

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pembangunan sumber daya manusia (SDM) unggul merupakan kunci kemajuan suatu bangsa dan prioritas utama pembangunan nasional yang salah satu prasyaratnya adalah terpenuhinya kebutuhan nutrisi masyarakat. Hasil Survei Status Gizi Indonesia (SSGI) 2024 mencatat prevalensi stunting sebesar 19,8% (Muhawarman, 2025), hal ini mencerminkan bahwa permasalahan gizi di Indonesia belum sepenuhnya teratasi. Respon pemerintah terhadap kondisi tersebut ialah menginisiasi Program Makan Bergizi Gratis (MBG) dengan target penurunan angka stunting menjadi 14,2% pada tahun 2029 (Muhawarman, 2025).

Program MBG menargetkan 82,9 juta penerima manfaat hingga tahun 2029 (Sidik, 2025) dari berbagai kelompok sasaran yaitu peserta didik jenjang PAUD sampai SMA/SMK dan pesantren, anak balita, ibu hamil dan ibu menyusui (Kemensetneg, 2024). Dukungan anggaran untuk program ini senilai 71 triliun pada APBN 2025 (Kemenkeu, 2024) dan diperkirakan akan terus meningkat hingga tahun 2029. Skala program yang luas menimbulkan berbagai tantangan implementasi, termasuk laporan dugaan keracunan makanan dari MBG di beberapa provinsi yang mencapai 12.658 anak pada tahun 2025 (KPAI, 2026). Kondisi tersebut memunculkan beragam respon dan kekhawatiran masyarakat terhadap efektivitas dan keamanan program MBG (Hendarto, 2025).

Perkembangan media sosial memperluas ruang pembentukan opini publik yang memungkinkan masyarakat dapat menyampaikan pandangannya secara terbuka dan spontan (Zhou, 2024). Media sosial yang relevan adalah TikTok dengan jumlah

pengguna di Indonesia mencapai 180,11 juta per Oktober 2025 (Maheswara, 2025). Pengguna TikTok mengekspresikan respon terhadap konten yang di unggah termasuk konten terkait program MBG melalui kolom komentar berupa dukungan, kritik, maupun kekhawatiran dengan bahasa yang digunakan cenderung informal dan beragam sehingga menghadirkan tantangan tersendiri dalam proses analisis teks (Wijaya dan Suwandhi, 2024).

Karakteristik dalam komentar TikTok mencakup variasi kosakata, penggunaan istilah *slang*, emoji, serta perbedaan konteks makna antar pengguna berpotensi meningkatkan dimensi fitur dan dapat memengaruhi kinerja metode klasifikasi (Ganie, 2023). Data teks seperti komentar perlu direpresentasikan ke dalam bentuk numerik melalui teknik pembobotan kata seperti *Term Frequency-Inverse Document Frequency* (TF-IDF) sebelum diklasifikasi. TF-IDF memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen serta tingkat kelangkaannya pada keseluruhan dokumen, sehingga memungkinkan data dapat diproses oleh algoritma klasifikasi.

Metode klasifikasi yang sering digunakan dalam analisis teks adalah *K-Nearest Neighbor* (KNN) karena mampu menangani data multikelas dengan pendekatan yang sederhana. KNN merupakan metode klasifikasi yang menentukan label suatu objek baru berdasarkan k objek tetangga terdekatnya (Ridho *et al.*, 2024). Kinerja KNN sangat dipengaruhi oleh jumlah dan kualitas fitur yang digunakan. Meskipun pembobotan TF-IDF mampu mempresentasikan kepentingan suatu kata dalam dokumen, jumlah fitur yang besar tetap dapat memengaruhi stabilitas dan akurasi model (Subandono dan Ariatmanto, 2025). Penerapan teknik seleksi fitur menjadi tahap krusial dalam meningkatkan performa klasifikasi.

Metode seleksi fitur yang umum digunakan dalam analisis sentimen adalah *Chi-Square* dan *Information Gain*. *Chi-Square* mengukur tingkat ketergantungan statistik antara suatu fitur dan kelas sentimen, sehingga mampu mengidentifikasi fitur yang memiliki hubungan signifikan terhadap kategori sentimen. *Information Gain* mengukur penurunan entropi atau ketidakpastian kelas ketika suatu fitur diketahui, sehingga mencerminkan kontribusi fitur terhadap kelas target (Subandono dan Ariatmanto, 2025). Kedua metode ini banyak digunakan untuk mereduksi dimensi fitur teks dan meningkatkan efektivitas proses klasifikasi.

Penelitian terdahulu menunjukkan bahwa efektivitas metode seleksi fitur dan algoritma klasifikasi sangat dipengaruhi oleh karakteristik data. Subandono dan Ariatmanto (2025) menemukan bahwa *Chi-Square* lebih efektif dalam meningkatkan akurasi model dan performa *Support Vector Machine* (SVM) lebih unggul dibandingkan *Random Forest* pada data ulasan aplikasi. Yousef dan ALali (2022) menemukan bahwa kombinasi SVM dengan *Information Gain* menghasilkan performa terbaik dibandingkan KNN dan *Decision Tree* pada data Twitter. Hasil tersebut menunjukkan bahwa SVM sering kali unggul pada klasifikasi teks, namun performa algoritma tetap bergantung pada konteks data yang digunakan.

Penelitian Mutiarachim *et al.* (2025) membandingkan *Chi-Square* dan *Information Gain* pada KNN menunjukkan hasil bahwa penerapan *Information Gain* mampu meningkatkan akurasi serta menurunkan tingkat kesalahan klasifikasi. Hasil penelitian ini mengindikasikan bahwa seleksi fitur memiliki peran penting dalam meningkatkan kinerja algoritma berbasis jarak seperti KNN. KNN juga memiliki performa yang lebih optimal dibandingkan SVM pada klasifikasi data numerik penerima bantuan sosial (Ridho *et al.*, 2024). Hal ini semakin menegaskan bahwa

efektivitas algoritma klasifikasi tidak bersifat universal, melainkan dipengaruhi oleh karakteristik dan representasi data. Berdasarkan beberapa penelitian tersebut, dapat disimpulkan bahwa pemilihan metode pembobotan dan seleksi fitur berperan penting dalam meningkatkan performa klasifikasi. Pada penelitian ini penulis melakukan analisis sentimen terhadap komentar publik di video TikTok tentang Program MBG dengan judul “Perbandingan Seleksi Fitur *Chi-Square* dan *Information Gain* untuk Klasifikasi Sentimen Program MBG Menggunakan *K-Nearest Neighbor*”.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, penulis merumuskan masalah yang akan dibahas dalam penelitian ini, yaitu:

1. Menganalisis dan membandingkan hasil kinerja metode seleksi fitur *Chi-Square* dan *Information Gain* dalam klasifikasi komentar masyarakat terkait program MBG menggunakan algoritma *K-Nearest Neighbor* (KNN),
2. Menguji apakah perbedaan kinerja yang dihasilkan oleh kedua metode tersebut signifikan secara statistik berdasarkan metode evaluasi yang digunakan.

1.3 Batasan Masalah

Pada penelitian ini diberikan batasan masalah yaitu data yang digunakan terdiri dari komentar pengguna TikTok pada 10 video tentang program MBG yang diunggah pada bulan Agustus dan September 2025. Video tersebut terdiri dari tiga kategori yaitu video positif, negatif, dan video netral atau informatif. Dari masing-masing video diambil 150 komentar dengan total 1.500 komentar berbahasa Indonesia yang relevan terhadap topik menggunakan teknik *web scrapping* melalui *Jupyter Notebook*.

1.4 Tujuan Penelitian

Berdasarkan latar belakang yang telah diuraikan, tujuan dari penelitian ini adalah :

1. Menganalisis dan membandingkan kinerja metode seleksi fitur *Chi-Square* dan *Information Gain* dalam klasifikasi komentar masyarakat terkait program MBG menggunakan algoritma *K-Nearest Neighbor* (KNN)
2. Mengetahui perbedaan kinerja yang dihasilkan oleh kedua metode tersebut sehingga dapat memberikan dasar-dasar pengambilan keputusan yang objektif dalam pemilihan metode seleksi fitur yang optimal.