

BAB I

PENDAHULUAN

1.1 Latar Belakang

Program Makan Bergizi Gratis (MBG) merupakan salah satu kebijakan pemerintah yang bertujuan meningkatkan kualitas gizi masyarakat, khususnya anak-anak [1]. Program ini juga menunjukkan berbagai dampak positif, seperti peningkatan kehadiran siswa di sekolah serta motivasi belajar setelah kebutuhan gizi terpenuhi [2]. Selain itu, program ini turut berkontribusi terhadap perekonomian lokal melalui pemberdayaan UMKM dalam penyediaan makanan [3]. Namun, dalam implementasinya, program ini menghadapi berbagai permasalahan terkait kualitas makanan yang disajikan. Beberapa kasus yang menjadi sorotan publik antara lain ditemukannya makanan dalam kondisi tidak layak konsumsi seperti sayur basi, keterlambatan distribusi yang memengaruhi kualitas makanan, serta lemahnya pengawasan dalam proses penyediaan makanan [4], [5]. Selain itu, kasus keracunan massal yang dilaporkan menimpa lebih dari 9.000 anak semakin memperkuat kekhawatiran terhadap kualitas dan keamanan makanan dalam program tersebut [6]. Permasalahan ini banyak dibahas di media sosial, khususnya TikTok, yang memiliki tingkat interaksi yang tinggi sehingga menghasilkan komentar dalam jumlah besar dan beragam [7]. TikTok juga menerapkan penyaringan komentar berdasarkan Pedoman Komunitas untuk membatasi komentar yang melanggar kebijakan platform, seperti spam dan ujaran kebencian [9]. Oleh karena itu, penelitian ini menggunakan komentar TikTok sebagai sumber data karena komentar yang tersedia telah melalui penyaringan awal oleh platform dan masih memuat tanggapan pengguna terhadap isu yang dibahas. Komentar di media sosial merupakan sumber data penting untuk memahami opini publik, namun karakteristiknya yang tidak terstruktur serta banyak menggunakan singkatan dan bahasa slang menjadikan analisis sentimen sebagai pendekatan yang efektif untuk mengkajinya secara sistematis [1].

Relevansi analisis sentimen dalam mengolah opini publik telah menarik perhatian banyak peneliti, sehingga berbagai penelitian sebelumnya telah dilakukan. Peneliti [10] melakukan penelitian terkait analisis prediktif respons pengguna terhadap implementasi program makan bergizi gratis pada data X menggunakan IndoBERT. Hasil dari penelitian menunjukkan akurasi sebesar 85%, namun performa model pada kelas netral masih belum optimal karena sering terjadi tumpang tindih dengan kelas lainnya. Peneliti [11] juga melakukan penelitian terkait analisis

sentimen pada isu pemutusan hubungan kerja di Indonesia pada data twitter dengan membandingkan beberapa model seperti, IndoBERT, SVM, *Random Forest*, *Decision Tree*. Hasil dari penelitian ini menunjukkan bahwa akurasi model IndoBERT sebesar 89,6%, SVM 88%, *Random Forest* 78,4%, dan *Decision Tree* 70,40%. Hasil ini menunjukkan bahwa model IndoBERT mampu mengungguli model lainnya. Peneliti [12] melakukan penelitian terkait analisis sentimen program MBG pada data TikTok menggunakan IndoBERTa dan *Random Forest*. Hasil dari penelitian menunjukkan akurasi sebesar 82%, namun pada dataset penelitian merupakan data *imbalanced*, sehingga menyebabkan akurasi pada kelas minor atau positif mendapatkan nilai presisi 59%. Hal ini menunjukkan bahwa permasalahan ketidakseimbangan distribusi kelas merupakan kondisi yang umum terjadi dalam data opini publik [13].

Untuk mengatasi permasalahan *imbalanced* dataset, penelitian sebelumnya telah menerapkan beberapa teknik augmentasi data seperti *Synonym Replacement*, *Back Translation*, dan *Random Swap*. Teknik *Synonym Replacement* (SR) merupakan metode yang optimal pada data teks informal dengan akurasi 69,67% dan *F1-score* 0,6714, namun memiliki kelemahan pada kelas netral karena dapat menggeser konotasi kalimat sehingga performanya menurun drastis (*F1-score* 0,1373). Sementara itu, *Back Translation* mampu mempertahankan makna melalui proses *paraphrase* dan menunjukkan performa yang lebih seimbang, khususnya pada kelas netral (*F1-score* 0,2593) [14]. Di sisi lain, *Random Swap* dapat meningkatkan variasi struktur kalimat tanpa mengubah konteks utama, meskipun berpotensi merusak tata bahasa jika digunakan secara berlebihan [15]. Meskipun masing-masing teknik mampu membantu mengatasi ketidakseimbangan distribusi kelas, setiap teknik memiliki kelebihan dan keterbatasan dalam menghasilkan variasi data serta mempertahankan makna kalimat. Oleh karena itu, penelitian ini menerapkan kombinasi *Synonym Replacement*, *Back Translation*, dan *Random Swap* agar kelebihan dari masing-masing teknik dapat saling melengkapi. Kombinasi tersebut diharapkan mampu menghasilkan data latih yang lebih beragam dan seimbang sehingga dapat meningkatkan performa model IndoBERT dalam mengklasifikasikan sentimen.

Berdasarkan permasalahan tersebut, penelitian ini menggunakan IndoBERT dalam melakukan analisis sentimen pada komentar TikTok terhadap program Makan Bergizi Gratis (MBG) dengan menerapkan *Back Translation*, *Random Swap*, dan *Synonym Replacement*. Pada penelitian ini, kategori sentimen disesuaikan menjadi pujian, kritik, dan saran untuk memberikan representasi opini yang lebih spesifik. Selain itu, penelitian ini juga menggunakan metode CRISP-

DM sebagai kerangka kerja dalam mengelola seluruh tahapan analisis data, mulai dari proses pemahaman permasalahan hingga tahap evaluasi akhir [16]. Metode CRISP-DM dipilih karena memiliki tahapan yang sistematis dan fleksibel serta banyak digunakan dalam berbagai penelitian data mining, dengan tingkat penggunaan yang dominan dibandingkan metode lainnya. Oleh karena itu, metode ini dinilai mampu mendukung proses analisis sentimen secara menyeluruh dan terarah [17], [18]. Dengan demikian, penelitian ini diharapkan mampu memberikan gambaran mengenai opini yang terkandung dalam komentar pengguna TikTok pada konten program Makan Bergizi Gratis (MBG) yang menjadi sampel penelitian, berdasarkan kategori pujian, kritik, dan saran. Selain itu, penerapan model IndoBERT yang dioptimalkan dengan teknik augmentasi data diharapkan dapat membantu menghasilkan analisis sentimen yang lebih baik sebagai informasi tambahan dalam evaluasi program. Sehingga pada tugas akhir ini diangkat topik penelitian yang berjudul **“ANALISIS SENTIMEN PENGGUNA TIKTOK TERHADAP KONTEN MAKAN BERGIZI GRATIS DALAM ASPEK KUALITAS MAKANAN DENGAN ALGORITMA INDOBERT”**.

2.1 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana distribusi opini pengguna TikTok terhadap aspek kualitas makanan pada Program Makan Bergizi Gratis (MBG) berdasarkan hasil analisis sentimen?
2. Bagaimana kemampuan model IndoBERT dalam mengklasifikasikan opini pengguna TikTok yang mengandung kalimat tidak terstruktur seperti penggunaan singkatan dan slang?
3. Bagaimana mengatasi ketidakseimbangan distribusi kelas pada klasifikasi opini pengguna TikTok terkait program Makan Bergizi Gratis (MBG) guna meningkatkan performa IndoBERT?

3.1 Tujuan

Tujuan penelitian ini adalah sebagai berikut:

1. Mengetahui distribusi opini pengguna TikTok terhadap aspek kualitas makanan pada Program Makan Bergizi Gratis (MBG) berdasarkan hasil analisis sentimen.

2. Menganalisis kemampuan model IndoBERT dalam mengklasifikasikan opini pengguna TikTok yang mengandung kalimat tidak terstruktur seperti penggunaan singkatan dan bahasa slang.
3. Meningkatkan performa model IndoBERT dalam klasifikasi opini pengguna TikTok terkait program Makan Bergizi Gratis (MBG) melalui penerapan teknik augmentasi data untuk mengatasi ketidakseimbangan distribusi kelas.
4. Menghasilkan hasil analisis sentimen yang dapat menggambarkan opini pengguna TikTok terhadap aspek kualitas makanan Program Makan Bergizi Gratis (MBG).

4.1 Manfaat

Beberapa manfaat pada penelitian ini adalah sebagai berikut:

1. Memberikan informasi mengenai distribusi opini pengguna TikTok terhadap aspek kualitas makanan pada Program Makan Bergizi Gratis (MBG) berdasarkan hasil analisis sentimen.
2. Memberikan informasi mengenai kemampuan model IndoBERT dalam mengklasifikasikan komentar yang mengandung singkatan dan bahasa slang.
3. Memberikan kontribusi dalam pengembangan penelitian analisis sentimen berbahasa Indonesia melalui penerapan teknik augmentasi data pada model IndoBERT untuk menangani ketidakseimbangan distribusi kelas.
4. Memberikan gambaran mengenai opini pengguna TikTok terhadap aspek kualitas makanan Program Makan Bergizi Gratis (MBG) berdasarkan hasil analisis sentimen.

5.1 Ruang Lingkup

Beberapa ruang lingkup pada tugas akhir ini:

1. Data yang digunakan merupakan hasil *scraping* menggunakan *Apify* terhadap komentar dari 28 konten TikTok terkait program Makan Bergizi Gratis (MBG) yang dipilih sebagai sampel data penelitian, dengan total sebanyak 61.906 komentar. Data dikumpulkan pada rentang waktu 13 Agustus 2025 hingga 02 Juni 2026.
2. Data yang digunakan hanya berupa komentar berbahasa Indonesia pada platform media sosial TikTok yang membahas program Makan Bergizi Gratis (MBG) dan berkaitan dengan aspek kualitas makanan, seperti rasa, kebersihan, porsi dan kondisi makanan.
3. Penelitian ini menggunakan tiga label sentimen yang terdiri dari pujian, saran, dan kritik.

4. Penelitian ini menggunakan metode CRISP-DM sebagai kerangka kerja dalam proses pengolahan data, yang meliputi tahapan *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *evaluation*, dan *Deployment*.
5. Mengukur kinerja model IndoBERT menggunakan evaluasi *Confusion Matrix*, akurasi, presisi, *recall*, dan *F1-score*.
6. Seluruh proses pengolahan data, pelatihan model, serta evaluasi dilakukan menggunakan bahasa pemrograman *Python* dengan memanfaatkan lingkungan komputasi Google Colab.
7. Proses pelabelan data dilakukan secara manual oleh peneliti berdasarkan pedoman kategorisasi sentimen pujian, kritik, dan saran yang disusun berdasarkan kajian literatur.
8. Teknik augmentasi data yang digunakan dibatasi pada *Back Translation*, *Synonym Replacement*, dan *Random Swap*, serta diterapkan hanya pada data latih (*training data*).
9. *Prototype* aplikasi berbasis *web* dikembangkan menggunakan *Streamlit* dan dijalankan pada lingkungan lokal sebagai media simulasi klasifikasi sentimen, sehingga belum diimplementasikan sebagai *website* yang dapat diakses secara publik.