Jurnal Algoritma, Logika dan Komputasi *Vol. IX (No. 02) : 01 - 13. Th. 2026*
Versi Online: https://journal.ubm.ac.id/index.php/alu *p-ISSN: 2620-620X*
DOI: http://dx.doi.org/10.30813/j-alu.v2i2.10017 *e-ISSN: 2621-9840*

PERBANDINGAN *INDOBERT* DAN *BERT* DALAM ANALISIS SENTIMEN KOMENTAR MEDIA SOSIAL TERHADAP PROGRAM MAKAN BERGIZI GRATIS

Comparison of IndoBERT and BERT for Sentiment Analysis of Free Nutritious Meal Program Comments

Baso Hamzah,105841106922@student.unismuh.ac.id[1)], Desi Anggreani, desianggreani@unismuh.ac.id [2)],Rizki Yusliana Bakti, rizkiyusliana@unismuh.ac.id [3)], Chyquitha Danuputri, chyquithadanuputri@unismuh.ac.id [4)] ,Irawadi, irawadilatif94@gmail.com [5)]

[1)]Informatika /Fakultas Teknik, Universitas Muhammadiyah Makassar
[2)]Informatika / Fakultas Teknik, Universitas Muhammadiyah Makassar
[3)]Informatika / Fakultas Teknik, Universitas Muhammadiyah Makassar
[4)]Informatika / Fakultas Teknik, Universitas Muhammadiyah Makassar
[5)]Pendidikan Agama Islam/Program Pascarjana, Universitas Islam Negri Palopo

Diterima 23 Mei 2026 / Disetujui 24 Agustus 2026

ABSTRACT

The Free Nutritious Meal Program (MBG) has become one of the government policies that has generated various public responses on social media. This study aims to analyze public sentiment tendencies toward the MBG program and compare the performance of IndoBERT and Multilingual BERT models in sentiment classification tasks. The dataset consisted of 3,099 comments collected from TikTok, Instagram, and X, which were processed through preprocessing stages, sentiment labeling, and tokenization. Model evaluation was conducted using the 5-Fold Cross Validation method with accuracy, precision, recall, and F1-score as evaluation metrics. The results showed that IndoBERT achieved better performance than Multilingual BERT, with an accuracy of 0.8209% and an F1-score of 0.8207%, while Multilingual BERT obtained an accuracy of 0.8103% and an F1-score of 0.8101%. Sentiment classification results using the best-performing model showed a distribution of 867 negative sentiments, 841 neutral sentiments, and 836 positive sentiments. In addition, language detection results indicated that the dataset was dominated by Indonesian-language comments at 83.35%. Overall, IndoBERT demonstrated more optimal and stable performance in analyzing sentiment toward comments related to the Free Nutritious Meal Program (MBG).
Keywords: Sentiment Analysis, IndoBERT, BERT, NLP, MBG

ABSTRAK

Program Makan Bergizi Gratis (MBG) merupakan salah satu kebijakan pemerintah yang menimbulkan beragam tanggapan publik di media sosial. Penelitian ini bertujuan untuk menganalisis kecenderungan sentimen masyarakat terhadap program MBG serta membandingkan performa model *IndoBERT* dan *BERT Multilingual* dalam klasifikasi sentimen. Dataset terdiri dari 3.099 komentar dari TikTok, Instagram, dan X yang diproses melalui tahapan *preprocessing*, pelabelan sentimen, serta tokenisasi. Evaluasi dilakukan menggunakan metode K-Fold Cross Validation 5-Fold dengan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa IndoBERT memperoleh performa lebih baik dibandingkan *BERT Multilingual* dengan *accuracy* sebesar 0,8209% dan *F1-score* 0,8207%, sedangkan BERT Multilingual memperoleh *accuracy* 0,8103% dan *F1-score* 0,8101%. Hasil klasifikasi sentimen menggunakan model terbaik menunjukkan distribusi sebanyak 867 sentimen negatif, 841 sentimen netral, dan 836 sentimen positif. Selain itu, hasil deteksi bahasa menunjukkan bahwa dataset didominasi oleh bahasa Indonesia sebesar 83,35%. Secara keseluruhan, IndoBERT menunjukkan

*Korespondensi Penulis:
E-mail: **105841106922@student.unismuh.ac.id**

performa yang lebih optimal dan stabil dalam melakukan analisis sentimen terhadap komentar terkait Program Makan Bergizi Gratis (MBG).
Kata Kunci: Analisis Sentimen, IndoBERT, BERT, NLP, MBG

PENDAHULUAN

Program Makan Bergizi Gratis (MBG) merupakan salah satu kebijakan strategis pemerintah yang bertujuan untuk menanggulangi stunting, meningkatkan status gizi anak usia sekolah, serta mendukung pemerataan kesejahteraan melalui penyediaan makanan bergizi di satuan pendidikan [1]. Hasil studi mengenai implementasi uji coba MBG di sekolah dasar menunjukkan bahwa program ini berpotensi meningkatkan tingkat kehadiran dan konsentrasi belajar siswa, khususnya anak-anak dari keluarga kurang mampu. Dengan demikian, MBG dinilai relevan sebagai instrumen pembangunan sumber daya manusia dalam jangka Panjang[2].

Dalam pelaksanaannya, Program MBG tidak terlepas dari berbagai narasi terarah dan opini publik yang berkembang di masyarakat, terutama di media sosial. Narasi tersebut meliputi klaim keberhasilan yang berlebihan, tuduhan pemborosan anggaran, hingga berbagai isu negatif lain yang belum tentu dapat dibuktikan kebenarannya[3]. Kondisi ini menunjukkan bahwa media sosial menjadi ruang yang signifikan dalam pembentukan persepsi publik terhadap kebijakan MBG. Oleh karena itu, diperlukan pendekatan analitis yang mampu memetakan kecenderungan opini publik secara objektif, sistematis, dan terukur, salah satunya melalui analisis sentimen.

Dalam analisis sentimen, komentar masyarakat di media sosial umumnya diklasifikasikan ke dalam tiga kelas utama, yaitu positif, negatif, dan netral [4]. Pendekatan tiga kelas ini banyak digunakan dalam penelitian terkait Program MBG dengan berbagai metode, baik algoritma machine learning maupun pendekatan berbasis leksikon. Hasil klasifikasi selanjutnya dianalisis untuk melihat distribusi dan dominasi opini publik, sehingga memberikan gambaran kuantitatif mengenai kecenderungan sikap masyarakat serta menjadi dasar evaluasi dan penyempurnaan kebijakan MBG[5]

Perkembangan teknologi pemrosesan bahasa alami menunjukkan bahwa metode analisis sentimen berbasis kamus serta algoritma machine learning konvensional seperti *Naïve Bayes* dan *Support Vector Machine (SVM)* semakin kurang efektif dalam menangani karakteristik bahasa media sosial. Opini warganet yang cenderung tidak baku, sarat singkatan, bahasa gaul, dan nuansa sarkasme menyulitkan model konvensional dalam menangkap konteks dan makna tersirat, sehingga *accuracy*nya menurun ketika diterapkan pada teks berbahasa Indonesia[6]. Sebagai alternatif, dikembangkan model berbasis *Transformer* seperti *BERT* yang memanfaatkan representasi konteks dua arah *(bidirectional)* untuk memahami hubungan antarkata secara lebih mendalam dan menghasilkan klasifikasi sentimen yang lebih akurat[7]

Dalam konteks bahasa Indonesia, *IndoBERT* dikembangkan sebagai varian *BERT* yang dilatih menggunakan korpus berbahasa Indonesia sehingga mampu menangkap karakteristik linguistik warganet dengan lebih baik [8]. Sejumlah penelitian menunjukkan bahwa *IndoBERT* memiliki kinerja yang lebih unggul dalam klasifikasi sentimen teks berbahasa Indonesia, yang tercermin dari nilai *accuracy* dan *F1-score* yang lebih tinggi[9]. Meskipun demikian, kajian yang secara langsung membandingkan kinerja *IndoBERT* dan *BERT* dalam analisis sentimen komentar media sosial terkait Program MBG masih terbatas. Kondisi ini menunjukkan adanya celah penelitian yang perlu dikaji lebih lanjut.

Penelitian ini dirancang untuk mengisi celah tersebut dengan melakukan evaluasi komparatif terhadap kinerja *IndoBERT* dan *BERT* dalam mengklasifikasikan sentimen komentar media sosial mengenai Program MBG. Analisis difokuskan pada perbandingan hasil evaluasi model serta pemetaan kecenderungan sentimen publik yang terbentuk di media sosial. Temuan penelitian ini diharapkan dapat memberikan kontribusi empiris dalam pemilihan model analisis sentimen yang sesuai untuk teks berbahasa Indonesia sekaligus menjadi masukan berbasis data dalam evaluasi kebijakan MBG.

Berdasarkan uraian latar belakang yang telah dijelaskan, maka identifikasi masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana kecenderungan sentimen komentar media sosial terhadap program (MBG) yang diklasifikasikan ke dalam kategori positif, negatif, dan netral?
2. Bagaimana perbandingan kinerja model IndoBERT dan BERT dalam analisis sentimen komentar media sosial terhadap Program Makan Bergizi Gratis

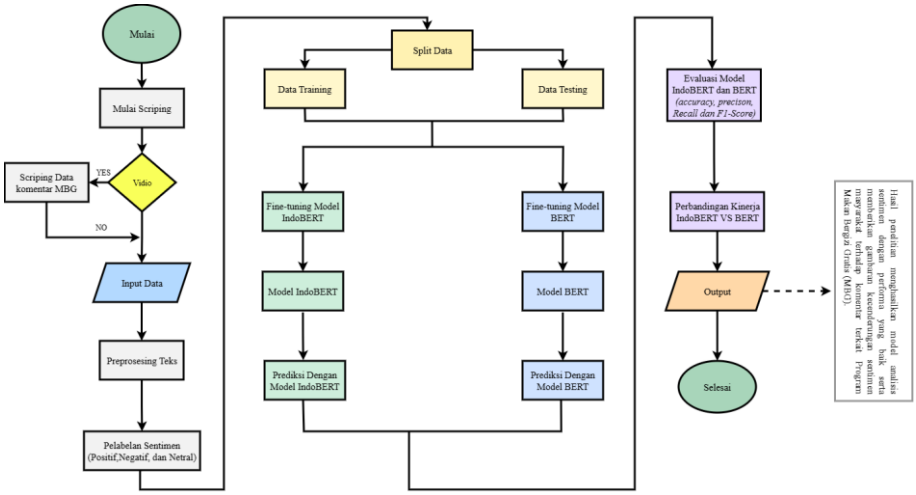
c. Tujuan dan Manfaat Penelitian

Berdasarkan rumusan masalah yang telah dijelaskan, penelitian ini bertujuan untuk mengetahui kecenderungan sentimen komentar masyarakat di media sosial terhadap Program Makan Bergizi Gratis (MBG) yang diklasifikasikan ke dalam kategori positif, negatif, dan netral. Selain itu, penelitian ini juga bertujuan untuk menganalisis dan membandingkan kinerja model *IndoBERT* dan *BERT* dalam melakukan klasifikasi sentimen komentar media sosial terkait Program MBG berdasarkan metrik evaluasi seperti *confusion matrix, accuracy, precision, recall, dan F1-score.*

Adapun manfaat dari penelitian ini adalah membantu peneliti dalam mengaplikasikan pengetahuan mengenai *Natural Language Processing* (NLP), analisis sentimen, serta pemodelan berbasis Transformer seperti *BERT* dan *IndoBERT.* Penelitian ini juga diharapkan dapat memperkaya referensi ilmiah di bidang kecerdasan buatan dan analisis kebijakan publik sehingga dapat menjadi rujukan bagi mahasiswa maupun peneliti lain dalam pengembangan penelitian selanjutnya. Selain itu, penelitian ini diharapkan dapat membantu masyarakat memahami dinamika opini publik secara lebih terukur serta meningkatkan literasi digital dalam menyaring informasi dan membedakan opini faktual dengan propaganda di media sosial.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen berbasis Natural Language Processing (NLP). Penelitian dilakukan untuk membandingkan kinerja model BERT dan IndoBERT dalam mengklasifikasikan sentimen komentar media sosial terhadap Program Makan Bergizi Gratis (MBG). Tahapan penelitian meliputi pengumpulan data, preprocessing data, pelabelan sentimen, pelatihan model, pengujian menggunakan K-Fold Cross Validation, serta evaluasi performa model menggunakan confusion matrix, accuracy, precision, recall, dan F1-score. Pengumpulan data dilakukan melalui teknik scraping pada platform media sosial TikTok menggunakan layanan Apify, serta pengambilan data secara manual dari Instagram dan X. Data yang diperoleh berupa komentar pengguna yang berkaitan dengan Program MBG. Alur tahapan penelitian yang digunakan dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1.Flowchart Perancangan System

Dataset dan Pembagian Data

Dataset yang digunakan dalam penelitian ini merupakan data komentar media sosial terkait Program Makan Bergizi Gratis (MBG) yang diperoleh dari platform TikTok, Instagram, dan X melalui proses scraping dan pengumpulan manual. Data yang telah dikumpulkan kemudian melalui beberapa tahap preprocessing untuk meningkatkan kualitas data sebelum dilakukan proses pelatihan model. Tahapan preprocessing meliputi *cleaning text* untuk menghapus karakter yang tidak diperlukan, normalisasi kata untuk menyeragamkan penulisan kata tidak baku, tokenisasi untuk memecah teks menjadi token-token kata, serta pelabelan sentimen ke dalam kategori positif, negatif, dan netral. Setelah proses preprocessing selesai, dataset digunakan dalam proses pelatihan dan pengujian model analisis sentimen.

Komposisi dataset berdasarkan kategori sentimen setelah melalui tahap preprocessing dapat dilihat pada Tabel 1.

Tabel 1. Komposisi Dataset Sentimen

No.	Kategori Sentimen	Jumlah Data
1.	Negatif	1034
2.	Netral	1034
3.	Positif	1031
	Total	3099

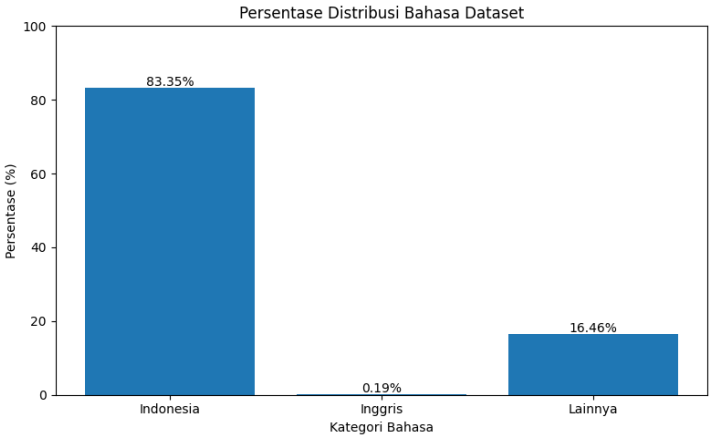
Berdasarkan Tabel 1 Komposisi Dataset Sentimen, dataset penelitian terdiri dari tiga kategori sentimen, yaitu positif, negatif, dan netral. Perbedaan jumlah data pada masing-masing kategori dapat memengaruhi performa model dalam proses klasifikasi sentimen. Oleh karena itu, penelitian ini menggunakan metode K-Fold Cross Validation dalam proses pengujian model untuk memperoleh hasil evaluasi yang lebih stabil dan konsisten. Meskipun jumlah data pada setiap kategori sentimen tidak sepenuhnya seimbang, *metode K-Fold Cross Validation* membantu menjaga distribusi data training dan testing agar tetap proporsional pada setiap fold pengujian [10]. Dengan demikian, proses evaluasi model dapat dilakukan secara lebih adil dan mampu mengurangi potensi bias terhadap kategori sentimen tertentu. Evaluasi performa model dilakukan menggunakan *confusion matrix, accuracy, precision, recall, dan F1-score* pada model *BERT dan IndoBERT.*

Distribusi Bahasa Dataset

Untuk mengetahui karakteristik bahasa pada dataset, dilakukan proses deteksi bahasa menggunakan library langdetect. Hasil deteksi menunjukkan bahwa mayoritas komentar pada dataset menggunakan bahasa Indonesia. Selain itu, terdapat sebagian kecil komentar berbahasa Inggris dan beberapa data lain yang terdeteksi sebagai bahasa selain Indonesia maupun Inggris akibat penggunaan slang, singkatan, kata tidak baku, serta komentar pendek. Distribusi bahasa pada dataset dapat dilihat pada Gambar 2. Distribusi Bahasa.

Berdasarkan hasil deteksi bahasa, dataset didominasi oleh bahasa Indonesia sebesar 83,35%, sedangkan bahasa Inggris hanya sebesar 0,19%, dan sisanya termasuk ke dalam kategori bahasa lainnya sebesar 16,46%

Cross Validation (CV) merupakan teknik evaluasi model *machine learning* yang bertujuan memberikan estimasi performa yang lebih robust, stabil, dan tidak bias dengan memanfaatkan seluruh dataset untuk pelatihan maupun pengujian secara bergantian [11]. Berbeda dengan *train-test split* tunggal yang rentan terhadap variasi pembagian data acak, CV memecah dataset menjadi K subset *(fold)* yang sama besar, melatih model K kali dengan setiap *fold* sebagai data test, dan menghitung rata-rata metrik performa[12]



Gambar 2. Distribusi Bahasa

Secara matematis, untuk dataset ukuran N, setiap *fold* berukuran N/K. Proses iterasi ke-*i* menggunakan *fold-i* sebagai *test* set dan K-1 *fold* lainnya sebagai *training* set. Estimasi performa final adalah rata-rata K skor evaluasi:

$$\widehat{\mathrm{CV}} = \frac{1}{K}\sum_{i=1}^{K} \mathrm{Score}_i \quad (1)$$

Dengan menggunakan metode *5-Fold Cross Validation*, dataset dibagi menjadi 5 bagian yang masing-masing terdiri dari 620 data. Pada setiap iterasi, sebanyak 4 *fold* (2479 data atau 80%) digunakan sebagai data pelatihan (*training*), sedangkan 1 *fold* (620 data atau 20%) digunakan sebagai data pengujian (*validation*). Proses ini dilakukan sebanyak 5 kali hingga seluruh data pernah menjadi data uji seperti yang di tuliskan pada tabel 2 *Pebagian Data* Training *dan* Testing *K-fold*:

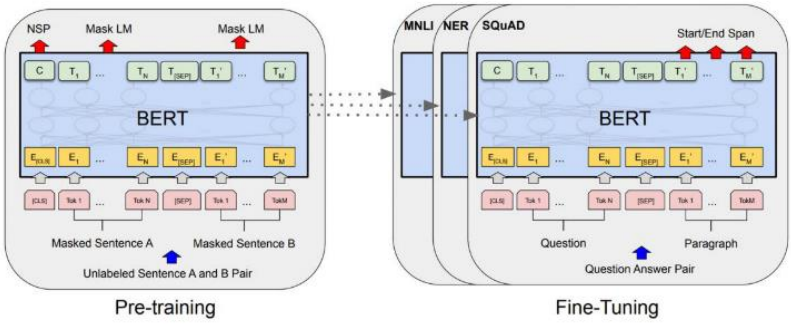
Tabel 2. Pebagian Data

Iterasi	Data Training (*Fold*)	Data Testing (*Fold*)	Jumlah Training	Jumlah Testing
1	*Fold* 2, 3, 4, 5	*Fold* 1	2479	620
2	*Fold* 1, 3, 4, 5	*Fold* 2	2479	620
3	*Fold* 1, 2, 4, 5	*Fold* 3	2479	620
4	*Fold* 1, 2, 3, 5	*Fold* 4	2479	620
5	*Fold* 1, 2, 3, 4	*Fold* 5	2480	619

Berdasarkan Tabel 2 *Pebagian Data* Training *dan* Testing *K-fold* di atas, setiap *fold* digunakan secara bergantian sebagai data pengujian, sedangkan *fold* lainnya digunakan sebagai data pelatihan. Dengan demikian, seluruh data akan digunakan sebagai data uji pada setiap iterasi, sehingga hasil evaluasi lebih representatif.

Model *BERT* dan *IndoBERT* yang digunakan dalam penelitian ini memiliki arsitektur *Transformer Encoder* yang serupa. Karakteristik utama model *BERT* adalah kemampuannya dalam memahami kalimat secara dua arah (*bidirectional*), sehingga setiap kata dapat dipahami berdasarkan konteks kata di sebelah kiri dan kanannya secara bersamaan. Untuk mencapai kemampuan tersebut, *BERT* dilatih menggunakan dua tugas pra-latih utama, yaitu *Masked Language Modeling* (MLM) yang bertujuan memprediksi kata yang disembunyikan dalam kalimat, serta *Next Sentence Prediction* (NSP) untuk memahami hubungan antar kalimat[13].

Pada tahap pemrosesan, *BERT* melakukan tokenisasi teks menjadi unit kata atau subword, kemudian setiap token diubah menjadi representasi vektor *(embedding)* dan diproses melalui beberapa lapisan *encoder* yang terdiri atas mekanisme *multi-head self-attention* dan *feed forward network* [14]. Mekanisme *self-attention* memungkinkan model memberikan bobot yang lebih besar pada kata-kata yang relevan dalam sebuah kalimat sehingga hubungan semantik dan sintaksis dapat dipahami secara lebih mendalam. Berkat kemampuan tersebut, berbagai penelitian analisis sentimen menunjukkan bahwa *BERT* mampu menghasilkan nilai *accuracy* dan *F1-score* yang lebih tinggi dibandingkan metode klasik seperti *Naïve Bayes* dan SVM, terutama pada teks media sosial yang memiliki variasi bahasa dan konteks yang kompleks. Ilustrasi arsitektur serta proses *pre-training* dan *fine-tuning* pada model *BERT* dan *IndoBERT* ditunjukkan pada Gambar 2 *Arsitektur BERT dan IndoBERT*.



Gambar 3. Arsitektur BERT dan IndoBERT[9]

Gambar 3 menunjukkan dua tahap utama dalam penggunaan model *BERT*, yaitu *pre-training* dan *fine-tuning*. Pada tahap *pre-training*, model dilatih menggunakan data teks tidak berlabel melalui pendekatan *Masked Language Modeling* (MLM) dan *Next Sentence Prediction* (NSP) untuk membangun pemahaman bahasa yang bersifat kontekstual dan dua arah. Selanjutnya, pada tahap *fine-tuning*, model yang telah dipra-latih disesuaikan dengan tugas analisis sentimen menggunakan dataset berlabel dengan menambahkan lapisan klasifikasi pada bagian keluaran, sehingga model mampu mengklasifikasikan komentar ke dalam kelas sentimen positif, negatif, atau netral secara lebih akurat [15].

Pada penelitian ini, performa model BERT dan IndoBERT dievaluasi menggunakan confusion matrix. *Confusion matrix* merupakan tabel evaluasi yang digunakan untuk membandingkan hasil prediksi model dengan label aktual pada data pengujian. Dalam analisis sentimen, *confusion matrix* menunjukkan hasil klasifikasi sentimen positif, negatif, dan netral, baik yang diprediksi dengan benar maupun yang mengalami kesalahan, sehingga dapat digunakan untuk mengetahui performa dan pola kesalahan klasifikasi model[16].
empat komponen utama:

a. *True Positive (TP)* yaitu data yang benar-benar positif dan diprediksi positif.
b. *True Negative (TN)* data negatif yang diprediksi negatif.
c. *False Positive (FP)* data negatif yang keliru diprediksi positif.
d. *False Negative (FN)* data positif yang keliru diprediksi negatif

Berdasarkan nilai-nilai yang terdapat pada *Confusion matrix*, selanjutnya dapat dihitung berbagai metrik *evaluasi* kinerja model, antara lain *accuracy*, *precision*, *recall*, dan *F1-score*. Perhitungan *Confusion matrix* beserta metrik *evaluasi* tersebut dilakukan menggunakan rumus-rumus (2), Rumus (3), rumus (4) dan rumus (5).

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

Versi Online: https://journal.ubm.ac.id/index.php/alu *p-ISSN: 2620-620X*
DOI: http://dx.doi.org/10.30813/j-alu.v2i2.10017 *e-ISSN: 2621-9840*

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

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

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

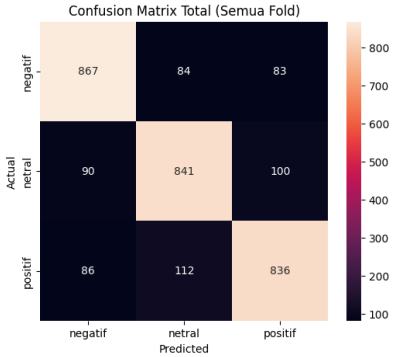
HASIL DAN PEMBAHASAN

Dalam penelitian ini, serangkaian pengujian komprehensif dilakukan dengan memvariasikan jumlah *fold* pada *Cross-Validation* dan durasi pelatihan *(epoch)*. Tujuan utama dari eksperimen ini adalah untuk menganalisis serta mengomparasikan kinerja model pada berbagai skenario, guna mengidentifikasi konfigurasi parameter yang menghasilkan performa klasifikasi paling optimal. Rekapitulasi hasil komparasi dari seluruh skenario pengujian tersebut disajikan pada tabel 6.

Tabel 3. Analisis Penggunaan Jumlah *Fold* Dan *Epoch*

IndoBERT					
Jumlah *fold*	Jumlah epoch	Accuracy	Precision	Recall	F1-score
5	10	0.812	0.813	0.812	0.812
5	20	0.819	0.810	0.819	0.819
5	30	0.821	0.822	0.821	0.821
10	5	0.810	0.811	0.810	0.810
5	5	0.808	0.809	0.808	0.818
IndoBERT					
Jumlah *fold*	Jumlah epoch	Accuracy	Precision	Recall	F1-score
5	10	0.802	0.803	0.802	0.802
5	20	0.804	0.805	0.804	0.804
5	30	0.810	0.811	0.810	0.810
10	5	0.796	0.798	0.796	0.796
5	5	0.790	0.791	0.790	0.789

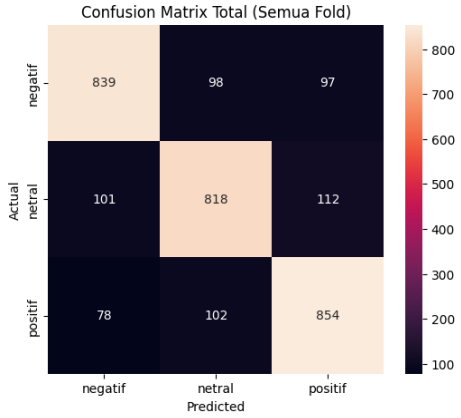
Berdasarkan hasil analisis komparatif yang telah dilakukan, model *IndoBERT* dengan konfigurasi *5-Fold Cross Validation* dan 30 *epoch* dapat direkomendasikan sebagai model terbaik dalam penelitian ini. Hal ini dikarenakan model mampu mencapai kinerja yang stabil serta memiliki kemampuan generalisasi yang baik, tanpa mengalami underfitting maupun penurunan metrik evaluasi selama proses pelatihan.



Gambar 4. Total *Confusion matrix* IndoBERT

Untuk melihat performa klasifikasi sentimen pada masing-masing model, dilakukan analisis menggunakan total *confusion matrix* terhadap model *IndoBERT* dan *BERT Multilingual* yang diperoleh dari hasil pengujian 5-Fold Cross Validation. *Confusion matrix* digunakan untuk membandingkan hasil prediksi model dengan label aktual pada data pengujian, sehingga dapat diketahui jumlah prediksi yang benar maupun kesalahan klasifikasi pada setiap kategori sentimen, yaitu negatif, netral, dan positif. Melalui visualisasi confusion matrix, performa model dapat dianalisis secara lebih detail karena tidak hanya menunjukkan tingkat ketepatan klasifikasi, tetapi juga memperlihatkan distribusi kesalahan prediksi antar kelas sentimen pada masing-masing model. Selain itu, *confusion matrix* juga membantu dalam mengidentifikasi kategori sentimen yang paling mudah maupun paling sulit diklasifikasikan oleh model. Dengan demikian, hasil total *confusion matrix* dari 5-fold pengujian dapat memberikan gambaran yang lebih komprehensif mengenai kemampuan model dalam memahami pola sentimen pada komentar media sosial terkait Program Makan Bergizi Gratis (MBG). Visualisasi total *confusion matrix* model IndoBERT ditunjukkan pada Gambar 3 Total *Confusion matrix* IndoBERT

Berdasarkan Gambar 3 *Confusion matrix* IndoBERT, model IndoBERT menunjukkan hasil klasifikasi yang baik pada seluruh kategori sentimen. Jumlah prediksi benar tertinggi terdapat pada kelas sentimen negatif sebanyak 867 data, diikuti sentimen netral sebanyak 841 data, dan sentimen positif sebanyak 836 data. Hasil tersebut menunjukkan bahwa model IndoBERT mampu melakukan klasifikasi sentimen secara cukup konsisten dan seimbang pada setiap kategori. Adapun hasil total *Confusion matrix* pada model BERT Multilingual ditampilkan pada Gambar 4 Total *Confusion matrix* BERT.



Gambar 5.Total *Confusion matrix BERT*

Berdasarkan Gambar 4 Total *Confusion matrix* BERT, model BERT Multilingual juga menunjukkan performa klasifikasi yang cukup baik. Jumlah prediksi benar tertinggi terdapat pada kelas sentimen positif sebanyak 854 data, diikuti sentimen negatif sebanyak 839 data, dan sentimen netral sebanyak 818 data. Namun, model BERT masih menunjukkan jumlah kesalahan klasifikasi yang lebih tinggi dibandingkan IndoBERT pada beberapa kategori sentimen.

Berdasarkan hasil *confusion matrix* pada kedua model, IndoBERT menunjukkan performa klasifikasi yang lebih baik dan stabil dibandingkan BERT Multilingual dalam melakukan analisis sentimen komentar media sosial terkait Program Makan Bergizi Gratis (MBG).

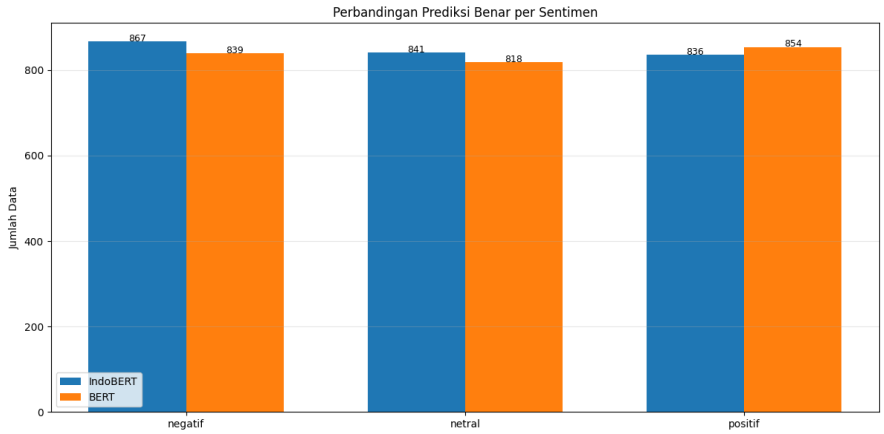
Distribusi Sentimen Komentar MBG

Untuk melihat perbandingan jumlah prediksi benar pada setiap kelas sentimen, dilakukan analisis terhadap hasil klasifikasi menggunakan model *IndoBERT* dan *BERT.* Perbandingan tersebut pada gambar 20 berikut untuk memberikan gambaran visual terkait performa masing-

masing model pada setiap kategori sentiment. Dapat dilihat pada gambar 5 hasil perediksi benar indobert dan bert.

Berdasarkan gambar 5 hasil perediksi benar IndoBERT dan BERT, grafik di atas terlihat bahwa model *IndoBERT* cenderung memiliki jumlah prediksi benar yang lebih tinggi pada kelas negatif dan netral. Sementara itu, model *BERT* menunjukkan performa yang lebih baik pada kelas positif. Perbedaan ini menunjukkan bahwa masing-masing model memiliki kecenderungan yang berbeda dalam mengklasifikasikan sentimen.

Untuk menganalisis konsistensi performa model dalam proses evaluasi, dilakukan pengujian menggunakan metode *K-Fold Cross Validation* sebanyak 5-fold dengan 30 epoch. Hasil *accuracy* pada setiap fold untuk model IndoBERT dan BERT kemudian divisualisasikan dalam bentuk tabel dan grafik performa model per-fold. Tabel 4 menunjukkan nilai *accuracy* yang diperoleh masing-masing model pada setiap fold pengujian, sedangkan Gambar 7 memperlihatkan perbandingan pola performa kedua model dalam bentuk grafik garis sehingga perubahan nilai *accuracy* pada setiap fold dapat diamati dengan lebih jelas..



Gambar 6.Hasil Perediksi Benar

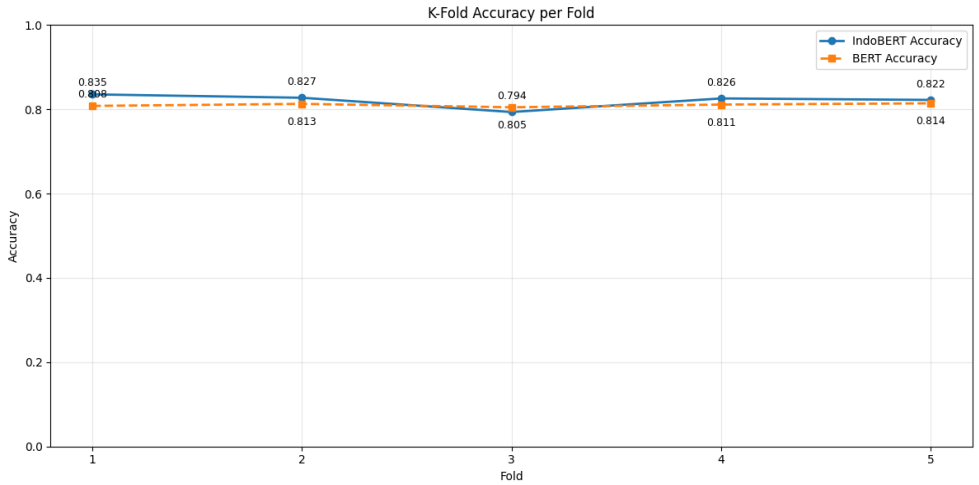
Tabel 4. Performa Model PerFold

IndoBERT	
Fold	Accuracy
1	0.835
2	0.827
3	0.794
4	0.826
5	0.822
BERT	
Fold	Accuracy
1	0.808
2	0.813
3	0.805
4	0.811
5	0.814

Berdasarkan Tabel 4 dan Gambar 7, terlihat bahwa kedua model memiliki nilai *accuracy* yang relatif stabil pada setiap fold. Model IndoBERT cenderung memiliki nilai *accuracy* sedikit lebih tinggi dibandingkan dengan model BERT pada sebagian besar fold pengujian. Meskipun demikian, selisih performa antara kedua model tidak terlalu signifikan, yang menunjukkan bahwa keduanya memiliki tingkat konsistensi yang baik dalam melakukan klasifikasi sentimen.

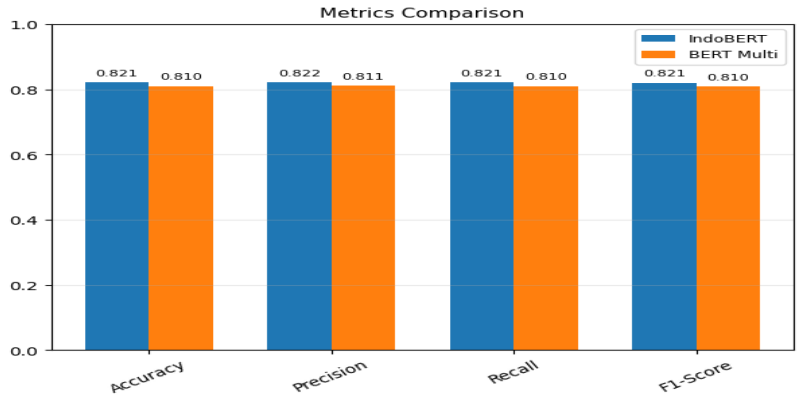
Selain itu, pada fold ke-3 terjadi penurunan *accuracy* pada model IndoBERT yang mengindikasikan adanya variasi distribusi data pada fold tersebut. Namun, performa kembali meningkat pada fold berikutnya sehingga model dapat dikatakan memiliki kestabilan performa yang baik selama proses *k-fold cross validation*.

Selain itu, pada *fold* ke-3 terjadi penurunan *accuracy* pada model *IndoBERT*, yang mengindikasikan adanya variasi distribusi data pada *fold* tersebut. Namun, performa kembali meningkat pada *fold* berikutnya, sehingga model dapat dikatakan memiliki kestabilan performa yang baik dalam proses *k-fold cross validation.*



Gambar 7.Performa Model Per-Fold

Untuk membandingkan kinerja model secara keseluruhan, dilakukan analisis terhadap nilai metrik evaluasi yang meliputi *accuracy, precision, recall, dan F1-score.* Perbandingan performa antara model *IndoBERT* dan *BERT* Multilingual disajikan pada gambar 7 performa akhir.



Gambar 8.Performa Akhir

Berdasarkan grafik di atas, terlihat bahwa model *IndoBERT* memiliki nilai performa yang sedikit lebih tinggi dibandingkan dengan model *BERT* Multilingual pada seluruh metrik evaluasi. Nilai *accuracy* 0,821%, *precision* 0,822%, *recall* 0,821%, *dan F1-score* 0,82% pada *IndoBERT* sedangkan *BERT* Multilingual Nilai *accuracy* 0,810%, *precision* 0,811%, *recall* 0,810%, *dan F1-score* 0,810%

Perbedaan ini menunjukkan bahwa *IndoBERT* memiliki kemampuan yang lebih baik dalam memahami konteks bahasa Indonesia, sehingga menghasilkan prediksi yang lebih akurat

dan konsisten. Meskipun selisih nilai antar model tidak terlalu besar, *IndoBERT* tetap menunjukkan keunggulan yang stabil pada seluruh metrik evaluasi. Dengan demikian, dapat disimpulkan bahwa *IndoBERT* memiliki performa yang lebih unggul dibandingkan *BERT Multilingual* dalam tugas klasifikasi sentimen pada dataset yang digunakan.

Untuk mendukung penelitian ini, dilakukan kajian terhadap beberapa penelitian terdahulu yang berkaitan dengan analisis sentimen menggunakan model berbasis *Transformer*. Penelitian-penelitian tersebut digunakan sebagai referensi dalam memahami metode, teknik evaluasi, serta performa model yang digunakan pada klasifikasi sentimen berbahasa Indonesia. Selain itu, kajian penelitian terdahulu juga membantu dalam mengidentifikasi persamaan, perbedaan, serta kontribusi penelitian yang dilakukan dibandingkan penelitian sebelumnya. Ringkasan penelitian terdahulu yang digunakan sebagai referensi dalam penelitian ini ditunjukkan pada Tabel 5 Penelitian Terdahulu.

Tabel 5. Penelitian Terdahulu

Penelitian Sebelumnya	Hasil	Perbedaan Penelitian yang telah di lakukan
sentimen komentar publik terhadap film dokumenter Dirty Vote di YouTube menggunakan metode K-Fold Cross Validation.	memperoleh 94%. IndoBERT dinilai lebih stabil dalam klasifikasi sentimen positif, negatif, dan netral.	Makan Bergizi Gratis (MBG) menggunakan data dari TikTok, Instagram, dan X, serta mengevaluasi performa model berdasarkan variasi *fold* dan *epoch* untuk melihat kestabilan model.
Fachry et al. (2025) mengimplementasikan kombinasi BiLSTM dan IndoBERT untuk analisis sentimen ulasan aplikasi TikTok di Google Play Store dengan kategori positif, negatif, dan netral menggunakan *10-fold cross-validation*.	Model terbaik adalah kombinasi BiLSTM+IndoBERT dengan *accuracy* 81% pada klasifikasi laporan dan rata-rata *accuracy* 92,03% pada pengujian *10-fold cross-validation*.	Penelitian ini tidak menggunakan model hybrid, melainkan membandingkan performa model transformer secara langsung antara IndoBERT dan BERT Multilingual untuk mengetahui efektivitas masing-masing model dalam klasifikasi sentimen komentar terkait Program MBG.
Sholihah et al. (2024) menggunakan model BERT untuk klasifikasi sentimen publik dengan metode *10-fold cross-validation* menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.	Model BERT memperoleh *accuracy*, *precision*, *recall*, dan *F1-score* tertinggi sebesar 96%, namun pada fold terendah (k=9) nilai *accuracy* turun menjadi 86%, sehingga performa model belum konsisten pada seluruh fold.	Penelitian ini membandingkan model BERT Multilingual dengan IndoBERT serta menambahkan variasi skenario pelatihan untuk menganalisis kestabilan performa model pada setiap fold pengujian.
Dedes et al. (2025) melakukan *fine-tuning* BERT Multilingual (*mBERT*) untuk analisis sentimen ulasan film berbahasa Indonesia dengan fokus pada pengaruh augmentasi data terhadap kinerja model.	Model *mBERT* yang di-*fine-tune* memperoleh *accuracy* sebesar 82,5% dengan *precision* kelas positif 76,0%, *recall* 95,0%, dan *F1-score* 84,44% pada data uji.	Penelitian ini tidak berfokus pada augmentasi data, melainkan membandingkan performa IndoBERT dan BERT Multilingual menggunakan variasi *epoch* dan *K-Fold Cross Validation* untuk menganalisis kestabilan model dalam klasifikasi sentimen
Geni et al. (2024) melakukan analisis sentimen tweet masyarakat Indonesia terkait Pemilu 2024 menggunakan metode *lexicon-based*, *machine learning* seperti Multinomial Naïve Bayes dan SVM, serta model Transformer IndoBERT dengan *fine-tuning*.	Hasil penelitian menunjukkan mayoritas tweet bersentimen netral sebesar ±83,7%, dengan model terbaik yaitu IndoBERT large-p1 yang memperoleh *accuracy* sebesar 83,5%.	Penelitian ini tidak hanya menggunakan satu model transformer, tetapi membandingkan IndoBERT dan BERT Multilingual menggunakan variasi *epoch* dan *K-Fold Cross Validation*, serta menganalisis distribusi sentimen dan kestabilan performa model secara lebih mendalam.

Berdasarkan Tabel 5 Penelitian Terdahulu, dapat diketahui bahwa penelitian sebelumnya umumnya menggunakan model berbasis *Transformer* seperti IndoBERT, IndoRoBERTa, dan

mBERT dalam tugas analisis sentimen berbahasa Indonesia. Sebagian besar penelitian menggunakan metode evaluasi *K-Fold Cross Validation* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengukur performa model. Hasil penelitian terdahulu menunjukkan bahwa model berbasis *Transformer* mampu memberikan performa klasifikasi yang baik pada berbagai jenis data teks dan topik pembahasan.

Penelitian ini memiliki perbedaan dibandingkan penelitian sebelumnya karena tidak hanya menggunakan satu model, tetapi membandingkan performa *IndoBERT* dan *BERT Multilingual* secara langsung menggunakan variasi *epoch* dan *fold* pengujian. Selain itu, penelitian ini juga melakukan analisis menggunakan *confusion matrix* total untuk melihat distribusi kesalahan klasifikasi pada setiap kategori sentimen. Dengan demikian, penelitian ini tidak hanya berfokus pada tingkat akurasi model, tetapi juga pada kestabilan performa model dalam berbagai skenario pelatihan.

SIMPULAN

Berdasarkan hasil klasifikasi sentimen menggunakan model dengan performa terbaik, yaitu IndoBERT, diperoleh distribusi sentimen komentar terhadap Program Makan Bergizi Gratis (MBG) sebanyak 867 sentimen negatif, 841 sentimen netral, dan 836 sentimen positif. Dataset yang digunakan terdiri dari 3.099 data komentar dengan distribusi kelas yang tidak seimbang, yaitu 1.034 data negatif, 1.034 data netral, dan 1.031 data positif. Selain itu, hasil deteksi bahasa menunjukkan bahwa dataset didominasi oleh bahasa Indonesia sebesar 83,35%, sedangkan sisanya terdiri dari bahasa Inggris dan bahasa lainnya.

Berdasarkan hasil pengujian menggunakan metode K-Fold Cross Validation, model IndoBERT menunjukkan performa yang lebih baik dibandingkan BERT Multilingual. Hasil rata-rata evaluasi 5-fold cross validation menunjukkan bahwa IndoBERT memperoleh *accuracy* sebesar 0,8209%, *precision* 0,8210%, *recall* 0,8209%, dan *F1-score* sebesar 0,8207%. Sementara itu, BERT Multilingual memperoleh *accuracy* sebesar 0,8103%, *precision* 0,8111%, *recall* 0,8103%, dan *F1-score* sebesar 0,8101%. Dengan demikian, dapat disimpulkan bahwa model IndoBERT memiliki performa yang lebih optimal dan stabil dalam melakukan klasifikasi sentimen pada komentar terkait Program Makan Bergizi Gratis (MBG) dibandingkan model BERT Multilingual.

DAFTAR PUSTAKA

[1] 2025 Kiftiyah et al., “PANCASILA : Jurnal Keindonesiaan,” 2025.

[2] N. Azzahra, A. D. Dharmawan, A. F. Mardatilah, and M. I. Habibi, “Pelaksanaan Uji Coba Program Makan Bergizi Gratis di SMP Negeri 4 Tangerang,” vol. 3, no. 4, pp. 5036–5044, 2025.

[3] M. W. Arif, “Analisis Sentimen Kebijakan Makan Bergizi Gratis di Media Sosial Menggunakan Natural Language Processing Berbasis Python TextBlob di Indonesia Teknik Informatika , Universitas Ngudi Waluyo , Indonesia Sentiment Analysis of Free Nutritious Meal Policy on Social Media Using Python TextBlob-Based Natural Language Processing in Indonesia,” vol. 5, no. 9, pp. 2463–2471, 2025.

[4] J. J. Levett, A. Alnasser, L. M. Elkaim, J. Drager, and T. Pauyo, “Negative sentiments toward anterior cruciate ligament injury prevention outweigh positive awareness discussions on social media,” *J. ISAKOS*, vol. 9, no. 5, p. 100306, 2024, doi: 10.1016/j.jisako.2024.100306.

[5] F. Fatkhurrohman, B. I. Nugroho, and N. Fadillah, “Analisis Sentimen Program Makan Bergizi Gratis Pemerintah RI Melalui Twitter Menggunakan Metode SVM,” vol. 4, no. 3, pp. 3906–3917, 2025.

[6] A. A. Qolbu, N. Fitriyati, and N. Inayah, “Performa Naïve Bayes , SVM , dan IndoBERT pada Analisis Sentimen Twitter IndiHome dengan Strategi Penanganan Data Tidak Seimbang,” vol. 814, no. 1, pp. 29–44, 2025, doi: 10.14421/fourier.2025.141.29-44.

[7] R. Savitri, F. Rizki, and A. Sobri, "Implementation of BERT in Sentiment Analysis of National Digital Samsat (SIGNAL) User Reviews Based on Machine Learning," vol. 17, no. 2, pp. 67–75, 2025.

[8] J. Vincent *et al.*, "ScienceDirect ScienceDirect ScienceDirect Extractive Indonesian News Text Summarization using Extractive Indonesian News MBERT , Text Summarization using and RoBERTa Extractive Indonesian News Text Summarization using and," *Procedia Comput. Sci.*, vol. 269, pp. 1087–1096, 2025, doi: 10.1016/j.procs.2025.09.050.

[9] P. Sayarizki and H. Nurrahmi, "Implementation of IndoBERT for Sentiment Analysis of Indonesian Presidential Candidates," vol. 9, no. August, pp. 61–72, 2024, doi: 10.34818/indojc.2024.9.2.934.

[10] J. Xu, X. Liu, Z. Gu, and G. Xiao, "A rapid cross-validation computing for three-way decisions in imbalanced data," *Inf. Sci. (Ny).*, vol. 707, no. February, p. 122016, 2025, doi: 10.1016/j.ins.2025.122016.

[11] E. G. Dandamudi *et al.*, "Biosensors and Bioelectronics : X Performance of transformer-convolutional neural network ensemble for melanoma diagnosis on segmented 3D total body photography data : Cross-Validation stratified K-fold," *Biosens. Bioelectron. X*, vol. 27, no. November, p. 100714, 2025, doi: 10.1016/j.biosx.2025.100714.

[12] K. Dedes *et al.*, "BERT Sentimen : Fine-Tuning Multibahasa untuk Ulasan Bahasa," vol. 4, no. 2, pp. 1080–1084, 2025.

[13] 2024 Sriyanti et al, "implementasi model bert pada analisis sentimen pengguna twitter terhadap aksi," vol. 12, no. 3, pp. 2335–2342, 2024.

[14] B. Shivakumar, U. Sandeep, V. Varshini, and P. Kumaran, "ScienceDirect ScienceDirect Multi-Head for Text2Text Text2Text Translation Multi-Head Attention Attention Transformer Transformer for Translation," *Procedia Comput. Sci.*, vol. 258, pp. 1962–1971, 2025, doi: 10.1016/j.procs.2025.04.447.

[15] W. Christian *et al.*, "Leveraging Leveraging IndoBERT IndoBERT and and DistilBERT DistilBERT for for Indonesian Indonesian emotion emotion Leveraging IndoBERT and DistilBERT for Indonesian classification in reviews classification in e-commerce reviews emotion classification in e-commerce reviews ScienceDirect," *Procedia Comput. Sci.*, vol. 269, pp. 321–330, 2025, doi: 10.1016/j.procs.2025.08.284.

[16] 2023 Nurhidayat & Dewi, "KOMPUTA : Jurnal Ilmiah Komputer dan Informatika KOMPUTA : Jurnal Ilmiah Komputer dan Informatika," vol. 12, no. 1, pp. 91–100, 2023.