

BAB I

PENDAHULUAN

1.1.Latar Belakang

Perkembangan teknologi informasi dan komunikasi telah mengubah cara masyarakat memperoleh informasi, berinteraksi, serta menyampaikan pendapat di ruang digital. Berdasarkan hasil Survei Sosial Ekonomi Nasional (Susenas) tahun 2024, sebanyak 72,78% penduduk Indonesia telah mengakses internet, meningkat dibandingkan tahun 2023 yang sebesar 69,21%. Peningkatan penetrasi internet tersebut menunjukkan bahwa aktivitas masyarakat semakin bergeser menuju ruang digital, termasuk dalam menyampaikan opini terhadap berbagai isu sosial, ekonomi, maupun politik. Salah satu media yang paling banyak dimanfaatkan sebagai ruang diskusi publik adalah Platform media sosial *X*, yang memungkinkan pengguna menyampaikan pendapat secara terbuka, cepat, dan *real-time* melalui unggahan singkat (*tweet*)[1].

Sebagai salah satu media sosial berbasis *Social Network Service* (SNS), Platform *X* menghasilkan lebih dari 500 juta *tweet* setiap hari yang mencerminkan berbagai pandangan masyarakat terhadap isu-isu yang sedang berkembang. *Volume* data yang sangat besar tersebut menjadikan media sosial *X* sebagai sumber informasi yang kaya untuk memahami opini publik. Namun, besarnya jumlah data yang dihasilkan juga menimbulkan tantangan tersendiri karena analisis secara manual menjadi tidak lagi memungkinkan[2]. Oleh sebab itu, diperlukan pendekatan berbasis *Natural Language Processing* (NLP) yang mampu mengolah data teks dalam jumlah besar secara otomatis sehingga informasi penting yang terkandung di dalamnya dapat diekstraksi secara efektif [3].

Salah satu isu yang memperoleh perhatian luas masyarakat Indonesia adalah evaluasi terhadap kinerja pemerintahan [4], khususnya pada masa awal kepemimpinan Presiden Prabowo Subianto dan Wakil Presiden Gibran Rakabuming Raka periode 2024–2029. Tahun pertama masa pemerintahan merupakan fase yang sangat penting karena berbagai kebijakan strategis mulai diimplementasikan dan menjadi dasar terbentuknya persepsi masyarakat terhadap

pemerintahan baru. Pada periode ini, berbagai kebijakan, program kerja, maupun peristiwa nasional memicu munculnya diskusi publik yang sangat dinamis di media sosial. Oleh karena itu, memahami bagaimana masyarakat merespons berbagai isu tersebut menjadi penting sebagai salah satu bentuk evaluasi terhadap dinamika opini publik [5].

Dalam praktiknya, pengukuran persepsi masyarakat terhadap pemerintah umumnya masih dilakukan melalui survei opini publik menggunakan kuesioner atau *polling* yang melibatkan jumlah responden terbatas. Meskipun pendekatan tersebut telah lama digunakan, metode survei memiliki beberapa keterbatasan, antara lain hanya merepresentasikan sebagian kecil populasi, bergantung pada jawaban yang bersifat *self-reported*, serta kurang mampu menangkap perubahan opini masyarakat yang berlangsung secara cepat di ruang digital. Di sisi lain, media sosial menyediakan jutaan opini yang muncul secara spontan tanpa intervensi, sehingga berpotensi memberikan gambaran yang lebih luas mengenai persepsi publik. Akan tetapi, karakteristik data media sosial yang tidak terstruktur, berukuran sangat besar, serta menggunakan bahasa informal menyebabkan proses analisis menjadi jauh lebih kompleks dibandingkan data survei konvensional[6].

Untuk mengatasi permasalahan tersebut, berbagai teknik *text mining* telah dikembangkan, salah satunya adalah *topic modeling*, yaitu metode yang digunakan untuk mengidentifikasi kelompok topik atau isu utama yang tersembunyi di dalam kumpulan dokumen teks. Dibandingkan metode konvensional seperti *Latent Dirichlet Allocation* (LDA) yang hanya mengandalkan representasi *bag-of-words*, metode BERTopic memanfaatkan representasi semantik berbasis *transformer* sehingga mampu menghasilkan topik yang lebih koheren, lebih mudah diinterpretasikan, dan lebih sesuai untuk karakteristik teks pendek seperti *tweet* di media sosial. Dengan kemampuan tersebut, BERTopic menjadi salah satu pendekatan yang potensial untuk mengekstraksi isu-isu utama yang berkembang dalam percakapan publik[7].

Meskipun demikian, mengetahui topik yang sedang dibicarakan saja belum cukup untuk memahami persepsi masyarakat secara menyeluruh. Informasi mengenai bagaimana sikap masyarakat terhadap setiap topik juga diperlukan

sehingga analisis sentimen menjadi tahap lanjutan yang penting. Berbagai penelitian telah menunjukkan bahwa model *transformer* seperti IndoBERT, IndoBERTweet, dan RoBERTa mampu memberikan performa yang baik dalam klasifikasi sentimen. Akan tetapi, hasil penelitian juga menunjukkan bahwa tidak terdapat satu model yang secara konsisten memberikan performa terbaik pada seluruh domain data. Perbedaan karakteristik bahasa, distribusi kelas, maupun konteks data menyebabkan setiap model memiliki keunggulan yang berbeda sehingga pemilihan model yang tepat menjadi aspek yang sangat penting dalam membangun sistem analisis sentimen[8].

Berbagai penelitian terdahulu telah menunjukkan bahwa BERTopic merupakan salah satu metode *topic modeling* modern yang mampu menghasilkan topik dengan kualitas lebih baik dibandingkan metode konvensional. Penelitian yang dilakukan oleh Tang dkk. berhasil memanfaatkan BERTopic untuk menganalisis 17.593 surat masyarakat pada Platform pemerintah *People Call Me* di Tiongkok sehingga diperoleh berbagai isu utama yang berkaitan dengan transportasi, jaminan sosial, layanan kesehatan, serta perlindungan konsumen. Penelitian tersebut menunjukkan bahwa BERTopic mampu mengidentifikasi tema-tema penting dari data teks dalam jumlah besar secara efektif [9].

Penelitian lain yang dilakukan oleh Farea dkk. menunjukkan bahwa kualitas *topic modeling* sangat dipengaruhi oleh proses representasi dokumen dan optimasi jumlah topik. Hasil penelitian tersebut mengindikasikan bahwa pendekatan berbasis representasi semantik menghasilkan topik yang lebih koheren dibandingkan pendekatan *Bag-of-Words* konvensional. Temuan serupa juga oleh Kumi dkk. yang memanfaatkan kombinasi LDA, NMF, dan BERTopic untuk menganalisis lebih dari 114.000 komentar *Reddit*. Penelitian tersebut berhasil mengidentifikasi dua belas tema utama yang menjadi perhatian masyarakat serta membuktikan bahwa BERTopic mampu menghasilkan representasi topik yang lebih mudah diinterpretasikan dibandingkan metode klasik [10].

Selain perkembangan pada *topic modeling*, penelitian mengenai analisis sentimen berbasis *transformer* juga berkembang sangat pesat. Berbagai model

seperti IndoBERT, IndoBERTweet, maupun RoBERTa telah banyak digunakan dalam berbagai tugas klasifikasi teks berbahasa Indonesia.

Penelitian yang dilakukan oleh Koto dkk. menunjukkan bahwa IndoBERT mampu mencapai *state-of-the-art performance* pada sebagian besar tugas NLP bahasa Indonesia karena dilatih menggunakan korpus bahasa Indonesia berskala besar yang mencakup berbagai domain teks[3]. Selanjutnya, Koto dkk. mengembangkan IndoBERTweet sebagai model yang secara khusus dilatih menggunakan korpus *Twitter* berbahasa Indonesia sehingga lebih mampu merepresentasikan karakteristik bahasa informal yang umum ditemukan pada media sosial[11]. Di sisi lain, penelitian Nur dkk. menunjukkan bahwa RoBERTa mampu memberikan performa terbaik dibandingkan BERT, DistilBERT, dan XLNet pada tugas klasifikasi emosi, sehingga menunjukkan potensinya untuk diterapkan pada berbagai tugas analisis sentiment [8] .

Meskipun demikian, hasil penelitian terdahulu menunjukkan bahwa belum terdapat satu model *transformer* yang secara konsisten memberikan performa terbaik pada seluruh domain data. Temuan-temuan tersebut mengindikasikan bahwa performa model sangat dipengaruhi oleh karakteristik dataset, distribusi kelas, serta konteks bahasa yang digunakan. Oleh karena itu, pemilihan model terbaik tidak dapat ditentukan hanya berdasarkan hasil penelitian sebelumnya, tetapi perlu dievaluasi secara empiris pada dataset yang digunakan dalam penelitian.

Penelitian mengenai *topic modeling* umumnya hanya berfokus pada proses identifikasi topik tanpa mengaitkan hasilnya dengan kecenderungan sentimen masyarakat terhadap setiap isu yang ditemukan. Di sisi lain, penelitian analisis sentimen lebih banyak diarahkan pada proses klasifikasi polaritas dokumen tanpa mempertimbangkan konteks topik yang melatarbelakangi terbentuknya opini tersebut. Kondisi ini menyebabkan hubungan antara isu yang sedang berkembang dan respons masyarakat terhadap isu tersebut belum dapat dijelaskan secara utuh.

Selain itu, pendekatan analisis sentimen pada penelitian terdahulu umumnya menggunakan satu model *transformer* yang telah ditentukan sejak awal. Padahal, berbagai hasil penelitian menunjukkan bahwa performa IndoBERT,

IndoBERTweet, maupun RoBERTa sangat dipengaruhi oleh karakteristik dataset, distribusi kelas, serta domain data yang dianalisis. Dengan demikian, pemilihan model berdasarkan asumsi atau hasil penelitian pada domain lain belum tentu menghasilkan performa optimal ketika diterapkan pada opini publik mengenai pemerintahan di media sosial *X*.

Keterbatasan lain terlihat pada aspek implementasi sistem. Sebagian besar penelitian masih berhenti pada tahap eksperimen model sehingga menghasilkan keluaran berupa evaluasi performa atau daftar topik tanpa menyediakan suatu sistem yang mampu mengintegrasikan proses pengumpulan data, *preprocessing*, ekstraksi topik, pemilihan model sentimen terbaik, visualisasi hasil, hingga penyajian *insight* dalam satu alur kerja yang terpadu. Akibatnya, hasil penelitian belum sepenuhnya mudah dimanfaatkan sebagai media analisis maupun pendukung pengambilan keputusan.

Di samping itu, penelitian yang secara khusus mengkaji dinamika opini publik terhadap satu tahun pertama pemerintahan Presiden Prabowo Subianto dan Wakil Presiden Gibran Rakabuming Raka melalui integrasi *topic modeling* dan analisis sentimen berbasis *transformer* masih sangat terbatas. Padahal periode awal pemerintahan merupakan fase strategis yang mencerminkan respons masyarakat terhadap arah kebijakan pemerintah dan berpotensi menghasilkan dinamika percakapan yang sangat beragam di media sosial.

Berdasarkan kondisi tersebut, diperlukan suatu pendekatan yang mampu mengintegrasikan proses identifikasi topik dengan analisis sentimen secara komprehensif, memilih model klasifikasi berdasarkan hasil evaluasi empiris, serta menyajikan keseluruhan proses analisis ke dalam suatu sistem yang terintegrasi sehingga mampu menghasilkan informasi yang lebih objektif dan mudah diinterpretasikan.

Untuk menjawab kesenjangan penelitian tersebut, penelitian ini mengusulkan GovBERTic (*Government BERTopic*) sebagai suatu *framework* analisis opini publik berbasis *Natural Language Processing* yang mengintegrasikan BERTopic untuk mengidentifikasi topik utama dengan analisis sentimen berbasis *transformer*. Pendekatan ini tidak hanya bertujuan mengidentifikasi isu-isu yang

berkembang dalam opini masyarakat, tetapi juga menganalisis kecenderungan sentimen pada setiap topik sehingga mampu menghasilkan informasi yang lebih komprehensif. Penelitian ini juga menerapkan mekanisme *model selector* melalui evaluasi komparatif terhadap model IndoBERT, IndoBERTweet, dan RoBERTa untuk menentukan model klasifikasi sentimen dengan performa terbaik sesuai karakteristik dataset yang digunakan.

GovBERTic dikembangkan melalui alur kerja yang terintegrasi yang diawali dengan proses pengumpulan data opini publik dari media sosial X , dilanjutkan dengan *preprocessing* data untuk menghasilkan teks yang siap dianalisis. Tahap berikutnya adalah identifikasi topik menggunakan BERTopic, kemudian dilakukan *model selector* untuk mengevaluasi performa IndoBERT, IndoBERTweet, dan RoBERTa sehingga diperoleh model yang paling sesuai untuk melakukan klasifikasi sentimen. Hasil identifikasi topik dan analisis sentimen selanjutnya disajikan dalam bentuk visualisasi serta *insight* yang menggambarkan hubungan antara isu yang berkembang dan kecenderungan sentimen masyarakat terhadap isu tersebut.

Efektivitas pendekatan yang diusulkan dievaluasi melalui pengukuran performa model sentimen menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* dengan metode Stratified K-Fold Cross Validation. Sementara itu, hasil *topic modeling* dievaluasi berdasarkan kualitas topik yang dihasilkan serta kemampuannya dalam merepresentasikan isu-isu yang berkembang pada data opini publik. Melalui proses evaluasi tersebut, penelitian ini diharapkan mampu menunjukkan bahwa integrasi *topic modeling* dan analisis sentimen dalam GovBERTic dapat menghasilkan analisis opini publik yang lebih komprehensif, objektif, dan mudah diinterpretasikan.

Penelitian ini berfokus pada pengembangan GovBERTic sebagai *framework* analisis opini publik yang mengintegrasikan BERTopic dengan *model selector* berbasis *transformer* untuk menganalisis opini masyarakat terhadap satu tahun pertama pemerintahan Presiden Prabowo Subianto dan Wakil Presiden Gibran Rakabuming Raka. Melalui pendekatan tersebut, penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan metode analisis opini publik

berbasis *Natural Language Processing* sekaligus menghasilkan informasi mengenai isu dominan dan kecenderungan sentimen masyarakat yang dapat dimanfaatkan sebagai referensi bagi penelitian selanjutnya maupun sebagai bahan pertimbangan dalam memahami dinamika opini publik.

1.2.Rumusan Masalah

Rumusan masalah yang menjadi fokus utama pada penelitian ini adalah sebagai berikut:

1. Bagaimana *topic modeling* BERTopic dapat diimplementasikan dalam kerangka GovBERTic untuk mengekstraksi isu-isu utama opini publik dari data *tweet* satu tahun pemerintahan Prabowo-Gibran?
2. Bagaimana proses integrasi dan seleksi model klasifikasi sentimen (IndoBERT, IndoBERTweet, dan RoBERTa) dilakukan dalam kerangka GovBERTic?
3. Bagaimana antar muka pengguna grafis (*Graphical User Interface/GUI*) sistem GovBERTic dapat dirancang dan diimplementasikan untuk menampilkan daftar isu utama, hasil sentimen, dan hasil perbandingan antar metode secara informatif?

1.3.Batasan Masalah

Batasan yang ditetapkan pada masalah yang diamati pada penelitian ini adalah sebagai berikut:

1. Data penelitian terbatas pada unggahan (*tweet*) berbahasa Indonesia yang berasal dari Platform media sosial *X* dengan isu satu tahun pertama masa pemerintahan Presiden dan Wakil Presiden baru Prabowo–Gibran, dengan kata kunci yang telah ditentukan oleh peneliti.
2. Periode pengambilan data dibatasi pada rentang waktu tertentu yang mewakili tahun pertama pemerintahan tersebut yaitu Oktober 2024 – Oktober 2025. Data di luar periode tersebut tidak termasuk dalam ruang lingkup analisis.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk menjawab rumusan masalah serta memberikan arahan yang lebih jelas bagi penelitian ini. Adapun tujuan penelitian ini adalah sebagai berikut:

1. Mengimplementasikan BERTopic sebagai model utama untuk melakukan *topic modeling* pada data media sosial *X* berbahasa Indonesia terkait isu-isu utama selama satu tahun pertama pemerintahan Prabowo-Gibran.
2. Menerapkan analisis sentimen dengan menggunakan model *selector* (IndoBERT, IndoBERTweet, dan RoBERTa) untuk menentukan model terbaik dalam klasifikasi sentimen *tweet* publik pada masing-masing topik.

1.5. Manfaat Penelitian

Manfaat penelitian yang dapat diperoleh dari hasil penelitian ini adalah sebagai berikut:

1. Bagi bidang keilmuan

Penelitian ini diharapkan dapat berkontribusi terhadap pengembangan ilmu pengetahuan di bidang *data science*, khususnya pada penerapan *topic modeling* berbasis *transformer* untuk analisis isu dan opini publik pada media sosial terkait pemerintahan. Penelitian ini juga diharapkan memperkaya literatur mengenai pemanfaatan BERTopic yang didukung representasi semantik GovBERTic dalam mengidentifikasi topik-topik diskusi publik pada data teks berbahasa Indonesia yang bersifat dinamis dan kontekstual.

Selain itu, penelitian ini diharapkan memberikan kontribusi dalam kajian analisis sentimen berbasis *transformer* melalui perbandingan kinerja IndoBERT, IndoBERTweet, dan RoBERTa sebagai model *selector* untuk menentukan model terbaik dalam klasifikasi sentimen publik ke dalam kategori positif, negatif, dan netral. Integrasi antara hasil *topic modeling* dan analisis sentimen pada penelitian ini diharapkan dapat menghasilkan pendekatan yang lebih komprehensif dalam memahami isu-isu utama yang dibicarakan masyarakat beserta kecenderungan sentimen yang menyertainya. Dengan demikian, penelitian ini dapat menjadi referensi bagi pengembangan metode

analisis opini publik berbasis kecerdasan buatan, khususnya pada konteks evaluasi persepsi masyarakat terhadap kinerja pemerintahan melalui media sosial.

2. Bagi masyarakat

Tersedianya instrumen kontrol sosial pemerintah dan sumber rujukan terpercaya untuk memahami persepsi publik serta dinamika opini serta isu utama yang mengemuka terkait kinerja pemerintahan secara lebih objektif dan partisipatif.

3. Bagi Pemerintah

Memberikan informasi berbasis data kepada pemerintah, khususnya bagi para penyusun kebijakan. Hasil penelitian ini bermanfaat sebagai rekomendasi yang dapat dipertimbangkan dalam pengambilan keputusan dan perumusan kebijakan yang sensitif terhadap dinamika opini publik serta isu-isu utama yang beredar di media sosial mengenai satu tahun pertama pemerintahan.