

BAB I

PENDAHULUAN

Bab pendahuluan ini membahas mengenai latar belakang, rumusan masalah, tujuan dan manfaat penelitian, ruang lingkup penelitian, serta sistematika penulisan yang digunakan dalam skripsi yang berjudul “Analisis Sentimen Berbasis Aspek Menggunakan *BERTopic* serta Perbandingan *IndoBERTweet* dan *DistilBERTIndo* pada Opini Kebijakan PBB Kabupaten Pati”.

1.1 Latar Belakang

Perkembangan media sosial mendorong masyarakat untuk secara aktif menyampaikan opini dan tanggapan terhadap berbagai isu publik. Platform digital tidak hanya berfungsi sebagai sarana komunikasi, tetapi juga menjadi ruang bagi masyarakat untuk mengekspresikan pandangan, kritik, maupun dukungan terhadap kebijakan yang diterapkan oleh pemerintah. Dalam konteks kebijakan publik, percakapan di media sosial dapat mencerminkan persepsi dan respons masyarakat terhadap suatu kebijakan (Purwandari dkk., 2024). Salah satu isu yang menarik perhatian publik adalah kebijakan Pajak Bumi dan Bangunan (PBB) di Kabupaten Pati pada tahun 2025, yang memicu berbagai diskusi dan tanggapan masyarakat di media sosial.

Kebijakan kenaikan PBB pada tahun 2025 di Kabupaten Pati memicu munculnya protes warga (Asmani, 2025). Pada era komunikasi digital, respons tersebut tidak lagi terbatas pada ruang fisik, tetapi juga berkembang melalui berbagai platform media sosial. Salah satu platform yang sering digunakan untuk menyampaikan opini publik adalah X, yang memungkinkan pengguna membangun jejaring komunikasi yang luas dalam menyebarkan informasi dan pendapat. Pahtoni & Jati (2024) menjelaskan bahwa karakteristik ini menjadikan X sebagai ruang yang aktif bagi pengguna untuk berbagi emosi dan pandangan sehari-hari. Selain itu, diskusi publik juga banyak muncul pada kolom komentar YouTube yang sering digunakan masyarakat untuk memberikan tanggapan terhadap berbagai konten informasi dan pemberitaan. Purwandari dkk. (2024) menjelaskan bahwa analisis komentar YouTube dapat memberikan wawasan mengenai persepsi publik terhadap isu sosial dan politik yang sedang berkembang. Oleh karena itu, pemanfaatan data dari kedua platform ini

berpotensi memberikan gambaran yang lebih komprehensif mengenai opini masyarakat terhadap kebijakan yang dibahas.

Percakapan publik di media sosial sering kali memuat informasi yang lebih rinci mengenai sumber ketidakpuasan masyarakat terhadap suatu kebijakan. Dalam konteks kebijakan PBB di Kabupaten Pati, kritik masyarakat dapat diarahkan pada berbagai aspek, seperti besaran nominal pajak, penetapan Nilai Jual Objek Pajak, sikap kepala daerah, maupun tindakan aparat dalam menangani demonstrasi. Keragaman target kritik tersebut menunjukkan bahwa opini publik terhadap kebijakan bersifat multi-aspek. Namun demikian, analisis sentimen konvensional yang hanya mengklasifikasikan opini ke dalam kategori positif, negatif, atau netral memiliki keterbatasan dalam menangkap kompleksitas tersebut. W. Zhang dkk. (2023) menjelaskan bahwa pendekatan ini sering kali gagal mengidentifikasi perbedaan polaritas sentimen pada setiap entitas atau aspek yang muncul dalam satu teks. Akibatnya, pendekatan tersebut tidak mampu memetakan aspek mana yang sebenarnya menjadi fokus utama kritik masyarakat, sehingga hasil analisis menjadi kurang operasional untuk evaluasi kebijakan secara lebih spesifik.

Untuk mengatasi keterbatasan tersebut, pendekatan *Aspect-Based Sentiment Analysis (ABSA)* menjadi relevan karena kemampuannya dalam menganalisis opini pada level aspek atau entitas yang lebih mendalam, bukan sekadar level kalimat secara umum (W. Zhang dkk., 2023). Dalam konteks penelitian ini, kemampuan tersebut dimanfaatkan untuk memetakan sentimen publik ke aspek yang terkandung dalam data, sehingga hasilnya lebih operasional bagi proses evaluasi. Meskipun demikian, penerapan *ABSA* memiliki tantangan tersendiri, terutama terkait penyebutan aspek yang sering kali bersifat implisit (W. Zhang dkk., 2023), serta keberadaan teks yang tidak baku di media sosial (Siino dkk., 2024).

Salah satu tantangan dalam *ABSA* adalah penentuan aspek secara otomatis dari data teks yang tidak terstruktur. Penentuan aspek secara manual cenderung memakan waktu dan berpotensi menimbulkan bias peneliti. Oleh karena itu, penelitian ini memanfaatkan metode *topic modeling* modern, yaitu *BERTopic*, untuk mengekstraksi aspek secara *data-driven*. *BERTopic* memanfaatkan representasi teks berbasis *Transformer* untuk mengelompokkan dokumen yang memiliki kemiripan semantik, kemudian membentuk topik yang lebih mudah ditafsirkan menggunakan pendekatan *class-based TF-IDF (c-TF-IDF)* (Grootendorst, 2022).

Setelah aspek berhasil diidentifikasi, diperlukan model klasifikasi yang mampu memahami konteks bahasa media sosial. Pemilihan pendekatan berbasis *Deep Learning*, khususnya arsitektur *Transformer*, didasari oleh keterbatasan metode *Machine Learning* konvensional seperti *Support Vector Machine (SVM)* atau *Naive Bayes* dalam menangkap makna kontekstual. Purwandari dkk. (2024) menjelaskan bahwa metode konvensional memiliki performa yang terbatas pada data teks media sosial (Youtube), di mana algoritma seperti *SVM* dan *Naive Bayes* hanya mampu menghasilkan akurasi berkisar antara 59% hingga 69%. Sementara itu, model berbasis *Transformer* pada studi yang sama terbukti mampu menghasilkan performa yang jauh lebih unggul, dengan akurasi mencapai 81% untuk *IndoBERT* dan 83% untuk *RoBERTa*. Keunggulan ini dikarenakan arsitektur *BERT (Bidirectional Encoder Representations from Transformers)* mampu memahami konteks kalimat dalam dua arah, sehingga menghasilkan representasi semantik yang lebih akurat (Devlin dkk., 2019).

Dalam pengembangan model berbasis *Transformer* untuk bahasa Indonesia, terdapat beberapa varian yang memiliki karakteristik berbeda. *IndoBERTtweet* dikembangkan menggunakan data Twitter sehingga lebih mampu menangkap penggunaan bahasa informal dan slang pada media sosial (Koto dkk., 2021). Sementara itu, *DistilBERTIndo* merupakan model hasil proses *knowledge distillation* yang dirancang dengan ukuran lebih kecil dan proses komputasi yang lebih efisien tanpa menurunkan performa secara signifikan (Sanh dkk., 2019). Penelitian terdahulu menunjukkan bahwa kedua arsitektur tersebut memiliki performa yang kompetitif pada tugas klasifikasi teks. Sebagai contoh, pemanfaatan *IndoBERTtweet* pada data platform X menghasilkan *weighted F1-score* sebesar 0,7822 (Sihombing & Widiyaningtyas, 2025), sedangkan penggunaan varian *DistilBERT* mencatatkan akurasi sebesar 92,11% dan *F1-score* 92,04% (Darren & Oetama, 2025). Mengingat performa kedua model yang komparatif tersebut, penelitian ini membandingkan penggunaan *IndoBERTtweet* dan *DistilBERTIndo* dalam proses klasifikasi aspek dan sentimen pada opini publik terkait kebijakan PBB di Kabupaten Pati.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan mengembangkan sistem *Aspect-Based Sentiment Analysis (ABSA)* untuk menganalisis opini publik terkait kebijakan kenaikan PBB di Kabupaten Pati. Penelitian ini mengintegrasikan *BERTopic* untuk penemuan aspek secara *data-driven* serta melakukan analisis komparatif antara performa *IndoBERTtweet* dan *DistilBERTIndo* dalam klasifikasi aspek dan sentimen. Hasil penelitian

ini diharapkan dapat memberikan wawasan bagi pemerintah daerah dalam mengevaluasi kebijakan publik secara lebih objektif berdasarkan aspirasi masyarakat yang terekam di media digital, serta memperkaya kajian analisis sentimen kebijakan publik berbasis ABSA di Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, rumusan masalah dalam penelitian ini dapat diuraikan sebagai berikut:

1. Bagaimana penerapan metode *BERTopic* dalam mengekstraksi topik dari data media sosial X dan YouTube untuk mendapatkan label aspek final terkait isu kebijakan PBB?
2. Bagaimana perbandingan kinerja model *IndoBERTweet* dan *DistilBERTIndo* untuk klasifikasi aspek pada data opini publik terkait kebijakan PBB?
3. Bagaimana perbandingan kinerja model *IndoBERTweet* dan *DistilBERTIndo* untuk klasifikasi sentimen pada data opini publik terkait kebijakan PBB?

1.3 Tujuan dan Manfaat

Berdasarkan rumusan masalah yang telah ditetapkan, penelitian ini bertujuan untuk mengembangkan sistem Analisis Sentimen Berbasis Aspek terkait kebijakan PBB di Kabupaten Pati, yang meliputi:

1. Menerapkan metode *topic modeling BERTopic* untuk mengekstraksi topik pada data media sosial X dan YouTube untuk menetapkan label aspek terkait kebijakan PBB di Kabupaten Pati.
2. Membandingkan kinerja model *IndoBERTweet* dan *DistilBERTIndo* dalam melakukan klasifikasi aspek pada data opini publik terkait kebijakan PBB.
3. Membandingkan kinerja model *IndoBERTweet* dan *DistilBERTIndo* dalam mengklasifikasikan sentimen pada data opini publik terkait kebijakan PBB.

Manfaat yang diharapkan dari penelitian ini adalah tersedianya model Analisis Sentimen Berbasis Aspek untuk menganalisis opini masyarakat terhadap kebijakan PBB di Kabupaten Pati berdasarkan data unggahan pada platform X dan YouTube. Hasil analisis tersebut dapat memberikan gambaran mengenai kecenderungan opini publik serta aspek yang banyak menjadi sorotan masyarakat. Selain itu, penelitian ini juga dapat menjadi

referensi bagi pengembangan penelitian analisis sentimen pada data media sosial dengan memanfaatkan model bahasa berbasis *transformer*.

1.4 Ruang Lingkup

Ruang lingkup penelitian ini dibatasi pada enam poin utama agar proses penelitian berjalan secara terarah dan konsisten dengan tujuan penelitian, yaitu sebagai berikut:

1. Data yang digunakan merupakan data teks opini publik yang diperoleh dari platform X dan komentar YouTube. Data difokuskan pada topik kebijakan kenaikan PBB dan isu terkait di Kabupaten Pati pada rentang waktu 1 Mei 2025 – 31 Agustus 2025.
2. Penentuan label aspek dilakukan menggunakan metode *topic modeling BERTopic*. Penelitian ini tidak menggunakan daftar aspek yang ditentukan secara manual sejak awal, melainkan membiarkan aspek terbentuk berdasarkan pola dominan yang muncul dalam korpus data, yang kemudian diinterpretasikan secara manual oleh peneliti.
3. Pemodelan klasifikasi dilakukan dengan pendekatan *fine-tuning* pada dua model *Pre-trained Language Model (PLM)* berbasis *transformer*, yaitu *IndoBERTweet* (model skala penuh) dan *DistilBERTIndo* (model terdistilasi). Penelitian ini tidak membangun arsitektur model baru dari awal maupun membandingkan dengan model *transformer* lain di luar kedua arsitektur tersebut.
4. Optimasi konfigurasi model dilakukan melalui pencarian hyperparameter otomatis menggunakan algoritma *Tree-structured Parzen Estimator (TPE)* pada kerangka kerja *Optuna* dengan jumlah percobaan terbatas sesuai ketersediaan sumber daya komputasi.
5. Evaluasi model difokuskan pada beberapa metrik utama yang mencakup *Accuracy*, *Precision*, *Recall*, *F1-Score*, *Macro-F1 Score* untuk klasifikasi sentimen, serta tambahan metrik *Hamming Loss* dan *Subset Accuracy* yang dikhususkan untuk evaluasi klasifikasi aspek *multi-label*.
6. Proses pelabelan data aspek dan sentimen dilakukan secara manual, sehingga data berlabel yang digunakan dalam pemodelan berjumlah 1.234 dokumen dari keseluruhan data yang tersedia.

1.5 Sistematika Penulisan

Sistematika penulisan digunakan untuk memberikan gambaran dan alur yang runtut terkait penyusunan laporan skripsi yang berjudul “Analisis Sentimen Berbasis Aspek Menggunakan *BERTopic* serta Perbandingan *IndoBERTweet* dan *DistilBERTIndo* pada Opini Kebijakan PBB Kabupaten Pati”. Sistematika penulisan laporan dijelaskan sebagai berikut:

BAB I PENDAHULUAN

Bab pendahuluan menyajikan uraian mengenai latar belakang penelitian, rumusan masalah, tujuan dan manfaat penelitian, ruang lingkup yang menjadi batasan penelitian, serta sistematika penulisan laporan secara keseluruhan terkait penerapan *BERTopic* dan komparasi model *IndoBERTweet* serta *DistilBERTIndo*.

BAB II TINJAUAN PUSTAKA

Bab tinjauan pustaka memuat penelitian terdahulu serta landasan teori yang menjadi dasar dan acuan dalam pelaksanaan penelitian. Pembahasannya berfokus pada konsep-konsep utama yang relevan dan mendukung arah penelitian antara lain terkait analisis sentimen berbasis aspek, *data scraping*, cara pra-pemrosesan data, penerapan *BERTopic* dan komparasi model *IndoBERTweet* serta *DistilBERTIndo* dalam klasifikasi aspek dan sentimen.

BAB III METODOLOGI PENELITIAN

Bab metodologi penelitian menjelaskan alur dan langkah langkah yang digunakan dalam pelaksanaan penelitian. Gambaran umum dari penelitian ini adalah pengumpulan data, pra-pemrosesan data, ekstraksi aspek menggunakan *BERTopic*, pemodelan aspek serta pemodelan sentimen menggunakan *IndoBERTweet* dan *DistilBERTIndo*. Tahapan penelitian diakhiri dengan proses evaluasi model untuk menilai kinerja dan efektivitas metode yang digunakan.

BAB IV HASIL DAN PEMBAHASAN

Bab hasil dan pembahasan menyajikan hasil yang diperoleh dari seluruh rangkaian eksperimen, disertai analisis dan pembahasan terhadap temuan penelitian terkait penerapan *BERTopic* dan komparasi model *IndoBERTweet* serta *DistilBERTindo*.

BAB V PENUTUP

Bab ini memuat kesimpulan yang diperoleh dari keseluruhan proses penelitian terkait penerapan *BERTopic* dan komparasi model *IndoBERTweet* serta *DistilBERTindo* serta saran untuk pengembangan penelitian selanjutnya.