bergizi Gratis (MBG). Topik 0 membahas lapangan kerja, Topik 1 manfaat gizi dan kesehatan, Topik 2 transparansi anggaran dan korupsi, Topik 3 dukungan terhadap Presiden Prabowo dan program MBG, Topik 4 kondisi ekonomi masyarakat, Topik 5 peningkatan kesejahteraan, Topik 6 peran sekolah dan pendidikan, serta Topik 7 harapan dan respons positif masyarakat terhadap keberlanjutan program. Secara umum, diskusi masyarakat mencakup aspek gizi, ekonomi, ketenagakerjaan, pendidikan, transparansi anggaran, dan dukungan terhadap pemerintah [26]. Distribusi Kata dan Distribusi Persentimen ditunjukkan pada Gambar 8 dan 9.

3.9. Pembahasan

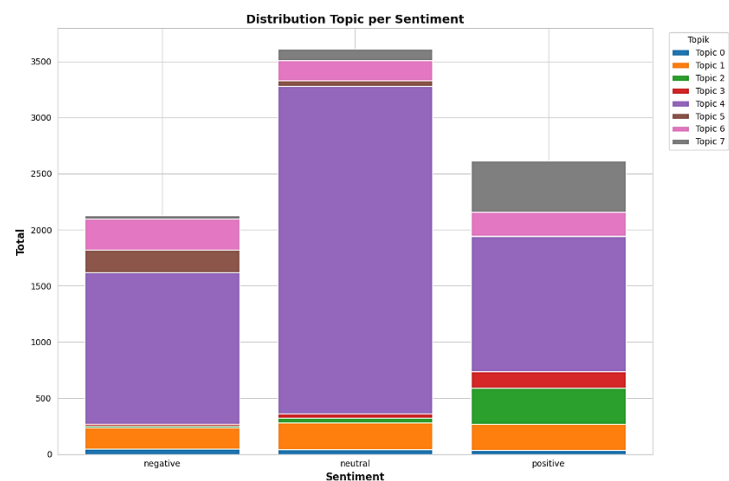
Hasil penelitian menunjukkan bahwa sentimen netral menjadi kategori yang paling dominan dibandingkan dengan sentimen positif dan negatif. Dominasi sentimen netral ini menunjukkan bahwa sebagian besar pengguna TikTok belum memberikan penilaian yang jelas terhadap Program Makan Bergizi Gratis (MBG). Kondisi tersebut diduga karena pengambilan data dilakukan pada saat program masih dalam tahap implementasi awal, sehingga masyarakat cenderung memberikan komentar berupa tanggapan, pertanyaan, atau opini yang bersifat informatif daripada menunjukkan dukungan maupun penolakan secara tegas. Selain itu, banyak komentar yang ditemukan mengandung makna ambigu atau tidak memiliki ekspresi emosi yang kuat. Dalam kondisi tersebut, model *IndoBERT* cenderung mengklasifikasikan komentar ke dalam kategori netral karena model melakukan analisis berdasarkan konteks keseluruhan kalimat, bukan hanya berdasarkan

kemunculan kata-kata tertentu. Dengan demikian, komentar yang bersifat deskriptif atau tidak menunjukkan kecenderungan sentimen yang jelas akan lebih banyak teridentifikasi sebagai sentimen netral.

Hasil pemodelan topik menggunakan BERTopic menunjukkan bahwa pembahasan masyarakat tidak hanya berfokus pada manfaat program MBG, tetapi juga mencakup isu lapangan kerja, pendidikan, kondisi ekonomi, serta transparansi pengelolaan anggaran. Keberadaan topik yang berkaitan dengan korupsi dan pengawasan dana menunjukkan bahwa masyarakat turut memperhatikan aspek akuntabilitas dalam pelaksanaan program tersebut. Sementara itu, topik yang berkaitan dengan dukungan terhadap pemerintah menunjukkan adanya respons positif dari sebagian masyarakat terhadap kebijakan yang dijalankan. Dari sisi performa model, IndoBERT memperoleh akurasi sebesar 83%, yang menunjukkan kemampuan yang baik dalam memahami konteks bahasa Indonesia pada komentar TikTok yang umumnya menggunakan bahasa tidak baku, singkatan, dan variasi penulisan. Hasil ini menunjukkan bahwa kombinasi BERTopic dan IndoBERT mampu memberikan analisis yang lebih komprehensif karena tidak hanya mengidentifikasi kecenderungan sentimen masyarakat, tetapi juga mengungkap isu-isu utama yang menjadi perhatian publik terkait Program Makan Bergizi Gratis di media sosial TikTok.



Gambar 8. Distribusi Kata



Gambar 9. Distribusi Persentimen

Gambar 9 menunjukkan distribusi topik berdasarkan sentimen negatif, netral, dan positif. Sentimen netral memiliki jumlah komentar terbanyak, disusul positif dan negatif. Pada ketiga kategori sentimen, Topik 4 tampak paling dominan, sehingga topik tersebut menjadi pembahasan utama dalam keseluruhan komentar.

4. KESIMPULAN

Melalui metode BERTopic dan BERT, analisis terhadap Program Makan Bergizi Gratis di TikTok berhasil memetakan topik dan distribusi sentimen pengguna. Hasilnya menunjukkan kecenderungan publik yang netral sebagai kategori dominan, yakni sebesar 43,29% atau sebanyak 3.614 komentar. Selanjutnya, sentimen positif berjumlah 2.611 komentar atau 31,27%, sedangkan sentimen negatif sebanyak 2.124 komentar atau 25,44%. Hal ini menunjukkan bahwa tanggapan masyarakat terhadap Program MBG cenderung netral, meskipun dukungan dan kritik juga cukup banyak muncul. Pada tahap evaluasi, data dibagi menjadi 6.557 data latih atau 79,99% dan 1.640 data uji atau 20,01%. Model BERT memperoleh akurasi terbaik sebesar 82,68% pada *epoch* ketiga, sedangkan hasil *classification report* menunjukkan akurasi sebesar 83%. Model berhasil mengklasifikasikan 472 komentar positif, 556 netral, dan 328 negatif berdasarkan pengujian *confusion matrix*. Temuan ini menunjukkan reliabilitas metode BERTopic dan BERT dalam menganalisis diskursus Program Makan Bergizi Gratis. Saran untuk penelitian berikutnya mencakup penambahan volume data dan komparasi metode guna mencapai tingkat akurasi yang lebih tinggi.

REFERENCES

- [1] R. Y. Go, H. Sarah, D. Gaspersz, and R. H. Kusumawardhana, "Analisis Sentimen Publik Program Makan Bergizi Gratis (MBG) di Youtube : Perbandingan Kinerja Algoritma Support Vector Machine (SVM) dan Random Forest," vol. 7, no. 1, pp. 24–31, 2026.
- [2] B. Ilyas and A. Sharifi, "International Journal of Disaster Risk Reduction A systematic review of social media-based sentiment analysis in disaster risk management," *Int. J. Disaster Risk Reduct.*, vol. 123, no. April, p. 105487, 2025, doi: 10.1016/j.ijdr.2025.105487.
- [3] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 4, p. 102048, Apr. 2024, doi: 10.1016/J.JKSUCI.2024.102048.
- [4] A. N. Hidayat, A. L. Ramdani, and L. Muthoharoh, "Aspect-Based Sentiment Analysis of Tourism Reviews Using Variants of the K-Nearest Neighbor Algorithm Analisis Sentimen Berbasis Aspek pada Ulasan Pariwisata Menggunakan Varian Algoritma K-Nearest Neighbor," vol. 6, no. April, pp. 534–545, 2026.
- [5] O. Salami and T. M. Fagbola, "Topic modelling and sentiment analysis for public opinion mining of the #BBNaija reality TV show: a critical analysis," *Soc. Netw. Anal. Min.*, vol. 15, no. 1, pp. 1–17, 2025, doi: 10.1007/s13278-025-01529-5.
- [6] E. Supriyadi and P. N. Makatita, "Sentiment Analysis of TikTok User Comments on QRIS Adoption in Indonesia Using IndoBERT," *Procedia Comput. Sci.*, vol. 269, pp. 121–130, 2025, doi: 10.1016/j.procs.2025.08.265.
- [7] K. F. Yuen, R. Barretta, X. Wang, J. Y. M. Nip, and B. Berthelie, "Social Media Sentiment Analysis," *Encycl. 2024, Vol. 4, Pages 1590-1598*, vol. 4, no. 4, pp. 1590–1598, Oct. 2024, doi: 10.3390/ENCYCLOPEDIA4040104.
- [8] U. K. Das *et al.*, "Enhancing sentiment analysis accuracy on social media comments using a tuned BERT model," *Discov. Comput.*, vol. 28, no. 1, 2025, doi: 10.1007/s10791-025-09599-x.
- [9] Z. Liu, T. Yang, W. Chen, J. Chen, Q. Li, and J. Zhang, "Sentiment analysis of social media comments based on multimodal attention fusion network," *Appl. Soft Comput.*, vol. 164, p. 112011, Oct. 2024, doi: 10.1016/J.ASOC.2024.112011.
- [10] M. Singh, A. K. Jakhar, and S. Pandey, "Sentiment analysis on the impact of coronavirus in social life using the BERT model," *Soc. Netw. Anal. Min.*, vol. 11, no. 1, pp. 1–11, 2021, doi: 10.1007/s13278-021-00737-z.
- [11] Z. Li and Z. Zou, "Punctuation and lexicon aid representation: A hybrid model for short text sentiment analysis on social media platform," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 3, p. 102010, Mar. 2024, doi: 10.1016/J.JKSUCI.2024.102010.
- [12] H. Gomez-Adorno, G. Bel-Enguix, G. Sierra, J. C. Barajas, and W. Álvarez, "Machine Learning and Deep Learning Sentiment Analysis Models: Case Study on the SENT-COVID Corpus of Tweets in Mexican Spanish," *Informatics*, vol. 11, no. 2, 2024, doi: 10.3390/informatics11020024.
- [13] J. A. Josen Limbong, I. Sembiring, and K. Dwi Hartomo, "Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes Dan K-Nearest Neighbor," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, pp. 347–356, 2019, doi: 10.25126/jtiik.202294960.
- [14] U. A. Pariamalina *et al.*, "Classification of Online Gambling Spam Comments on YouTube Using Support Vector Machine," vol. 9, no. 1, pp. 153–163, 2026.
- [15] M. A. Imbiri, L. Y. Baisa, and J. J. A. Limbong, "Sentiment Classification and Influential Actor Detection on Twitter (Case Study : The Raja Ampat Mining Conflict)," vol. 7, no. 1, pp. 15–31, 2026.
- [16] J. N. Anggiyanti and F. S. Papilaya, "Implementasi pemodelan topik bertopic dan analisis sentimen bert terhadap ulasan aplikasi satushehat berbasis natural language processing lanjutan," vol. 11, no. 1, pp. 165–177, 2026.

- [17] R. Egger and J. Yu, "A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts," *Front. Sociol.*, vol. 7, no. May, pp. 1–16, 2022, doi: 10.3389/fsoc.2022.886498.
- [18] E. Sutisna, R. Dafitri, and D. Daryani, "Lexicon-Based Sentiment Analysis for Stock Price Prediction in Islamic Fashion Industry," vol. 6, no. April, pp. 591–598, 2026.
- [19] N. Cavus, B. Oktekin, and M. Goksu, "A Cloud-Based Sentiment Analysis System with a BERT Algorithm for Fake News on Twitter," *Appl. Sci.*, vol. 15, no. 20, pp. 1–23, 2025, doi: 10.3390/app152011046.
- [20] N. Darraz, I. Karabila, A. El-Ansari, N. Alami, and M. El Mallahi, "Integrated sentiment analysis with BERT for enhanced hybrid recommendation systems," *Expert Syst. Appl.*, vol. 261, p. 125533, Feb. 2025, doi: 10.1016/J.ESWA.2024.125533.
- [21] J. Zhang, G. Jin, Y. Liu, and X. Xue, "Attention and sentiment of Chinese public toward rural landscape based on Sina Weibo," *Sci. Rep.*, vol. 14, no. 1, pp. 1–14, 2024, doi: 10.1038/s41598-024-64527-1.
- [22] A. F. Anugrah and R. D. Agatha, "Utilizing IndoBERT and BERTopic to Explore Public Opinion on BPS Instagram Posts," *J. Appl. Informatics Comput.*, vol. 9, no. 5, pp. 2836–2844, 2025, doi: 10.30871/jaic.v9i5.10327.
- [23] W. M. Sihombing and T. Widiyaningtyas, "Sentiment Analysis and Topic Modelling Using IndoBERTweet and BERTopic for Public Health Issues," *Indones. J. Innov. Stud.*, vol. 26, no. 4, 2025, doi: 10.21070/ijins.v26i4.1833.
- [24] S. Reports, "Scientific Reports Article in Press Public feedback analysis on multi-stage emergency management policies using BERTopic-SKEP integrated model IN AR IN," 2025.
- [25] W. Wahyuni, T. P. Lestari, M. Apriliana, and R. Gumelta, "Implementation of BERTopic for Topic Modeling Analysis of the Free Nutritious Meal Program Based on YouTube Comments," *J. Appl. Informatics Comput.*, vol. 9, no. 4, pp. 1964–1971, 2025, doi: 10.30871/jaic.v9i4.9754.
- [26] M. Uncovska, B. Freitag, S. Meister, and L. Fehring, "Rating analysis and BERTopic modeling of consumer versus regulated mHealth app reviews in Germany," *npj Digit. Med.*, vol. 6, no. 1, 2023, doi: 10.1038/s41746-023-00862-3.