

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Bagaimana orang mengakses, menyebarkan, dan memahami informasi di dunia digital telah berubah karena kemajuan dalam teknologi informasi dan komunikasi [1]. Kehadiran media sosial turut memperluas ruang diskusi publik sehingga berbagai isu sosial, ekonomi, maupun kebijakan pemerintah dapat dengan cepat memperoleh perhatian masyarakat [2]. Salah satu *platform* yang umum digunakan untuk menyampaikan pendapat adalah X (dahulu dikenal sebagai Twitter). *Platform* ini memungkinkan pengguna membagikan pandangan dan tanggapan secara *real-time* melalui unggahan teks singkat yang bersifat terbuka [3]. Berdasarkan data DemandSage, jumlah pengguna X di Indonesia diperkirakan mencapai sekitar 25,2 juta pada tahun 2025 [4, 5]. Besarnya jumlah pengguna tersebut menjadikan X sebagai salah satu sumber data yang relevan untuk mengamati persepsi masyarakat terhadap suatu isu publik.

Data yang berasal dari media sosial umumnya berbentuk tidak terstruktur dan memiliki karakteristik yang berbeda dengan dokumen formal [6]. Pengguna sering menggunakan singkatan, *slang*, ironi, maupun variasi bahasa lain yang dapat mempersulit proses interpretasi secara manual [7]. Kondisi tersebut menyebabkan analisis data dalam jumlah besar menjadi kurang efisien apabila dilakukan tanpa bantuan komputasi. Untuk mengatasi permasalahan tersebut, pendekatan *Natural Language Processing* (NLP) banyak dimanfaatkan dalam proses pengolahan dan analisis teks secara otomatis [8]. Melalui NLP, komputer dapat mempelajari pola bahasa manusia serta mengekstraksi informasi yang terkandung dalam kumpulan data teks berukuran besar [9].

Salah satu implementasi NLP yang banyak digunakan dalam kajian opini publik adalah analisis sentimen. Teknik ini digunakan untuk mengelompokkan opini ke dalam kategori positif, negatif, atau netral [2]. Analisis sentimen telah diterapkan pada berbagai isu kebijakan publik dan fenomena sosial, di antaranya kebijakan kenaikan Pajak Pertambahan Nilai (PPN) [10], Revisi Undang-Undang Tentara Nasional Indonesia (RUU TNI) [11], serta kebijakan Tabungan Perumahan Rakyat (TAPERA) [12]. Hasil dari berbagai penelitian tersebut memperlihatkan bahwa data media sosial dapat dimanfaatkan untuk memperoleh gambaran

mengenai respons dan persepsi masyarakat terhadap kebijakan pemerintah.

Program Makan Bergizi Gratis (MBG) merupakan salah satu kebijakan pemerintah Indonesia yang ditujukan untuk meningkatkan kualitas gizi masyarakat sekaligus mendukung pembangunan sumber daya manusia menuju visi Generasi Emas 2045 [13]. Program ini dikelola oleh Badan Gizi Nasional (BGN) dengan dukungan anggaran sebesar Rp71 triliun pada APBN tahun 2025 [14]. Skala pelaksanaan yang luas serta besarnya alokasi anggaran menjadikan MBG sebagai salah satu kebijakan yang banyak mendapat perhatian dan menjadi topik diskusi di media sosial. Beragam tanggapan masyarakat muncul seiring implementasi program tersebut, mulai dari dukungan terhadap tujuan program hingga kritik yang berkaitan dengan distribusi, kesiapan infrastruktur, transparansi pelaksanaan, dan keberlanjutan anggaran [15, 16]. Sejumlah pemberitaan nasional juga memperlihatkan adanya perbedaan pandangan masyarakat terhadap program MBG, baik terkait kualitas distribusi makanan, efektivitas pelaksanaan, maupun kemungkinan munculnya kendala dalam implementasi kebijakan.

Jumlah opini yang terus bertambah di media sosial membuat proses analisis secara manual menjadi kurang praktis untuk dilakukan. Karena itu, berbagai pendekatan berbasis *machine learning* maupun *deep learning* mulai digunakan untuk mendukung proses analisis sentimen [17]. Algoritma *machine learning* seperti Naive Bayes dan Support Vector Machine (SVM) telah banyak digunakan dalam penelitian sebelumnya dan menunjukkan kemampuan klasifikasi yang cukup baik [11, 12]. Meskipun demikian, metode tersebut masih menghadapi keterbatasan dalam menangkap konteks bahasa secara menyeluruh, khususnya pada teks media sosial yang sering mengandung bahasa informal dan makna ambigu [17].

Perkembangan *deep learning* kemudian menghadirkan model berbasis *Transformer*, salah satunya *Bidirectional Encoder Representations from Transformers* (BERT), yang dirancang untuk memahami hubungan antarkata secara dua arah dalam sebuah kalimat [18]. Untuk kebutuhan bahasa Indonesia, tersedia IndoBERT sebagai model pra-latih yang dikembangkan agar lebih sesuai dengan karakteristik bahasa Indonesia [19]. Sejumlah penelitian menunjukkan bahwa IndoBERT mampu menghasilkan performa yang lebih baik dibandingkan pendekatan *machine learning* konvensional pada tugas analisis sentimen berbahasa Indonesia [20, 21, 10]. Kemampuan tersebut juga terlihat pada pemrosesan teks media sosial yang banyak mengandung bahasa informal [1].

Meskipun analisis sentimen dapat memberikan informasi mengenai polaritas opini masyarakat, pendekatan ini belum mampu menjelaskan topik-topik

yang menjadi fokus pembahasan dalam suatu diskusi publik. Oleh sebab itu, diperlukan metode tambahan berupa *topic modeling* untuk menemukan pola topik yang muncul dari kumpulan data teks [22].

Metode *topic modeling* yang banyak digunakan sebelum berkembangnya pendekatan berbasis *Transformer* adalah *Latent Dirichlet Allocation* (LDA). LDA membentuk topik berdasarkan distribusi probabilitas kata dalam dokumen. Namun, karena menggunakan pendekatan *bag-of-words*, metode ini masih memiliki keterbatasan dalam menangkap hubungan semantik antarkata, terutama ketika diterapkan pada data media sosial yang umumnya singkat dan tidak terstruktur [22, 7].

Untuk mengatasi keterbatasan tersebut, BERTopic memanfaatkan representasi *embedding* berbasis *Transformer* sehingga mampu menghasilkan topik yang lebih koheren secara semantik [23]. Penelitian sebelumnya menunjukkan bahwa BERTopic lebih sesuai untuk menganalisis data media sosial dibandingkan pendekatan *topic modeling* tradisional. Selain itu, kombinasi IndoBERT dan BERTopic telah digunakan untuk memperoleh gambaran opini publik yang lebih komprehensif [7, 1, 3]. Meskipun demikian, penerapan kedua metode tersebut untuk menganalisis opini masyarakat mengenai Program Makan Bergizi Gratis pada media sosial X masih belum banyak ditemukan.

Berdasarkan uraian tersebut, penelitian ini menggunakan IndoBERT untuk melakukan klasifikasi sentimen dan BERTopic untuk mengidentifikasi topik pembahasan pada data *tweet* mengenai Program Makan Bergizi Gratis. Kombinasi kedua metode tersebut diharapkan dapat memberikan gambaran yang lebih utuh mengenai opini masyarakat, baik dari sisi sentimen maupun topik yang mendominasi diskusi di media sosial.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, penelitian ini dirumuskan ke dalam beberapa pertanyaan sebagai berikut:

1. Bagaimana hasil analisis sentimen masyarakat terhadap Program Makan Bergizi Gratis berdasarkan data dari media sosial X menggunakan model IndoBERT?
2. Bagaimana kinerja model IndoBERT dalam mengklasifikasikan sentimen pada data *tweet* berbahasa Indonesia yang membahas Program Makan Bergizi

Gratis?

3. Topik apa saja yang mendominasi diskusi masyarakat mengenai Program Makan Bergizi Gratis berdasarkan hasil identifikasi menggunakan BERTopic?
4. Bagaimana distribusi topik yang terbentuk pada setiap kategori sentimen hasil klasifikasi?

1.3 Batasan Permasalahan

Adapun batasan penelitian yang diterapkan adalah sebagai berikut:

1. Data penelitian berupa *tweet* publik dari media sosial X yang diperoleh melalui proses *scraping* menggunakan kata kunci yang berkaitan dengan Program Makan Bergizi Gratis.
2. Data dikumpulkan pada periode 6 Januari 2025 hingga 1 Mei 2026.
3. Penelitian hanya menggunakan *tweet* yang ditulis dalam bahasa Indonesia.
4. Klasifikasi sentimen dibatasi ke dalam tiga kategori, yaitu positif, negatif, dan netral.
5. Ruang lingkup penelitian difokuskan pada analisis data berbentuk teks, sehingga tidak mencakup gambar, video, maupun analisis terhadap perilaku akun pengguna.
6. *Tweet* yang memiliki tingkat kemiripan tinggi diproses untuk mengurangi duplikasi data sebelum tahap analisis.
7. Analisis sentimen dilakukan menggunakan model IndoBERT, sedangkan identifikasi topik memanfaatkan metode BERTopic.
8. Kinerja model klasifikasi dievaluasi menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengklasifikasikan sentimen masyarakat terhadap Program Makan Bergizi Gratis menggunakan pendekatan *deep learning* berbasis IndoBERT.
2. Mengevaluasi performa model IndoBERT dalam melakukan klasifikasi sentimen terhadap data *tweet* berbahasa Indonesia.
3. Mengidentifikasi topik diskusi publik terkait Program Makan Bergizi Gratis menggunakan metode BERTopic.
4. Menganalisis hubungan antara sentimen dan topik diskusi yang terbentuk dari data media sosial X.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Manfaat Akademis
Memberikan kontribusi dalam pengembangan penelitian di bidang *Natural Language Processing* (NLP), khususnya analisis sentimen dan *topic modeling* pada data media sosial berbahasa Indonesia menggunakan pendekatan berbasis *Transformer*.
2. Manfaat Praktis
Memberikan gambaran mengenai persepsi dan opini masyarakat terhadap Program Makan Bergizi Gratis yang dapat digunakan sebagai bahan evaluasi kebijakan pemerintah.
3. Manfaat Teknis
Menjadi referensi dalam penerapan model IndoBERT dan BERTopic untuk analisis data teks tidak terstruktur pada media sosial.
4. Manfaat Sosial
Membantu memahami isu dan topik utama yang berkembang di masyarakat terkait implementasi Program Makan Bergizi Gratis.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian disusun sebagai berikut:

1. Bab 1 PENDAHULUAN

Memuat latar belakang penelitian, rumusan masalah, batasan penelitian, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

2. Bab 2 LANDASAN TEORI

Menyajikan landasan teori yang menjadi dasar penelitian, meliputi analisis sentimen, *machine learning*, *deep learning*, *Transformer*, BERT, IndoBERT, *topic modeling*, dan BERTopic.

3. Bab 3 METODOLOGI PENELITIAN

Menjelaskan metodologi penelitian, mulai dari proses pengumpulan data, *preprocessing*, pelabelan data, pelatihan model, hingga tahap evaluasi dan analisis.

4. Bab 4 HASIL DAN DISKUSI

Menguraikan hasil penelitian yang diperoleh beserta pembahasan terhadap hasil analisis sentimen, evaluasi model, dan identifikasi topik menggunakan BERTopic.

5. Bab 5 KESIMPULAN DAN SARAN

Memuat kesimpulan yang diperoleh berdasarkan hasil penelitian serta saran yang dapat menjadi masukan bagi penelitian selanjutnya.

