

BAB I PENDAHULUAN

1.1 Latar Belakang

Program Makanan Bergizi Gratis (MBG) yang diluncurkan oleh Prabowo Subianto telah menjadi salah satu kebijakan publik yang paling banyak dibicarakan dan diperdebatkan di ruang digital, terutama di media sosial seperti Twitter. Volume percakapan publik yang tinggi menunjukkan bahwa masyarakat sangat peduli terhadap keberlanjutan, efektivitas, dan dampak program ini. Seiring dengan meningkatnya aliran informasi, opini publik dapat mempengaruhi persepsi publik, keputusan politik, dan arah kebijakan pemerintah. Namun, jumlah komentar dan opini yang besar membuat analisis manual menjadi tidak mungkin dilakukan dengan cepat dan akurat [1].

Masalah utama yang muncul adalah bagaimana memproses secara efisien jumlah besar opini publik sehingga informasi yang diperoleh dapat digunakan sebagai dasar untuk pembuat kebijakan atau evaluasi program. Dalam konteks ini, analisis sentimen menjadi sangat penting, karena dapat mengidentifikasi apakah publik merespons secara positif, negatif, atau netral terhadap Program MBG. Analisis sentimen tidak hanya menggambarkan tren opini publik tetapi juga memberikan wawasan yang dapat digunakan oleh pemangku kepentingan dalam mengembangkan strategi komunikasi atau mengevaluasi implementasi program.

Tantangan teknis muncul saat menentukan metode klasifikasi yang paling sesuai untuk analisis sentimen. *Support Vector Machine* (SVM) telah lama digunakan sebagai metode pembelajaran mesin klasik yang kuat dan berkinerja baik pada teks, namun kinerjanya sangat bergantung pada pemilihan parameter yang tepat. Para peneliti menyadari bahwa tantangan teknis terletak pada pemilihan metode klasifikasi. Pendekatan pembelajaran mesin klasik seperti *Support Vector Machine* (SVM) telah terbukti andal, tetapi kinerjanya sangat bergantung pada proses optimasi [2].

Di sisi lain, perkembangan teknologi Pemrosesan Bahasa Alami (*Natural*

Language Processing/NLP) telah menghasilkan pendekatan berbasis Transformer baru seperti BERT dan variannya. IndoBERT adalah model bahasa berbasis Transformer yang telah dilatih secara khusus menggunakan korpus bahasa Indonesia, sehingga memiliki kemampuan yang lebih baik dalam memahami konteks kalimat, struktur sintaksis, dan makna yang tersirat. Berbagai studi menunjukkan bahwa model berbasis Transformer outperfrom metode pembelajaran mesin tradisional dalam banyak tugas NLP, termasuk klasifikasi sentimen. Meskipun IndoBERT dan SVM memiliki perbedaan mendasar dalam arsitektur dan mekanisme pembelajaran, perbandingan antara keduanya tetap relevan dalam studi ini. SVM dengan representasi *TF-IDF* mewakili pendekatan pembelajaran mesin berbasis fitur klasik yang sering digunakan sebagai baseline dalam analisis sentimen, sementara IndoBERT mewakili pendekatan pembelajaran mendalam modern yang didasarkan pada model bahasa yang telah dilatih sebelumnya dan mampu membangun representasi kata kontekstual. Perbandingan ini bertujuan untuk mengevaluasi dampak pendekatan representasi bahasa yang berbeda terhadap kinerja analisis sentimen pada teks opini publik berbahasa Indonesia, khususnya dalam konteks kebijakan publik.

Berdasarkan studi literatur, perbandingan antara IndoBERT dan SVM merupakan topik yang relevan [3]. Namun, sebagian besar studi hanya menggunakan SVM standar tanpa optimasi mendalam. Selain itu, belum ada studi yang secara khusus membahas sentimen publik terhadap Program MBG, sehingga tidak ada hasil perbandingan yang komprehensif dan relevan terkait isu-isu terkini. Oleh karena itu, terdapat kesenjangan penelitian dalam bentuk kurangnya analisis perbandingan kinerja IndoBERT yang disesuaikan dengan SVM yang dioptimalkan menggunakan *Grid Search* dalam konteks analisis sentimen Program MBG.

Penelitian ini memiliki urgensi tinggi, karena hasil klasifikasi sentimen yang akurat dapat memberikan gambaran yang valid tentang persepsi publik terhadap program tersebut. Informasi ini berguna bagi pemerintah, analis kebijakan, akademisi, dan pihak lain yang tertarik pada proses evaluasi dan

pengembangan program. Melalui penelitian ini, diharapkan model yang memberikan kinerja terbaik, paling stabil, dan paling sesuai untuk diterapkan dalam analisis sentimen skala besar dapat diidentifikasi. Selain itu, penelitian ini juga memberikan kontribusi ilmiah terhadap pengembangan teknik NLP Indonesia, khususnya dalam bidang analisis sentimen kebijakan publik.

1.2 Rumusan Masalah

Berdasarkan deskripsi latar belakang, terdapat kesenjangan penelitian berupa ketidakhadiran studi yang secara khusus membandingkan kinerja IndoBERT yang telah dioptimalkan *fine-tuning* dengan SVM yang dioptimalkan menggunakan *Grid Search* dalam analisis sentimen Program Makanan Bergizi Gratis (MBG). Selain itu, tidak ada bukti faktual mengenai model mana yang memberikan hasil paling akurat dan stabil dalam memproses opini publik berbahasa Indonesia terkait isu ini.

Berdasarkan celah penelitian yang telah diidentifikasi, pertanyaan penelitian dalam studi ini dirumuskan sebagai berikut. Pertama, Bagaimana proses pengolahan dan prapemrosesan (Preprocessing) *Dataset* dilakukan sebelum digunakan untuk pemodelan dan optimasi IndoBERT dan SVM. Kedua, bagaimana kinerja model IndoBERT dan SVM yang dioptimalkan dalam mengklasifikasikan sentimen publik ke dalam kategori negatif, netral, dan positif. Ketiga, bagaimana perbandingan kinerja model IndoBERT dan SVM berdasarkan hasil evaluasi, seperti Akurasi, Presisi, Recall, F1-Score, dan matriks kebingungan (*Confusion*).

1.3 Batasan Masalah

Studi ini memiliki beberapa keterbatasan yang ditetapkan untuk mempertahankan fokus analisis dan memastikan cakupan studi tetap terfokus serta dapat diselesaikan secara sistematis. Keterbatasan-keterbatasan tersebut adalah sebagai berikut:

1. Data Penelitian: Data yang digunakan dalam studi ini hanya terdiri dari teks opini publik yang diperoleh melalui proses crawling dari platform X

(Twitter) selama periode tertentu menggunakan kata kunci yang terkait dengan Program Makanan Bergizi Gratis Prabowo (MBG). Data di luar platform X, seperti Facebook, Instagram, atau portal berita, tidak termasuk dalam penelitian ini.

2. Analisis Sentimen: Penelitian ini hanya mengklasifikasikan sentimen ke dalam tiga kategori, yaitu Sentimen Positif, Sentimen Negatif, dan Sentimen Netral. Analisis sentimen lanjutan seperti analisis sentimen berbasis aspek, klasifikasi emosi, atau analisis multi-Label tidak dilakukan.
3. Model yang Dibandingkan: Model yang digunakan dan dianalisis hanya terdiri dari *Support Vector Machine* (SVM) dengan optimasi *Grid Search* dan IndoBERT dengan metode *fine-tuning*. Model lain seperti Naïve Bayes, Random Forest, LSTM, RoBERTa, atau Large Language Model (LLM) tidak dibahas dalam studi ini.
4. Teknik Representasi Teks: Representasi teks yang digunakan hanya TF-IDF untuk model SVM dan Transformer Embedding dari IndoBERT untuk *fine-tuning*. Representasi lain seperti Word2Vec, FastText, atau GloVe tidak digunakan.
5. Aspek Prapemrosesan: Proses prapemrosesan berfokus pada case folding, pembersihan data, dan format tokenisasi dari IndoBERT. Teknik seperti penghapusan kata stop dan Stemming tidak diterapkan secara manual karena IndoBERT sudah menggunakan tokenisasi subkata.
6. Evaluasi Kinerja: Evaluasi model hanya menggunakan metrik Akurasi, Presisi, Recall, dan F1-Score. Visualisasi tambahan seperti ROC-AUC tidak digunakan karena penelitian ini merupakan klasifikasi multi-kelas.

1.4 Tujuan Penelitian

Berdasarkan formulasi masalah yang telah ditetapkan, penelitian ini memiliki tujuan sebagai berikut:

1. Menjelaskan dan menerapkan tahap pemrosesan data (prapemrosesan)

untuk membentuk *Dataset* sentimen Program Makanan Bergizi Gratis (MBG) yang siap digunakan dalam pemodelan. Proses ini meliputi pembersihan teks, tokenisasi, dan penandaan sentimen.

2. Membangun dan menganalisis kinerja model *Support Vector Machine* (SVM) yang telah dioptimalkan dengan Grid Search, serta model IndoBERT yang telah dioptimalkan dengan *fine-tuning*, dalam mengklasifikasikan sentimen terkait Program MBG.
3. Membandingkan kinerja kedua model berdasarkan metrik evaluasi Akurasi, Presisi, Recall, dan F1-Score untuk mengidentifikasi model yang memberikan hasil optimal dalam analisis sentimen Program Makanan Bergizi Gratis (MBG).

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat baik secara teoritis maupun praktis, sebagai berikut:

1.5.1 Manfaat Teoritis

Memberikan kontribusi ilmiah terhadap pengembangan studi analisis sentimen, khususnya dalam membandingkan kinerja model deep learning IndoBERT dan Machine Learning klasik SVM dalam konteks isu kebijakan publik di Indonesia. Menambah referensi akademik untuk penelitian terkait pemrosesan bahasa alami NLP berbasis bahasa Indonesia, terutama di bidang media sosial. Memperkuat literatur mengenai efektivitas model Transformer dibandingkan algoritma tradisional dalam tugas klasifikasi sentimen, sehingga memberikan dasar untuk penelitian lebih lanjut.

1.5.2 Manfaat Praktis

Bagi lembaga pemerintah atau pembuat kebijakan, hasil penelitian ini dapat digunakan sebagai acuan untuk memahami tren opini publik terkait Program Makanan Bergizi Gratis (MBG) secara lebih cepat, akurat, dan terukur.

Bagi institusi/organisasi atau departemen, penelitian ini dapat menjadi

contoh penerapan metode NLP modern dalam menyelesaikan masalah nyata, serta menambah koleksi penelitian berbasis data dan kecerdasan buatan.

Bagi pengembang sistem analitik media sosial, penelitian ini memberikan gambaran model mana yang lebih optimal untuk memproses teks berbahasa Indonesia, sehingga dapat diterapkan dalam sistem pemantauan opini publik.

Bagi peneliti di masa depan, studi ini dapat dijadikan acuan untuk pengembangan lebih lanjut, seperti menambahkan *Dataset*, mengeksplorasi model Transformer lainnya, atau menguji metode optimasi yang berbeda dalam algoritma pembelajaran mesin.

1.6 Sistematika Penulisan

- Bab I : Pendahuluan berisi deskripsi latar belakang yang menjelaskan pentingnya analisis sentimen Program Makanan Bergizi Gratis (MBG) sebagai salah satu isu publik yang banyak dibahas di media sosial. Bab ini juga menjelaskan formulasi masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistem penulisan tesis. Bab ini memberikan gambaran awal agar pembaca dapat memahami arah penelitian dan konteks isu yang diteliti.
- Bab II : Tinjauan Pustaka berisi teori dan konsep yang menjadi dasar penelitian. Bab ini membahas Pemrosesan Bahasa Alami (*Natural Language Processing/NLP*), analisis sentimen, metode klasifikasi, konsep dan karakteristik *Support Vector Machine* (SVM) beserta teknik optimasi *Grid Search*, serta penjelasan tentang model *IndoBERT* dan arsitektur *Transformer*. Selain itu, bab ini juga meninjau penelitian sebelumnya yang relevan untuk memperkuat dasar teoritis penelitian dan menunjukkan posisi penelitian ini di antara studi-studi sebelumnya.
- Bab III : Metodologi penelitian menjelaskan langkah-langkah dan prosedur penelitian secara sistematis. Bab ini menguraikan desain penelitian, teknik pengumpulan data melalui *scraping* media sosial,

pembersihan dan pemrosesan data teks, pembentukan *Dataset* sentimen, tahap penerapan model SVM yang dioptimalkan dengan Grid Search, dan tahap penyempurnaan model IndoBERT. Bab ini juga menjelaskan metrik evaluasi yang digunakan, seperti Akurasi, Presisi, Recall, dan F1-Score, serta perangkat dan alat yang digunakan. Bab ini berfungsi sebagai panduan rinci untuk melakukan eksperimen dan analisis.

Bab IV : Hasil dan Pembahasan memaparkan hasil penelitian yang diperoleh dari pengolahan data dan pengujian model. Bab ini memaparkan hasil pra-pengolahan data, karakteristik *Dataset*, kinerja model SVM yang dioptimalkan, kinerja model IndoBERT yang disesuaikan, serta perbandingan kedua model berdasarkan metrik evaluasi yang telah ditentukan. Pembahasan dilakukan secara analitis untuk menjelaskan alasan perbedaan kinerja, interpretasi hasil evaluasi, dan relevansinya terhadap sentimen publik terhadap Program MBG. Bab ini merupakan inti dari seluruh penelitian.

Bab V : Kesimpulan dan Rekomendasi berisi kesimpulan yang ditarik berdasarkan hasil penelitian dan secara langsung menjawab pertanyaan dan tujuan penelitian yang telah ditetapkan. Selain itu, bab ini juga berisi rekomendasi untuk penelitian lebih lanjut guna mengembangkan metode, *Dataset*, atau pendekatan analisis sentimen yang lebih baik. Bab ini juga menyimpulkan seluruh seri penelitian.