

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Pemenuhan kebutuhan gizi anak usia sekolah merupakan fondasi utama dalam pembangunan sumber daya manusia yang berkualitas. Anak-anak yang memperoleh asupan nutrisi seimbang cenderung mengalami perkembangan fisik dan kognitif yang optimal serta menunjukkan prestasi akademik yang lebih baik. Di Indonesia, tantangan kekurangan gizi dan *stunting* masih menjadi isu serius, terutama di wilayah dengan tingkat kemiskinan tinggi. Menurut *World Health Organization* (WHO), sekitar 144 juta anak di bawah usia lima tahun mengalami *stunting*, dan 45% dari kematian anak usia dini berkaitan dengan kekurangan gizi [1]. Sebagai respons terhadap permasalahan tersebut, pemerintah Indonesia merancang Program Makan Bergizi Gratis (MBG) sebagai bagian dari kebijakan prioritas nasional untuk meningkatkan status gizi peserta didik secara merata.

Program MBG bertujuan menyediakan makanan bergizi kepada peserta didik, terutama di daerah yang rentan terhadap masalah gizi. Penelitian menunjukkan bahwa penerapan program makanan bergizi di sekolah dasar dapat meningkatkan kehadiran siswa dan memperbaiki kebiasaan makan harian peserta didik [2]. Jika dijalankan secara konsisten dan didukung dengan edukasi gizi yang terintegrasi, program ini berpotensi menurunkan angka *stunting* [3].

Namun, implementasi Program Makan Bergizi Gratis (MBG) memicu masalah baru berupa polarisasi opini yang tajam di kalangan masyarakat. Kebijakan ini mengundang perdebatan sengit terkait transparansi anggaran yang besar, kesiapan infrastruktur distribusi, hingga kekhawatiran mengenai kualitas dan kebersihan makanan yang disajikan. Ketidakpastian sentimen publik ini menjadi tantangan serius bagi pemerintah, karena opini negatif yang berkembang secara liar dan masif tanpa evaluasi berbasis data dapat menurunkan kepercayaan publik terhadap efektivitas program [4]. Masalah mendasarnya adalah bagaimana mengolah ribuan opini subjektif yang tersebar secara acak di media sosial menjadi wawasan objektif yang dapat digunakan sebagai landasan evaluasi kebijakan.

Untuk menangkap dinamika opini tersebut, penelitian ini memilih X sebagai sumber data utama. Pemilihan X didasarkan pada karakteristik uniknya sebagai *platform* berbasis teks yang memfasilitasi diskusi opini secara lebih eksplisit dan

argumentatif dibandingkan media sosial berbasis visual seperti Instagram atau TikTok [5]. Selain itu, basis pengguna X di Indonesia dikenal sangat responsif dan kritis terhadap isu-isu politik serta kebijakan publik, menjadikannya barometer yang sensitif untuk mengukur sentimen sosial. Sifat penyebaran informasi yang *real-time* melalui fitur *trending topic* dan *hashtag* juga memungkinkan peneliti untuk merekam reaksi publik secara instan sesaat setelah isu terkait MBG mencuat, memberikan data yang relevan secara temporal [6].

Analisis opini masyarakat di *platform* ini dapat memberikan wawasan berharga mengenai persepsi publik terhadap kebijakan MBG. Hasil analisis tersebut dapat digunakan sebagai dasar evaluasi untuk menilai efektivitas komunikasi kebijakan pemerintah secara lebih objektif dan berbasis data. Untuk menganalisis opini publik secara sistematis, penelitian ini menggunakan pendekatan analisis sentimen, yaitu proses komputasional untuk mengidentifikasi dan mengkategorikan emosi dalam teks sebagai positif, negatif, atau netral. Dalam konteks bahasa Indonesia, pendekatan ini semakin akurat berkat kemajuan teknologi *Natural Language Processing* (NLP) berbasis *deep learning*. Salah satu model NLP yang terbukti efektif adalah IndoBERT, versi BERT yang dilatih pada korpus besar bahasa Indonesia [6].

Penelitian sebelumnya menunjukkan IndoBERT mampu melakukan klasifikasi emosi pada *tweet* berbahasa Indonesia dengan *F1-score* yang kompetitif. Selain itu, Juarto dan Yulianto (2023) melaporkan bahwa IndoBERT mengungguli metode tradisional seperti TF-IDF yang dikombinasikan dengan XGBoost, LightGBM, dan *Random Forest* pada tugas klasifikasi berita berbahasa Indonesia, dengan akurasi hingga 94% [7].

IndoBERTtweet adalah varian yang diadaptasi untuk korpus Twitter berbahasa Indonesia. Model ini dikembangkan melalui *pretraining* adaptif domain dari IndoBERT dengan penambahan kosakata khusus media sosial. Eksperimen pada beberapa *dataset* Twitter Indonesia menunjukkan IndoBERTtweet unggul pada tugas analisis sentimen, klasifikasi emosi, dan deteksi ujaran kebencian, sehingga relevan untuk analisis opini publik di *Platform X* [5].

Berdasarkan uraian di atas, penelitian ini menerapkan IndoBERTtweet sebagai model pemrosesan bahasa alami untuk menganalisis sentimen publik terhadap Program Makan Bergizi Gratis (MBG) di *Platform X*, dengan tujuan utama memberikan masukan berbasis data yang komprehensif bagi pembuat kebijakan dan pemangku kepentingan terkait, sehingga hasil analisis tidak hanya mampu menggambarkan persepsi masyarakat secara lebih akurat, tetapi juga dapat

digunakan sebagai landasan dalam merumuskan strategi komunikasi, evaluasi efektivitas program, serta penyusunan kebijakan lanjutan yang lebih responsif terhadap kebutuhan dan aspirasi publik.

1.2 Rumusan Masalah

1. Bagaimana performa model IndoBERTweet dalam mengklasifikasikan sentimen publik (positif, negatif, netral) terhadap Program MBG?
2. Bagaimana perbandingan dampak dua strategi pelabelan *heuristic labeling* dan *pseudo labeling* terhadap kualitas label dan performa model IndoBERTweet pada *dataset* yang sama?

1.3 Batasan Permasalahan

- Periode pengambilan data dibatasi pada rentang 20 Oktober 2024 sampai dengan 1 Oktober 2025 untuk menjaga relevansi temporal data.
- *Dataset* yang digunakan berjumlah 17.480 data.
- Analisis difokuskan pada teks berbahasa Indonesia yang dipublikasikan di *Platform X*; media sosial lain tidak termasuk dalam cakupan penelitian ini.
- Pada penelitian ini menggunakan dua teknik *labeling* yaitu *Heuristic* dan *Pseudo*.

1.4 Tujuan Penelitian

- Mengklasifikasikan *tweet* terkait MBG menjadi tiga kelas (negatif, netral, positif), dan melaporkan performa terbaik (*accuracy*, *precision*, *recall*, *F1-macro*) serta analisis *confusion matrix* pada *testing set*.
- Membandingkan dua strategi pelabelan *heuristic labeling* dan *pseudo labeling* dalam hal kualitas label dan dampaknya terhadap performa dan dievaluasi berdasarkan perubahan *F1-Score* pada *validation set* dan *testing set* model.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini meliputi:

- **Akademis:** Menyediakan analisis performa model IndoBERTweet sebagai referensi bagi penelitian lanjutan di bidang pemrosesan bahasa alami dan analisis sentimen bahasa Indonesia.
- **Praktis:** Memberikan masukan berbasis data kepada pembuat kebijakan dan pemangku kepentingan mengenai persepsi publik terhadap Program MBG, sehingga dapat mendukung evaluasi dan perbaikan implementasi program.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian ini disusun sebagai berikut:

- Bab 1 Pendahuluan: Latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, dan sistematika penulisan.
- Bab 2 Landasan Teori: Tinjauan pustaka dan teori yang mendasari penelitian.
- Bab 3 Metodologi Penelitian: Metode pengumpulan data, praproses, arsitektur model, dan prosedur pelatihan.
- Bab 4 Hasil dan Diskusi: Penyajian hasil eksperimen, evaluasi model, dan pembahasan.
- Bab 5 Kesimpulan dan Saran: Kesimpulan penelitian, kontribusi, keterbatasan, dan rekomendasi untuk penelitian selanjutnya.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A