

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi (TIK) di era digital telah merevolusi pola komunikasi publik, menjadikan media sosial X sebagai arena utama penyampaian opini terhadap kebijakan sosial dan pemerintahan di Indonesia. Media sosial X atau yang dahulu dikenal sebagai Twitter merupakan media sosial yang bersifat terbuka (platform publik), Alasan pengambilan data dari X dikarenakan menyediakan opini masyarakat dapat diakses secara langsung tanpa harus berteman atau bergabung dalam grup tertentu, berbeda dengan platform seperti Facebook atau Instagram yang cenderung bersifat privat. Media sosial X berfungsi sebagai ekosistem interaktif tempat masyarakat mengekspresikan dukungan, kritik, dan analisis terhadap kebijakan sosial seperti Program Makan Bergizi Gratis. Program Makan Bergizi Gratis merupakan kebijakan sosial nasional yang bertujuan meningkatkan kualitas gizi peserta didik dan menekan ketimpangan akses pangan, namun pelaksanaannya memicu beragam reaksi publik yang terekam di media sosial X. Program ini dirancang untuk meningkatkan gizi anak-anak sekolah dan mengurangi kesenjangan akses pangan, sehingga menjadi perhatian luas di masyarakat. Tingginya volume interaksi publik di media sosial X menimbulkan urgensi analisis sentimen publik berbasis data guna memahami arah persepsi masyarakat terhadap kebijakan sosial secara ilmiah dan terukur. Analisis manual terhadap opini publik dalam jumlah besar tidaklah efisien, sehingga dibutuhkan metode otomatis yang mampu mengklasifikasikan sentimen dengan

tingkat akurasi yang baik. Hal inilah yang menjadikan *sentiment analysis* berbasis *machine learning* menjadi pendekatan yang relevan untuk penelitian ini.

Identifikasi terhadap masalah utama terletak pada kemampuan sistem untuk menafsirkan opini publik secara tepat di tengah keragaman bahasa yang digunakan di media sosial. Cuitan pada platform X umumnya menggunakan bahasa campuran, singkatan, hingga ekspresi emosional seperti sarkasme dan ironi, sehingga menimbulkan ambiguitas makna yang menantang bagi sistem klasifikasi sentimen. Kondisi linguistik ini menimbulkan tantangan bagi model klasifikasi karena konteks emosional tidak selalu merepresentasikan struktur sintaksis yang terbaca secara literal oleh algoritma. Selain itu, data media sosial umumnya tidak terstruktur, memiliki distribusi kelas yang tidak seimbang antara sentimen positif, negatif, dan netral, serta rentan terhadap *noise* akibat keberadaan *spam* atau konten berulang. Permasalahan lain yang muncul adalah menentukan algoritma paling efektif dalam memproses teks pendek dengan struktur bahasa kompleks seperti cuitan di media sosial. Algoritma yang berbeda dapat menghasilkan performa yang bervariasi tergantung pada representasi fitur dan kualitas data. Oleh karena itu, penting untuk mengidentifikasi model yang paling efektif dan efisien dalam melakukan klasifikasi sentimen terhadap topik kebijakan tertentu seperti Program Makan Bergizi Gratis.

Untuk menjawab tantangan tersebut, penelitian ini membandingkan dua algoritma machine learning populer dalam klasifikasi teks, yaitu *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB), yang memiliki pendekatan berbeda dalam membangun model prediksi. Kedua algoritma ini memiliki pendekatan yang

berbeda secara fundamental. *Support Vector Machine* (SVM) memisahkan data berdasarkan margin optimal pada ruang vektor berdimensi tinggi, sedangkan *Naïve Bayes* (NB) memprediksi probabilitas kelas dengan asumsi independensi antarfitur. Perbandingan keduanya penting karena meskipun SVM dikenal unggul dalam akurasi dan generalisasi, *Naïve Bayes* sering kali lebih ringan secara komputasi dan mampu bekerja baik pada teks pendek. Penelitian ini memberikan solusi melalui penerapan kedua metode tersebut pada dataset opini publik tentang Program Makan Bergizi Gratis yang diperoleh dari media sosial X. Hasil dari proses klasifikasi ini diharapkan dapat menunjukkan algoritma mana yang lebih efektif dalam menangkap kecenderungan sentimen publik dengan tingkat kesalahan minimal, sekaligus menjadi referensi empiris untuk penelitian lanjutan di bidang analisis opini kebijakan publik.

Keterbaruan penelitian ini terletak pada fokus analisis sentimen terhadap kebijakan sosial pemerintah yang jarang dikaji secara empiris dengan pendekatan perbandingan algoritma berbasis machine learning. Sebagian besar penelitian terdahulu mengenai klasifikasi sentimen di Indonesia masih berfokus pada bidang komersial, seperti ulasan produk, layanan pelanggan, atau promosi merek. Sementara itu, penelitian yang menyoroiti kebijakan sosial pemerintah, khususnya Program Makan Bergizi Gratis, masih sangat terbatas. Penelitian ini menghadirkan nilai tambah dengan mengarahkan *machine learning* pada konteks kebijakan publik, sehingga tidak hanya menilai performa algoritma, tetapi juga memberikan gambaran empiris mengenai persepsi masyarakat terhadap program pemerintah yang berdampak luas. Keterbaruan lain muncul pada pemanfaatan data media sosial

X yang dinamis dan aktual, serta penerapan prapemrosesan teks khusus bahasa Indonesia untuk mengurangi *noise* dan meningkatkan akurasi model. Penelitian ini memberikan kontribusi signifikan dalam memperluas kajian computational social science melalui penerapan analisis sentimen pada kebijakan publik berbasis data media sosial Indonesia.

Keseluruhan penelitian ini diharapkan dapat memberikan pemahaman yang lebih komprehensif mengenai efektivitas algoritma *Support Vector Machine* dan *Naïve Bayes* dalam mengklasifikasikan sentimen publik pada isu kebijakan sosial. Hasil perbandingan kinerja antara SVM dan NB diharapkan tidak hanya menunjukkan model dengan akurasi tertinggi, tetapi juga menjadi dasar metodologis bagi lembaga pemerintah dalam mengembangkan sistem pemantauan opini publik secara berkelanjutan dan berbasis data. Dengan pendekatan ilmiah dan berbasis data, analisis sentimen dapat membantu pembuat kebijakan memahami persepsi masyarakat secara objektif, sehingga kebijakan yang dihasilkan dapat lebih responsif terhadap aspirasi publik. Oleh karena itu, penelitian ini memiliki relevansi akademik dan praktis sekaligus, dengan tetap berfokus pada integritas ilmiah serta konsistensi dalam ruang lingkup kajian yang ditetapkan.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana efektivitas algoritma Support Vector Machine dan Naïve Bayes dalam mengklasifikasikan sentimen publik terhadap Program Makan Bergizi Gratis di media sosial X?

2. Bagaimana tahapan text preprocessing dan metode representasi TF-IDF mempengaruhi kinerja klasifikasi kedua algoritma tersebut?
3. Model algoritma manakah yang paling konsisten dan efisien untuk digunakan dalam analisis sentimen kebijakan sosial berbasis data media sosial?

1.3 Batasan Masalah

Berdasarkan rumusan masalah diatas, ada beberapa batasan masalah ini adalah sebagai berikut:

1. Penelitian ini dibatasi pada analisis sentimen publik terhadap Program Makan Bergizi Gratis yang diperoleh dari cuitan pengguna *media sosial X*, tanpa melibatkan platform media sosial lainnya atau sumber berita daring.
2. Proses klasifikasi sentimen hanya dilakukan terhadap tiga kategori utama, yaitu sentimen positif, negatif, dan netral, dengan data teks yang telah melalui tahapan prapemrosesan menggunakan metode *case folding*, *tokenizing*, *stopword removal*, dan *stemming*.
3. Analisis performa algoritma difokuskan pada perbandingan antara *Support Vector Machine* dan *Naïve Bayes*, dengan representasi fitur menggunakan pembobotan *Term Frequency–Inverse Document Frequency (TF-IDF)* serta evaluasi berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

1.4 Tujuan dan Manfaat Penelitian

Penelitian ini difokuskan pada batasan-batasan yang telah ditentukan berdasarkan perumusan masalah sebelumnya. Adapun tujuan dan manfaat penelitian adalah sebagai berikut:

1.4.1 Tujuan Penelitian

1. Mengidentifikasi bentuk dan kecenderungan sentimen publik (positif, negatif, netral) terhadap Program Makan Bergizi Gratis pada *media sosial X* berdasarkan data teks yang telah melalui prapemrosesan (*case folding, tokenizing, stopword removal, dan stemming*).
2. Membangun dan menerapkan model klasifikasi sentimen berbasis *Support Vector Machine* dan *Naïve Bayes* dengan representasi fitur *Term Frequency–Inverse Document Frequency (TF–IDF)* pada korpus cuitan terkait Program Makan Bergizi Gratis.
3. Membandingkan kinerja kedua algoritma tersebut menggunakan metrik evaluasi *accuracy, precision, recall, dan F1-score* untuk menilai efektivitas klasifikasi dalam ruang lingkup data dan prosedur yang dibatasi pada penelitian ini.

1.4.2 Manfaat Penelitian

1. Penelitian ini dapat memberikan kontribusi nyata terhadap pengembangan ilmu pengetahuan, khususnya pada bidang *machine learning, text mining, dan sentiment analysis*.
2. Memberikan informasi yang komprehensif mengenai kecenderungan sentimen masyarakat terhadap Program Makan Bergizi Gratis. Melalui analisis sentimen terhadap data yang dikumpulkan, penelitian ini mampu menggambarkan persepsi publik apakah cenderung positif, negatif, atau netral terhadap program tersebut.

3. Penelitian dapat menambah literatur ilmiah pada bidang pengolahan teks berbahasa Indonesia, khususnya dalam konteks analisis sentimen dan penerapan algoritma klasifikasi.

1.5 Tinjauan Umum Objek Penelitian

Program Makan Bergizi Gratis merupakan inisiatif kebijakan sosial yang ditujukan untuk meningkatkan kualitas gizi peserta didik dan mengurangi kesenjangan akses pangan. Program ini menempatkan pemenuhan gizi sebagai prasyarat penting bagi tumbuh kembang anak dan peningkatan capaian belajar, sehingga memunculkan atensi publik yang luas. Respons masyarakat terhimpun secara intens di *media sosial X* yang memiliki karakter percakapan cepat, terbuka, dan terdokumentasi. Cuitan yang membahas program ini hadir dalam ragam bentuk, mulai dari dukungan terhadap manfaat program hingga kritik atas pelaksanaan teknis di lapangan. Variasi gaya bahasa, penggunaan istilah nonbaku, dan kemunculan ekspresi emosional seperti ironi atau sarkasme menjadi ciri yang mewarnai diskursus publik di platform tersebut.

Objek penelitian pada studi ini adalah himpunan cuitan terkait Program Makan Bergizi Gratis yang diambil dari *media sosial X* dan dianalisis melalui pendekatan *sentiment analysis*. Teks diproses melalui tahap prapemrosesan untuk memperoleh representasi fitur *Term Frequency Inverse Document Frequency (TF-IDF)* sebagai dasar klasifikasi. Perbandingan kinerja algoritma difokuskan pada *Support Vector Machine* dan *Naïve Bayes* mengingat keduanya memiliki landasan yang berbeda, yakni pemisahan margin maksimum dan pendekatan probabilistik. Penilaian kinerja dilakukan terhadap tiga kategori sentimen positif, negatif, dan

netral dengan tolok ukur *accuracy*, *precision*, *recall*, dan *F1-score*. Tinjauan ini menegaskan batas objek pada isi cuitan yang relevan dengan Program Makan Bergizi Gratis di *media sosial X*, tanpa memperluas ke platform lain atau variabel nonteks, agar hasil perbandingan algoritma bersandar pada ruang lingkup yang terdefinisi dengan jelas dan konsisten.

1.6 Sistematika Penulisan

Naskah disusun untuk menjaga alur pemikiran yang utuh dari pengenalan masalah hingga penarikan simpulan, dengan fokus tunggal pada perbandingan algoritma *Support Vector Machine* dan *Naïve Bayes* dalam klasifikasi sentimen publik terhadap Program Makan Bergizi Gratis di *media sosial X*. Sistematika penulisan dirancang agar setiap bab saling terkait, konsisten, dan tidak meluas ke luar ruang lingkup penelitian.

BAB I: PENDAHULUAN

Bab ini menyajikan gambaran umum penelitian yang meliputi latar belakang masalah, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, tinjauan umum objek penelitian, serta sistematika penulisan. Fokus bab mengukuhkan konteks dan batas penelitian pada analisis sentimen di *media sosial X* dengan perbandingan algoritma *Support Vector Machine* dan *Naïve Bayes*.

BAB II: LANDASAN TEORI

Bab ini menguraikan konsep teoretis yang relevan dan diperlukan untuk menjawab perumusan masalah, meliputi pengertian dan ruang lingkup *sentiment analysis*, karakteristik teks pendek di *media sosial X*,

prapemrosesan teks bahasa Indonesia, representasi fitur *Term Frequency–Inverse Document Frequency (TF–IDF)*, prinsip dasar *Support Vector Machine* dan *Naïve Bayes*, serta metrik evaluasi klasifikasi mencakup *accuracy*, *precision*, *recall*, dan *F1-score*. Uraian teori dibatasi pada konsep yang langsung mendukung perbandingan kinerja kedua algoritma.

BAB III: METODOLOGI PENELITIAN

Bab ini menjelaskan rancangan penelitian secara operasional, mencakup sumber dan kriteria data cuitan, tahapan prapemrosesan (*case folding*, *tokenizing*, *stopword removal*, *stemming*), pembentukan representasi *TF–IDF*, perancangan dan implementasi model *Support Vector Machine* dan *Naïve Bayes*, serta prosedur evaluasi dengan *k-fold cross-validation* dan pengukuran *accuracy*, *precision*, *recall*, dan *F1-score*. Batasan penelitian ditegaskan pada platform, kategori sentimen, dan ruang lingkup pengujian.

BAB IV: IMPLEMENTASI DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi dan analisis perbandingan kinerja model berdasarkan metrik evaluasi yang telah ditetapkan. Penyajian mencakup deskripsi dataset hasil prapemrosesan, keluaran klasifikasi pada tiga kategori sentimen, perbandingan hasil *Support Vector Machine* dan *Naïve Bayes*, serta pembahasan yang menautkan temuan dengan tujuan penelitian. Penekanan diarahkan pada konsistensi interpretasi dalam ruang lingkup kebijakan Program Makan Bergizi Gratis.

BAB V: KESIMPULAN DAN SARAN

Bab ini merangkum temuan utama terkait kinerja relatif kedua algoritma pada data sentimen Program Makan Bergizi Gratis di *media sosial X*, menyatakan keterbatasan yang melekat pada penelitian, serta memberikan saran teknis yang selaras dengan batasan ruang lingkup studi untuk pengembangan kerja selanjutnya. Rumusan kesimpulan dan saran dijaga ringkas, terarah, dan konsisten dengan hasil yang diperoleh di bab sebelumnya.