

Analisis Dinamika Sentimen Publik terhadap Program Makan Bergizi Gratis di Platform X Menggunakan Model IndoBERTweet

DAFFA FAZLY RASHIDAN, ATIK WINTARTI*

Program Studi Sarjana Sains Data, Universitas Negeri Surabaya, Indonesia

Abstrak

Program Makan Bergizi Gratis (MBG) merupakan kebijakan publik yang memunculkan diskursus dan respons beragam di Platform X. Kompleksitas bahasa informal, penggunaan slang, serta ekspresi evaluatif dalam cuitan menjadi tantangan dalam klasifikasi sentimen. Penelitian ini menganalisis dinamika sentimen publik terhadap Program MBG menggunakan model IndoBERTweet pada periode 28 November 2023 hingga 31 Maret 2025 yang dibagi ke dalam tiga fase, yaitu fase kampanye, fase transisi dan uji coba, serta fase implementasi awal. Pendekatan yang digunakan mencakup strategi *minimalist filtering* untuk mempertahankan elemen teks yang berpotensi mengandung informasi kontekstual dan penerapan *class weighting* untuk mengurangi pengaruh ketidakseimbangan distribusi kelas pada proses *fine-tuning*. Evaluasi pada data uji menghasilkan akurasi sebesar 85,50% dan *Weighted F1-Score* sebesar 0,8563. Berdasarkan klasifikasi terhadap 126.365 cuitan aktual, sentimen negatif merupakan kategori dominan dengan proporsi 60,6%, diikuti sentimen positif sebesar 26,6% dan netral sebesar 12,8%. Analisis temporal menunjukkan bahwa proporsi sentimen negatif menurun dari 72,0% pada fase kampanye menjadi 63,9% pada fase transisi dan uji coba, kemudian menjadi 54,8% pada fase implementasi awal. Hasil analisis N-gram menunjukkan bahwa perubahan pola sentimen berkaitan dengan pergeseran fokus diskursus dari aspek politik dan pendanaan menuju pembahasan mengenai pelaksanaan dan konsekuensi operasional program. Temuan ini menunjukkan adanya perubahan karakter diskursus publik selama periode pengamatan, meskipun interpretasi terhadap perubahan sentimen perlu mempertimbangkan karakteristik data media sosial dan keterbatasan metode klasifikasi.

Kata Kunci *Analisis Sentimen; IndoBERTweet; Kebijakan Publik; Makan Bergizi Gratis; Platform X*

Abstract

The Free Nutritious Meal (MBG) program is a public policy that has generated diverse discussions and responses on Platform X. The complexity of informal language, the use of slang, and evaluative expressions in tweets present challenges for sentiment classification. This study analyzes public sentiment dynamics toward the MBG program using the IndoBERTweet model over the period from 28 November 2023 to 31 March 2025, divided into three phases: the campaign phase, the transition and pilot phase, and the early implementation phase. The approach incorporates a *minimalist filtering* strategy to preserve textual elements that may contain contextual information and *class weighting* to reduce the influence of class imbalance during *fine-tuning*. Evaluation on the test set yielded an accuracy of 85.50% and a *Weighted F1-Score* of 0.8563. Based on classification of 126,365 actual tweets, negative sentiment was the dominant category at 60.6%, followed by

*Corresponding author. Email: daffa.22006@mhs.unesa.ac.id or atikwintarti@unesa.ac.id.

positive sentiment at 26.6% and neutral sentiment at 12.8%. Temporal analysis showed that the proportion of negative sentiment decreased from 72.0% during the campaign phase to 63.9% during the transition and pilot phase, and subsequently to 54.8% during the early implementation phase. N-gram analysis indicates that changes in sentiment patterns were associated with a shift in public discourse from political and funding-related issues toward discussions concerning the implementation and operational consequences of the program. These findings indicate a change in the characteristics of public discourse during the observation period, although the interpretation of sentiment changes should consider the characteristics of social media data and the limitations of the classification method.

Keywords *Sentiment Analysis; IndoBERTweet; Public Policy; Free Nutritious Meal; Platform X*

1 Pendahuluan

Kualitas Sumber Daya Manusia (SDM) merupakan elemen fundamental dalam menyongsong visi Indonesia Emas 2045, di mana pemenuhan gizi anak usia sekolah memegang peranan penting bagi perkembangan kognitif dan produktivitas jangka panjang. Berdasarkan data Survei Status Gizi Indonesia (SSGI) 2024, prevalensi stunting nasional telah turun menjadi 19,8%, namun angka tersebut masih berada di sekitar ambang batas 20% yang digunakan oleh *World Health Organization* (WHO) of BKPK (2025). Pemerintah merespons kondisi tersebut melalui inisiasi Program Makan Bergizi Gratis (MBG) yang diproyeksikan sebagai salah satu kebijakan untuk mendukung peningkatan kualitas sumber daya manusia Setpres (2025). Efektivitas kebijakan publik tidak hanya berkaitan dengan pelaksanaan teknis, tetapi juga dengan penerimaan dan persepsi masyarakat terhadap kebijakan tersebut. Oleh karena itu, pemetaan opini publik dapat memberikan informasi tambahan mengenai respons masyarakat selama proses perkembangan suatu kebijakan.

Platform X menjadi salah satu ruang digital yang memungkinkan masyarakat menyampaikan respons terhadap isu kebijakan secara langsung. Data dari platform tersebut memiliki karakteristik bahasa yang beragam, termasuk penggunaan bahasa tidak baku, singkatan, slang, ekspresi emosional, serta bentuk komunikasi yang bergantung pada konteks Ardiani dkk. (2020). Kondisi tersebut menjadi tantangan bagi analisis sentimen karena makna suatu teks tidak selalu dapat ditentukan hanya berdasarkan keberadaan kata-kata tertentu. Oleh karena itu, pendekatan berbasis *machine learning* dan *deep learning* dapat digunakan untuk mempelajari hubungan kontekstual dalam teks Islam dkk. (2024).

Sejumlah penelitian telah mengkaji sentimen masyarakat terhadap Program MBG dengan pendekatan yang berbeda. Arif dan Kustiyono (2025) menggunakan pendekatan berbasis leksikon melalui TextBlob dan menemukan dominasi sentimen negatif. Pendekatan berbasis leksikon memiliki keterbatasan ketika berhadapan dengan kata atau frasa yang maknanya bergantung pada konteks. Penelitian Munir (2025) menerapkan *Support Vector Machine* (SVM) dan memperoleh kinerja klasifikasi yang memadai. Meskipun demikian, pendekatan tersebut masih bergantung pada representasi fitur yang ditentukan sebelum proses klasifikasi dan belum secara eksplisit memodelkan hubungan kontekstual antarkata dalam struktur kalimat.

Perkembangan *Natural Language Processing* (NLP) selanjutnya mendorong penggunaan arsitektur *Transformer*, termasuk keluarga BERT, untuk memperoleh representasi teks yang mempertimbangkan konteks dua arah Vaswani dkk. (2017); Devlin dkk. (2019). IndoBERT merupakan salah satu model yang dikembangkan untuk teks berbahasa Indonesia, sedangkan IndoBERTweet

dikembangkan dengan karakteristik data media sosial sebagai salah satu pertimbangan utama dalam pembentukan model Koto dkk. (2021). Karakteristik tersebut menjadikan IndoBERTweet relevan untuk digunakan dalam klasifikasi teks pada Platform X.

Penelitian ini berfokus pada dua aspek utama. Aspek pertama adalah evaluasi kemampuan IndoBERTweet dalam melakukan klasifikasi sentimen pada teks berbahasa Indonesia yang berasal dari beragam sumber data pelatihan. Aspek kedua adalah analisis perubahan sentimen publik terhadap Program MBG selama 17 bulan, yaitu sejak 28 November 2023 hingga 31 Maret 2025. Periode tersebut dibagi menjadi tiga fase, yaitu fase kampanye (28 November 2023–31 Januari 2024), fase transisi dan uji coba (1 Februari–31 Desember 2024), serta fase implementasi awal (1 Januari–31 Maret 2025). Pembagian tersebut memungkinkan perubahan distribusi sentimen dibandingkan berdasarkan tahapan perkembangan kebijakan, bukan hanya berdasarkan satu periode pengamatan.

2 Metode

2.1 Dataset

Penelitian ini menggunakan dua kelompok *dataset* dengan tujuan berbeda, yaitu *dataset* pelatihan untuk membangun model klasifikasi dan *dataset* objek penelitian untuk menganalisis sentimen publik terhadap Program MBG. *Dataset* pelatihan digunakan untuk proses *fine-tuning* model IndoBERTweet. Data pelatihan berasal dari beberapa sumber dengan topik dan karakteristik bahasa yang berbeda. Penggabungan beberapa sumber dilakukan untuk memperoleh variasi data yang lebih luas sebelum proses pelatihan model. Rincian jumlah data mentah dari masing-masing sumber ditampilkan pada Tabel 1.

Tabel 1: Rincian Sumber *Dataset* Pelatihan

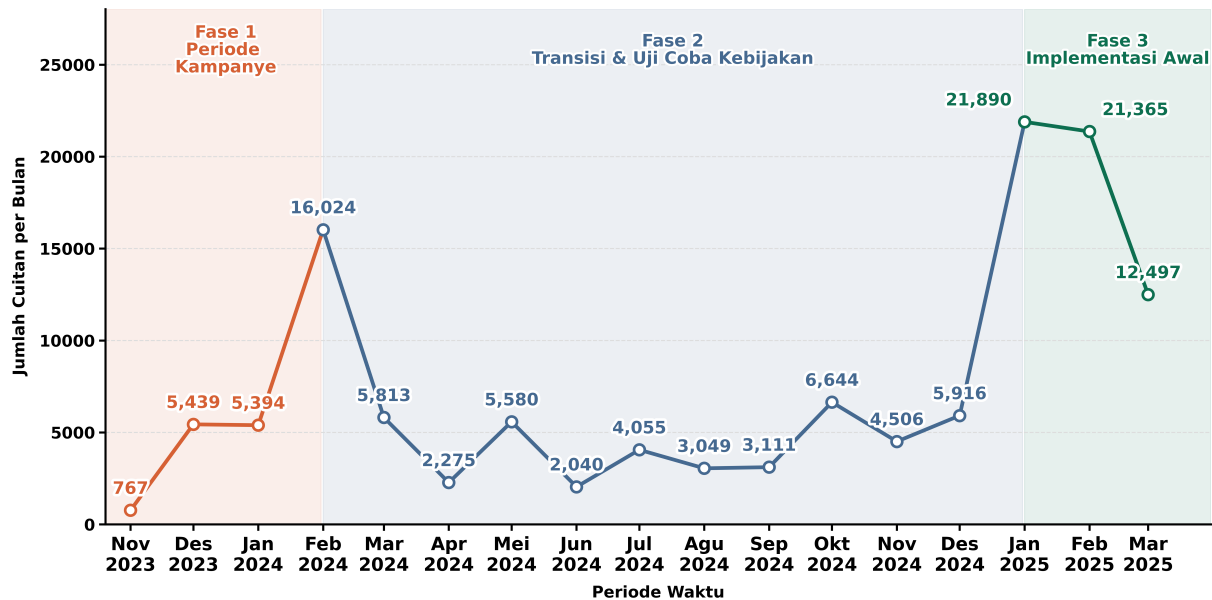
Judul Penelitian	Topik	Metode Utama	Jumlah
Twitter Emotion <i>dataset</i> Saputri dkk. (2018)	Emosi & Psikologi	Konversi label emosi multi-kelas menjadi sentimen Positif dan Negatif.	4.401
Instagram Cyberbullying Luqyana dkk. (2018)	Komentar Sosial	Pengambilan kolom teks komentar dan label sentimen.	400
Cellular Service Provider Rofiqoh dkk. (2017)	Layanan Publik	Pengambilan kolom teks cuitan dan label sentimen.	300
Pilkada DKI 2017 Rossi dkk. (2017)	Politik	Pengambilan kolom teks cuitan dan label sentimen.	900
PPKM Tweets Widiarta (2023)	Kebijakan Publik	Pengambilan kolom teks cuitan dan label sentimen.	23.644
<i>dataset</i> Makan Siang Gratis	Makan Siang Gratis	Pengambilan kolom teks cuitan dan label sentimen.	656
Total Keseluruhan			30.301

Dataset berasal dari sumber dengan karakteristik dan konvensi pelabelan yang berbeda. Oleh karena itu, dilakukan harmonisasi label dan integrasi data agar seluruh korpus memiliki struktur kelas yang konsisten sebelum digunakan dalam proses *fine-tuning*. Tahapan pra-pemrosesan diterapkan dengan mempertahankan elemen teks yang berpotensi memberikan informasi kontekstual

bagi model. Setelah melalui tahapan pra-pemrosesan, diperoleh *dataset* pelatihan bersih sebanyak 26.250 data. Distribusi label menunjukkan ketidakseimbangan kelas, dengan sentimen netral sebesar 53,5% (14.049 data), sentimen negatif sebesar 29,9% (7.849 data), dan sentimen positif sebesar 16,6% (4.352 data).

Dataset objek penelitian merupakan kumpulan cuitan yang digunakan untuk menganalisis sentimen publik terhadap Program MBG. Data dihimpun melalui teknik *web scraping* pada Platform X dan dibatasi pada cuitan berbahasa Indonesia dalam periode 28 November 2023 hingga 31 Maret 2025. Kueri yang digunakan adalah “makan siang gratis” OR “makan bergizi gratis” OR “#mbg” OR “mbg”. Kueri tersebut digunakan untuk mencakup variasi istilah yang muncul selama perkembangan wacana kebijakan. Selain itu, digunakan parameter eksklusi untuk mengurangi kemunculan konten yang tidak berkaitan dengan topik penelitian.

Dari proses pengumpulan tersebut diperoleh 137.452 cuitan mentah. Setelah proses penyaringan dan pra-pemrosesan, sebanyak 126.365 cuitan digunakan sebagai data objek penelitian.

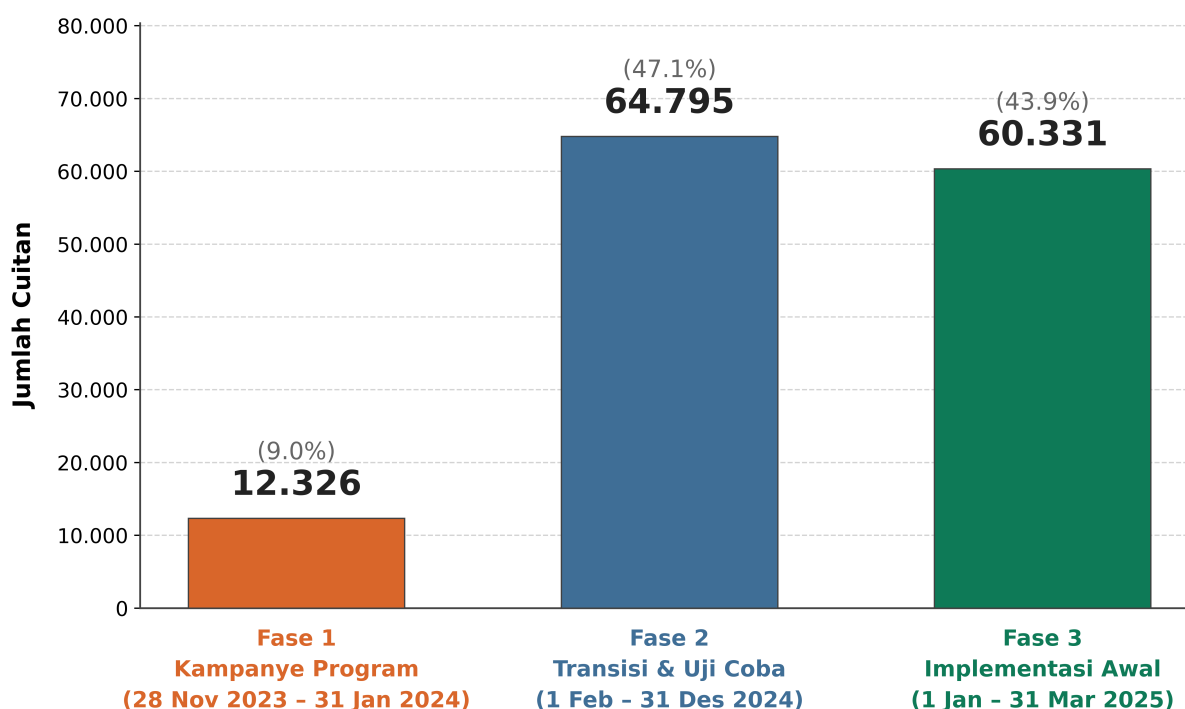


Gambar 1: Tren Volume Cuitan Terkait Program Makan Bergizi Gratis (November 2023–Maret 2025)

Dataset objek penelitian dibagi menjadi tiga fase temporal:

1. **Fase 1 (Kampanye):** 28 November 2023 hingga 31 Januari 2024.
2. **Fase 2 (Transisi dan Uji Coba):** 1 Februari 2024 hingga 31 Desember 2024.
3. **Fase 3 (Implementasi Awal):** 1 Januari 2025 hingga 31 Maret 2025.

Pembagian tersebut digunakan sebagai kerangka analisis temporal untuk membandingkan perubahan distribusi sentimen pada tahapan perkembangan kebijakan.



Gambar 2: Distribusi Kuantitas Data Cuitan Berdasarkan Pembagian Fase Temporal

2.2 Alur Penelitian

Penelitian ini menggunakan alur kerja yang terdiri atas pengolahan *dataset* pelatihan dan pengolahan *dataset* objek penelitian. *dataset* pelatihan melalui tahapan pra-pemrosesan, pembagian data, penyeimbangan kelas, tokenisasi, dan *fine-tuning*. Model yang telah dilatih kemudian digunakan untuk melakukan klasifikasi terhadap *dataset* objek penelitian.

Pada tahap pra-pemrosesan, elemen teks yang berpotensi memberikan informasi kontekstual dipertahankan. Pendekatan ini berbeda dari strategi pembersihan teks yang menghapus sebagian besar elemen non-leksikal. Pertimbangan tersebut didasarkan pada karakteristik arsitektur *Transformer* yang memanfaatkan hubungan kontekstual antarelemen dalam suatu urutan teks.

Data pelatihan kemudian dibagi menjadi data latih, validasi, dan uji dengan rasio 70:15:15. Ketidakseimbangan kelas ditangani menggunakan *class weighting* pada proses pelatihan. Setelah proses *fine-tuning*, model dievaluasi menggunakan data uji yang tidak digunakan selama pelatihan.

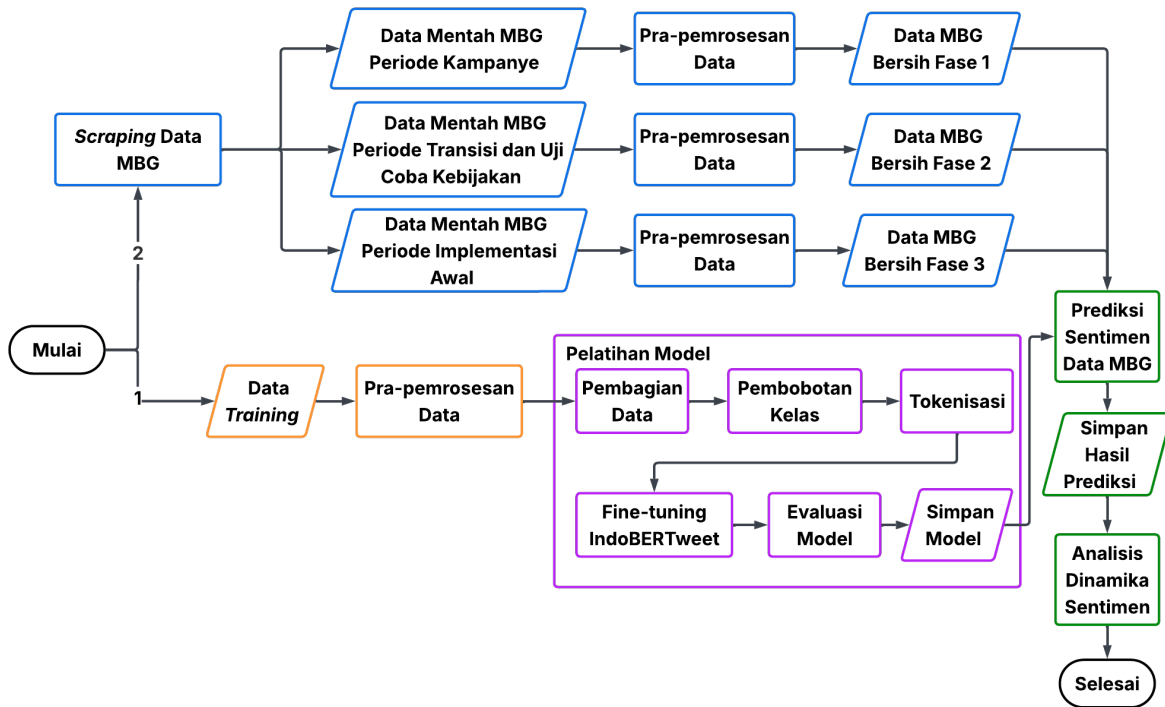
dataset objek penelitian diproses menggunakan tahapan pra-pemrosesan yang sama dan kemudian diklasifikasikan menggunakan model yang telah dilatih. Hasil klasifikasi diagregasi berdasarkan waktu dan fase kebijakan untuk memperoleh distribusi sentimen serta pola perubahan sentimen.

2.3 Arsitektur IndoBERTweet

Pendekatan klasifikasi sentimen dalam penelitian ini menggunakan arsitektur berbasis *Transformer* Vaswani dkk. (2017). Salah satu komponen utama arsitektur tersebut adalah mekanisme *Scaled Dot-Product Attention* yang digunakan untuk memperoleh representasi berdasarkan hubungan antar-token. Mekanisme tersebut dirumuskan sebagai:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (1)$$

di mana Q (*Query*), K (*Key*), dan V (*Value*) merupakan representasi yang digunakan dalam proses atensi, sedangkan d_k merupakan dimensi vektor *Key*.



Gambar 3: Diagram Alur Penelitian

BERT (*Bidirectional Encoder Representations from Transformers*) menggunakan representasi kontekstual dua arah sehingga hubungan antarkata dapat dipelajari berdasarkan konteks sebelum dan sesudah token [Devlin dkk. \(2019\)](#). Pendekatan ini berbeda dengan metode yang hanya mengandalkan frekuensi kata atau fitur leksikal yang telah ditentukan sebelumnya.

IndoBERTweet merupakan model yang dikembangkan untuk karakteristik teks media sosial berbahasa Indonesia dan menggunakan arsitektur RoBERTa (*Robustly Optimized BERT Pretraining Approach*) [Koto dkk. \(2021\)](#). Penggunaan model yang dilatih pada korpus cuitan memberikan dasar yang relevan untuk penelitian yang menggunakan data dari Platform X. Dalam penelitian ini, IndoBERTweet digunakan sebagai model dasar yang kemudian disesuaikan melalui proses *fine-tuning* menggunakan *dataset* pelatihan yang telah disiapkan.

2.4 Metrik Evaluasi

Evaluasi kinerja bertujuan untuk mengukur kemampuan model dalam mengklasifikasikan sentimen pada data yang tidak digunakan selama proses pelatihan. Evaluasi dilakukan menggunakan *Confusion Matrix* yang terdiri atas *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) [Grandini dkk. \(2020\)](#).

True Positive menunjukkan jumlah data suatu kelas yang berhasil diprediksi sebagai kelas tersebut. *True Negative* menunjukkan data dari kelas lain yang berhasil tidak diklasifikasikan sebagai kelas yang sedang dievaluasi. Sebaliknya, *False Positive* menunjukkan data dari kelas lain yang keliru diprediksi sebagai kelas tersebut, sedangkan *False Negative* menunjukkan data yang sebenarnya termasuk kelas tersebut tetapi diprediksi sebagai kelas lain.

Kinerja model kemudian diukur menggunakan empat metrik utama:

1. **Akurasi (*Accuracy*)**: Mengukur proporsi seluruh prediksi yang benar terhadap keseluruhan data pengujian.

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

2. **Presisi (*Precision*)**: Mengukur proporsi prediksi positif yang benar terhadap seluruh prediksi yang dikategorikan sebagai positif.

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

3. **Recall**: Mengukur proporsi data positif aktual yang berhasil dikenali oleh model.

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

4. **F1-Score**: Merupakan rata-rata harmonik antara *precision* dan *recall*.

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

3 Hasil dan Pembahasan

3.1 Evaluasi Performa Model

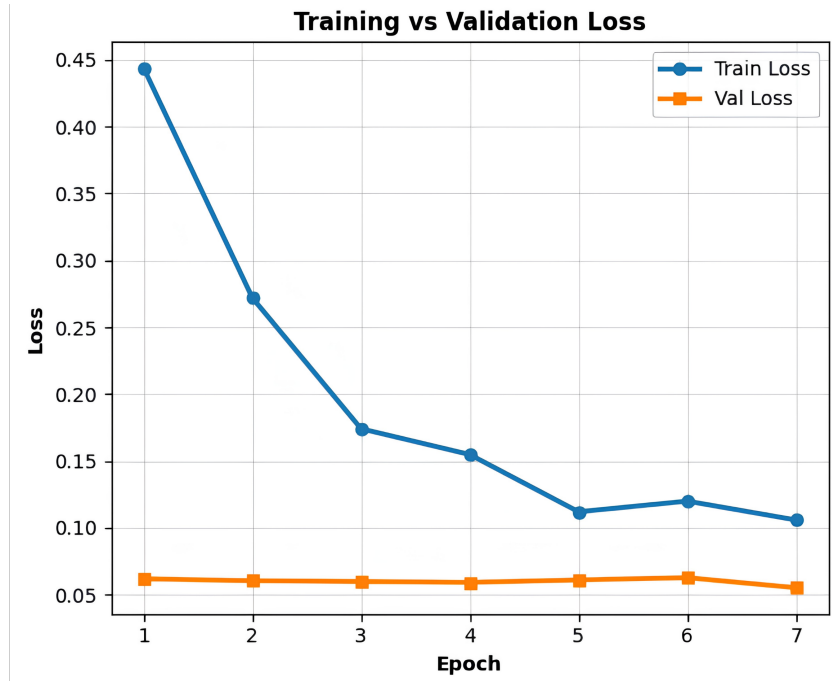
Stabilitas proses *fine-tuning* dievaluasi melalui perubahan nilai *training loss* dan *validation loss* pada setiap *epoch*. Hasil proses tersebut ditampilkan pada Gambar 4.

Berdasarkan Gambar 4, kedua kurva menunjukkan kecenderungan menurun selama proses pelatihan dan tidak memperlihatkan peningkatan jarak yang tajam pada tahap akhir. Pola tersebut menunjukkan bahwa proses pelatihan berlangsung relatif stabil pada konfigurasi yang digunakan. Namun, pola kurva *loss* saja belum cukup untuk menyimpulkan bahwa model sepenuhnya bebas dari *overfitting*. Oleh karena itu, penilaian generalisasi model dilakukan melalui evaluasi terhadap data uji yang tidak digunakan selama proses pelatihan.

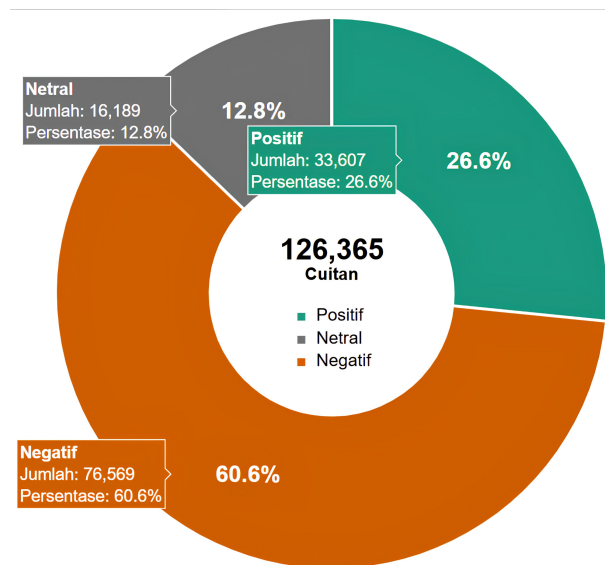
Evaluasi akhir dilakukan menggunakan 3.937 data uji. Hasil klasifikasi divisualisasikan melalui *Confusion Matrix* pada Gambar 5, sedangkan nilai metrik evaluasi disajikan pada Tabel 2. Berdasarkan ini, model menghasilkan akurasi sebesar 85,50% dan *Weighted F1-Score* sebesar 0,8563. Nilai F1-Score tertinggi terdapat pada kelas netral sebesar 0,9048, sedangkan kelas positif memiliki nilai F1-Score sebesar 0,7384. Perbedaan tersebut perlu dipertimbangkan bersama distribusi data pengujian, karena kelas netral memiliki jumlah sampel yang lebih besar dibandingkan kelas positif.

Pada kelas positif, model memperoleh *recall* sebesar 0,7565. Nilai tersebut menunjukkan bahwa sebagian besar data positif pada data uji berhasil dikenali, meskipun masih terdapat data yang diklasifikasikan ke kelas lain. Penggunaan *class weighting* dalam penelitian ini dirancang untuk memberikan bobot lebih besar terhadap kelas dengan jumlah data lebih sedikit. Namun,

berdasarkan hasil penelitian ini saja, efektivitas *class weighting* tidak dapat dipisahkan secara empiris dari kontribusi komponen pelatihan lainnya karena penelitian tidak melakukan eksperimen *ablation* dengan dan tanpa *class weighting*.

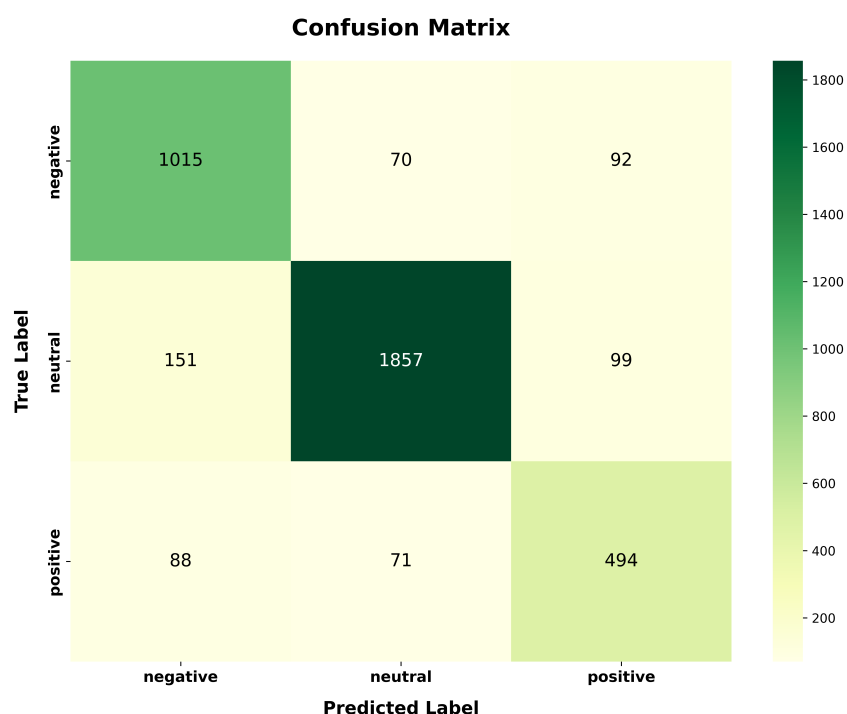


Gambar 4: Kurva *Training Loss* dan *Validation Loss* pada Proses *Fine-Tuning*



Gambar 6: Kecenderungan Sentimen Keseluruhan Program MBG (November 2023–Maret 2025)

Confusion Matrix juga menunjukkan adanya kesalahan klasifikasi antar-kelas. Sebagai contoh,



Gambar 5: *Confusion Matrix* Hasil Klasifikasi Model pada Data Uji

terdapat 92 data negatif yang diprediksi sebagai positif. Kesalahan tersebut menunjukkan bahwa batas antarkelas tidak sepenuhnya terpisah berdasarkan representasi yang dipelajari model. Faktor linguistik seperti konteks, ironi, atau penggunaan ekspresi evaluatif dapat menjadi salah satu kemungkinan penyebab, tetapi penelitian ini tidak melakukan anotasi khusus terhadap sarkasme sehingga penyebab tersebut tidak dapat dikonfirmasi secara langsung.

3.2 Kecenderungan Sentimen Publik

Model yang telah dievaluasi kemudian digunakan untuk mengklasifikasikan 126.365 cuitan aktual yang dikumpulkan selama periode November 2023 hingga Maret 2025. Distribusi sentimen secara keseluruhan ditampilkan pada Gambar 6.

Tabel 2: Hasil Metrik Evaluasi Klasifikasi Model IndoBERTweet

Kelas Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.8094	0.8624	0.8350	1.177
Netral	0.9294	0.8813	0.9048	2.107
Positif	0.7212	0.7565	0.7384	653
Akurasi Keseluruhan			0.8550	3.937
Weighted Avg	0.8590	0.8550	0.8563	3.937

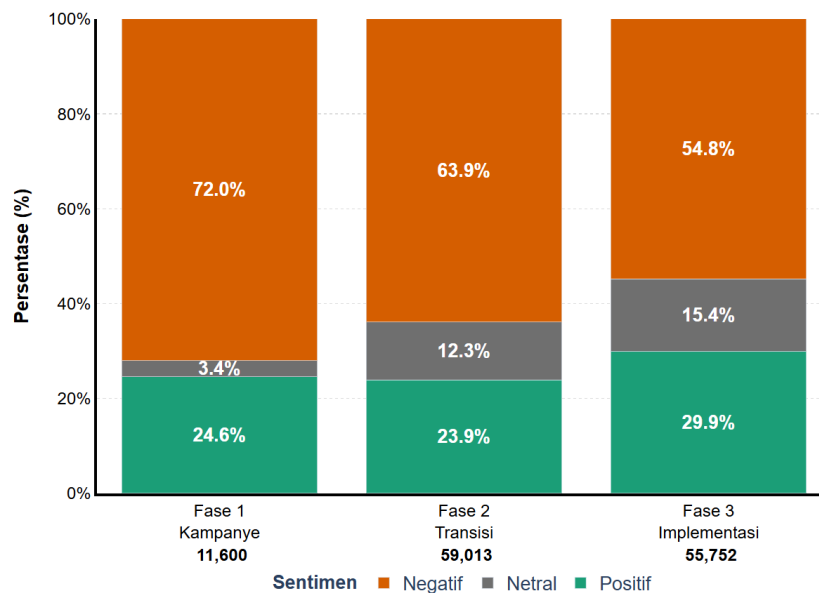
Hasil klasifikasi menunjukkan bahwa sentimen negatif merupakan kategori dominan dengan proporsi 60,6% atau 76.569 cuitan. Sentimen positif mencapai 26,6% atau 33.607 cuitan, sedangkan sentimen netral sebesar 12,8% atau 16.189 cuitan.

Dominasi sentimen negatif menunjukkan bahwa selama periode pengamatan terdapat proporsi ekspresi evaluatif yang lebih besar yang dikategorikan negatif terhadap program. Analisis terhadap frasa yang muncul dalam data menunjukkan bahwa sebagian diskursus negatif berkaitan dengan isu anggaran, kapasitas fiskal, dan kemungkinan pengalihan sumber daya dari sektor lain. Namun, hasil klasifikasi sentimen tidak dapat secara langsung digunakan untuk menyimpulkan bahwa seluruh sentimen negatif disebabkan oleh satu faktor tertentu. Interpretasi tersebut perlu dikaitkan dengan hasil analisis diskursus pada tingkat frasa.

Sentimen positif menunjukkan adanya kelompok percakapan yang menekankan manfaat program, terutama terkait pemenuhan gizi anak, pencegahan stunting, serta kemungkinan dampak ekonomi bagi pelaku usaha pangan lokal. Sementara itu, sentimen netral cenderung muncul pada teks yang bersifat informatif atau deskriptif.

3.3 Dinamika dan Komparasi Sentimen Antarfase

Analisis temporal dilakukan untuk mengetahui perubahan distribusi sentimen pada tiga fase perkembangan kebijakan. Perbandingan proporsi sentimen antarfase ditampilkan pada Gambar 7, sedangkan dinamika volume sentimen harian ditampilkan pada Gambar 8.



Gambar 7: Komparasi Distribusi Sentimen Antarfase Kebijakan

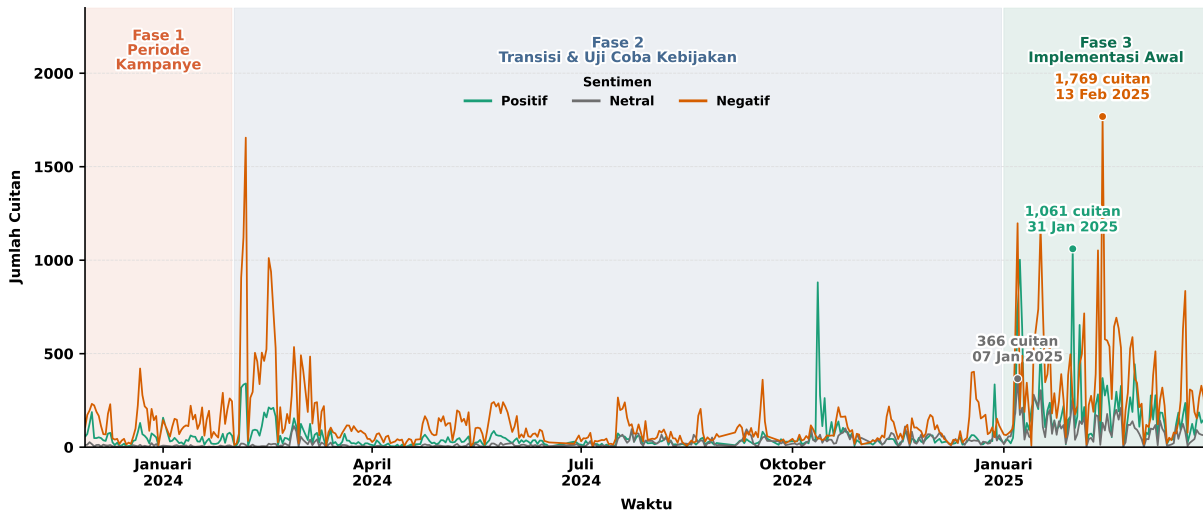
Jumlah cuitan meningkat dari 11.600 cuitan pada Fase 1 menjadi 59.013 cuitan pada Fase 2 dan 55.752 cuitan pada Fase 3. Peningkatan volume tersebut menunjukkan bahwa intensitas diskusi mengenai program bertambah setelah periode kampanye. Namun, volume diskusi tidak secara langsung menunjukkan tingkat dukungan atau penolakan masyarakat sehingga perlu dianalisis bersama distribusi sentimen.

Secara proporsional, terdapat perubahan distribusi sentimen antarfase. Sentimen negatif menurun dari 72,0% pada Fase 1 menjadi 63,9% pada Fase 2 dan 54,8% pada Fase 3. Pada periode yang sama, proporsi sentimen netral meningkat dari 3,4% menjadi 15,4%. Sentimen positif juga mengalami peningkatan dan mencapai 29,9% pada Fase 3.

Penurunan proporsi sentimen negatif menunjukkan adanya perubahan komposisi opini yang

terklasifikasi selama periode pengamatan. Salah satu interpretasi yang memungkinkan adalah bahwa ketika program bergerak dari tahap wacana menuju implementasi awal, pembahasan masyarakat menjadi lebih beragam sehingga proporsi ekspresi negatif berkurang. Akan tetapi, perubahan tersebut tidak dapat secara langsung diinterpretasikan sebagai peningkatan penerimaan masyarakat karena distribusi sentimen juga dapat dipengaruhi oleh perubahan topik, intensitas pemberitaan, peristiwa tertentu, serta karakteristik pengguna yang aktif berdiskusi pada setiap periode.

Pola tersebut terlihat pada grafik *time-series*. Fase 1 menunjukkan beberapa lonjakan diskusi yang berdekatan dengan momentum politik. Fase 2 memperlihatkan volume diskusi yang lebih tinggi dan relatif berkelanjutan. Pada Fase 3, sejumlah lonjakan terjadi bersamaan dengan periode implementasi awal program. Pola ini menunjukkan bahwa diskursus mengenai kebijakan bersifat dinamis dan dapat merespons perkembangan informasi atau peristiwa yang terjadi selama pelaksanaan program.



Gambar 8: Dinamika Sentimen Harian Program MBG (November 2023–Maret 2025)

3.4 Analisis Diskursus Berdasarkan N-gram

Untuk melengkapi analisis kuantitatif sentimen, dilakukan pemetaan frasa menggunakan pendekatan N-gram dan pembobotan TF-IDF. Hasil representasi frasa dominan ditampilkan pada Tabel 3.

Frasa yang muncul menunjukkan bahwa aspek ekonomi dan anggaran menjadi salah satu tema penting dalam diskursus negatif. Frasa seperti “duit dari mana” dan “buat bayar utang” menunjukkan adanya perhatian terhadap sumber pembiayaan dan konsekuensi fiskal. Sebaliknya, frasa seperti “sehat dan cerdas” serta “gizi anak anak” menunjukkan bahwa sebagian diskursus positif berorientasi pada tujuan program.

Kemunculan frasa “indonesia emas 2045” pada kelompok positif maupun negatif juga menunjukkan bahwa suatu frasa dapat digunakan dalam konteks sentimen yang berbeda. Kondisi tersebut menunjukkan pentingnya konteks dalam interpretasi hasil klasifikasi. Meskipun demikian, keberadaan frasa tersebut tidak dapat digunakan sebagai bukti tunggal bahwa model berhasil mendeteksi sarkasme atau ironi, karena penelitian ini tidak melakukan anotasi khusus maupun

pengujian terkontrol terhadap fenomena tersebut.

Tabel 3: Representasi Frasa Dominan (N-gram) pada Diskursus Publik

Tema Diskursus	Representasi Frasa Positif	Representasi Frasa Negatif
Ekonomi & Anggaran	“prabowo peduli umkm”	“duit dari mana”, “rp 10 000”, “buat bayar utang”
Pendidikan & Kesehatan	“sehat dan cerdas”, “gizi anak anak”	“pendidikan gratis malah”, “ga masuk akal”
Aspirasi vs Ironi	“masa depan cerah”, “indonesia emas 2045”	“indonesia emas 2045”, “seluruh rakyat indonesia”

3.5 Komparasi dengan Penelitian Terdahulu

Posisi penelitian ini dibandingkan dengan beberapa penelitian terdahulu ditampilkan pada Tabel 4.

Tabel 4: Komparasi Pendekatan Penelitian terhadap Studi Terdahulu

Peneliti (Tahun)	Pendekatan	Fokus Penelitian Ini
Arif & Kustiyono (2025); Munir (2025)	Pendekatan leksikon dan <i>Machine Learning</i> menggunakan SVM.	Menggunakan representasi kontekstual IndoBERTweet untuk klasifikasi sentimen.
Ilham & Priambodo (2024); Roihan dkk. (2024)	Menggunakan model BERT pada data berbahasa Indonesia.	Menggunakan IndoBERTweet yang dikembangkan dengan karakteristik korpus media sosial.
Khairani dkk. (2024)	Membandingkan model <i>Transformer</i> pada data komentar.	Memperluas analisis pada data Platform X dengan pendekatan temporal selama tiga fase kebijakan.

Berdasarkan perbandingan tersebut, kontribusi penelitian ini terutama terletak pada kombinasi penggunaan IndoBERTweet untuk klasifikasi sentimen dan analisis temporal terhadap diskursus Program MBG. Penelitian ini tidak mengklaim bahwa IndoBERTweet secara umum lebih unggul dibandingkan seluruh model sebelumnya karena tidak dilakukan eksperimen *baseline comparison* yang menguji beberapa model pada *dataset* dan konfigurasi yang sama.

Penggunaan *class weighting* merupakan strategi untuk menangani distribusi kelas yang tidak seimbang pada *dataset* pelatihan. Hasil evaluasi menunjukkan bahwa kelas positif tetap memperoleh *recall* sebesar 0,7565. Namun, kontribusi spesifik *class weighting* terhadap nilai tersebut belum dapat dipisahkan karena penelitian tidak melakukan eksperimen *ablation* yang membandingkan model dengan dan tanpa strategi tersebut.

Demikian pula, *minimalist filtering* digunakan dengan pertimbangan bahwa penghapusan elemen teks secara berlebihan dapat menghilangkan informasi kontekstual. Hasil penelitian menunjukkan bahwa pendekatan tersebut dapat digunakan dalam konfigurasi model yang diterapkan. Akan tetapi, efektivitasnya secara kausal belum dapat dipastikan tanpa eksperimen pembandingan yang secara khusus membandingkan beberapa konfigurasi pra-pemrosesan.

Kontribusi utama penelitian ini dengan demikian berada pada penerapan model IndoBERTweet pada *dataset* Program MBG serta pemetaan perubahan distribusi sentimen selama tiga fase

kebijakan. Analisis temporal memberikan perspektif tambahan dibandingkan analisis yang hanya menggunakan satu periode pengamatan, sementara analisis N-gram digunakan untuk membantu memberikan konteks terhadap perubahan distribusi sentimen yang ditemukan.

4 Kesimpulan

Penelitian ini menggunakan IndoBERTweet untuk mengklasifikasikan sentimen pada teks berbahasa Indonesia dan menganalisis dinamika opini publik terhadap Program Makan Bergizi Gratis di Platform X selama 17 bulan, yaitu dari 28 November 2023 hingga 31 Maret 2025. Evaluasi pada 3.937 data uji menghasilkan akurasi sebesar 85,50% dan *Weighted F1-Score* sebesar 0,8563. Nilai F1-Score tertinggi diperoleh pada kelas netral sebesar 0,9048, sedangkan kelas positif memperoleh nilai 0,7384. Hasil tersebut menunjukkan bahwa model mampu melakukan klasifikasi dengan kinerja yang relatif konsisten pada konfigurasi penelitian yang digunakan.

Penerapan *class weighting* digunakan sebagai strategi untuk menangani ketidakseimbangan kelas pada *dataset* pelatihan. Model memperoleh *recall* sebesar 0,7565 pada kelas positif. Meskipun hasil tersebut menunjukkan bahwa sebagian besar data positif berhasil dikenali, penelitian ini belum melakukan eksperimen *ablation* sehingga kontribusi spesifik *class weighting* terhadap kinerja model belum dapat ditentukan secara terpisah.

Berdasarkan klasifikasi terhadap 126.365 cuitan aktual, sentimen negatif mendominasi dengan proporsi 60,6%, diikuti sentimen positif sebesar 26,6% dan sentimen netral sebesar 12,8%. Analisis N-gram menunjukkan bahwa diskursus negatif banyak berkaitan dengan persoalan anggaran, sumber pembiayaan, dan konsekuensi kebijakan, sedangkan diskursus positif banyak berkaitan dengan pemenuhan gizi, kesehatan anak, dan manfaat program.

Analisis temporal menunjukkan adanya penurunan proporsi sentimen negatif dari 72,0% pada fase kampanye menjadi 63,9% pada fase transisi dan uji coba, kemudian menjadi 54,8% pada fase implementasi awal. Pada saat yang sama, proporsi sentimen positif dan netral meningkat. Perubahan tersebut menunjukkan adanya pergeseran komposisi diskursus selama perkembangan kebijakan. Namun, perubahan proporsi sentimen tidak secara langsung dapat disamakan dengan peningkatan penerimaan masyarakat karena distribusi tersebut dapat dipengaruhi oleh perubahan topik, peristiwa, pemberitaan, serta karakteristik pengguna Platform X.

Penelitian selanjutnya dapat memperkuat temuan ini melalui eksperimen *baseline comparison* dengan model lain pada *dataset* dan konfigurasi yang sama, eksperimen *ablation* untuk mengukur kontribusi *class weighting* dan *minimalist filtering*, serta anotasi khusus terhadap sarkasme dan ironi. Pendekatan tersebut dapat memberikan bukti empiris yang lebih kuat mengenai faktor-faktor yang berkontribusi terhadap kinerja klasifikasi dan interpretasi perubahan sentimen publik.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Program Studi Sarjana Sains Data Universitas Negeri Surabaya serta seluruh pihak yang telah memberikan dukungan dan masukan selama pelaksanaan penelitian ini.

References

Ardiani, L., Sujaini, H., dan Tursina, T. (2020). Implementasi sentiment analysis tanggapan masyarakat terhadap pembangunan di kota pontianak. *Jurnal Sistem dan Teknologi Informasi*

- (Justin), 8:183.
- Arif, M. W. dan Kustiyono, K. (2025). Analisis sentimen kebijakan makan bergizi gratis di media sosial menggunakan natural language processing berbasis python textblob di indonesia. *Jurnal Pendidikan dan Teknologi Indonesia*, 5:2463–2471.
- Devlin, J., Chang, M.-W., Lee, K., dan Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North*, pages 4171–4186. Association for Computational Linguistics.
- Grandini, M., Bagli, E., dan Visani, G. (2020). Metrics for multi-class classification: an overview. *arXiv*.
- Islam, T., Sheakh, M. A., Sadik, M. R., Tahosin, M. S., Foysal, M. M. R., Ferdush, J., dan Begum, M. (2024). Lexicon and deep learning-based approaches in sentiment analysis on short texts. *Journal of Computer and Communications*, 12:11–34.
- Koto, F., Lau, J. H., dan Baldwin, T. (2021). Indobertweet: A pretrained language model for indonesian twitter with effective domain-specific vocabulary initialization. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10660–10668. Association for Computational Linguistics.
- Luqyana, W. A., Cholissodin, I., dan Perdana, R. S. (2018). Analisis sentimen cyberbullying pada komentar instagram dengan metode klasifikasi support vector machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2:4704–4713.
- Munir, R. A. (2025). Analisis sentimen cuitan di media sosial x tentang program makan bergizi gratis dengan metode nlp. *Jurnal Informatika dan Teknik Elektro Terapan*, 13.
- of BKPK, P. R. (2025). Ssgi 2024: National stunting prevalence drops to 19,8
- Rofiqoh, U., Perdana, R. S., dan Fauzi, M. A. (2017). Analisis sentimen tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler indonesia pada twitter dengan metode support vector machine dan lexicon based features. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1:1725–1732.
- Rossi, A., Lestari, T., Perdana, R. S., dan Fauzi, M. A. (2017). Analisis sentimen tentang opini pilkada dki 2017 pada dokumen twitter berbahasa indonesia menggunakan naïve bayes dan pembobotan emoji. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1:1718–1724.
- Saputri, M. S., Mahendra, R., dan Adriani, M. (2018). Emotion classification on indonesian twitter dataset. In *2018 International Conference on Asian Language Processing (IALP)*, pages 90–95. IEEE.
- Setpres, B. (2025). Presiden prabowo: Program mbg investasi untuk masa depan bangsa.
- Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Łukasz Kaiser, dan Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, pages 5998–6008.
- Widiarta, A. (2023). Indonesian twitter sentiment analysis dataset-ppkm.