

BAB I PENDAHULUAN

1.1 Latar Belakang Masalah

Sebagai salah satu langkah strategis pemerintah untuk mendorong status gizi warga terutama anak usia sekolah dan kelompok yang rentan secara sosial-ekonomi kebijakan Makan Bergizi Gratis (MBG) diluncurkan di tengah publik yang sudah cukup akrab dengan polemik semacam ini. Program makan siang gratis, misalnya, sebelumnya lebih dulu memicu perdebatan pro-kontra, khususnya menyangkut konsekuensi kesehatan dan beban anggaran negara (Pratama & Wibowo, 2024). Cakupannya yang nasional, ditambah alokasi dana besar dan efek sosial yang meluas, membuat pelaksanaan MBG pun tak lepas dari beragam tanggapan publik.

Di era digital, ruang untuk menuangkan pendapat, kritik, maupun dukungan kian terbuka lebar lewat berbagai kanal jejaring sosial, dan media sosial menjadi salah satu yang paling banyak dimanfaatkan masyarakat. Arus informasi di sana menyebar dalam hitungan singkat berkat sifatnya yang dinamis, kendati data yang dihasilkan tak jarang sarat ungkapan tidak baku atau *slang* yang justru menyulitkan proses analisis (Cahyani & Abdillah, 2024). Beragamnya pandangan warganet terhadap MBG pun tercermin di sana, mulai dari pujian atas pemerataan asupan gizi hingga sorotan tajam terhadap pola distribusi dan variasi menu.

Bila diolah secara tepat, tumpukan data berbasis teks semacam ini sesungguhnya adalah aset berharga yang dapat dijadikan dasar evaluasi kebijakan hanya saja, ukurannya yang besar dan sifatnya yang tidak terstruktur membuat analisis manual menjadi kurang praktis. Kebutuhan inilah yang mendorong penggunaan pendekatan komputasional melalui *Natural Language Processing* (NLP) dan *Machine Learning* untuk melakukan penambangan opini (*opinion mining*), guna menangkap kecenderungan sikap penulis terhadap suatu isu (Ryandi, Pratiwi, & Sari, 2025). Keberhasilan proses tersebut sangat ditentukan oleh salah satu tahapnya, yaitu *text preprocessing*, yang mengolah data mentah menjadi

bentuk siap pakai agar akurasi model dapat ditingkatkan (Albab, Yuniar, & Fawaiq, 2023).

Dua algoritma paling sering diadu performanya dalam klasifikasi teks untuk keperluan analisis sentimen adalah *Naïve Bayes* dan *Support Vector Machine* (SVM). Keunggulan SVM terletak pada kemampuan menggeneralisasi ruang fitur berdimensi tinggi, dan terbukti lebih tangguh memisahkan *hyperplane* pada data teks politik yang rumit (Lestari & Hartono, 2024), sedangkan *Naïve Bayes* lebih dikenal lewat kecepatan komputasi serta efektivitasnya menangani *dataset* berukuran besar meski mengandalkan asumsi independensi antar fitur yang tergolong sederhana (Mauliza & Sipayung, 2024).

Guna mendongkrak kemampuan SVM mengolah karakteristik teks media sosial yang beraneka ragam, penelitian ini memadukan dua strategi ekstraksi fitur berbasis N-gram, yakni *Word N-gram* dan *Character N-gram*. Bila *Word N-gram* berperan menangkap konteks frasa dan susunan kata bermuatan sentimen, *Character N-gram* justru unggul mengenali kata tidak baku, singkatan, maupun variasi ejaan yang lazim dijumpai pada teks informal berbahasa Indonesia di media sosial (Iriananda et al., 2024). Harapannya, perpaduan kedua pendekatan tersebut mampu mendorong SVM mengklasifikasikan sentimen dengan akurasi yang lebih tinggi (Nugroho, 2021).

Meski sudah lazim dipakai, keberhasilan kedua algoritma tersebut tetap sangat bergantung pada karakteristik *dataset* serta domain topik yang dikaji, dan sejauh ini belum banyak kajian yang secara khusus membandingkan kinerjanya dengan pendekatan gabungan *Word & Character N-gram* pada opini publik berbahasa Indonesia terkait kebijakan MBG. Celah inilah yang coba dijawab oleh penelitian ini, dengan membandingkan kinerja SVM berfitur *Word & Character N-gram* dan *Naïve Bayes* untuk menemukan metode paling akurat dalam mengklasifikasikan sentimen masyarakat terhadap Program MBG.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana penerapan algoritma *Naïve Bayes* serta *Support Vector Machine* (SVM) yang dilengkapi fitur *Word N-gram* dan *Character N-gram* dalam mengklasifikasikan sentimen masyarakat terhadap Program MBG?
2. Bagaimana hasil perbandingan kinerja kedua metode tersebut bila ditinjau dari metrik *accuracy*, *precision*, *recall*, dan *f1-score*?

1.3 Batasan Masalah

Agar pembahasan tidak melebar dan tetap terarah, penulis menetapkan sejumlah batasan penelitian berikut ini:

1. Data yang dipakai bersumber dari dataset terbuka di Kaggle, berisi opini masyarakat seputar Program Makan Bergizi Gratis yang dihimpun dari media sosial.
2. Data yang diambil mencakup rentang tahun 2024 hingga 2025.
3. Analisis hanya mencakup teks berbahasa Indonesia, sementara data berbahasa asing atau daerah akan disingkirkan atau diterjemahkan bila memungkinkan.
4. Algoritma klasifikasi yang diperbandingkan meliputi *Multinomial Naïve Bayes* dan *Support Vector Machine* berfitur *Word N-gram* serta *Character N-gram*, dengan hasil klasifikasi sentimen dikelompokkan menjadi tiga kelas: Positif, Negatif, dan Netral.

1.4 Tujuan penelitian

Adapun tujuan yang hendak dicapai melalui penelitian ini adalah:

1. Menerapkan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) berfitur *Word N-gram* serta *Character N-gram* untuk mengelompokkan

sentimen masyarakat terhadap Program Makan Bergizi Gratis (MBG) ke dalam kategori positif, negatif, atau netral.

2. Mengukur sekaligus membandingkan performa algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) berfitur *Word N-gram* serta *Character N-gram*, melalui nilai *accuracy*, *precision*, *recall*, dan *f1-score*, guna menentukan metode dengan kinerja terbaik.

1.5 Manfaat penelitian

1. Manfaat Teoritis: Penelitian ini diharapkan dapat menyumbangkan wawasan ilmiah bagi bidang ilmu komputer dan *data science*, khususnya menyangkut perbandingan performa algoritma klasifikasi pada data teks berbahasa Indonesia yang mengangkat isu kebijakan publik.
2. Manfaat Praktis: Bagi para pemangku kepentingan, penelitian ini diharapkan dapat menyediakan *insight* berbasis data mengenai persepsi publik, yang dapat dimanfaatkan sebagai salah satu bahan evaluasi Program MBG.

1.6 Sistematika Penulisan

Untuk memberi gambaran menyeluruh, skripsi ini dibagi ke dalam beberapa bab dengan rincian sebagai berikut:

BAB I : PENDAHULUAN

Bab ini memuat latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, sistematika penulisan, dan penelitian-penelitian terkait.

BAB II : TINJAUAN PUSTAKA

Bab ini menguraikan landasan teori seputar Analisis Sentimen, *Machine Learning*, *Natural Language Processing* (NLP), TF-IDF, *Naïve Bayes*, SVM, hingga evaluasi model.

BAB III : METODOLOGI PENELITIAN

Bab ini memaparkan jenis penelitian, sumber data, rancangan sistem, tahapan *preprocessing*, pembentukan model, dan proses evaluasinya.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi dan analisis perbandingan performa model yang telah dibangun.

BAB V : KESIMPULAN

Bab penutup ini merangkum simpulan dari keseluruhan pembahasan sekaligus saran yang berguna bagi pengembangan penelitian selanjutnya.

1.7 Penelitian Terkait

Tabel 1.1 Penelitian Terkait

Penulis dan Tahun	Judul	Hasil / Pencapaian
Pratama, A., & Wibowo, A. T. (2024)	Analisis Sentimen Terhadap Program Makan Siang Gratis Menggunakan Algoritma Naïve Bayes Classifier	Menghasilkan akurasi 84%. Sentimen positif mendominasi, namun terdapat kekhawatiran terkait anggaran.
Lestari, D. P., & Hartono, R. (2024)	Komparasi Kinerja Algoritma SVM dan Naive Bayes dalam Analisis Sentimen Pemilu 2024	Algoritma SVM unggul dengan akurasi 89%, sedangkan Naive Bayes memperoleh akurasi 85%. SVM lebih baik pada data teks politik kompleks.

Cahyani, P., & Abdillah, L. A. (2024)	Perbandingan Performa Algoritma Naïve Bayes, SVM dan Random Forest: Studi Kasus Analisis Sentimen Pengguna Sosial Media X	SVM menghasilkan akurasi tertinggi sebesar 87%. Normalisasi kata (<i>slang</i>) sangat mempengaruhi akurasi akhir.
Susanto, L. A. (2024)	Komparasi Model SVM dan K-Nearest Neighbor Pada Analisis Sentimen Aplikasi Polri Super App	SVM memiliki performa lebih baik (akurasi 88%) dibandingkan KNN (82%) untuk data ulasan aplikasi.
Ramadhan, M. F. (2024)	Analisis Sentimen Kebijakan Publik pada Platform TikTok Menggunakan Support Vector Machine	Model SVM menghasilkan akurasi 86%. Tantangan utama adalah emoji dan singkatan ekstrem.
Sari, N., & Nugroho, E. (2024)	Penerapan Metode SMOTE untuk Mengatasi Imbalanced Data pada Analisis Sentimen Layanan	Penerapan SMOTE berhasil meningkatkan akurasi signifikan dari 72% menjadi 83% pada data tidak seimbang.

	Masyarakat Menggunakan Naive Bayes	
Badri, B. (2023)	Sentiment Analisis Untuk Identifikasi Kepuasan Masyarakat Terhadap Kenaikan BBM Menggunakan Algoritma Naïve Bayes	Akurasi sebesar 81%. Sentimen negatif dominan di awal, namun pola berubah seiring waktu
Iriananda, S. W., et al. (2024)	Optimasi Klasifikasi Sentimen Komentar Pengguna Game Menggunakan SVM, Grid Search dan Kombinasi N-Gram	Kombinasi Word N-gram (Unigram+Bigram+Trigram) pada SVM terbukti meningkatkan akurasi klasifikasi sentimen teks berbahasa Indonesia secara signifikan.
Wijaya, K. (2023)	Implementasi Support Vector Machine Pada Analisis Sentimen Penerima Bantuan Sosial Tunai	Penggunaan Kernel RBF pada SVM memberikan hasil pemisahan kelas terbaik dengan akurasi mencapai 88.5%.

Putri, S. R., et al. (2023)	Analisis Sentimen Masyarakat Terhadap Kebijakan Kurikulum Merdeka Menggunakan Metode Naive Bayes	Akurasi sebesar 79%. Penggunaan daftar <i>stopword</i> yang tepat sangat mempengaruhi hasil akurasi.
Hidayat, T. (2023)	Analisis Sentimen Kebijakan Kenaikan PPN 11% Menggunakan Algoritma K-Nearest Neighbor	Akurasi model mencapai 76%. KNN terbukti memiliki keterbatasan pada data teks berdimensi tinggi dibandingkan SVM/NB.
Iriananda, S. W., et al. (2022)	Kinerja Seleksi Fitur N-Gram Pada Analisis Sentimen Ulasan Mobile Game di Google Playstore	Seleksi fitur berbasis N-gram terbukti meningkatkan nilai akurasi algoritma klasifikasi pada teks berbahasa Indonesia.