BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Program Makan Bergizi Gratis (MBG) merupakan salah satu kebijakan strategis pemerintah yang saat ini menjadi fokus utama dalam memperkuat ketahanan pangan dan meningkatkan kualitas nutrisi masyarakat Indonesia [1, 2]. Secara operasional, program ini dirancang untuk mengatasi permasalahan krusial seperti *stunting* dan malnutrisi pada anak sekolah melalui penyediaan asupan makanan dengan standar gizi yang terkontrol [3, 4]. Sebagai sebuah kebijakan berskala nasional yang berdampak langsung pada masyarakat luas, implementasi MBG memicu diskursus publik yang sangat dinamis [5]. Kebijakan ini tidak hanya dipandang dari sudut pandang pemenuhan gizi anak semata, melainkan juga memicu perbincangan hangat di tengah masyarakat mengenai efektivitas jalur distribusi, kesiapan infrastruktur di berbagai daerah, hingga dampak ekonominya terhadap anggaran negara [6, 7].

Di era digital saat ini, platform media sosial seperti YouTube telah bertransformasi menjadi salah satu wadah terbesar bagi masyarakat untuk mengekspresikan opini, kritik, maupun aspirasi secara terbuka [8]. Narasi video terkait sosialisasi, berita, hingga debat mengenai program MBG di YouTube secara konsisten menarik interaksi yang sangat tinggi, yang terlihat dari menumpuknya komentar warganet dari berbagai lapisan masyarakat [9]. Opini-opini tersebut tersebar dalam bentuk teks non-formal, bahasa kasual, hingga istilah *slang* khas internet [10]. Pola komunikasi digital yang masif ini menghasilkan data tekstual yang sangat kaya namun tidak terstruktur [11]. Jika data ini diolah dengan metode yang tepat, kolom komentar YouTube dapat menjadi cerminan riil untuk membaca bagaimana persepsi, kekhawatiran, dan harapan masyarakat yang sebenarnya terhadap program MBG ini [12].

Untuk mengekstrak makna dari ribuan opini tersebut, analisis sentimen menjadi teknik komputasi yang sangat efektif untuk mengidentifikasi dan mengklasifikasikan kecenderungan emosional publik [13]. Dalam beberapa tahun terakhir, penelitian analisis sentimen untuk teks berbahasa Indonesia berkembang pesat berkat teknologi *Natural Language Processing* (NLP) berbasis *deep learning* [14]. Salah satu model representasi bahasa yang terbukti unggul saat ini adalah

IndoBERT [15]. Model ini dikembangkan secara spesifik menggunakan korpus bahasa Indonesia yang masif, sehingga memiliki kemampuan yang sangat baik dalam memahami konteks semantik yang mendalam pada teks media sosial, bahkan untuk kalimat yang sifatnya non-formal atau tidak baku jika dibandingkan dengan model *machine learning* tradisional [16, 17].

Meskipun IndoBERT menawarkan akurasi yang tinggi, proses klasifikasi sentimen pada data media sosial berskala besar sering kali menghadapi kendala pada tahap pelabelan (*labeling*) data [18]. Jika dilakukan secara manual, proses ini akan memakan waktu yang lama dan biaya yang besar [19]. Untuk mengatasi kendala tersebut, tren penelitian terkini mulai memanfaatkan model *transformer* lain seperti RoBERTa (*Robustly Optimized BERT Pretraining Approach*) untuk melakukan pelabelan otomatis (*automated labeling*) berdasarkan nilai probabilitas tertinggi yang dihasilkan model [20, 21]. Selain itu, untuk memberikan hasil analisis yang lebih mendalam, klasifikasi sentimen juga kerap dikombinasikan dengan metode *Topic Modeling* seperti *Latent Dirichlet Allocation* (LDA) [22, 23]. Metode LDA ini terbukti efektif untuk mengelompokkan teks ke dalam beberapa klaster topik tanpa perlu pelabelan manual, sehingga peneliti tidak hanya mengetahui sentimennya saja, tetapi juga tahu substansi masalah apa yang sedang dibahas [24].

Namun demikian, sebagian besar penelitian analisis sentimen terkait kebijakan pemerintah sebelumnya masih cenderung berfokus pada klasifikasi sentimen secara global (positif dan negatif saja) [25]. Akibatnya, pembuat kebijakan kesulitan untuk memetakan aspek mana dari program MBG—apakah masalah anggaran, kualitas gizi, atau kendala distribusi—yang paling banyak menuai respons negatif atau positif [26]. Di sisi lain, masalah ketidakseimbangan kelas (*imbalanced dataset*) pada data media sosial juga sering diabaikan, padahal hal ini dapat membuat model menjadi bias saat proses pelatihan [27]. Selain itu, dinamika opini mengenai program MBG ini sangat fluktuatif, khususnya pada masa transisi dan awal implementasinya dari periode Januari 2025 hingga Mei 2026, yang mana topik spesifik ini belum pernah diteliti secara komprehensif menggunakan kombinasi model arsitektur modern [28].

Berdasarkan permasalahan tersebut, penelitian ini bermaksud untuk melakukan analisis sentimen yang lebih mendalam terhadap program MBG menggunakan 9205 komentar YouTube dalam rentang periode Januari 2025 hingga Mei 2026. Penelitian ini mengimplementasikan model IndoBERT sebagai algoritma klasifikasi utama, dengan tambahan metode *Random Over Sampling*

(ROS) untuk menangani ketidakseimbangan data teks [29]. Untuk efisiensi data, proses *labeling* awal akan dioptimalkan secara otomatis menggunakan model RoBERTa, sedangkan metode LDA digunakan untuk mengekstraksi klaster topik utama di balik opini masyarakat [30]. Penelitian ini diharapkan dapat mengukur performa model IndoBERT melalui metrik *Accuracy, Precision, Recall*, dan *F1-score*, sekaligus menyajikan pemetaan sentimen berbasis topik yang terstruktur. Hasil dari penelitian ini nantinya dapat menjadi bahan pertimbangan dan evaluasi bagi pihak pemerintah dalam melihat respons riil masyarakat terhadap program Makan Bergizi Gratis (MBG) [31].

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, rumusan masalah dari penelitian ini adalah sebagai berikut.

1. Bagaimana cara mengimplementasikan model IndoBERT dan LDA untuk melakukan analisis sentimen terhadap MBG pada kolom komentar YouTube?
2. Berapa nilai *Accuracy, Precision, Recall,* dan *F1-score* dari model IndoBERT dalam analisis sentimen terhadap MBG pada kolom komentar YouTube?

1.3 Batasan Permasalahan

Adapun beberapa batasan masalah yang ada pada penelitian ini yaitu sebagai berikut.

1. Data komentar yang diambil dibatasi pada periode tertentu (dari Januari 2025 hingga Mei 2026) untuk menangkap dinamika opini yang paling relevan.
2. Dataset yang diambil dari kolom komentar YouTube yang berkaitan dengan MBG sebanyak 9205 komentar, dikarenakan diambil berdasarkan pembahasan yang relevan.

1.4 Tujuan Penelitian

Penelitian ini dilakukan dengan tujuan sebagai berikut.

1. Mengimplementasikan model IndoBERT dan LDA untuk melakukan analisis sentimen terhadap MBG pada kolom komentar YouTube.

2. Menghitung nilai *Accuracy, Precision, Recall,* dan *F1-score* dari model IndoBERT dalam analisis sentimen terhadap MBG pada kolom komentar YouTube.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini yaitu sebagai berikut.

1. Memberikan gambaran mengenai penerapan model IndoBERT yang dikombinasikan dengan LDA dalam menganalisis sentimen dan topik pembahasan masyarakat terhadap program MBG pada kolom komentar YouTube, sehingga dapat menjadi referensi bagi penelitian sejenis di bidang analisis sentimen berbahasa Indonesia.
2. Menghasilkan informasi performa model IndoBERT berupa nilai *Accuracy, Precision, Recall,* dan *F1-score*, serta pemetaan sentimen dan topik yang dapat menjadi bahan pertimbangan dan evaluasi bagi pemerintah dalam melihat respons riil masyarakat terhadap program Makan Bergizi Gratis.

1.6 Sistematika Penulisan

Berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian, dimulai dari Pendahuluan hingga Simpulan dan Saran.

Sistematika penulisan laporan adalah sebagai berikut.

1. Bab 1 PENDAHULUAN
   Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan skripsi.
2. Bab 2 LANDASAN TEORI
   Bab ini membahas landasan teori yang mendukung penelitian, meliputi konsep analisis sentimen, media sosial, Natural Language Processing (NLP), preprocessing teks bahasa Indonesia, model IndoBERT, serta metode evaluasi model seperti confusion matrix, precision, recall, dan F1-score.
3. Bab 3 METODOLOGI PENELITIAN
   Bab ini menjelaskan metodologi penelitian yang digunakan, mulai dari pengumpulan dan persiapan dataset, proses preprocessing, perancangan arsitektur model IndoBERT, hingga tahapan pelatihan dan evaluasi model.

4. Bab 4 HASIL DAN DISKUSI

   Bab ini menyajikan hasil implementasi dan pengujian model, analisis performa berdasarkan metrik evaluasi, serta pembahasan terhadap hasil yang diperoleh.

5. Bab 5 KESIMPULAN DAN SARAN

   Bab ini berisi kesimpulan dari hasil penelitian serta saran untuk pengembangan penelitian selanjutnya.