

BAB I

PENDAHULUAN

1.1 Latar Belakang

Era digital telah mentransformasi ruang publik, di mana platform media sosial seperti YouTube telah menjadi ekosistem utama untuk membentuk dan menyebarkan opini masyarakat [1]. Di Indonesia, dengan populasi pengguna internet yang masif, YouTube berperan tidak hanya sebagai platform hiburan, tetapi juga sebagai sumber informasi, edukasi, dan arena diskusi sosial politik keagamaan yang sangat berpengaruh [2]. Fenomena ini menjadikan kolom komentar YouTube sebagai *proxy* yang kaya akan data untuk memahami sentimen publik secara natural dan reflektif terhadap isu yang sedang berkembang [3].

Paralel dengan dinamika digital, Indonesia sebagai negara dengan keberagaman agama yang tinggi terus menghadapi tantangan dalam menjaga kerukunan, dengan isu intoleransi beragama yang masih kerap muncul dalam wacana publik [4]. Wacana ini turut bermigrasi dan berkembang pesat di ruang digital, termasuk di YouTube, melalui konten video dan interaksi di kolom komentarnya [5]. Analisis terhadap opini publik di ruang ini menjadi penting secara strategis untuk pemetaan awal (*early warning system*) potensi gesekan sosial, evaluasi efektivitas kebijakan, dan penguatan wacana toleransi [6].

Namun, menganalisis sentimen dari data *user-generated-content* seperti komentar YouTube menghadapi kendala teknis yang signifikan. Sifat teks yang informal, penuh dengan singkatan, slang, emotikon, dan kata tidak baku (*noisy text*) menyulitkan model *Natural Language Processing* (NLP) konvensional [7]. Pendekatan berbasis *lexicon* atau model *machine learning* tradisional sering kali gagal menangkap konteks dan nuansa semantik dalam bahasa yang dinamis ini [8].

Perkembangan mutakhir dalam NLP dihadirkan oleh arsitektur Transformer dan model pra-latihan (*pre-trained*) seperti BERT, yang merevolusi kemampuan pemahaman konteks kata secara mendalam [9]. Untuk bahasa Indonesia pada domain teks informal, IndoBERTweet hadir sebagai model yang dilatih pada korpus tweet, sehingga secara teoritis lebih adaptif menangani karakteristik bahasa di media sosial [10]. Meskipun Transformer unggul secara kontekstual, integrasi Bi-LSTM dapat memperkuat dependensi temporal.

Di sisi lain, arsitektur *Bidirectional Long Short-Term Memory* (Bi-LSTM) telah terbukti memiliki generalisasi yang baik dalam menangkap ketergantungan dan pola sekuensial dari dua arah (maju dan mundur) dalam suatu urutan data, seperti kalimat [11], [12]. Pada implementasi konvensional, model ini sering dipadukan dengan *Embedding* statis seperti Word2Vec, yang belum sepenuhnya merepresentasikan makna kontekstual dinamis suatu kata.

Oleh karena itu, muncul peluang untuk menggabungkan kekuatan kedua pendekatan dalam sebuah arsitektur *Hybrid*. Ide ini berangkat dari hipotesis bahwa representasi kontekstual yang kaya dari IndoBERTweet dapat berfungsi sebagai *feature extractor* tingkat lanjut, yang kemudian outputnya diproses lebih lanjut oleh lapisan Bi-LSTM untuk menangkap pola ketergantungan sekuensial yang kompleks dalam opini [13]. Sinergi ini diharapkan menghasilkan model yang lebih akurat untuk analisis sentimen pada teks yang penuh nuansa dan kompleksitas seperti pembahasan isu intoleransi.

Penelitian sejenis untuk isu sosial keagamaan di Indonesia masih banyak berfokus pada platform X (sebelumnya dikenal Twitter) [10], sementara eksplorasi mendalam menggunakan data komentar YouTube masih terbatas. Padahal, YouTube menawarkan konteks yang berbeda, dimana opini seringkali terikat pada konteks video spesifik dan memiliki pola interaksi yang unik [3]. Oleh karena itu, penelitian ini bertujuan mengisi celah (*gap*) tersebut dengan mengimplementasikan dan mengevaluasi arsitektur *Hybrid* IndoBERTweet dan Bi-LSTM untuk analisis sentimen (positif, negatif, netral) terhadap isu intoleransi beragama dengan sumber data primer dari komentar YouTube.

Hasil penelitian ini diharapkan memberikan kontribusi praktis sebagai alat bantu bagi pemangku kebijakan, organisasi masyarakat sipil, dan peneliti sosial untuk memantau sentimen publik secara otomatis, mendukung perumusan kebijakan atau kampanye yang lebih responsif. Secara akademis, penelitian ini memberikan eksperimen empiris dan wawasan baru pada bidang NLP bahasa Indonesia mengenai efektivitas arsitektur *Hybrid* IndoBERTweet dan Bi-LSTM pada domain isu sensitif dengan data dari sumber yang masih kurang dieksplorasi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka permasalahan dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana mengintegrasikan *Bidirectional Long Short-Term Memory* (Bi-LSTM) dan IndoBERTweet dalam menganalisis sentimen publik terhadap isu intoleransi beragama pada dataset komentar YouTube?
2. Model manakah di antara Bi-LSTM, IndoBERTweet, dan *Hybrid* (IndoBERTweet dan Bi-LSTM) yang menunjukkan performa paling optimal dalam mengklasifikasikan sentimen opini publik terhadap isu intoleransi beragama di platform YouTube?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengintegrasikan model *Bidirectional Long Short-Term Memory* (Bi-LSTM) dan IndoBERTweet dalam suatu pendekatan *Hybrid* untuk analisis sentimen publik terhadap isu intoleransi beragama berdasarkan dataset komentar YouTube.
2. Menganalisis dan membandingkan performa ketiga model (Bi-LSTM, IndoBERTweet, serta *Hybrid* IndoBERTweet dan Bi-LSTM) untuk menentukan konfigurasi model yang paling optimal dalam analisis sentimen isu intoleransi beragama.

1.4 Manfaat Penelitian

Berdasarkan tujuan penelitian di atas, maka manfaat dari penelitian ini adalah sebagai berikut:

1. Memberikan kontribusi terhadap pengembangan ilmu di bidang *Natural Language Processing* (NLP), khususnya dalam analisis sentimen berbasis bahasa Indonesia.
2. Menjadi referensi empiris mengenai penerapan IndoBERTweet, Bi-LSTM, dan pendekatan *Hybrid* IndoBERTweet dan Bi-LSTM dalam analisis sentimen komentar media sosial, terutama pada isu intoleransi beragama.
3. Memperkaya kajian akademis tentang intoleransi beragama dengan memanfaatkan pendekatan komputasi dan kecerdasan buatan sebagai alat analisis terhadap dinamika opini publik di ruang digital.

1.5 Batasan Penelitian

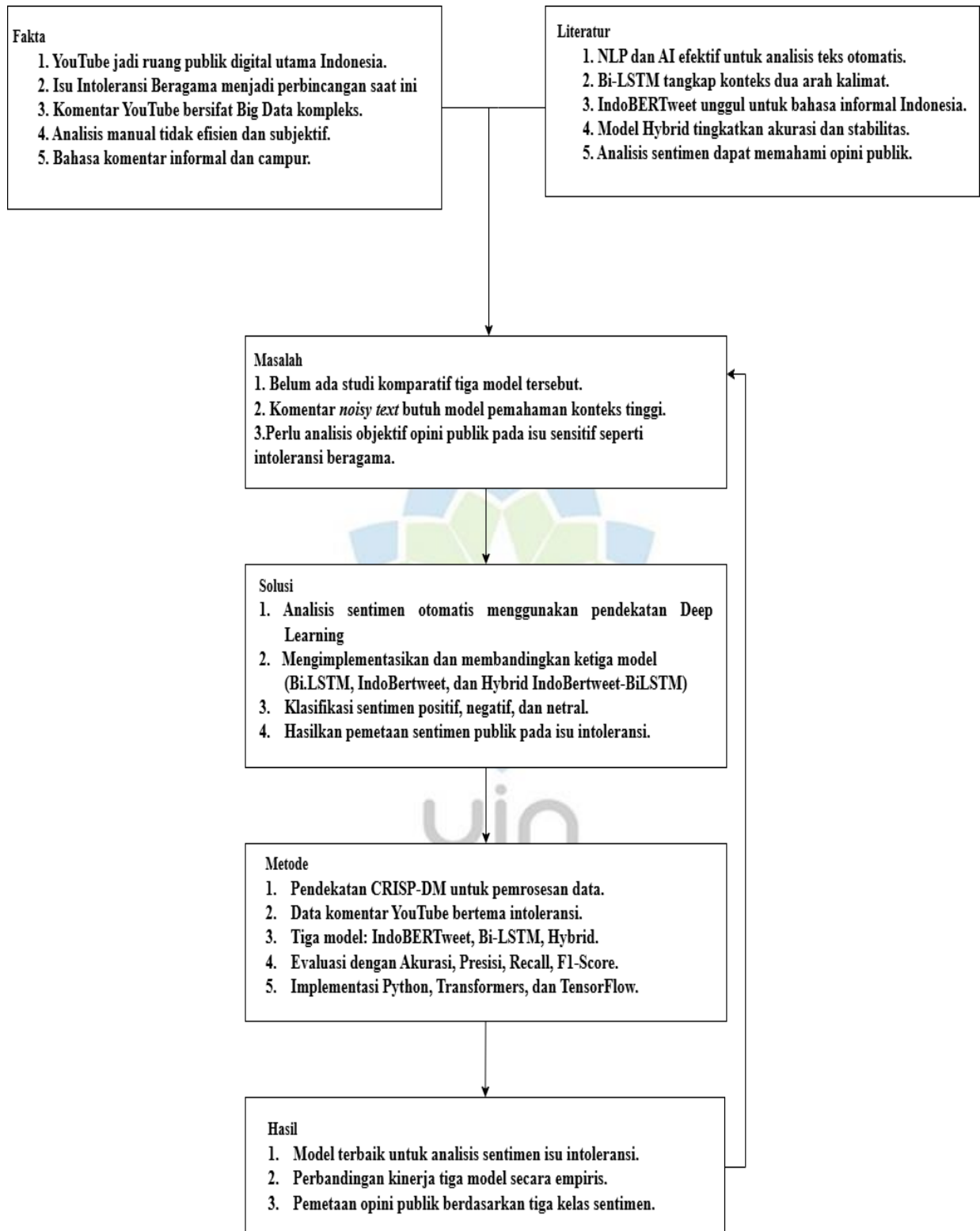
Agar penelitian ini lebih terarah dan tidak menyimpang dari tujuan yang telah ditetapkan, maka diberikan batasan-batasan sebagai berikut:

1. Fokus penelitian ini hanya pada sentimen terkait isu intoleransi beragama pada tahun 2025, yang terjadi di Padang, Sukabumi, dan Indonesia secara umum.
2. Dataset yang digunakan pada penelitian ini adalah komentar pada video YouTube yang membahas isu intoleransi beragama di Sukabumi pada chanel guru gembul.
3. Klasifikasi opini publik dibagi ke dalam tiga kategori atau kelas sentimen, yaitu positif, negatif, dan netral.
4. Penelitian ini hanya sampai pada tahap pengembangan dan evaluasi model. Tidak termasuk implementasi *deployment* model ke dalam sistem produksi, pembuatan *dashboard*, atau pengembangan aplikasi *end-user*.

1.6 Kerangka Pemikiran

Penelitian ini dilatarbelakangi oleh pentingnya memahami opini publik terhadap isu intoleransi beragama di Indonesia, yang banyak diperdebatkan pada platform YouTube. Analisis manual terhadap komentar yang bersifat *Big Data*, kompleks, dan informal tidak lagi efektif, sehingga diperlukan pendekatan otomatis berbasis *deep learning*. Berdasarkan studi literatur, model *Bidirectional LSTM* (Bi-LSTM) dan IndoBERTweet telah terbukti efektif untuk menangkap konteks dan bahasa informal, sementara arsitektur *Hybrid* berpotensi meningkatkan akurasi analisis sentimen.

Untuk mengisi celah penelitian yang belum ada studi komparatif ketiga model tersebut pada isu intoleransi beragama, penelitian ini mengusulkan implementasi dan evaluasi model Bi-LSTM, IndoBERTweet, serta *Hybrid* IndoBERTweet dan Bi-LSTM. Tujuannya adalah menentukan model terbaik untuk klasifikasi sentimen positif, negatif, dan netral, serta memetakan opini publik secara objektif. Alur lengkap kerangka pemikiran penelitian ini disajikan dalam Gambar 1. 1.



Gambar 1. 1 Kerangka Pemikiran.

1.7 Sistematika Penulisan

Pada bagian sistematika penulisan dijelaskan susunan laporan penelitian yang mencakup isi setiap bab, urutan penulisan, serta keterkaitan antarbagian dalam laporan. Adapun sistematika penulisan dalam penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bagian ini mencakup latar belakang, rumusan masalah, tujuan penelitian, manfaat yang diharapkan, batasan masalah, kerangka pemikiran dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini membahas landasan teori yang digunakan dalam penelitian, meliputi tinjauan literatur, penelitian terdahulu, konsep-konsep, teori, model, dan rumus yang relevan sebagai dasar analisis terhadap permasalahan penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini menguraikan kerangka kerja penelitian, metode pengumpulan dan pengolahan data, prosedur pengembangan sistem, serta teknik evaluasi yang digunakan untuk menjawab rumusan masalah secara sistematis.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian berdasarkan tahapan yang telah dilaksanakan, serta pembahasan yang mengaitkan hasil tersebut dengan rumusan masalah dan landasan teori.

BAB V PENUTUP

Bab ini berisi simpulan dari penelitian yang menjawab rumusan masalah, serta memberikan saran yang dapat dijadikan acuan bagi penelitian selanjutnya agar kualitas penelitian dapat ditingkatkan.