

BAB I

PENDAHULUAN

1.1 Latar Belakang

Stunting merupakan masalah penting yang harus diatasi di Indonesia. Stunting disebabkan oleh malnutrisi sejak dalam kandungan serta asupan gizi yang tidak memadai atau infeksi di masa kanak-kanak dapat menghambat pertumbuhan fisik dan perkembangan kognitif anak, melemahkan sistem kekebalan tubuh, serta meningkatkan risiko infeksi, keterlambatan perkembangan jangka panjang, dan kematian (UNICEF, 2022). Menurut statistik UNICEF (2022) Di antara negara-negara ASEAN, Indonesia mencatat prevalensi bayi lahir dengan berat badan rendah (*low birthweight*) sebesar 6,2%, menjadikannya negara dengan angka terendah dalam daftar tersebut. Bayi dengan berat lahir kurang dari 2.500 gram dikategorikan sebagai *low birthweight*, yang merupakan indikator penting dari kondisi kesehatan ibu dan janin selama kehamilan. Meskipun angka ini lebih baik dibandingkan negara ASEAN lain seperti Filipina (14,5%) dan Laos (17,3%), Indonesia tetap menghadapi tantangan serius karena bayi dengan berat badan rendah memiliki risiko lebih tinggi untuk kematian neonatal (dalam 28 hari pertama), pertumbuhan terhambat, IQ yang lebih rendah, serta kerentanan terhadap penyakit kronis seperti diabetes dan obesitas di usia dewasa. Sebagai respons terhadap permasalahan tersebut, pemerintah Indonesia meluncurkan program makan bergizi gratis yang ditargetkan kepada anak sekolah dan ibu hamil.

Program Makan Bergizi Gratis merupakan salah satu program prioritas nasional yang diinisiasi oleh Presiden Prabowo Subianto dengan tujuan utama meningkatkan kualitas gizi masyarakat, khususnya anak-anak sekolah dan ibu hamil, guna menurunkan angka stunting serta memperkuat sumber daya manusia Indonesia di masa mendatang (Sarjito, 2024). Program ini Bertujuan untuk meningkatkan asupan harian dengan anggaran yang terjangkau, sekitar Rp10.000 per individu per target (Kemenkes, 2023). Program ini dirancang untuk memberikan makan siang dan susu secara gratis kepada jutaan siswa dari jenjang SD hingga SMA/SMK, serta kelompok rentan lainnya, dengan harapan dapat mengubah kemampuan dan masa depan anak-anak Indonesia (Esa dkk., 2024). Namun hal ini menimbulkan berbagai pro dan kontra dari masyarakat melalui media sosial.

Media sosial, khususnya platform seperti *Twitter* yang sekarang dikenal sebagai X, telah menjadi wadah utama bagi masyarakat untuk menyampaikan opini, berbagi informasi,

dan membentuk opini publik secara luas. X mencapai rekor tertinggi sepanjang masa dengan 541,56 juta pengguna aktif bulanan pada tahun 2023. Jumlah ini meningkat 47% dari tahun 2022, yang hanya memiliki 368 juta pengguna aktif bulanan. Dikutip dalam famewall.io, 2024 oleh Amelia & Yusuf (2025), Indonesia menduduki peringkat keempat di dunia dengan pengguna aktif di X sebanyak 24,85 juta dari populasi masyarakat Indonesia yang berjumlah 278.696.200 jiwa. Tingginya jumlah pengguna tersebut menjadikan X sebagai sumber data yang sangat potensial untuk analisis sentimen, karena volume unggahan yang dihasilkan setiap harinya sangat besar, beragam, dan mencerminkan opini publik secara *real-time*. Selain itu, sifat X yang berbasis teks singkat membuat informasi yang dibagikan lebih padat dan langsung pada inti opini, sehingga memudahkan proses pengolahan data untuk tujuan penelitian. Karakteristik ini menjadikan platform X sebagai pilihan yang relevan dan strategis untuk mengkaji kecenderungan sentimen masyarakat terhadap topik tertentu.

Analisis sentimen merupakan metode yang efektif untuk mengkategorikan opini publik ke dalam sentimen positif, negatif, atau netral. Proses ini memungkinkan data teks yang tidak terstruktur, seperti unggahan di platform X, diubah menjadi informasi terukur yang dapat dianalisis secara sistematis. Dengan demikian, analisis sentimen tidak hanya memberikan gambaran yang jelas mengenai kecenderungan opini masyarakat terhadap platform X, tetapi juga membantu dalam mengidentifikasi tren, isu populer, serta perubahan persepsi publik dari waktu ke waktu. Banyak penelitian sebelumnya telah memanfaatkan analisis sentimen untuk memahami bagaimana opini publik terbentuk dan berubah terhadap suatu topik, peristiwa, atau entitas tertentu.

Penelitian yang dilakukan oleh Hermansyah & Hasibuan (2025) menganalisis sentimen masyarakat terhadap Program Makan Bergizi Gratis melalui media sosial X dengan memanfaatkan algoritma *machine learning* klasik yaitu *Linear Regression* dan *Random Forest*. Dalam studi ini, peneliti mengumpulkan data dari media sosial, melakukan *preprocessing* teks seperti normalisasi dan *stemming*, serta memberikan label sentimen untuk kemudian diklasifikasikan menggunakan kedua algoritma tersebut. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* memiliki performa lebih unggul dibandingkan *Linear Regression*, dengan akurasi mencapai 85% dan F1-Score sebesar 90%, terutama dalam mendeteksi sentimen positif. Sementara itu, *Linear Regression* menunjukkan keunggulan dalam aspek presisi, meskipun secara keseluruhan *Random Forest* memberikan hasil klasifikasi yang lebih seimbang. Mayoritas sentimen publik terhadap

program ini bersifat positif, yakni sebesar 69,5%, meskipun tetap terdapat kritik terkait aspek implementasi dan pendanaan. Penelitian Hermansyah & Hasibuan (2025) menyoroti pentingnya pemanfaatan media sosial sebagai sumber informasi untuk memahami persepsi publik secara langsung dan sebagai dasar perumusan kebijakan publik yang lebih responsif dan efektif.

Penelitian lain dilakukan oleh Khemani dkk. (2024) berbeda dari penelitian Hermansyah & Hasibuan (2025) yang menggunakan algoritma *machine learning* klasik, Penelitian Khemani dkk. (2024) menggunakan metode *deep learning* untuk menguji efektivitas model dalam klasifikasi sentimen dibandingkan model *machine learning* klasik seperti *Linear Regression* dan *Random Forest*. Hasil penelitian membuktikan bahwa model *deep learning* yang dipakai yaitu *Graph Convolutional Network* (GCN) yang merupakan salah satu model *Graph Neural Network* (GNN) mendapatkan akurasi lebih baik daripada model - model *machine learning* pada umumnya. GCN terbukti memiliki akurasi tertinggi yaitu 93,86% dan lebih baik dibandingkan model lain seperti *passive aggressive classifier*, *random forest*, *decision tree*, *logistic regression*, *light GBM*.

Dalam bidang pemrosesan bahasa alami, pemilihan metode ekstraksi fitur yang tepat sangat penting untuk menghasilkan representasi teks yang informatif dan akurat. Salah satu metode yang banyak digunakan adalah *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF merupakan teknik statistik yang menonjolkan kata-kata penting dalam dokumen dengan mempertimbangkan frekuensi kemunculan kata di dokumen tertentu dan keseluruhan korpus, sehingga efektif untuk menyoroti kata kunci unik yang relevan dengan konteks. Studi terbaru oleh Zhou, dkk. (2024) yang membandingkan antara TF-IDF dan word2vec menunjukkan bahwa TF-IDF efektif dalam mengekstraksi perilaku respons utama dari data proses, sedangkan Word2Vec lebih unggul dalam menangkap fitur urutan perilaku yang kompleks. Penelitian Zhou dkk. (2024) juga menegaskan bahwa TF-IDF lebih cocok digunakan pada dataset kecil atau kasus sederhana karena sifatnya yang cepat, mudah, dan tetap memberikan hasil yang kompetitif, sementara Word2Vec lebih direkomendasikan untuk dataset besar dan kebutuhan generalisasi yang tinggi karena kemampuannya menangkap makna dan konteks kata serta ketahanannya terhadap *overfitting*, untuk hasil yang lebih optimal, kombinasi antara TF-IDF dan Word2Vec dapat dipertimbangkan guna memaksimalkan pemahaman model terhadap kata kunci sekaligus konteks. Temuan ini

menegaskan pentingnya pemilihan metode ekstraksi fitur yang sesuai dengan karakteristik data dan tujuan analisis dalam proyek-proyek analisis teks dan sentimen.

Berdasarkan temuan-temuan tersebut, penelitian skripsi ini akan berfokus pada pemanfaatan model *Graph Neural Network* (GNN), yaitu *Graph Convolutional Network* (GCN) untuk tugas klasifikasi sentimen terhadap opini masyarakat mengenai Program Makan Bergizi Gratis melalui media sosial X (*Twitter*). Untuk merepresentasikan teks ke dalam bentuk numerik, digunakan pendekatan umum dalam pemrosesan bahasa alami, yaitu *Term Frequency-Inverse Document Frequency* (TF-IDF). Pemilihan kombinasi antara representasi teks dan model GNN didasarkan pada kebutuhan untuk menangkap baik informasi semantik lokal maupun struktur relasional antar kata dalam dokumen dan jumlah dataset yang terbatas. Dengan pendekatan ini, penelitian ini bertujuan mengeksplorasi efektivitas metode berbasis graf dalam memahami dan mengklasifikasikan sentimen publik yang tersebar melalui platform digital secara lebih kontekstual dan terstruktur.

1.2 Rumusan Masalah

Penelitian ini berangkat dari kebutuhan untuk memahami bagaimana performa model *Graph Convolutional Network* (GCN) dengan representasi teks berbasis TF-IDF dalam mengklasifikasikan sentimen opini publik terhadap Program Makan Bergizi Gratis di media sosial X dengan cara membandingkan dengan model-model *machine learning* klasik seperti *Logistic Regression*, *Support Vector Machine* (SVM), *Random Forest*, *XGboost*, *Naïve bayes*. Selain itu, penelitian ini juga mengevaluasi sejauh mana performa menggunakan strategi *undersampling* berbasis pemilihan kalimat terpanjang dari masing-masing kelas mayoritas yang akan di *undersampling*. Pendekatan ini didasarkan pada asumsi bahwa kalimat yang lebih panjang cenderung mengandung informasi yang lebih kaya dan kontekstual, sehingga berpotensi meningkatkan kualitas representasi teks dalam proses pelatihan model klasifikasi.

1.3 Batasan Masalah

Penelitian ini dibatasi pada klasifikasi sentimen publik terhadap Program Makan Bergizi Gratis yang disampaikan melalui platform media sosial X (*Twitter*). Pengambilan data dari media sosial X karena media sosial ini karena beberapa faktor antara lain, banyaknya pengguna indonesia di X, relevansi dan kecepatan dalam menangkap opini

publik, X juga menyediakan akses data lebih mudah dibanding media lainnya. Data dikumpulkan secara otomatis menggunakan sistem pencarian berbasis kata kunci, yaitu “makan bergizi gratis” dan “makan siang gratis”, kata kunci “MBG” tidak digunakan karena waktu itu masih belum jelas informasi tentang nama program yang digunakan pemerintah, antara “makan bergizi gratis” atau “makan siang gratis”. Rentang waktu pengambilan data dibatasi antara tanggal 1 November 2024 hingga 31 Maret 2025, dengan asumsi bahwa periode ini mencerminkan masa awal perbincangan publik terkait program kebijakan tersebut. Total data yang dikumpulkan sebanyak 9249 baris. Model hanya fokus pada GCN dan perbandingan hanya dilakukan dengan model *machine learning classic*, hal ini karena keterbatasan waktu. Label sentimen dilakukan secara manual dan evaluasi kinerja model dibatasi pada empat metrik utama, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk mengklasifikasi sentimen publik terhadap program makan bergizi gratis dan membandingkan performa model *Graph Convolutional Network* (GCN) yang menggunakan representasi teks berbasis TF-IDF dalam mengklasifikasikan sentimen publik terhadap Program Makan Bergizi Gratis di media sosial X, dengan beberapa model *machine learning* klasik seperti *Logistic Regression*, *SVM*, *Random Forest*, *XGBoost*, dan *Naïve Bayes*. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi efektivitas strategi *undersampling* berbasis pemilihan kalimat terpanjang dari masing-masing kelas.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1.5.1 Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan ilmu pengetahuan, khususnya dalam bidang pemrosesan bahasa alami (*Natural Language Processing/NLP*) dan analisis sentimen. Dengan memanfaatkan model *Graph Convolutional Network* (GCN) berbasis representasi TF-IDF, penelitian ini berkontribusi dalam memperluas kajian penerapan model berbasis graf dalam konteks analisis teks berbahasa Indonesia. Selain itu, penelitian ini juga memberikan pemahaman lebih lanjut mengenai

efektivitas teknik penyeimbangan data melalui strategi *undersampling* dalam pengembangan model klasifikasi di masa mendatang.

1.5.2 Manfaat Praktis

Secara praktis, hasil penelitian ini diharapkan dapat dimanfaatkan oleh:

1. Pemerintah dan pembuat kebijakan, sebagai acuan untuk memahami persepsi publik terhadap kebijakan Program Makan Bergizi Gratis, sehingga dapat dijadikan sebagai dasar evaluasi dan penyempurnaan program secara lebih adaptif dan berbasis data.
2. Peneliti dan praktisi di bidang data dan teknologi informasi, sebagai rujukan dalam memilih model klasifikasi sentimen yang tepat berdasarkan karakteristik data, serta eksplorasi metode baru untuk penanganan ketidakseimbangan kelas pada data sosial media.
3. Mahasiswa dan akademisi, sebagai literatur tambahan dalam pengembangan kajian ilmiah di bidang NLP, *deep learning*, serta penerapan GNN dalam klasifikasi data teks berbahasa Indonesia.