

BAB I

PENDAHULUAN

1.1 Latar Belakang

Proyek Kereta Cepat Jakarta-Bandung (Whoosh) merupakan salah satu proyek infrastruktur strategis nasional yang menarik perhatian publik sejak diresmikan pada Oktober 2023. Proyek ini tidak hanya menjadi simbol kemajuan teknologi transportasi di Indonesia, tetapi juga memunculkan berbagai opini terkait beban utang pembiayaan yang ditanggung pemerintah dan PT Kereta Api Indonesia (KAI). Penelitian sebelumnya menunjukkan bahwa analisis sentimen berbasis aspek (aspect-based) menggunakan *IndoBERT* yang telah di *fine-tune* mampu mengidentifikasi opini spesifik terhadap layanan hotel dengan akurasi tinggi [1]. Pendekatan serupa juga berhasil diterapkan pada ulasan bintang lima aplikasi Instagram di *Google Play Store*, di mana 24,1% ulasan ternyata mengandung sentimen negatif meskipun diberi rating tertinggi [2]. Media sosial telah menjadi ruang utama bagi masyarakat untuk menyampaikan opini secara cepat dan masif. Salah satu kasus menarik adalah analisis sentimen terhadap Badan Pengelola Investasi Danantara, yang menunjukkan 45% opini negatif di platform X meskipun isu investasi strategis sedang digencarkan [3]. Pada konteks layanan publik, performa *IndoBERT* juga terbukti unggul dibandingkan *Naïve Bayes* dan *SVM* dalam menganalisis opini di X terhadap layanan IndiHome, dengan akurasi mencapai 96% [4]. Hal serupa terjadi pada isu kenaikan PPN 12%, di mana *IndoBERT* berhasil mengklasifikasikan dominasi sentimen negatif dari data *Multi-platform* (X, Instagram) dengan akurasi 84,94% [5].

Penelitian lebih lanjut menunjukkan bahwa *IndoBERT* efektif digunakan untuk menganalisis ulasan aplikasi *Mobile JKN* di *Google Play Store* [6], serta sentimen pengguna terhadap Kereta Cepat Whoosh di X menggunakan pendekatan *Naïve Bayes* [7]. Dalam kasus *ChatGPT*, *fine-tuning* model *XLNet* pada data X juga menghasilkan performa tinggi [8], sementara klasifikasi emosi berbahasa Indonesia dengan *IndoBERT* mencapai akurasi 73% [9]. Analisis sentimen komentar *YouTube* terhadap Whoosh menggunakan *Bidirectional-LSTM* juga menunjukkan potensi besar dari pendekatan *Deep Learning* [10]. Berdasarkan laporan resmi, total biaya

proyek Whoosh mencapai Rp 114,24 triliun, dengan porsi utang sebesar 75% dari *China Development Bank* [11]. Isu ini menjadi viral di media sosial, terutama ketika muncul narasi bahwa utang Whoosh membebani keuangan negara dan PT KAI. Penelitian sebelumnya juga membuktikan bahwa *IndoBERT* mampu menganalisis sentimen masyarakat terhadap Whoosh di platform X dengan konfigurasi hyperparameter optimal. Prediksi emosi dalam teks berbahasa Indonesia menggunakan IndoBERT juga menunjukkan kemampuan model dalam menangkap nuansa kompleks [12].

Peningkatan kinerja IndoBERT melalui teknik confident learning berhasil memperbaiki akurasi dari 85,15% menjadi 86,03% pada kasus sentimen politik [13], sementara studi empiris membandingkan BERT dan DistilBERT pada pemilu presiden Indonesia menunjukkan efisiensi komputasi DistilBERT hingga 84% lebih cepat [14]. Analisis sentimen terkait migrasi tenaga kerja Indonesia di platform X menggunakan SVM-IndoBERT mencapai akurasi 90,5% setelah hyperparameter tuning [15], dan sentimen publik terhadap Pemilu 2024 di X menunjukkan Naïve Bayes sebagai model terbaik dengan akurasi 64% [16]. Fine-grained sentiment analysis pada big data *Multi-platform* di Indonesia menggunakan IndoBERT mencapai akurasi 96% dengan lima kategori sentimen [17], sementara perbandingan NBC dan SVM terhadap PP No.82 Tahun 2021 menunjukkan SVM lebih stabil dengan akurasi hingga 100% [18]. Analisis sentimen pengguna kendaraan listrik di X dengan Naïve Bayes, KNN, dan Decision Tree juga memberikan wawasan penting [19], begitu pula penerapan IndoBERT untuk deteksi sarkasme menggunakan Random Forest [20].

Sebagian besar penelitian analisis sentimen terdahulu umumnya terbatas pada pengujian satu platform tunggal (single-platform), yang rentan memicu bias representasi demografis serta fenomena ruang gema (echo chamber). Oleh karena itu, pengintegrasian multi-platform (X, YouTube, dan TikTok) menjadi sangat urgen untuk membangun korpus opini publik yang komprehensif dan heterogen. Platform X mengakomodasi respons spontan dan diskursus kebijakan yang padat, YouTube menyediakan ruang diskusi mendalam berupa teks panjang (long-form) yang argumentatif terhadap rincian data finansial, sedangkan TikTok

merepresentasikan dinamika percakapan generasi muda dan masyarakat luas dengan karakteristik bahasa gaul (slang) yang ekspresif.

Meskipun diskursus mengenai beban utang proyek Kereta Cepat Whoosh terkesan memiliki kecenderungan negatif, analisis empiris berbasis data tetap krusial untuk membuktikan sebaran opini secara objektif, memetakan keberadaan narasi tandingan yang rasional (seperti potensi manfaat konektivitas jangka panjang dan kinerja pendapatan operasional PT KAI), serta mengidentifikasi kluster-kluster spesifik kekhawatiran masyarakat, mulai dari beban fiskal APBN, akuntabilitas tata kelola proyek, hingga dependensi geopolitik pendanaan asing.

Urgensi penelitian ini bertujuan untuk menganalisis sentimen publik terhadap beban utang proyek Kereta Cepat Whoosh pada platform X, YouTube, dan TikTok selama periode 2025–2026 dengan mengintegrasikan metode IndoBERT dan Support Vector Machine (SVM). Kombinasi kedua metode ini diterapkan karena model IndoBERT standar terkadang kurang optimal jika langsung digunakan untuk tahap klasifikasi akhir [21]. Oleh karena itu, penelitian ini membagi proses menjadi dua tahapan: IndoBERT dioptimalkan sebagai feature extractor untuk merepresentasikan makna semantik dan kontekstual dari bahasa informal serta singkatan khas media sosial, sementara SVM berperan sebagai pengklasifikasi (classifier) yang menentukan kategori sentimen ke dalam label positif, negatif, atau netral [22]. Melalui pendekatan hibrida ini, akurasi model dalam mengenali opini publik diharapkan meningkat secara signifikan dibandingkan hanya menerapkan metode tunggal.

1.2 Rumusan Masalah

Berdasarkan latar belakang, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana mengimplementasikan model IndoBERT-Support Vector Machine dalam menganalisis sentimen publik terhadap beban utang proyek Kereta Cepat Whoosh pada data *Multi-platform* X, YouTube dan TikTok?
2. Bagaimana kinerja IndoBERT- Support Vector Machine pada sentimen publik terhadap isu beban utang proyek Kereta Cepat Whoosh berdasarkan analisis data viral pada platform X, YouTube dan TikTok?

1.3 Tujuan

Tujuan dari penelitian ini adalah :

1. Mengimplementasikan model *Hybrid* IndoBERT-Support Vector Machine untuk menganalisis sentimen publik terhadap beban utang proyek Kereta Cepat Whoosh menggunakan data *Multi-platform* dari X, YouTube dan TikTok.
2. Mengevaluasi kinerja model IndoBERT- Support Vector Machine dalam mengklasifikasikan sentimen publik terkait isu beban utang proyek Kereta Cepat Whoosh berdasarkan data viral dari platform X, YouTube dan TikTok.

1.4 Batasan Masalah

Agar penelitian ini lebih terarah dan terfokus, batasan masalah yang ditetapkan adalah sebagai berikut:

1. Penelitian ini menganalisis data opini publik berbasis teks berbahasa Indonesia yang bersumber secara multi-platform dari tiga media sosial utama, yaitu X (Twitter), YouTube, dan TikTok, terkait topik beban utang proyek Kereta Cepat Whoosh.
2. Analisis sentimen dilakukan menggunakan pendekatan *Hybrid* IndoBERT sebagai *Feature Extractor* dan Support Vector Machine (SVM) sebagai model klasifikasi dengan tiga kategori sentimen, yaitu positif, negatif, dan netral.
3. Data yang digunakan terbatas pada konten yang memiliki tingkat interaksi tinggi (viral), seperti jumlah like, share, komentar, dan view, dalam periode pengambilan data tahun 2025-2026.
4. Penelitian ini hanya berfokus pada analisis data berbasis teks dan tidak mencakup analisis data dalam bentuk lain seperti audio maupun video.
5. Cakupan media sosial dibatasi secara spesifik pada platform X, YouTube, dan TikTok karena ketiganya mewakili tiga karakteristik teks yang saling melengkapi (teks ringkas-spontan, teks panjang-argumentatif, dan teks bahasa gaul/slang video pendek), sehingga tidak memperluas ke platform lain seperti Facebook atau Instagram.
6. Pelabelan data dilakukan menggunakan pendekatan pseudo labeling dengan model IndoBERT SMSA, sehingga kualitas label yang dihasilkan sangat

bergantung pada performa model awal dalam memahami konteks sentimen bahasa Indonesia.

7. Penelitian ini tidak mencakup pengembangan sistem atau *deployment* dalam bentuk aplikasi, melainkan difokuskan pada analisis hasil model serta interpretasi sentimen menggunakan visualisasi seperti grafik dan wordcloud.

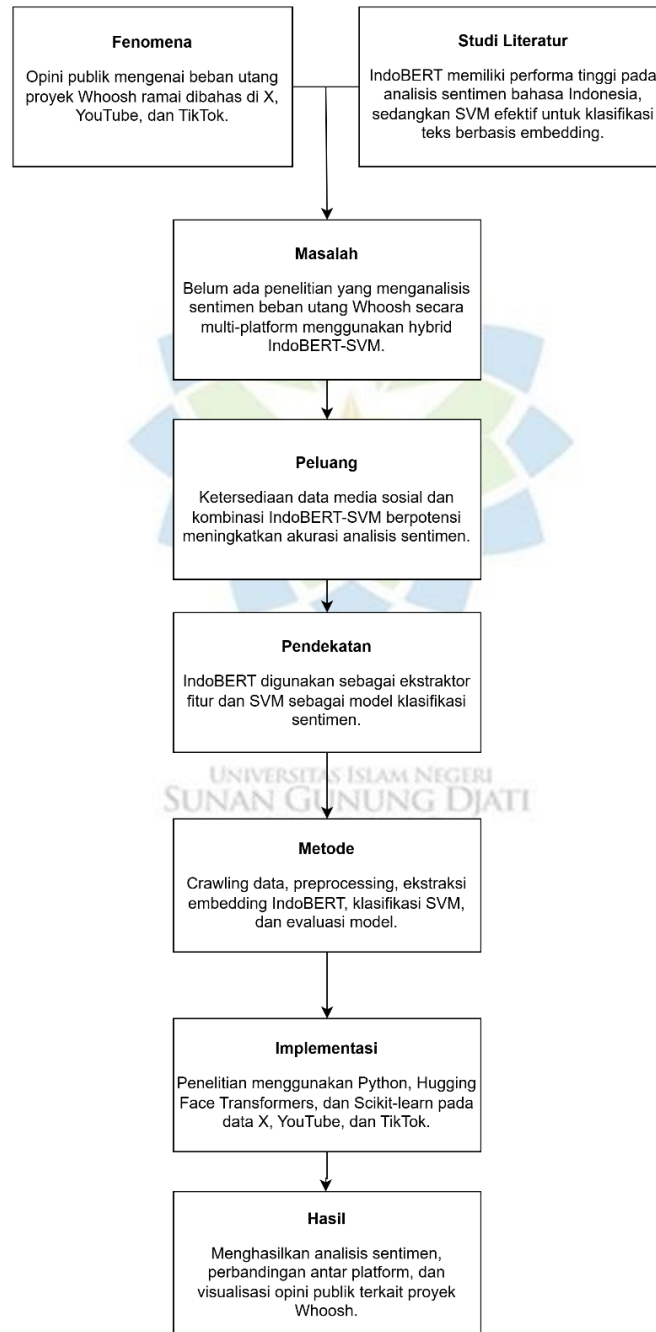
1.5 Manfaat

Hasil penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan kontribusi dalam pengembangan analisis sentimen berbasis *Natural Language Processing (NLP)* dengan memanfaatkan pendekatan *Hybrid IndoBERT-SVM* pada data *Multi-platform*, sehingga dapat digunakan sebagai referensi dalam penelitian serupa.
2. Menyediakan hasil analisis sentimen publik terhadap isu beban utang proyek Kereta Cepat Whoosh yang disajikan dalam bentuk visualisasi seperti grafik dan WordCloud, sehingga memudahkan dalam memahami pola opini masyarakat di platform X, YouTube, dan TikTok.
3. Menjadi referensi akademis bagi peneliti selanjutnya dalam mengembangkan metode analisis sentimen, khususnya yang mengombinasikan model *Deep Learning* dan *Machine learning* konvensional.
4. Memberikan gambaran umum mengenai persepsi publik terhadap proyek strategis nasional, yang dapat digunakan sebagai bahan pertimbangan dalam kajian ilmiah terkait komunikasi publik dan opini masyarakat.

1.6 Kerangka Pemikiran

Berdasarkan penjelasan sebelumnya, disusun kerangka pemikiran yang menggambarkan alur penelitian secara menyeluruh, mulai dari identifikasi fenomena hingga tahap implementasi penelitian. Kerangka pemikiran tersebut dapat dilihat pada Gambar 1.1.



Gambar 1. 1 Kerangka Pemikiran

1.7 Sistematika Penulisan

Sistematika penulisan disusun untuk memberikan gambaran yang jelas mengenai alur dan struktur penelitian ini. Penelitian disajikan dalam lima bab dengan susunan sebagai berikut:

BAB I PENDAHULUAN

Bab ini membahas latar belakang mengenai urgensi analisis opini publik terhadap isu beban utang proyek Kereta Cepat Whoosh di media sosial. Selain itu, bab ini memuat rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dalam konteks akademis, serta kerangka pemikiran yang telah diperbarui sebagai landasan konseptual pelaksanaan penelitian.

BAB II KAJIAN LITERATUR

Bab ini memuat landasan teori dan tinjauan teoretis yang mendukung penelitian, meliputi konsep Analisis Sentimen, *Natural Language Processing (NLP)*, arsitektur model IndoBERT, dan algoritma Support Vector Machine (SVM). Selain itu, bab ini memuat analisis komparatif mengenai performa model *Transformers* terhadap metode konvensional, serta landasan konseptual mengenai kombinasi model hibrida (*Hybrid model*) IndoBERT dan SVM. Dibahas pula teori penanganan ketidakseimbangan data melalui SMOTE, pendekatan Semi-Supervised Learning, metodologi CRISP-DM, karakteristik *Multi-platform* media sosial (X, YouTube, dan TikTok), serta pemetaan celah riset (research gap) terhadap penelitian terdahulu yang relevan.

BAB III METODOLOGI

Bab ini menjelaskan metode penelitian yang digunakan dengan mengacu pada metodologi CRISP-DM yang terdiri dari tahapan Business Understanding, Data Understanding, Data Preparation (praproses teks dan penanganan ketidakseimbangan kelas), Modeling (pembangunan model *Hybrid*), Evaluation, dan Deployment. Pada penelitian ini, tahap deployment difokuskan pada penyajian hasil analisis dalam bentuk visualisasi data. Bab ini juga memaparkan kriteria data viral dari platform X, YouTube, dan TikTok, alur pengolahan flow *Hybrid* model, serta perangkat yang digunakan dalam proses pemodelan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil eksperimen dan analisis mendalam dari seluruh tahapan penelitian yang dilakukan. Pembahasan meliputi gambaran umum dataset *Multi-platform*, hasil penyusutan data pada praproses teks, perbandingan pendekatan pelabelan (InSet Lexicon versus IndoBERT Pseudo-Labeling), serta spesifikasi matriks fitur IndoBERT *CLS Embeddings*. Bab ini juga memuat evaluasi perbandingan tiga strategi sampling (Original, Undersampling, dan SMOTE Oversampling) melalui skenario multi-split dan multi-kernel SVM, validasi dengan 5-Fold Cross Validation, eksperimen ekspansi dataset menggunakan Semi-Supervised Learning, serta interpretasi pola opini publik memanfaatkan WordCloud dan grafik distribusi sentimen.

BAB V SIMPULAN DAN SARAN

Bab ini memuat simpulan akhir penelitian sebagai jawaban atas rumusan masalah yang telah ditetapkan, termasuk kontribusi metodologis mengenai efisiensi pendekatan pseudo-labeling jika dibandingkan dengan pelabelan manual pada penelitian terdahulu. Selain itu, bab ini menyajikan saran-saran akademis dan praktis yang bersifat konstruktif sebagai rekomendasi untuk pengembangan penelitian analisis sentimen media sosial selanjutnya.