

BAB I

PENDAHULUAN

1.1 Latar Belakang

Era digital membuat arus informasi berkembang dengan sangat cepat, termasuk dalam penyebaran opini publik terkait kebijakan pemerintah. Salah satu kebijakan yang banyak mendapat sorotan adalah program makan bergizi gratis, yang memunculkan beragam tanggapan dari masyarakat di ruang digital [1]. Media sosial X menjadi salah satu platform utama bagi masyarakat untuk menyampaikan pandangan, baik dalam bentuk dukungan maupun kritik terhadap program tersebut. X juga dimanfaatkan untuk menyampaikan opini mengenai isu politik, sosial, dan ekonomi. Perkembangan ini menunjukkan bahwa media sosial berperan penting dalam membentuk opini publik yang berkembang di masyarakat [2].

Makan bergizi berperan penting bagi tumbuh kembang manusia, terutama generasi muda. Pemerintah dan berbagai institusi melaksanakan program makan bergizi gratis untuk meningkatkan kualitas kesehatan masyarakat [3]. Program ini memunculkan perdebatan di ruang publik. Media sosial, khususnya X, menjadi sarana utama penyampaian kritik dan dukungan. Isu utama yang dipersoalkan adalah besarnya anggaran program makan bergizi gratis pada skala nasional. Publik juga memperdebatkan efektivitas distribusi, sasaran penerima, dan potensi penyalahgunaan dana. Kondisi ini memperlihatkan adanya perbedaan persepsi publik terhadap manfaat dan tantangan implementasi program [4].

Analisis sentimen merupakan cabang dari *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini dalam teks [5]. Analisis sentimen bertujuan mengelompokkan opini berbentuk teks. Data diperoleh melalui *crawling* komentar media sosial. Data diklasifikasikan ke dalam sentimen positif atau negatif. *Wordcloud* digunakan untuk menampilkan kata-kata yang dominan muncul pada masing-masing kelas sentimen sebagai visualisasi pendukung analisis [6]. Analisis sentimen digunakan untuk mengidentifikasi kecenderungan opini publik. Penelitian ini bertujuan mengukur distribusi sentimen positif dan negatif masyarakat terhadap Program Makan Bergizi Gratis.

Sebelum proses klasifikasi dilakukan, teks harus melalui tahap *preprocessing* agar lebih terstruktur dan siap dianalisis. *Text preprocessing* merupakan proses pembersihan teks agar data siap dieksekusi. Proses ini penting karena kondisi teks berpengaruh langsung terhadap hasil akurasi [7][8]. Tahapan *preprocessing* mencakup *cleaning*, *case folding*, *tokenization*, *normalisasi*, *stopword removal*, dan *stemming*. Dengan *preprocessing* yang tepat, data teks menjadi lebih terstruktur dan siap digunakan untuk ekstraksi fitur maupun proses klasifikasi pada analisis sentimen [7].

Setelah tahap *preprocessing*, teks direpresentasikan dalam bentuk numerik agar dapat diproses oleh algoritma *machine learning*. Metode *Term Frequency–Inverse Document Frequency* (TF-IDF) digunakan untuk memberikan bobot pada kata berdasarkan frekuensi kemunculannya dalam dokumen dan tingkat keunikannya dalam keseluruhan kumpulan dokumen. Pembobotan TF-IDF menghasilkan fitur numerik yang relevan untuk klasifikasi sentimen, dengan kontribusi fitur yang berbeda terhadap kelas sentimen [9]. Seleksi fitur dengan metode *Chi-Square* dilakukan untuk mengukur keterkaitan antara fitur dan kelas sentimen, sehingga hanya fitur dengan kontribusi signifikan yang dipertahankan dalam proses klasifikasi [10].

Data sentimen dari media sosial menunjukkan ketidakseimbangan antara sentimen positif dan negatif di kelas. Karena model klasifikasi cenderung lebih berpihak pada kelas mayoritas, ketidakseimbangan ini dapat memengaruhi proses pembelajaran model klasifikasi. Metode *Synthetic Minority Over-sampling Technique* (SMOTE) digunakan untuk menyeimbangkan data dengan menghasilkan sampel sintetis pada kelas minoritas berdasarkan kedekatan antar data. Distribusi kelas yang lebih seimbang memungkinkan proses pelatihan model dilakukan secara lebih proporsional [11].

Analisis sentimen memerlukan algoritma klasifikasi yang mampu mengolah data teks secara efektif dan mengelompokkannya ke dalam kategori sentimen yang relevan. *Random Forest* digunakan sebagai algoritma klasifikasi dalam penelitian ini dan termasuk ke dalam metode *ensemble learning*. Pendekatan ini mengombinasikan beberapa model agar kinerja meningkat dan akurasi prediksi lebih baik daripada penerapan model tunggal [12]. *Random Forest* berkaitan erat

dengan pohon keputusan dan bekerja dengan menggabungkan suara mayoritas dari banyak pohon. Algoritma ini mampu mengurangi masalah *overfitting* dengan menekan variasi model dan meningkatkan generalisasi [13].

Beberapa penelitian sebelumnya menunjukkan bahwa algoritma *Random Forest* mampu memberikan performa tinggi dalam klasifikasi sentimen. Hadinata dan Purbolaksono membuktikan efektivitas *Random Forest* dalam menganalisis sentimen dengan akurasi di atas 90% [14]. Namun, penelitian tersebut belum mengombinasikan *Random Forest* dengan seleksi fitur *Chi-Square* dan teknik penyeimbangan data SMOTE. Kombinasi tersebut berpotensi meningkatkan akurasi model sekaligus mengurangi bias akibat distribusi data yang tidak seimbang. Berdasarkan kondisi tersebut, penelitian ini difokuskan pada pengembangan model klasifikasi sentimen dengan optimalisasi tahapan *praproses* dan pemodelan. Optimalisasi dilakukan melalui pembobotan fitur menggunakan TF-IDF, seleksi fitur *Chi-Square*, serta penyeimbangan data menggunakan SMOTE.

Kebaruan penelitian ini terletak pada rancangan eksperimen yang mengevaluasi secara sistematis pengaruh seleksi fitur *Chi-Square* dengan rasio bertingkat serta penerapan SMOTE terhadap kinerja *Random Forest* dalam analisis sentimen kebijakan publik. Pendekatan ini diterapkan pada konteks Program Makan Bergizi Gratis menggunakan data media sosial X berbahasa Indonesia, sehingga menghasilkan analisis sentimen yang lebih seimbang dan terukur berdasarkan berbagai metrik evaluasi klasifikasi.

1.2 Rumusan Masalah

Adapun rumusan penelitian yakni:

1.2.1 Permasalahan

Program pemerintah makan bergizi secara gratis memicu beragam reaksi masyarakat di media sosial, khususnya X. Sebagian masyarakat memberikan dukungan, sementara lainnya menyampaikan kritik. Opini tersebut tersebar acak dan tidak terstruktur sehingga menyulitkan dalam memahami persepsi publik secara menyeluruh.

1.2.2 Metode Usulan

Penelitian ini menggunakan metode *Random Forest* berbasis TF-IDF dengan penerapan SMOTE untuk penyeimbangan kelas serta seleksi fitur *Chi-Square* guna meningkatkan akurasi klasifikasi opini publik ke dalam sentimen positif dan negatif.

1.2.3 Pertanyaan Penelitian

1. Berapa perbandingan performa model *Random Forest* setelah penerapan SMOTE dibandingkan tanpa SMOTE?
2. Berapa persentase rasio seleksi fitur *Chi-Square* yang menghasilkan fitur paling relevan untuk klasifikasi sentimen?
3. Berapa konfigurasi *hyperparameter Random Forest* yang paling optimal setelah dilakukan proses *hyperparameter tuning* menggunakan himpunan fitur terbaik hasil seleksi *Chi-Square*?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

1. Mendapatkan hasil perbandingan performa model *Random Forest* dengan dan tanpa penerapan SMOTE untuk mengetahui besaran peningkatan atau penurunan kinerja model.
2. Mendapatkan hasil seleksi fitur *Chi-Square* pada beberapa rasio pemilihan fitur sehingga diperoleh persentase rasio yang menghasilkan fitur paling relevan untuk klasifikasi sentimen.
3. Mendapatkan konfigurasi *hyperparameter Random Forest* yang optimal melalui proses *hyperparameter tuning* berdasarkan himpunan fitur terbaik yang diperoleh dari hasil seleksi *Chi-Square*.

1.3.2 Manfaat Penelitian

1. Menghasilkan analisis sentimen publik terhadap program makan bergizi gratis sebagai bahan informasi mengenai opini masyarakat di media sosial.
2. Mengetahui pengaruh penerapan kombinasi seleksi fitur *Chi-Square* untuk memaksimalkan kinerja *Random Forest* terhadap program makan bergizi gratis melalui media sosial X.

1.4 Batasan Masalah

Sehubungan dengan rumusan masalah yang telah disampaikan, penelitian ini dibatasi pada hal-hal berikut:

1. Data yang digunakan adalah data *tweets* mengenai program makan bergizi gratis mulai dari 01 Juni 2025 sampai dengan 30 Agustus 2025.
2. Data yang digunakan adalah data tweet yang berbahasa Indonesia.
3. Komentar dianalisis menggunakan pembobotan TF-IDF, seleksi fitur *Chi-Square*, penyeimbangan data SMOTE, dan klasifikasi *Random Forest*.
4. Data diperoleh melalui *crawling* X.
5. Kata kunci yang digunakan yaitu “makan bergizi gratis” dan “makan siang gratis”.
6. Analisis sentimen terbatas pada dua kategori positif dan negatif.
7. Bahasa pemrograman yang digunakan untuk implementasi metode yaitu *python*.

1.5 Sistematika Penulisan

Berikut merupakan beberapa aturan dalam penulisan yang digunakan dalam penyusunan skripsi. Adapun sistematika skripsi yang digunakan terdiri dari beberapa bagian, yaitu:

BAB I PENDAHULUAN

Bab ini memaparkan konteks awal penelitian yang mencakup latar belakang permasalahan, perumusan masalah, tujuan serta manfaat penelitian, batasan ruang lingkup, dan sistematika penulisan skripsi.

BAB II TINJAUAN PUSTAKA

Bab ini menguraikan landasan teoretis dan kajian pustaka yang relevan sebagai dasar konseptual dalam mendukung penelitian yang dilakukan.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan rancangan dan alur penelitian, tahapan *prapemrosesan* data, prosedur penerapan metode yang digunakan, serta skenario pengujian yang dirancang untuk mengevaluasi kinerja model.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini menyajikan hasil eksperimen dan pembahasan yang diperoleh dalam penelitian, mencakup proses pengolahan data, hasil klasifikasi menggunakan

algoritma *Random Forest* dengan seleksi fitur *Chi-Square*, evaluasi berdasarkan skenario pengujian, serta implementasi sistem dalam bentuk antarmuka berbasis *Streamlit*.

BAB V KESIMPULAN DAN SARAN

Bab terakhir berisi kesimpulan dari hasil penelitian berdasarkan tujuan yang telah ditetapkan, serta saran untuk pengembangan penelitian di masa yang akan datang.