

BAB I

PENDAHULUAN

Bab ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan skripsi yang berjudul *Analisis Sentimen Opini Publik di Platform Media Sosial X Mengenai Program Makan Bergizi Gratis Menggunakan BERT-CNN*.

1.1 Latar Belakang

Indonesia saat ini tengah berupaya mewujudkan visi “Indonesia Emas 2045” dengan menempatkan fokus utama pada peningkatan kualitas sumber daya manusia yang unggul dan berdaya saing global. Salah satu langkah strategis yang diambil pemerintah baru untuk mendukung visi tersebut adalah implementasi Program Makan Bergizi Gratis (MBG) yang mulai dilaksanakan secara bertahap sejak Januari 2025. Program ini dirancang tidak hanya untuk meningkatkan status gizi anak sekolah dan menekan angka *stunting*, tetapi juga untuk mendorong kehadiran siswa serta meringankan beban ekonomi rumah tangga (Aji, 2025). Dengan alokasi anggaran yang signifikan, yakni mencapai Rp71 triliun dalam Rancangan Anggaran Pendapatan dan Belanja Negara (RAPBN) 2025, program ini menjadi salah satu kebijakan fiskal dan sosial berskala nasional yang memerlukan pengawasan ketat dalam pelaksanaannya. Tantangan seperti kompleksitas distribusi logistik, standarisasi menu, hingga isu transparansi anggaran menjadikan program ini topik yang sangat dinamis dan krusial untuk dievaluasi selama periode awal pelaksanaannya. Di samping itu, keberhasilan program ini juga bergantung pada keterlibatan aktif masyarakat dalam memantau pelaksanaannya di lapangan, menjaga transparansi, serta memberikan umpan balik yang konstruktif.

Di era digital saat ini, partisipasi publik dalam mengawasi kebijakan pemerintah tidak lagi terbatas pada mekanisme formal, melainkan telah merambah ke ruang digital. Platform media sosial X (sebelumnya Twitter) menjadi salah satu wadah utama bagi masyarakat untuk menyuarakan aspirasi, kritik, maupun dukungan terhadap implementasi Program MBG secara *real-time*. Platform X memungkinkan pengguna untuk menyampaikan pendapat mengenai berbagai isu dengan fitur-fitur menarik seperti membagikan cerita, berita, atau twit antar pengguna (Feizollah dkk., 2019). Menurut Wildanu & Agasya (2024) tercatat

Indonesia menduduki peringkat kelima di dunia sebagai negara dengan jumlah pengguna X terbanyak, mencapai 24 juta per tahun 2024. Dengan jumlah pengguna yang sangat besar, platform X dapat menghasilkan data dalam jumlah yang masif namun umumnya bersifat tidak terstruktur dan sulit untuk dikelola.

Analisis sentimen digunakan sebagai pendekatan yang relevan untuk mengatasi permasalahan tersebut. Analisis sentimen merupakan proses komputasional yang bertujuan untuk mengevaluasi opini, sikap, dan emosi terhadap suatu topik, serta mengklasifikasikan polaritasnya ke dalam kategori positif, negatif, atau netral (Singh dkk., 2020). Pendekatan ini memungkinkan evaluasi otomatis terhadap sikap publik apakah cenderung mendukung, menolak, atau netral terhadap berbagai aspek Program MBG. Dalam perkembangannya, berbagai algoritma telah digunakan dalam analisis sentimen, mulai dari *machine learning* konvensional hingga *deep learning*. Pendekatan tradisional seperti *Support Vector Machine* (SVM), *Naïve Bayes*, dan *Random Forest* telah banyak digunakan dan terbukti cukup baik. Aprianti dkk. (2025) menggunakan *Naïve Bayes* dengan teknik pembobotan TF-IDF pada data terkait Program MBG dan mencapai akurasi optimal sebesar 86,78%. Di sisi lain, Tyas dkk. (2025) menunjukkan bahwa pendekatan berbasis *deep learning* seperti *IndoBERT* mampu memberikan pemahaman semantik yang lebih mendalam dalam memprediksi respons masyarakat yang cenderung kritis terhadap Program MBG.

Salah satu terobosan penting dalam *deep learning* untuk bidang *Natural Language Processing* (NLP) adalah *Bidirectional Encoder Representations from Transformers* (BERT), yang dikembangkan oleh Google pada tahun 2018 (Devlin dkk., 2019). BERT merupakan model bahasa *pre-trained* yang menghasilkan representasi kata berbasis konteks (*contextualized embeddings*) melalui pembacaan teks dua arah, sehingga lebih efektif untuk analisis sentimen, terutama pada teks yang kompleks seperti tweet kebijakan pemerintah (Cintaqia & Inoue, 2023; Mishra, 2020). Di sisi lain, *Convolutional Neural Network* (CNN) efektif dalam mengekstraksi fitur lokal melalui operasi konvolusi dengan berbagai ukuran *kernel* (Kim, 2014). Saputra & Isnain (2024) menunjukkan bahwa CNN berbasis *static word embeddings* mampu mencapai akurasi 91% pada klasifikasi sentimen tiga kelas. Meskipun demikian, *static embeddings* masih kurang optimal dalam menangkap makna kontekstual pada teks media sosial yang didominasi bahasa informal, singkatan, dan ambiguitas makna. Keterbatasan tersebut mendorong pengembangan arsitektur yang mengintegrasikan BERT dan CNN untuk menggabungkan keunggulan representasi semantik dan ekstraksi fitur lokal.

Sejumlah penelitian membuktikan bahwa integrasi BERT dan CNN mampu meningkatkan performa klasifikasi sentimen dibandingkan penggunaan CNN secara terpisah. Penelitian Mandhasiya dkk. (2024) menunjukkan bahwa CNN mampu meningkatkan akurasi BERT sebesar 12,97% dalam analisis sentimen politik Indonesia. Lebih lanjut, Mishra (2020) mengembangkan metode BE-CNN dengan arsitektur tiga lapis konvolusi yang menggunakan *multiple kernel sizes* untuk ekstraksi fitur multi-skala, dan mencapai akurasi 97% pada data twit X. Selain itu, Imron dkk. (2023) membuktikan bahwa kombinasi *IndoBERT* dan CNN mengungguli kombinasi *IndoBERT* dan *Long Short-Term Memory* (LSTM) sebesar 5,94% dalam tugas klasifikasi teks bahasa Indonesia, dengan akurasi mencapai 94,86%. Temuan ini mengindikasikan bahwa CNN lebih efektif dalam mengekstraksi fitur lokal dari teks dibandingkan dengan arsitektur sekuensial seperti LSTM, khususnya ketika dikombinasikan dengan representasi kontekstual BERT yang kaya akan informasi semantik.

Berdasarkan berbagai keunggulan tersebut, penelitian ini mengusulkan penggunaan arsitektur hibrida BERT-CNN untuk menganalisis sentimen publik mengenai Program Makan Bergizi Gratis di platform X. Penggunaan BERT sebagai *feature extractor* bertujuan untuk menangkap representasi kontekstual yang mendalam, sementara lapisan CNN difungsikan untuk mengekstraksi fitur lokal yang lebih diskriminatif untuk meningkatkan performa klasifikasi. Melalui skenario penanganan ketidakseimbangan data dan optimasi parameter yang ketat, penelitian ini diharapkan mampu menghasilkan pemetaan polaritas sentimen yang akurat dan reliabel. Hasil analisis ini nantinya dapat menjadi landasan evaluasi strategis bagi pemerintah dalam memahami respons dan ekspektasi masyarakat dalam menyempurnakan implementasi Program MBG di masa mendatang.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan, dapat dirumuskan permasalahan yang diajukan dalam penelitian ini sebagai berikut:

1. Bagaimana konfigurasi *hyperparameter* dan skenario pelatihan yang paling optimal pada model representasi BERT serta arsitektur ekstraksi fitur CNN untuk menghasilkan performa metrik klasifikasi terbaik?
2. Bagaimana performa model BERT-CNN dengan konfigurasi optimal tersebut dalam melakukan klasifikasi sentimen terhadap opini publik mengenai Program MBG?

1.3 Tujuan dan Manfaat

Bagian ini menjelaskan tujuan dan manfaat penelitian sebagai dasar arah serta kontribusi yang diharapkan. Tujuan penelitian ini adalah sebagai berikut:

1. Menentukan konfigurasi *hyperparameter* dan skenario pelatihan yang paling optimal pada model BERT serta arsitektur CNN guna mencapai performa klasifikasi sentimen yang maksimal.
2. Menganalisis dan mengevaluasi performa model BERT-CNN dengan konfigurasi optimal tersebut dalam mengklasifikasikan sentimen opini publik mengenai Program Makan Bergizi Gratis di platform X.

Adapun manfaat yang diharapkan dari penelitian ini adalah:

1. Memberikan kontribusi ilmiah mengenai implementasi integrasi arsitektur berbasis *Transformer* (BERT) dengan CNN dalam pengolahan teks bahasa Indonesia, khususnya pada data media sosial yang bersifat dinamis dan tidak terstruktur.
2. Menjadi referensi akademik terkait strategi eksperimen dan optimasi *hyperparameter* pada model hibrida tingkat lanjut untuk meningkatkan akurasi serta reliabilitas tugas klasifikasi sentimen.
3. Memberikan gambaran awal mengenai pola distribusi opini publik di platform X terhadap Program Makan Bergizi Gratis sebagai referensi empiris yang dapat mendukung penelitian lanjutan dan pengambilan keputusan berbasis data pada masa mendatang.
4. Menjadi rujukan bagi penelitian selanjutnya yang berfokus pada pengembangan sistem analisis opini performa tinggi dengan memanfaatkan kombinasi ekstraksi fitur kontekstual dan spasial.

1.4 Ruang Lingkup

Penelitian ini menetapkan sejumlah batasan untuk menjaga fokus pembahasan agar tetap terarah dan sesuai dengan ruang lingkup yang diteliti. Batasan-batasan tersebut diuraikan sebagai berikut:

1. Data penelitian berupa tweet berbahasa Indonesia dari platform X yang membahas Program Makan Bergizi Gratis dikumpulkan pada periode 6 Januari 2025 hingga 6

Januari 2026, dengan total 6.642 data yang telah melalui serangkaian proses validasi dan pengolahan sehingga memenuhi kriteria untuk dianalisis.

2. Klasifikasi sentimen difokuskan pada tiga kategori utama, yaitu positif, negatif, dan netral.
3. Proses pelabelan data dilakukan secara manual oleh peneliti sendiri sebagai anotator tunggal dikarenakan keterbatasan sumber daya peneliti, dengan berpedoman pada panduan anotasi yang telah ditetapkan berdasarkan konteks bahasa Indonesia dan isu Program Makan Bergizi Gratis.
4. Model dasar yang digunakan sebagai representasi teks difokuskan pada model pra-latih *IndoBERT* (*indobenchmark/indobert-base-p2*), yang diintegrasikan dengan lapisan klasifikasi *Convolutional Neural Network 1D*.
5. Optimasi model dibatasi pada pencarian *hyperparameter* pembelajaran seperti *learning rate*, *batch size*, *weight decay*, teknik penyeimbangan data, serta penyetelan arsitektur klasifikasi seperti ukuran *n-gram* atau *filter*, jumlah *filter*, *dropout*, dan fungsi aktivasi.
6. Seluruh proses komputasi, pelatihan, dan pengujian model menggunakan pendekatan *Stratified K-Fold Cross Validation* yang dieksekusi pada lingkungan *cloud computing* Google Colab dan Kaggle Notebook.

1.5 Sistematika Penulisan

Sistematika penulisan memberikan gambaran laporan dari penelitian ini secara urut dan jelas. Berikut adalah sistematika penulisan dari penelitian ini:

BAB I PENDAHULUAN

Bab ini menyajikan mengenai latar belakang, rumusan masalah, manfaat dan tujuan, ruang lingkup, dan sistematika penulisan skripsi mengenai analisis sentimen opini publik di platform media sosial X mengenai Program MBG menggunakan BERT-CNN.

BAB II LANDASAN TEORI

Bab ini menyajikan teori-teori yang berhubungan dengan topik atau masalah yang dibahas pada skripsi analisis sentimen opini publik di platform media sosial X mengenai Program MBG menggunakan BERT-CNN.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan informasi mengenai metodologi penelitian yang digunakan dalam analisis sentimen opini publik di platform media sosial X mengenai Program MBG menggunakan BERT-CNN.

BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas terkait analisis dan hasil pembahasan dari penulisan skripsi analisis sentimen opini publik di platform media sosial X mengenai Program MBG menggunakan BERT-CNN. Selain itu, bab ini juga membahas mengenai lingkungan dan perangkat yang digunakan dalam skripsi, skenario pengujian, dan hasil pengujian.

BAB V PENUTUP

Bab ini menyajikan kesimpulan dari penjelasan yang telah diuraikan pada bab-bab sebelumnya. Bab ini juga memberikan saran dari penulis sebagai bahan masukan untuk pengembangan dan penelitian lebih lanjut di masa yang akan datang.