

# Exploring Public Opinion on the 'Makan Bergizi Gratis' Program on X: A Comparative Analysis of IndoBERT-Large and NusaBERT-Large Models

Aurelya Prameswari Arunia<sup>1</sup>, Ramadhan Rakhmat Sani<sup>2</sup>, Ika Novita Dewi<sup>3</sup>, MY Teguh Sulistyono<sup>4</sup>

\* Sistem Informasi, Universitas Dian Nuswantoro

[aurelyaprameswari6@gmail.com](mailto:aurelyaprameswari6@gmail.com)<sup>1</sup>, [ramadhan\\_rs@dsn.dinus.ac.id](mailto:ramadhan_rs@dsn.dinus.ac.id)<sup>2</sup>, [ikadewi@dsn.dinus.ac.id](mailto:ikadewi@dsn.dinus.ac.id)<sup>3</sup>, [teguh.sulistyono@dsn.dinus.ac.id](mailto:teguh.sulistyono@dsn.dinus.ac.id)<sup>4</sup>

## Article Info

### Article history:

Received 2025-11-13

Revised 2026-01-07

Accepted 2026-01-13

### Keyword:

*Sentiment Analysis,  
Program Makan Bergizi Gratis,  
IndoBERT,  
NusaBERT,  
Social Media X*

## ABSTRACT

Program Makan Bergizi Gratis (MBG) has triggered extensive discourse on social media platform X, which serves as a primary space for public expression of opinions toward government policies. This study aims to analyze public sentiment toward the MBG program while simultaneously comparing the performance of two prominent Transformer-based models, namely IndoBERT-Large and NusaBERT-Large. This research adopts a quantitative approach employing supervised learning on 10,201 Indonesian-language posts (tweets) collected through web scraping from February 2024 to September 2025. A total of 2,000 samples were manually annotated as ground truth, achieving a high level of inter-annotator reliability (Cohen's Kappa,  $\kappa = 0.81$ ). The experimental results indicate that IndoBERT-Large outperforms NusaBERT-Large, achieving an accuracy of 83.00%, while NusaBERT-Large demonstrates competitive performance with an accuracy of 80.50%. Substantively, public discourse is dominated by negative sentiment, accounting for nearly 50% of the total data, reflecting public concerns regarding budgetary constraints and technical implementation issues. Positive sentiment ranges between 33% and 36%, indicating sustained and substantial public support for the program. These findings confirm the effectiveness of Transformer-based models in accurately capturing the dynamics of public opinion toward government policies using social media data.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

## I. PENDAHULUAN

Pada tanggal 6 Januari 2025, Pemerintahan Presiden Prabowo Subianto-Gibran Rakabuming Raka secara resmi meluncurkan Program Makan Bergizi Gratis atau MBG sebagai salah satu inisiatif strategis dalam mewujudkan visi Indonesia Emas 2045 [1]. Program yang dikelola oleh Badan Gizi Nasional ini mendapatkan alokasi anggaran sebesar Rp71 triliun dari APBN 2025, dengan proyeksi awal melayani 19,47 juta penerima manfaat yang mencakup anak sekolah, ibu hamil, dan ibu menyusui [2]. Target ambisius yang ditetapkan adalah mencapai 20 juta penerima manfaat sebelum peringatan kemerdekaan pada 17 Agustus 2025, kemudian diperluas hingga 82,9 juta penerima pada akhir tahun [3]. Selain bertujuan meningkatkan kualitas sumber daya manusia melalui perbaikan gizi anak sekolah, program MBG juga diharapkan dapat memberdayakan UMKM dan

menggerakkan ekonomi kerakyatan. Lebih dari 100 ribu lapangan kerja baru diproyeksikan akan tercipta dari implementasi program ini [4]. Namun, seperti kebijakan publik lainnya, program MBG telah memunculkan beragam respons dari masyarakat yang tercermin dalam diskusi-diskusi di berbagai platform digital, terutama media sosial, seperti platform X (sebelumnya Twitter).

Platform X sangat populer di Indonesia dengan lebih dari 33 juta pengguna aktif, menjadikannya barometer efektif untuk mengukur opini publik [5]. Platform ini kini menjadi arena utama masyarakat dalam menyampaikan pendapat terhadap kebijakan pemerintah. Karakteristik X yang membatasi jumlah karakter per postingan mendorong pengguna untuk mengekspresikan pandangan secara singkat namun langsung pada intinya. Respons publik terhadap program Makan Bergizi Gratis di platform X memperlihatkan variasi yang cukup signifikan. Sebagian pengguna

menyampaikan apresiasi dengan harapan program ini dapat berkontribusi pada peningkatan kualitas tumbuh kembang anak Indonesia. Di sisi lain, terdapat pula kritik yang menyoroti aspek transparansi pengelolaan anggaran serta kekhawatiran terhadap kemungkinan penyimpangan dana. Selain itu, muncul berbagai pertanyaan terkait mekanisme distribusi program, khususnya untuk menjangkau wilayah-wilayah terpencil. Keragaman respons tersebut menghasilkan data tekstual dalam jumlah besar yang berpotensi memberikan wawasan penting, namun sulit dianalisis secara manual.

Perkembangan teknologi *Natural Language Processing* (NLP), memungkinkan analisis opini publik dilakukan secara sistematis melalui pendekatan komputasional [6]. Salah satu pendekatan yang banyak digunakan dalam kajian opini publik adalah analisis sentimen, yaitu proses mengklasifikasikan teks berdasarkan kecenderungan sikap atau emosi yang terkandung di dalamnya, seperti positif, negatif, dan netral. Analisis sentimen telah banyak dimanfaatkan dalam berbagai domain, termasuk evaluasi kebijakan publik, karena mampu memberikan gambaran kuantitatif mengenai persepsi masyarakat terhadap suatu isu secara lebih objektif dan efisien [7].

Dalam beberapa tahun terakhir, model berbasis transformer, khususnya *Bidirectional Encoder Representations from Transformers* (BERT), telah membuka peluang baru dalam menganalisis teks berbahasa Indonesia secara lebih mendalam [8]. Untuk bahasa Indonesia, berbagai model BERT telah dikembangkan dan digunakan secara luas, di antaranya *IndoBERT* dan *NusaBERT*. *IndoBERT* merupakan model bahasa *pre-trained* yang dikembangkan secara khusus untuk bahasa Indonesia, termasuk mampu memahami variasi bahasa informal, singkatan, dan bahasa gaul yang sering digunakan dalam komunikasi media sosial [6][8]. Model ini hadir dalam dua varian, yaitu *IndoBERT-base* dan *IndoBERT-large*, di mana varian *large* memiliki kapasitas lebih besar dalam menangkap nuansa bahasa yang kompleks [9]. Sejumlah penelitian terdahulu menunjukkan efektivitas *IndoBERT* dalam berbagai konteks analisis sentimen. Penelitian [10] melaporkan bahwa *IndoBERT* mampu mencapai tingkat akurasi yang sangat tinggi dalam mengklasifikasikan sentimen pada ulasan aplikasi kesehatan di Indonesia. Dalam konteks politik, penelitian [11] berhasil menerapkan *IndoBERT* untuk analisis sentimen terhadap calon presiden Indonesia dengan hasil yang menunjukkan performa model yang stabil dalam menangani wacana politik yang kompleks. Selain itu, penelitian [12] membuktikan bahwa *fine-tuning IndoBERT* yang dikombinasikan dengan arsitektur *R-CNN* mampu menghasilkan performa yang lebih unggul dibandingkan metode konvensional dalam klasifikasi sentimen ulasan berbahasa Indonesia.

Di sisi lain, pengembangan model bahasa Indonesia juga melahirkan *NusaBERT*, yang dirancang untuk menangkap keragaman linguistik bahasa Indonesia melalui pelatihan pada korpus yang lebih luas dan multidomain. Penelitian komparatif yang dilakukan oleh [13] menunjukkan bahwa

model-model yang dipra-latih khusus pada data berbahasa Indonesia, termasuk *IndoBERT* dan *NusaBERT*, secara umum mengungguli model multilingual dalam tugas analisis sentimen lintas bahasa daerah di Indonesia. Studi tersebut juga menyoroti variasi performa *NusaBERT* dan *IndoBERT* pada berbagai bahasa regional, yang menunjukkan pentingnya evaluasi model dalam konteks data yang spesifik. Selain itu, penelitian [14] pada ulasan aplikasi *Access by KAI* membuktikan bahwa *NusaBERT* mampu menghasilkan akurasi dan *F1-score* yang lebih tinggi dibandingkan pendekatan tradisional berbasis *SVM*, menegaskan keunggulan model transformer dalam memahami konteks bahasa Indonesia. Studi lain [15] yang membandingkan *IndoBERT*, *IndoRoBERTa*, dan *NusaBERT* pada analisis sentimen isu *LGBT* di platform X juga menunjukkan bahwa *IndoBERT* memberikan performa terbaik, namun *NusaBERT* tetap menunjukkan kinerja yang kompetitif dalam menangani isu sensitif di media sosial.

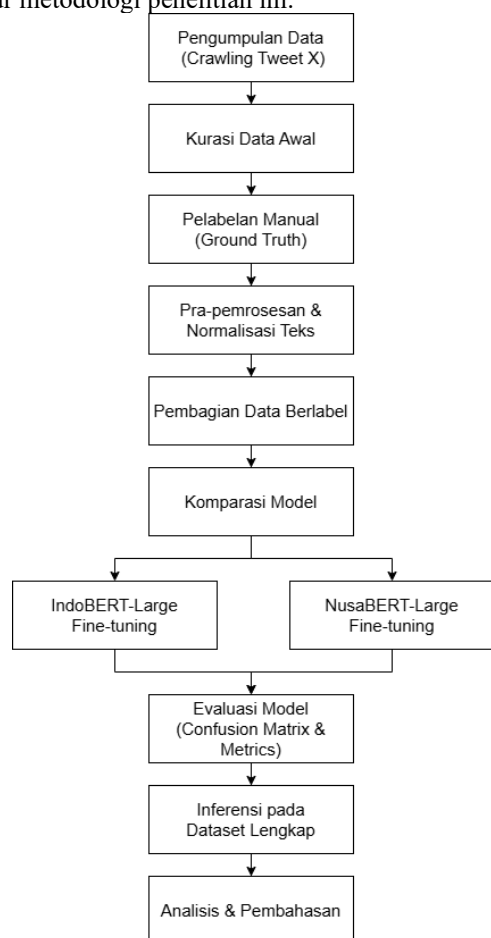
Meskipun berbagai penelitian tersebut telah membuktikan efektivitas *IndoBERT* dan *NusaBERT* secara terpisah maupun dalam konteks tertentu, kajian yang secara khusus membandingkan kedua model tersebut dalam analisis sentimen kebijakan publik di Indonesia masih relatif terbatas. Selain itu, sebagian besar penelitian terdahulu belum secara eksplisit mengevaluasi performa kedua model dalam konteks diskursus kebijakan dengan karakteristik bahasa yang kompleks dan sarat dengan kritik implisit, seperti pada program berskala nasional dengan implikasi fiskal yang besar.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis sentimen publik terhadap Program Makan Bergizi Gratis di platform X dengan menggunakan pendekatan komparatif antara dua model transformer berbahasa Indonesia, yaitu *IndoBERT* dan *NusaBERT*. Melalui pendekatan ini, penelitian ini mengevaluasi efektivitas kedua model dalam mengklasifikasikan opini masyarakat ke dalam kategori positif, negatif, dan netral, serta menelaah model yang menunjukkan kinerja paling *robust* dalam menangani karakteristik bahasa informal dan ketidakseimbangan kelas pada data opini kebijakan publik. Diharapkan, hasil penelitian ini dapat memberikan kontribusi empiris dalam pemilihan model NLP yang lebih tepat untuk analisis kebijakan publik di Indonesia, serta menjadi dasar bagi pengambil kebijakan dalam memahami respons masyarakat secara lebih akurat dan berbasis data.

## II. METODE

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan desain studi komparatif untuk mengevaluasi performa klasifikasi sentimen pada diskusi program Makan Bergizi Gratis di platform X. Kerangka metodologi disusun dengan membandingkan dua model transformer secara paralel. Alur penelitian berfokus pada *fine-tuning* dan evaluasi komparatif antara model *IndoBERT-*

*Large* dan *NusaBERT-Large*. Gambar 1 menunjukkan arsitektur metodologi penelitian ini.



Gambar 1. Arsitektur Metodologi Secara Keseluruhan

#### A. Pengumpulan Data

Data penelitian diperoleh melalui teknik web scraping pada platform X menggunakan pustaka tweet-harvest yang dijalankan dalam lingkungan Node.js. Proses pengumpulan data atau *crawling* dilakukan secara terotomatisasi dengan memanfaatkan *Twitter authentication token* guna memastikan akses data yang stabil dan valid.

Strategi pengumpulan data dirancang dengan dua fokus utama untuk menjamin keterwakilan dataset yang komprehensif. Fokus pertama bertujuan menangkap diskursus terkait aspek ekonomi dan anggaran program melalui kombinasi kueri (“makan gratis” OR MBG) AND (anggaran OR “71 triliun” OR APBN OR biaya OR ekonomi OR UMKM).

Fokus kedua diarahkan untuk menjangkau variasi sentimen publik dengan kombinasi kueri (“makan gratis” OR MBG) AND (gagal OR masalah OR kritik OR korupsi OR “tidak efektif” OR bagus OR hebat OR sukses OR mantap).

Pemilihan kata kunci yang berkaitan dengan isu fiskal dan kritik dilakukan secara sengaja sebagai strategi teknis untuk memastikan keberagaman sampel sentimen, khususnya pada kelas negatif dan netral. Pendekatan ini penting untuk

menghindari ketidakseimbangan kelas atau *imbalanced data* yang ekstrem serta membantu model mempelajari pola kritik publik secara lebih representatif.

Seluruh kueri dibatasi pada konten berbahasa Indonesia dengan aktivasi filter *LATEST* untuk menjaga relevansi dan kualitas data. Periode observasi ditetapkan secara *longitudinal* mulai 1 Februari 2024 hingga 1 September 2025. Rentang waktu ini dipilih untuk merekam dinamika opini publik secara utuh, mulai dari fase wacana politik pada masa kampanye, tahap perumusan kebijakan pasca-pemilu, hingga fase implementasi awal program pada Januari 2025.

Dataset mentah yang terkumpul selanjutnya melalui tahap kurasi data yang meliputi penghapusan *tweet* duplikat, eliminasi *retweet* tanpa komentar tambahan, serta penyaringan konten yang tidak relevan dengan topik penelitian. Proses ini menghasilkan 10.201 *tweet* unik sebagai dataset utama penelitian. Secara deskriptif, panjang teks *tweet* berada pada rentang 10–280 karakter dengan rata-rata 87 karakter. Dari sisi temporal, dataset menunjukkan lonjakan aktivitas diskusi pada periode awal implementasi program, diikuti pola yang relatif stabil dengan *fluktuasi* yang dipengaruhi oleh peristiwa atau pemberitaan tertentu terkait Program MBG.

#### B. Pembentukan Ground Truth melalui Pelabelan Manual

Proses pembentukan *ground truth* dilakukan melalui anotasi manual terhadap 2.000 *tweet* yang dipilih menggunakan teknik *simple random sampling* dari keseluruhan dataset yang telah dikurasi. Penentuan jumlah sampel ini didasarkan pada pertimbangan keseimbangan antara keterwakilan data dan kecukupan jumlah data untuk proses *fine-tuning model transformer*.

Pelabelan manual dilakukan setelah tahap kurasi data awal dan sebelum proses normalisasi teks yang bersifat agresif. Pendekatan ini dipilih agar anotator dapat menilai sentimen berdasarkan teks yang masih merepresentasikan ekspresi asli pengguna, termasuk penggunaan bahasa informal, singkatan, dan nuansa implisit yang berpotensi memengaruhi interpretasi sentimen. Dengan demikian, *ground truth* yang dihasilkan diharapkan lebih mencerminkan makna semantik asli dari opini publik.

Proses anotasi melibatkan dua anotator independen yang merupakan penutur asli bahasa Indonesia, memiliki latar belakang pendidikan minimal sarjana, serta memahami konteks kebijakan publik di Indonesia. Untuk menjaga objektivitas dan konsistensi, pelabelan dilakukan berdasarkan skema klasifikasi yang telah didefinisikan secara jelas. Label positif diberikan pada *tweet* yang menunjukkan dukungan, harapan, atau apresiasi terhadap Program MBG. Label negatif mencakup *tweet* yang berisi kritik terhadap pengelolaan anggaran, kekhawatiran akan potensi penyimpangan dana, maupun keraguan terhadap mekanisme distribusi program. Sementara itu, label netral diberikan pada *tweet* yang bersifat informatif, berupa pertanyaan teknis, atau tidak menunjukkan kecenderungan sentimen yang jelas.

Tingkat kesepakatan antar-anotator diukur menggunakan koefisien *Cohen's Kappa* sebagai indikator reliabilitas pelabelan. Hasil pengukuran menunjukkan nilai  $\kappa = 0.81$ , yang mengindikasikan tingkat kesepakatan substansial. Perbedaan label yang muncul diselesaikan melalui diskusi hingga tercapai konsensus untuk menentukan label akhir. Dataset berlabel yang telah divalidasi ini selanjutnya dibagi ke dalam subset pelatihan dan validasi untuk mendukung evaluasi performa komparatif antara model *IndoBERT-Large* dan *NusaBERT-Large*. Tabel 1 menyajikan contoh data yang telah diberi pelabelan.

TABEL 1  
CONTOH DATA YANG TELAH DIBERI LABEL

<i>Tweet</i>	<i>Label</i>
@ARSIPAJA Yaudah sih meskipun ga diblokir juga belum tentu bisa operasi. Kenapa sih semua harus salahin pemerintah. Prabowo gibran itu pasangan juru selamat jadi semua keputusannya ga mungkin salah. Iri bilang dek! nb: MBG program terbaik pemerintah sepan	Positif
Pak Bos Stop Omon Omon @prabowo Bukan Nyinyir tapi Fakta sudah berulang kali kejadian dan banyak korban MBG BASI BELATUNGAN BERACUN Stop MBG alihkan ke pendidikan saja @smindrawatii https://t.co/DE8XaPYjU8	Negatif
mbg di sekolah adik aku jadi makan makanan berat gitu..	Netral

### C. Pra-pemrosesan Data Teks

Dataset berlabel manual yang terdiri dari 2.000 tweet melalui serangkaian tahapan pra-pemrosesan data teks sebelum digunakan dalam proses *fine-tuning model*. Tahapan pra-pemrosesan ini dirancang untuk mempertahankan informasi semantik yang relevan sekaligus mengurangi noise yang berpotensi menurunkan performa model klasifikasi sentimen. *Pipeline* pra-pemrosesan yang sama juga diterapkan pada dataset lengkap pada tahap inferensi untuk menjaga konsistensi analisis. Prosedur pra-pemrosesan meliputi beberapa langkah utama sebagai berikut:

1. *Case Folding*: Dilakukan dengan mengonversi seluruh karakter teks menjadi huruf kecil untuk menghindari perbedaan perlakuan terhadap kata yang sama akibat variasi kapitalisasi.
2. *Text Cleaning*: Diterapkan untuk menghilangkan elemen non-tekstual yang tidak berkontribusi langsung terhadap muatan sentimen, mencakup penghapusan URL atau tautan web, penghapusan mention atau nama pengguna untuk menjaga privasi pengguna, penghapusan simbol hashtag dengan tetap mempertahankan teks hashtag karena sering merepresentasikan opini, serta penghapusan karakter khusus, emoji, dan simbol non-standar yang berpotensi menambah noise dalam pemrosesan model
3. *Normalization* : Tahap normalization bertujuan untuk mengubah varian kata menjadi bentuk standarnya,

sehingga kata-kata dengan makna yang sama namun bentuk yang berbeda dapat dianalisis secara konsisten. Proses-proses ini membantu dalam meningkatkan akurasi model NLP yang akan digunakan untuk menganalisis sentimen dari data teks yang dikumpulkan.

Pada penelitian ini tidak dilakukan penghapusan *stopword* maupun *stemming*. Keputusan tersebut diambil karena kata-kata seperti “tidak”, “sangat”, dan “kurang” memiliki peran penting dalam menentukan polaritas sentimen. Model berbasis *transformer* seperti *IndoBERT-large* dan *NusaBERT-large* dirancang untuk memahami konteks kata dalam struktur kalimat secara utuh, sehingga mempertahankan keutuhan kalimat dinilai lebih efektif dalam meningkatkan kinerja klasifikasi sentimen.

### D. Arsitektur Model dan Proses Fine-Tuning

#### 1. Arsitektur Model IndoBERT-large

Model *IndoBERT-large* yang digunakan dalam penelitian ini memiliki identifier *indobenchmark/indobert-large-pl* dan tersedia pada *Hugging Face Model Hub*. *IndoBERT-large* merupakan implementasi *BERT* berarsitektur *large* yang terdiri dari 24 *transformer layers*, *hidden size* 1024 dimensi, 16 *self-attention heads*, serta sekitar 345 juta parameter. Model ini telah dipra-latih pada korpus teks bahasa Indonesia berskala besar, sehingga mampu menangkap karakteristik linguistik bahasa Indonesia, termasuk struktur *sintaksis*, variasi *morfologi*, serta penggunaan bahasa informal yang lazim ditemukan pada media sosial.

Pemilihan varian *large* didasarkan pada kapasitas representasi model yang lebih tinggi dibandingkan varian *base*, khususnya dalam memahami nuansa semantik dan konteks implisit yang sering muncul dalam wacana kebijakan publik.

#### 2. Arsitektur Model NusaBERT-large

Sebagai pembanding, penelitian ini menggunakan *NusaBERT-large*, yaitu model *transformer* yang dikembangkan dengan tujuan menangkap keberagaman linguistik bahasa Indonesia dan bahasa daerah. *NusaBERT-large* memiliki arsitektur yang sebanding dengan *IndoBERT-large*, baik dari jumlah layer maupun dimensi representasi tersembunyi, sehingga memungkinkan perbandingan performa yang setara dari sisi kompleksitas model.

Perbedaan utama *NusaBERT* terletak pada strategi pre-training yang memanfaatkan korpus multidomain dan multibahasa, termasuk data berbahasa daerah. Karakteristik ini menjadikan *NusaBERT* relevan untuk diuji dalam konteks analisis sentimen kebijakan publik, di mana ekspresi opini masyarakat sering kali menggunakan variasi bahasa non-standar, campuran, atau kontekstual.

### 3. Strategi Fine-Tuning dan Konfigurasi Pelatihan

Proses fine-tuning dilakukan secara identik pada kedua model untuk menjaga konsistensi dan memastikan keadilan dalam perbandingan. Pada masing-masing model, ditambahkan *classification head* berupa *fully connected layer* di atas representasi token khusus [CLS]. Representasi vektor berdimensi 1024 dari token [CLS] kemudian dipetakan ke tiga kelas sentimen positif, negatif, dan netral menggunakan fungsi aktivasi *softmax*.

Konfigurasi *hyperparameter* yang digunakan dalam proses fine-tuning kedua model dirangkum pada Tabel 2 dan diterapkan secara konsisten untuk menjaga keadilan perbandingan. Pendekatan ini bertujuan untuk meminimalkan potensi bias eksperimental, sehingga perbedaan performa yang dihasilkan dapat lebih merefleksikan kemampuan intrinsik masing-masing model.

Implementasi pelatihan dilakukan menggunakan *framework PyTorch* dengan memanfaatkan *Trainer API* dari pustaka *Transformers Hugging Face*. Seluruh eksperimen dijalankan pada lingkungan *Google Colaboratory* dengan dukungan akselerator GPU untuk mempercepat proses pelatihan dan evaluasi model.

TABEL 2

KONFIGURASI HYPERPARAMETER FINE-TUNING INDOBERT-LARGE DAN NUSABERT-LARGE

Parameter	Nilai
Learning Rate	3e-5
Batch Size	16 untuk subset <i>training</i> , dan 16 untuk subset <i>validasi</i>
Maximum Sequence Length	128 token
Optimizer	AdamW dengan <i>weight decay</i> 0.01
Warmup Steps	500
Maximum Epochs	10
Early Stopping	Diaktifkan dengan <i>patience</i> 1 epoch, memonitor F1-score <i>weighted</i> pada subset <i>validasi</i> . Pelatihan dihentikan secara otomatis jika F1-score tidak menunjukkan peningkatan selama 1 epoch, dan model secara otomatis menyimpan <i>checkpoint</i> dari epoch dengan F1-score terbaik.

#### E. Integrasi dan Analisis Lintas Dimensi

Evaluasi kinerja model dilakukan secara sistematis pada setiap akhir epoch menggunakan subset *validasi*. Proses ini bertujuan untuk memantau stabilitas pelatihan serta mencegah terjadinya *overfitting*. Model dengan performa terbaik pada subset *validasi*, berdasarkan nilai *weighted F1-score*, dipilih sebagai model final untuk masing-masing arsitektur.

Evaluasi akhir terhadap model *IndoBERT-large* dan *NusaBERT-large* dilakukan menggunakan dataset *validasi* yang terdiri dari 400 *tweet* berlabel. Kinerja kedua model dianalisis menggunakan metrik evaluasi standar klasifikasi yang diturunkan dari *confusion matrix*. *Confusion matrix* digunakan untuk merepresentasikan distribusi prediksi model

terhadap kelas aktual, yang terdiri atas empat komponen utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

Berdasarkan komponen tersebut, penelitian ini menggunakan empat metrik evaluasi utama, yaitu *Akurasi*, *Presisi*, *Recall*, dan *F1-Score*. Metrik *F1-Score* digunakan sebagai acuan utama evaluasi mengingat distribusi kelas sentimen pada dataset cenderung tidak seimbang atau *imbalanced*, sehingga metrik ini lebih representatif dibandingkan akurasi semata. Secara matematis, metrik evaluasi yang digunakan dirumuskan sebagai berikut:

1. *Accuracy* : Mengukur proporsi total prediksi yang benar (baik positif maupun negatif) dari keseluruhan data uji. Formula akurasi adalah sebagai berikut:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

2. *Precision* : Mengukur tingkat ketepatan dari prediksi positif. Dari semua *tweet* yang diprediksi sebagai 'Positif' oleh model, berapa persen yang sebenarnya benar-benar 'Positif'. Formula presisi adalah sebagai berikut:

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (2)$$

3. *Recall* : Mengukur kemampuan model untuk menemukan kembali semua data positif yang ada. Dari semua *tweet* yang seharusnya 'Positif', berapa persen yang berhasil ditemukan oleh model. Formula *recall* adalah sebagai berikut:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

4. *F1-Score* : Merupakan rata-rata harmonik dari *Presisi* dan *Recall*, yang memberikan skor tunggal yang menyeimbangkan kedua metrik tersebut. *F1-Score* sangat berguna, terutama pada kasus di mana distribusi kelas sentimen mungkin tidak seimbang. Oleh karena itu, *F1-Score* akan menjadi acuan utama dalam menentukan model mana yang memiliki performa superior dalam studi komparatif ini. Formula *F1-Score* adalah sebagai berikut:

$$F1 - \text{Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (4)$$

Selain pelaporan metrik agregat, evaluasi juga dilakukan pada tingkat per kelas sentimen untuk mengidentifikasi perbedaan performa model dalam mengklasifikasikan kelas positif, negatif, dan netral. Analisis *confusion matrix* digunakan untuk memberikan gambaran yang lebih komprehensif mengenai pola kesalahan prediksi, khususnya dalam konteks ketidakseimbangan kelas dan karakteristik bahasa informal pada data opini kebijakan publik.

#### F. Integrasi Pada Dataset Lengkap

Setelah proses evaluasi selesai, model *IndoBERT-Large* dan *NusaBERT-Large* dengan performa terbaik pada subset *validasi* digunakan untuk melakukan inferensi terhadap

keseluruhan dataset yang telah dikurasi, yang terdiri atas 10.201 *tweet*. Tahapan ini bertujuan untuk memetakan distribusi sentimen publik secara menyeluruh sekaligus menguji konsistensi kinerja masing-masing model ketika diterapkan pada data yang lebih besar dan beragam.

Setiap unit teks diklasifikasikan ke dalam salah satu dari tiga kategori sentimen, yaitu positif, negatif, dan netral, berdasarkan probabilitas tertinggi yang dihasilkan oleh *softmax layer*. Proses inferensi dilakukan secara terpisah pada kedua arsitektur model guna memungkinkan perbandingan hasil prediksi secara langsung. Selanjutnya, analisis deskriptif diterapkan untuk mengidentifikasi kecenderungan dominan dalam respons masyarakat terhadap Program Makan Bergizi Gratis selama periode observasi.

Tahap inferensi ini juga dimanfaatkan untuk menelaah stabilitas prediksi model dalam menghadapi variasi bahasa informal, ekspresi implisit, serta dinamika diskursus kebijakan yang berkembang di media sosial. Perbedaan distribusi sentimen yang dihasilkan oleh *IndoBERT-Large* dan *NusaBERT-Large* kemudian dianalisis dengan mempertimbangkan karakteristik *pre-training* masing-masing model, khususnya dalam hal kemampuan menangkap nuansa semantik pada opini kebijakan publik. Temuan dari tahap ini selanjutnya menjadi dasar bagi analisis yang lebih mendalam pada bagian hasil dan pembahasan, terutama dalam mengaitkan pola sentimen publik dengan dinamika kebijakan di Indonesia.

### III. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil eksperimen serta pembahasan terhadap analisis sentimen publik mengenai Program Makan Bergizi Gratis atau MBG di platform X. Hasil yang dipaparkan mencakup distribusi sentimen masyarakat berdasarkan klasifikasi model, evaluasi performa komparatif antara *IndoBERT-Large* dan *NusaBERT-Large* menggunakan metrik akurasi, presisi, recall, dan *F1-score*, serta analisis kemampuan kedua model dalam menangani karakteristik bahasa informal dan ketidakseimbangan kelas data. Pembahasan difokuskan pada interpretasi hasil eksperimen dengan mengaitkannya pada karakteristik masing-masing model dan konteks kebijakan publik yang dianalisis.

#### A. Hasil Evaluasi Performa Model

Pada tahap ini, dilakukan evaluasi performa klasifikasi sentimen menggunakan dua model transformer berbahasa Indonesia, yaitu *IndoBERT-Large* dan *NusaBERT-Large*. Evaluasi dilakukan pada subset validasi yang terdiri dari 400 *tweet* berlabel manual atau *ground truth*. Kinerja kedua model diukur menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-score* berbobot atau *weighted average* untuk memperoleh gambaran performa yang komprehensif, khususnya dalam kondisi distribusi kelas yang tidak seimbang. Tabel 3 menyajikan perbandingan performa klasifikasi secara langsung antara *IndoBERT-Large* dan *NusaBERT-Large*.

TABEL 3  
PERBANDINGAN PERFORMA KLASIFIKASI MODEL

Metrik Evaluasi	IndoBERT-Large	NusaBERT-Large
Accuracy	83,00%	80,50%
Precision (Weighted)	83,00%	80,00%
Recall (Weighted)	83,00%	81,00%

Berdasarkan Tabel 3, *IndoBERT-Large* menunjukkan performa terbaik dengan tingkat akurasi sebesar 83,00% serta nilai *weighted F1-score* sebesar 83,00%. Konsistensi nilai *precision* dan *recall* berbobot pada model ini menunjukkan kemampuan yang seimbang dalam mengklasifikasikan seluruh kategori sentimen, termasuk kelas minoritas.

*NusaBERT-Large* memperoleh akurasi sebesar 80,50% dengan *weighted F1-score* sebesar 80,00%, yang menunjukkan performa yang tetap kompetitif meskipun berada sedikit di bawah *IndoBERT-Large*. Selisih performa sebesar 2,5% mengindikasikan adanya perbedaan efektivitas representasi semantik antara kedua model dalam menangani bahasa informal dan konteks kebijakan publik di media sosial. Meskipun demikian, hasil ini menegaskan bahwa kedua model *transformer* mampu mencapai kinerja yang stabil pada tugas klasifikasi sentimen berbasis data kebijakan publik berbahasa Indonesia.

#### B. Analisis Metrik Klasifikasi per Kelas Sentimen

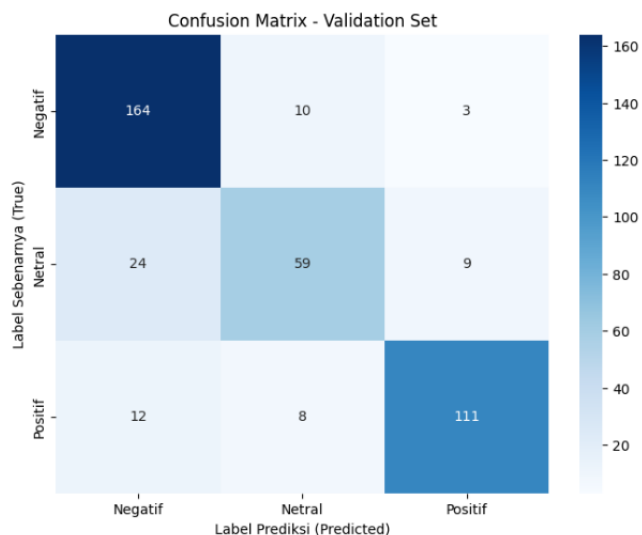
Setelah mengevaluasi performa model secara keseluruhan, analisis selanjutnya difokuskan pada kinerja masing-masing model dalam mengklasifikasikan setiap kategori sentimen. Pendekatan ini penting untuk mengidentifikasi potensi bias model terhadap kelas tertentu serta menilai konsistensi performa pada data yang memiliki distribusi kelas tidak seimbang. Tabel 4 menyajikan perbandingan nilai *F1-score* per kelas sentimen antara *IndoBERT-Large* dan *NusaBERT-Large*.

TABEL 4  
PERBANDINGAN F1-SCORE PER KELAS SENTIMEN

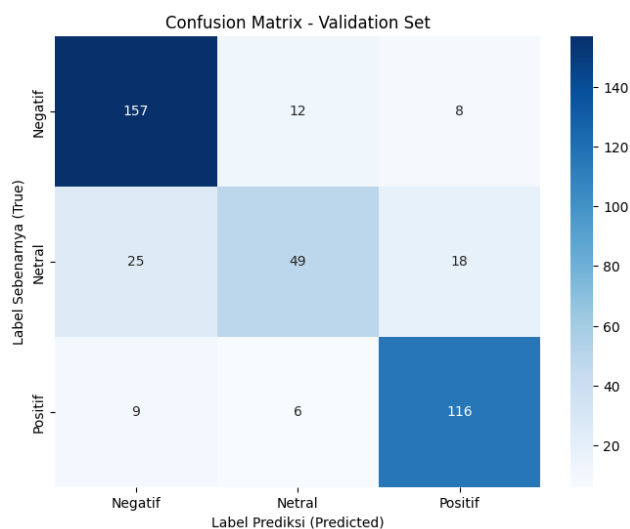
Kategori Sentimen	IndoBERT-Large (F1-Score)	NusaBERT-Large (F1-Score)
Negatif	0.87	0.85
Netral	0.70	0.62
Positif	0.87	0.85

Analisis terhadap performa per kelas serta visualisasi *confusion matrix* pada Gambar 2 dan Gambar 3 menunjukkan beberapa temuan penting. Pertama, kedua model menunjukkan performa yang sangat kuat dalam mengidentifikasi sentimen negatif. *IndoBERT-Large*, khususnya, memiliki tingkat *recall* yang tinggi pada kelas negatif, yang mengindikasikan kemampuannya dalam menangkap sebagian besar ekspresi kritik dan kekhawatiran masyarakat terhadap program MBG.

Kedua, pada kelas sentimen positif, *NusaBERT-Large* menunjukkan nilai *recall* yang sedikit lebih tinggi dibandingkan *IndoBERT-Large*. Temuan ini mengindikasikan bahwa *NusaBERT-Large* relatif lebih sensitif dalam mengenali ekspresi dukungan atau apresiasi masyarakat, meskipun secara keseluruhan performa agregatnya masih berada di bawah *IndoBERT-Large*.



Gambar 2. Confusion Matrix pada Validation Set IndoBERT



Gambar 3. Confusion Matrix pada Validation Set NusaBERT

Tantangan utama bagi kedua model terlihat pada Gambar 2 dan Gambar 3 pada klasifikasi sentimen netral. Kedua *Confusion matrix* menunjukkan bahwa sebagian tweet netral cenderung salah diklasifikasikan sebagai sentimen negatif. Fenomena ini dapat dijelaskan oleh karakteristik diskursus di platform X, di mana penyampaian informasi faktual sering kali menggunakan terminologi bernuansa ekonomi atau fiskal, seperti “anggaran” dan “biaya”, yang secara semantik berdekatan dengan wacana kritik kebijakan.

### C. Distribusi Sentimen Publik terhadap Program Makan Bergizi Gratis

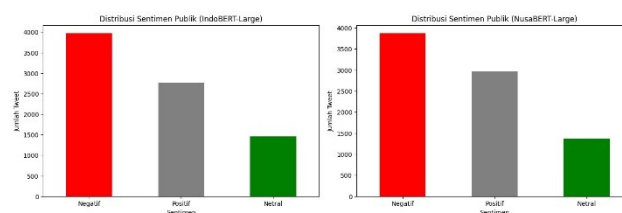
Setelah melalui tahap validasi, model *IndoBERT-Large* dan *NusaBERT-Large* digunakan untuk melakukan inferensi terhadap keseluruhan dataset yang telah dikurasi guna memetakan persepsi publik secara lebih luas. Tahapan ini bertujuan untuk menilai konsistensi prediksi model serta mengidentifikasi pola dominan opini masyarakat terhadap Program Makan Bergizi Gratis (MBG) di platform X. Tabel 5 menyajikan perbandingan distribusi sentimen yang dihasilkan oleh kedua model.

TABEL 5  
HASIL PENCARIAN K OPTIMAL

Kategori Sentimen	Prediksi IndoBERT-Large (%)	Prediksi NusaBERT-Large (%)
Negatif	48,42%	47,23%
Positif	33,81%	36,12%
Netral	17,77%	16,65%

Berdasarkan Tabel 5, terlihat adanya konsistensi yang relatif tinggi antara hasil prediksi *IndoBERT-Large* dan *NusaBERT-Large*. Kedua model menunjukkan bahwa sentimen negatif mendominasi diskursus publik mengenai Program MBG dengan proporsi mendekati 50%. Temuan ini mengindikasikan bahwa perbincangan di platform X lebih banyak diwarnai oleh ekspresi kritik, kekhawatiran terkait alokasi anggaran, serta isu implementasi kebijakan di lapangan.

Meskipun demikian, persentase sentimen positif tetap menunjukkan nilai yang signifikan, berada pada kisaran 33% hingga 36%. *NusaBERT-Large* menghasilkan proporsi sentimen positif yang sedikit lebih tinggi dibandingkan *IndoBERT-Large*. Pola ini sejalan dengan hasil analisis pada subbab sebelumnya, yang menunjukkan bahwa *NusaBERT-Large* memiliki sensitivitas yang lebih baik dalam mengenali ekspresi dukungan atau apresiasi masyarakat.



Gambar 4. Grafik Batang Distribusi Sentimen Publik Berdasarkan Model IndoBERT-Large dan NusaBERT-Large

Secara keseluruhan, distribusi sentimen pada gambar 4 ini menunjukkan bahwa opini publik terhadap Program MBG masih bersifat terpolarisasi, dengan kecenderungan resistensi yang lebih besar dibandingkan dukungan eksplisit. Proporsi sentimen netral yang berada pada kisaran 16–17% mencerminkan keberadaan kelompok masyarakat yang lebih berperan sebagai penyampai informasi atau pengamat kebijakan tanpa memberikan penilaian sentimen yang jelas. Temuan ini menjadi dasar penting untuk pembahasan lebih

lanjut mengenai dinamika opini publik dan implikasinya terhadap kebijakan publik.

#### D. Pembahasan dan Interpretasi Temuan

Bagian ini mengintegrasikan temuan teknis dari eksperimen pemodelan dengan konteks diskursus publik mengenai Program Makan Bergizi Gratis atau MBG. Pendekatan ini memungkinkan interpretasi yang lebih komprehensif mengenai alasan di balik perbedaan kinerja model serta implikasinya terhadap pemahaman opini masyarakat terhadap kebijakan publik.

Hasil eksperimen menunjukkan bahwa *IndoBERT-Large* memiliki performa yang lebih unggul secara agregat dibandingkan *NusaBERT-Large*, dengan tingkat akurasi mencapai 83,00%. Keunggulan ini diduga berkaitan erat dengan proses *pre-training* *IndoBERT-Large* pada korpus bahasa Indonesia formal dalam skala besar. Diskursus mengenai kebijakan MBG banyak memuat terminologi administratif dan teknokratis, seperti anggaran, stunting, dan implementasi, yang relatif lebih selaras dengan karakteristik data pelatihan *IndoBERT-Large*.

Sebaliknya, *NusaBERT-Large* menunjukkan sensitivitas yang lebih tinggi dalam mengidentifikasi sentimen positif, sebagaimana tercermin dari nilai recall pada kelas positif. Temuan ini mengindikasikan bahwa ekspresi dukungan masyarakat di platform X cenderung disampaikan menggunakan bahasa non-formal, ekspresif, dan kontekstual, yang lebih efektif ditangkap oleh *NusaBERT-Large* melalui keragaman korpus *pre-training* yang mencakup variasi bahasa dan dialek.

Dominasi sentimen negatif, yang mencapai sekitar 48% dari keseluruhan prediksi, mencerminkan adanya tingkat kehati-hatian dan kekhawatiran publik yang signifikan terhadap pelaksanaan Program MBG. Analisis kualitatif terhadap konten tweet menunjukkan bahwa sentimen negatif tersebut tidak selalu merepresentasikan penolakan terhadap tujuan program, melainkan lebih terfokus pada isu-isu implementatif. Isu tersebut meliputi kekhawatiran terhadap beban fiskal negara, keraguan terhadap mekanisme distribusi di wilayah terpencil, serta respons terhadap kasus-kasus insidental yang viral di media sosial.

Di sisi lain, proporsi sentimen positif yang berada pada kisaran 33–36% menunjukkan adanya basis dukungan publik yang cukup kuat. Kelompok ini umumnya memandang Program MBG sebagai investasi jangka panjang dalam peningkatan kualitas sumber daya manusia dan penguatan ekonomi lokal. Temuan ini mengindikasikan bahwa tingkat penerimaan publik terhadap program relatif tinggi, namun masih sangat dipengaruhi oleh kejelasan implementasi dan komunikasi kebijakan.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa model berbasis *Transformer*, khususnya *IndoBERT-Large* dan *NusaBERT-Large*, mampu memberikan representasi kuantitatif yang reliabel terhadap dinamika opini publik di media sosial. Temuan ini menegaskan pentingnya pemanfaatan analisis sentimen berbasis NLP sebagai

instrumen pendukung evaluasi kebijakan publik, terutama dalam merespons kekhawatiran masyarakat melalui peningkatan transparansi dan kualitas komunikasi publik.

#### IV. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap Program Makan Bergizi Gratis atau MBG di platform X serta membandingkan performa dua model *transformer* berbahasa Indonesia, yaitu *IndoBERT-Large* dan *NusaBERT-Large*. Berdasarkan hasil eksperimen dan analisis yang telah dilakukan, dapat disimpulkan bahwa kedua model mampu mengklasifikasikan sentimen publik secara efektif pada konteks diskursus kebijakan publik berbahasa Indonesia.

Hasil evaluasi menunjukkan bahwa *IndoBERT-Large* memiliki performa yang lebih unggul secara agregat dengan tingkat akurasi dan *weighted F1-score* sebesar 83,00%, sementara *NusaBERT-Large* mencapai performa yang kompetitif dengan akurasi 80,50%. Analisis per kelas sentimen mengungkap bahwa kedua model sangat andal dalam mendeteksi sentimen negatif, sedangkan *NusaBERT-Large* menunjukkan sensitivitas yang relatif lebih tinggi dalam mengidentifikasi sentimen positif. Tantangan utama bagi kedua model terletak pada klasifikasi sentimen netral, yang sering kali tumpang tindih secara semantik dengan sentimen negatif dalam konteks bahasa media sosial.

Inferensi terhadap keseluruhan dataset menunjukkan bahwa diskursus publik mengenai Program MBG didominasi oleh sentimen negatif, diikuti oleh sentimen positif dan netral. Dominasi sentimen negatif mencerminkan adanya kekhawatiran dan kritik publik yang signifikan, terutama terkait aspek implementasi dan fiskal, sementara keberadaan sentimen positif yang cukup besar menunjukkan adanya dukungan publik yang tidak dapat diabaikan.

Secara keseluruhan, penelitian ini menegaskan bahwa pendekatan analisis sentimen berbasis model *transformer*, khususnya *IndoBERT-Large* dan *NusaBERT-Large*, dapat digunakan sebagai instrumen yang andal untuk memetakan opini publik terhadap kebijakan pemerintah di media sosial. Temuan ini juga menunjukkan bahwa pemilihan model yang sesuai dengan karakteristik bahasa dan konteks data sangat berpengaruh terhadap hasil analisis sentimen.

#### DAFTAR PUSTAKA

- [1] “Presiden Prabowo Targetkan Seluruh Anak Dapat Akses Makan Bergizi Gratis pada Akhir 2025.” Accessed: Sep. 22, 2025. [Online]. Available: <https://www.presidentri.go.id/siaran-pers/presiden-prabowo-targetkan-seluruh-anak-dapat-akses-makan-bergizi-gratis-pada-akhir-2025/>
- [2] “Indonesia.go.id - Ini Tiga Skema Penyaluran Makan Bergizi Gratis.” Accessed: Sep. 22, 2025. [Online]. Available: <https://indonesia.go.id/kategori/editorial/8750/ini-tiga-skema-penyaluran-makan-bergizi-gratis?lang=1>
- [3] H. K. P. RI, “Menko Polkam: Program Makan Bergizi Gratis Terus Diperluas, Ketahanan Gizi Fondasi Ketahanan Nasional,” Kemenko Polkam R.I. Accessed: Sep. 22, 2025. [Online]. Available: <https://polkam.go.id/menko-polkam-program-makan-bergizi-gratis-terus-diperluas-ketahanan-gizi-fondasi-ketahanan-nasional/>

- [4] "Program Makan Bergizi Gratis (MBG): Dinamika dan Sorotan - Media Keuangan." Accessed: Sep. 22, 2025. [Online]. Available: <https://mediakeuangan.kemenkeu.go.id/article/show/program-makan-bergizi-gratis-mbg-dinamika-dan-sorotan>
- [5] "X Users, Stats, Data, Trends, and More," DataReportal – Global Digital Insights. Accessed: Sep. 04, 2025. [Online]. Available: <https://datareportal.com/essential-x-stats>
- [6] Y. A. Singgalen, "Performance Analysis of IndoBERT for Sentiment Classification in Indonesian Hotel Review Data," *J. Inf. Syst. Res. JOSH*, vol. 6, no. 2, pp. 976–986, Jan. 2025, doi: 10.47065/josh.v6i2.6505.
- [7] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press, 2020.
- [8] "Hybrid Models for Emotion Classification and Sentiment Analysis in Indonesian Language - Ahmadian - 2024 - Applied Computational Intelligence and Soft Computing - Wiley Online Library." Accessed: Sep. 22, 2025. [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1155/2024/2826773>
- [9] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," Nov. 02, 2020, arXiv: arXiv:2011.00677. doi: 10.48550/arXiv.2011.00677.
- [10] L. P. Mahadyta, "Analisis Sentiman Review Aplikasi Kesehatan Indonesia di Googele Play Store dengan Indobert," Thesis, Universitas Islam Indonesia, 2024. Accessed: Sep. 22, 2025. [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/50297>
- [11] P. Sayarizki, *Analisis Sentimen Tweets Terhadap Bakal Calon Presiden Indonesia 2024 Menggunakan Metode IndoBERT - Dalam bentuk pengganti sidang - Artikel Jurnal*. Universitas Telkom, S1 Informatika, 2024. Accessed: Sep. 22, 2025. [Online]. Available: <https://repositori.telkomuniversity.ac.id/pustaka/218165/analisis-sentimen-tweets-terhadap-bakal-calon-presiden-indonesia-2024-menggunakan-metode-indobert-dalam-bentuk-pengganti-sidang-artikel-jurnal.html>
- [12] H. Jayadianti, W. Kaswidjanti, A. Utomo, S. Saifullah, F. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," *Ilk. J. Ilm.*, vol. 14, pp. 348–354, Dec. 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.
- [13] T. D. Purnomo and J. Sutopo, "Comparison Of Pre-Trained Bert-Based Transformer Models For Regional Language Text Sentiment Analysis In Indonesia," *Int. J. Sci. Technol.*, vol. 3, no. 3, pp. 11–21, Nov. 2024, doi: 10.56127/ijst.v3i3.1739.
- [14] M. I. Arsyam, "Analisis Sentimen Dan Ringkasan Ulasan Aplikasi Access By Kai Menggunakan Nusabert Dan Llm Qwen3," other, Universitas Duta Bangsa Surakarta, 2025. Accessed: Dec. 22, 2025. [Online]. Available: <https://eprints.udb.ac.id/id/eprint/4783/>
- [15] D. Ramdani, "Perbandingan kinerja IndoBERT, IndoRoBERTa, dan NusaBERT dalam analisis sentimen isu LGBT di media sosial X," other, UIN Sunan Gunung Djati Bandung, 2025. doi: 10.10.%20DAFTAR%20PUSTAKA.pdf.