BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi telah mengubah cara masyarakat berkomunikasi dan berinteraksi di era digital. Media sosial menjadi platform utama dalam mengekspresikan pendapat dan sentimen masyarakat terhadap berbagai isu sosial. Berdasarkan laporan Digital dari *We Are Social*, jumlah pengguna aktif media sosial di Indonesia mencapai 139 juta pada tahun 2024 [1]. Di tahun yang sama, Indonesia menempati peringkat ketiga sebagai negara dengan rata-rata pengguna aktif platform X terbanyak di dunia [1]. Menurut DataReportal, jumlah pengguna aktif bulanan platform X di Indonesia mencapai 24,69 juta pada awal tahun 2024 [2]. Perkembangan platform media sosial di Indonesia terus menunjukkan pertumbuhan yang signifikan. Memasuki tahun 2025, jumlah pengguna aktif media sosial meningkat menjadi 143 juta [3]. Sementara itu, Indonesia naik ke peringkat kedua sebagai negara dengan rata-rata pengguna aktif platform X terbanyak di dunia setelah Jepang, dengan jumlah pengguna aktif bulanan mencapai 25,2 juta pada awal tahun 2025 [3][4]. Peningkatan ini menunjukkan bahwa X tetap menjadi salah satu platform media sosial yang berpengaruh di Indonesia. Melalui platform ini, pengguna sering kali mengungkapkan tanggapan atau opini mereka terhadap topik-topik yang sedang menjadi perhatian publik, baik dalam bentuk dukungan maupun penolakan.

Salah satu isu sosial yang kerap menjadi perdebatan di media sosial adalah fenomena LGBT (lesbian, *gay*, biseksual, dan *transgender*). Istilah LGBT secara khusus merujuk pada kelompok lesbian, yaitu perempuan yang memiliki ketertarikan emosional dan romantis terhadap sesama perempuan; *gay*, yaitu laki-laki yang memiliki ketertarikan terhadap sesama laki-laki; biseksual, yaitu individu yang menjalin hubungan asmara dengan baik laki-laki maupun perempuan; serta *transgender*, yaitu individu yang identitas gendernya tidak sesuai dengan jenis kelamin yang ditetapkan saat lahir [5]. Di Indonesia, isu LGBT masih dianggap tabu dan kontroversial, serta sering kali memunculkan sentimen negatif dari masyarakat luas. Survei menunjukkan bahwa mayoritas masyarakat Indonesia memandang LGBT sebagai ancaman terhadap norma sosial dan agama yang dianut

[6]. Meskipun diskusi tentang hak-hak LGBT mulai meningkat di ruang publik dan media sosial, penerimaan terhadap komunitas ini masih sangat terbatas, diperburuk oleh kerangka hukum yang belum mendukung secara komprehensif serta pemberitaan media yang cenderung negatif [7]. Oleh karena itu, diperlukan analisis sentimen untuk mengidentifikasi dan memahami pola persepsi serta kecenderungan opini publik terhadap isu LGBT di ruang digital.

Analisis sentimen merupakan bidang dalam *Natural Language Processing* (NLP) yang berfokus untuk mengkaji dan memahami opini atau ulasan masyarakat yang disampaikan dalam bentuk teks [8]. Proses pembelajaran dalam analisis ini dapat menentukan polaritas dari teks yang dianalisis ke dalam kategori positif, netral, atau negatif [8]. Sejalan dengan meningkatnya diskusi mengenai isu-isu sosial di media sosial, penelitian tentang analisis sentimen juga mengalami kemajuan yang cukup pesat dalam beberapa tahun terakhir. Berbagai pendekatan dan model telah digunakan untuk memahami opini publik terhadap berbagai topik. Salah satu penelitian dilakukan oleh Teddy Oswari, dkk. [9], yang menggunakan model IndoBERT untuk menganalisis sentimen terhadap isu LGBT berdasarkan komentar di YouTube, dengan akurasi mencapai 74% dan *F1-Score* sebesar 82% dalam klasifikasi pro dan kontra. Penelitian serupa dilakukan oleh Muhammad Ridwan, dkk. [5], yang memanfaatkan algoritma *Support Vector Machine* (SVM) untuk menganalisis sentimen terkait komunitas LGBT di Indonesia berdasarkan data dari Twitter, dengan akurasi klasifikasi tertinggi sebesar 74%.

Selain itu, penelitian oleh Alwi Jaya [10] menggunakan model RoBERTa Base Sentiment Classifier Indonesia untuk menganalisis sentimen terhadap profesi Pegawai Negeri Sipil (PNS) di Twitter, yang memperoleh tingkat akurasi sebesar 93,2% serta nilai *F1-Score* makro sebesar 91,02%. Studi lain oleh Nisrina Akbar Rizky Putri, dkk. [8] membandingkan model IndoBERT dan IndoRoBERTa dalam menganalisis sentimen terkait kemajuan kecerdasan buatan (AI) di Indonesia berdasarkan komentar YouTube. Hasil evaluasi menunjukkan bahwa IndoBERT mengungguli IndoRoBERTa dengan akurasi pengujian masing-masing sebesar 84% dan 83%, serta 83% dan 84% pada skema yang berbeda. Penelitian oleh Taufiq Dwi Purnomo, dkk. [11] juga memberikan kontribusi dengan membandingkan delapan model berbasis BERT dalam analisis sentimen multibahasa menggunakan dataset

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG

NusaX, di mana model IndoBERT-base (IndoNLU) mencatatkan kinerja dengan rata-rata *F1-Score* sebesar 0,93190, sedangkan untuk model NusaBERT-base mencapai 0,88878.

Berdasarkan tinjauan dari penelitian sebelumnya, dapat diidentifikasi bahwa masih terbatas penelitian yang secara spesifik menganalisis sentimen masyarakat Indonesia terhadap isu LGBT di *platform* X menggunakan pendekatan *deep learning* berbasis *Transformer* seperti IndoBERT, IndoRoBERTa, dan NusaBERT. Beberapa metode NLP yang dikembangkan dari model BERT dan secara khusus dilatih dalam korpus bahasa Indonesia seperti IndoBERT, IndoRoBERTa, dan NusaBERT telah memberikan kemajuan penting dalam analisis sentimen. IndoBERT merupakan varian dari model BERT yang secara khusus dikembangkan untuk menangani teks dalam bahasa Indonesia [12]. IndoRoBERTa adalah pengembangan dari model RoBERTa yang diadaptasi khusus untuk bahasa Indonesia [13]. Model ini melakukan sejumlah modifikasi pada proses *pre-training*, di antaranya menerapkan *dynamic masking*, menghilangkan fitur *Next Sentence Prediction* (NSP), menggunakan data dengan panjang kalimat yang lebih besar selama proses pelatihan [8]. Sementara itu, NusaBERT merupakan model bahasa yang dikembangkan dari IndoBERT untuk menangani keberagaman bahasa Indonesia [11]. Model ini dilatih menggunakan korpus multibahasa yang mencakup berbagai bahasa daerah di Indonesia [14].

Ketiga model tersebut dibangun berdasarkan arsitektur BERT dan dirancang untuk memahami bahasa Indonesia secara mendalam, namun masing-masing memiliki pendekatan dan karakteristik pelatihan yang berbeda. Dengan demikian, penting untuk melakukan studi komparatif guna mengetahui kinerja terbaik di antara ketiganya, khususnya dalam konteks analisis sentimen terhadap isu-isu kompleks. Kebaruan dari penelitian ini terletak pada perbandingan kinerja tiga model *Transformer* sekaligus implementasi model terbaik dalam menganalisis sentimen masyarakat Indonesia terhadap isu LGBT di media sosial X. Oleh karena itu, penelitian ini diharapkan mampu memberikan kontribusi ilmiah dalam pengembangan model analisis sentimen berbasis *Transformer* yang lebih mendalam terhadap opini publik masyarakat Indonesia mengenai isu LGBT di media sosial X,

serta menunjukkan efektivitas dan kinerja masing-masing model dalam analisis sentimen pada ranah media sosial berbahasa Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, penelitian ini akan berfokus pada rumusan masalah berikut:

1. Bagaimana mengimplementasikan model IndoBERT, IndoRoBERTa, dan NusaBERT untuk menganalisis sentimen masyarakat terhadap isu LGBT di media sosial X?
2. Bagaimana perbandingan kinerja model IndoBERT, IndoRoBERTa, dan NusaBERT dalam menganalisis sentimen masyarakat terhadap isu LGBT di media sosial X?

1.3 Tujuan

Berdasarkan rumusan masalah yang telah diuraikan, penelitian ini dilaksanakan dengan tujuan:

1. Mengimplementasikan model IndoBERT, IndoRoBERTa, dan NusaBERT dalam menganalisis sentimen masyarakat terhadap isu LGBT di media sosial X.
2. Mengetahui perbandingan kinerja model IndoBERT, IndoRoBERTa, dan NusaBERT dalam menganalisis sentimen masyarakat terhadap isu LGBT di media sosial X.

1.4 Manfaat

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Manfaat Teoritis
   a. Memperluas kajian ilmiah dalam bidang analisis sentimen menggunakan pendekatan *Deep Learning*, khususnya implementasi model BERT untuk isu-isu sensitif di media sosial.
   b. Memberikan pemahaman mendalam tentang pola sentimen masyarakat terhadap isu LGBT di media sosial X melalui pendekatan *Deep Learning*.

c. Menyumbang referensi akademis tentang efektivitas penggunaan model BERT dalam analisis sentimen berbahasa Indonesia.

2. Manfaat Praktis
   a. Bagi Komunitas Akademik
      1) Menyediakan dataset tervalidasi tentang sentimen masyarakat terhadap isu LGBT yang dapat digunakan untuk penelitian lanjutan.
      2) Memberikan kerangka kerja metodologi untuk analisis sentimen isu-isu sensitif di media sosial menggunakan model BERT.
   b. Bagi Pengembang AI
      1) Memberikan contoh implementasi model BERT untuk analisis sentimen dalam konteks bahasa Indonesia.
      2) Memberikan standar pembanding (*benchmark*) untuk pengembangan sistem analisis sentimen yang lebih baik.
   c. Bagi Peneliti
      1) Mengembangkan keahlian dalam penerapan teknologi *Deep Learning* untuk analisis data media sosial.
      2) Meningkatkan pemahaman tentang implementasi model BERT dalam konteks analisis sentimen.
   d. Bagi Masyarakat
      1) Memberikan gambaran objektif tentang persepsi publik terhadap isu LGBT berdasarkan data media sosial.
      2) Menyediakan wawasan berbasis data untuk pemahaman yang lebih baik tentang dinamika sosial terkait isu LGBT di Indonesia.

1.5 Batasan Masalah

Untuk memfokuskan penelitian ini, beberapa batasan masalah yang ditetapkan adalah:

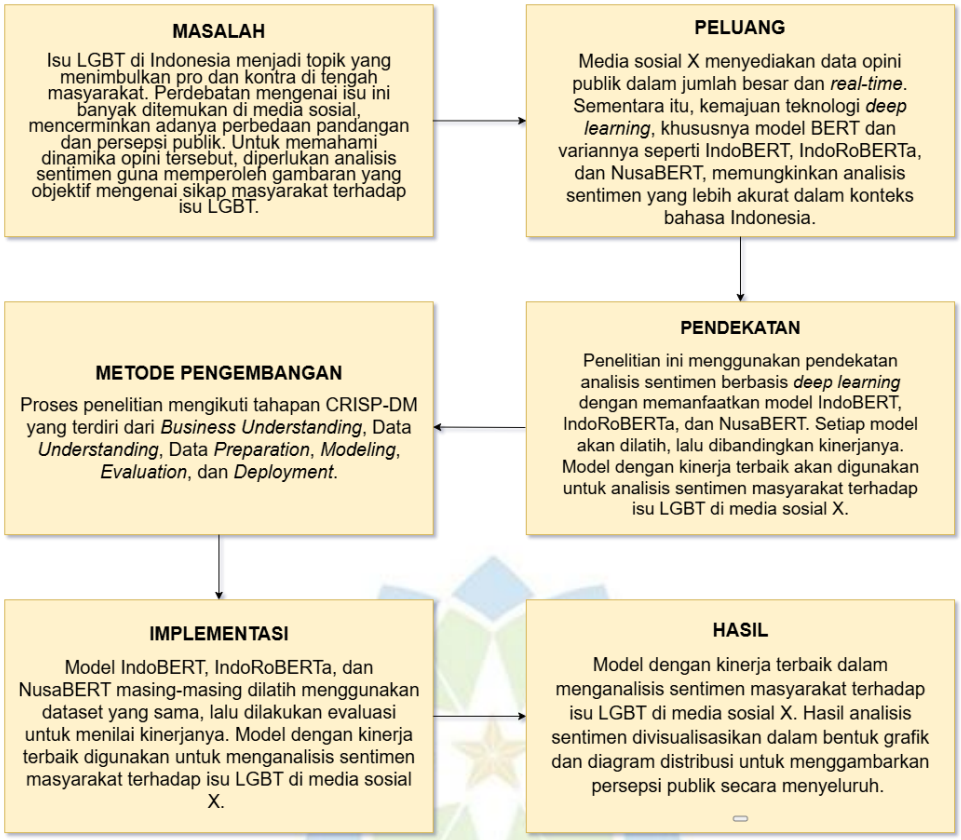
1. Data yang akan digunakan diambil dari media sosial X menggunakan *library Tweet-Harvest.*
2. Periode pengambilan data dibatasi dari bulan Juli 2024 hingga Juni 2025.
3. Jumlah *crawling* data dibatasi sebanyak 30.000 *tweet* yang mengandung kata kunci “lgbt”.

4. Penelitian berfokus pada komentar masyarakat terhadap isu LGBT berbahasa Indonesia, serta hanya menganalisis konten tekstual.
5. Pelabelan data dilakukan secara manual dan otomatis menggunakan model indobertweet-base-Indonesian-sentiment-analysis, yang dikembangkan oleh Aardiiiiy
6. Klasifikasi sentimen dibagi menjadi tiga kategori, Positif (mendukung/menerima), Negatif (menolak/tidak setuju) dan Netral (objektif/tidak memihak).
7. Model BERT yang digunakan meliputi IndoBERT, IndoRoBERTa, dan NusaBERT, yang merupakan varian model BERT yang secara khusus dilatih pada dataset berbahasa Indonesia.
8. Kinerja model yang akan akan diukur meliputi nilai akurasi, presisi, *recall*, dan *F1-score*.

1.6 Kerangka Pemikiran

Dalam upaya memetakan proses dan tahapan penelitian secara sistematis, diperlukan suatu alur konseptual yang menggambarkan tahapan dan proses analisis yang akan dilakukan. Kerangka pemikiran dalam penelitian ini bertujuan untuk memberikan gambaran yang jelas tentang alur analisis sentimen terhadap isu LGBT di media sosial X menggunakan pendekatan model *Deep Learning*. Visualisasi dari kerangka pemikiran penelitian ini dapat dilihat pada Gambar 1.1.

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG



Gambar 1.1 Kerangka Pemikiran

1.7 Sistematika Penulisan

Bagian ini berisi uraian mengenai susunan dan alur penulisan laporan tugas akhir yang memberikan gambaran umum isi dari setiap bab. Selain itu, dijelaskan juga keterkaitan antar bab dalam menyusun laporan tugas akhir secara terstruktur dan logis. Adapun sistematika penulisan laporan ini dijabarkan sebagai berikut:

BAB I PENDAHULUAN

Bab ini memuat uraian mengenai latar belakang masalah, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, kerangka pemikiran, dan sistematika penulisan sebagai gambaran awal mengenai ruang lingkup penelitian.

BAB II KAJIAN LITERATUR

Pada bab ini dijelaskan hasil penelitian terdahulu, teori-teori, serta konsep dasar yang menjadi dasar ilmiah dan referensi dalam melakukan penelitian ini.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan metode serta tahapan-tahapan yang digunakan dalam pelaksanaan penelitian, mulai dari proses perencanaan hingga pemrosesan dan analisis data.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi pemaparan hasil penelitian yang diperoleh dari implementasi dan pengujian model yang dikembangkan. Selanjutnya, hasil tersebut dianalisis dan dibahas untuk menjawab rumusan masalah yang telah disusun sebelumnya.

BAB V SIMPULAN DAN SARAN

Bab terakhir memuat kesimpulan dari hasil penelitian sebagai jawaban dari rumusan masalah, serta memberikan saran-saran yang relevan untuk pengembangan penelitian selanjutnya atau penerapan di masa mendatang

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG