

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi telah memberikan dampak besar terhadap cara manusia berinteraksi dan mengekspresikan opini. Salah satu wujud nyata dari perkembangan tersebut adalah meningkatnya penggunaan media sosial seperti YouTube sebagai sarana untuk menyampaikan pendapat, kritik, dan emosi secara bebas[1]. Melalui media sosial, masyarakat dapat berpartisipasi aktif dalam diskursus publik dan membentuk persepsi kolektif terhadap berbagai isu sosial. Oleh karena itu, data yang bersumber dari media sosial menjadi objek penting dalam penelitian berbasis kecerdasan buatan, khususnya dalam bidang *sentiment analysis*[1].

Salah satu isu sosial yang menarik perhatian publik dan mendapat respons dari pemerintah Indonesia adalah fenomena yang berkaitan dengan Warga Negara Indonesia (WNI) di Jepang. Berdasarkan data yang disampaikan dalam siaran pers resmi Kedutaan Besar Republik Indonesia (KBRI) Tokyo dan Konsulat Jenderal Republik Indonesia (KJRI) Osaka per Juli 2025, jumlah WNI yang berada di Jepang pada Desember 2024 mencapai 199.824 orang, sekitar 5% dari total warga asing yang tinggal di Jepang, dan sekitar 7.000 di antaranya merupakan pelajar dan mahasiswa [2]. Hal ini menunjukkan bahwa komunitas WNI di Jepang merupakan bagian yang nyata dan signifikan dari populasi warga asing di negara tersebut.

Beberapa peristiwa yang melibatkan warga negara Indonesia (WNI) di Jepang juga menjadi sorotan publik dan media internasional. Menurut pemberitaan yang mengutip laporan BBC Indonesia, pada akhir Juni 2025 aparat penegak hukum Jepang menangkap tiga WNI yang diduga melakukan perampokan terhadap warga lokal di Prefektur Ibaraki, Jepang. Ketiga WNI tersebut telah diberikan pendampingan hukum oleh Kedutaan Besar Republik Indonesia (KBRI) di Tokyo, meskipun mereka diketahui tinggal di Jepang melebihi masa izin tinggal yang berlaku. Selain itu, laporan tersebut mencatat bahwa sebelum insiden kriminal tersebut viral di ruang digital, telah beredar informasi dan video di media sosial mengenai aktivitas WNI di Jepang yang menimbulkan respons beragam dari publik, seperti pada awal hingga pertengahan bulan Juni 2025 adanya pemasangan atribut komunitas pada fasilitas umum yang kemudian memicu opini masyarakat terkait

perilaku komunitas diaspora[3].

Dalam konteks tersebut, berbagai isu yang melibatkan atau berkaitan dengan Warga Negara Indonesia (WNI) di Jepang kerap memicu beragam respons dari masyarakat, khususnya di media sosial seperti YouTube. Kolom komentar pada video yang membahas isu tersebut menjadi ruang bagi publik untuk mengekspresikan opini, baik dalam bentuk dukungan, kritik, maupun sentimen negatif yang bersifat emosional dan subjektif. Pola opini yang terbentuk di ruang digital ini berpotensi memengaruhi persepsi publik secara lebih luas, terutama ketika informasi yang beredar tidak selalu diiringi dengan klarifikasi yang memadai.

Namun, Pada kasus tertentu, satu peristiwa dapat menghasilkan ribuan komentar dalam waktu singkat, sehingga membentuk wacana publik yang beragam dan berpotensi menimbulkan stigma sosial. Besarnya volume komentar tersebut menjadikan analisis manual tidak efisien dan berpotensi subjektif. Oleh karena itu, diperlukan pendekatan analisis sentimen yang objektif, terukur, dan berbasis data untuk mengidentifikasi kecenderungan sentimen publik secara sistematis. Pendekatan tersebut diharapkan mampu memberikan gambaran yang lebih komprehensif mengenai pola sentimen masyarakat terhadap isu WNI di Jepang, sekaligus mendukung pemahaman yang lebih mendalam terhadap dinamika opini publik di media sosial.

Analisis sentimen sendiri merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mengetahui informasi sentimen yang terkandung dalam suatu kalimat. Berhubungan dengan banyaknya opini yang disampaikan oleh masyarakat, analisis sentimen ini menjadi berguna untuk menilai dan mengetahui polaritas sentimen terkait kasus tersebut[4].

Beberapa penelitian terdahulu telah membahas analisis sentimen teks berbahasa Indonesia dengan berbagai pendekatan. Penelitian[5] menemukan bahwa model LSTM dengan konfigurasi terbaik mampu mencapai akurasi sebesar 81,28% dalam menganalisis sentimen komentar YouTube terkait pemblokiran TikTok Shop. Dalam penelitian oleh Khoirunnisa & Setiawan (2025) ditemukan bahwa model LSTM dengan ekspansi fitur Word2Vec dan optimisasi Adam pada dataset Twitter berbahasa Indonesia berhasil mencapai akurasi 83,04% dan F1-score 82,62% dalam analisis sentimen publik[6]. Meskipun hasilnya cukup baik, kedua pendekatan tersebut memiliki keterbatasan dalam menangani konteks semantik yang kompleks, terutama dalam kalimat informal khas media sosial.

Kemunculan arsitektur *transformer*, khususnya Bidirectional Encoder Representations from Transformers (BERT), membawa kemajuan signifikan dalam

bidang *Natural Language Processing* (NLP). Penelitian yang dilakukan oleh [7] berjudul "*Sentiment Analysis on Twitter Social Media towards Climate Change in Indonesia Using IndoBERT Model*" menunjukkan bahwa model IndoBERT memiliki kemampuan yang baik dalam memahami konteks bahasa Indonesia pada media sosial. Penelitian ini menggunakan data dari platform Twitter untuk menganalisis opini publik mengenai isu perubahan iklim di Indonesia. Hasilnya menunjukkan bahwa IndoBERT mampu mencapai tingkat akurasi yang tinggi sebesar 95,3% dalam mengklasifikasikan sentimen positif, negatif, dan netral, dibandingkan dengan pendekatan berbasis lexicon maupun metode klasik seperti Naive Bayes. Namun, seperti model *Transformer* pada umumnya, BERT termasuk IndoBERT memiliki keterbatasan optimal dalam menangkap *long-term dependency* pada urutan teks dikarenakan terbatas oleh *fixed-length context* dalam permodelan bahasa [8].

Untuk mengatasi kekurangan tersebut, berbagai studi mulai mengembangkan pendekatan hibrida yang menggabungkan keunggulan model *transformer* dan *recurrent neural network* (RNN). Salah satu penelitian yang dilakukan oleh Kamal (2025) [9], yang menerapkan metode *ensemble stacking* antara IndoBERT dan LSTM untuk identifikasi berita palsu di portal berita online, dengan capaian akurasi mencapai 95,41%. Hasil ini menunjukkan bahwa penggabungan model *IndoBERT–LSTM* dapat menghasilkan kinerja yang lebih baik dibandingkan penggunaan salah satu model secara tunggal. Kombinasi tersebut memungkinkan sistem untuk memanfaatkan keunggulan *IndoBERT* dalam memahami konteks semantik dua arah serta kemampuan *LSTM* dalam menangkap urutan temporal dan dependensi jangka panjang antar token dalam teks.

Berdasarkan penelitian terdahulu, analisis sentimen berbahasa Indonesia telah banyak dilakukan menggunakan *LSTM* maupun *IndoBERT* secara terpisah. Namun, penerapan model *hybrid IndoBERT–LSTM* untuk menganalisis komentar YouTube terkait isu sosial spesifik seperti WNI di Jepang masih sangat terbatas. Selain itu, sebagian besar penelitian sebelumnya berfokus pada data Twitter atau topik umum, sehingga belum menggambarkan karakteristik opini publik pada platform YouTube yang bersifat lebih panjang dan emosional.

Oleh karena itu, penelitian ini mengusulkan penerapan model *hybrid IndoBERT–LSTM* untuk melakukan analisis sentimen pada komentar *YouTube* berbahasa Indonesia yang membahas isu Warga Negara Indonesia (WNI) di Jepang. Pemilihan model ini didasarkan pada kemampuan IndoBERT sebagai *pre-trained language model* yang dirancang khusus untuk bahasa Indonesia dalam menangkap

konteks semantik dan makna kata secara kontekstual. Dengan memanfaatkan representasi embedding kontekstual dari IndoBERT, model mampu memahami variasi bahasa informal, istilah sehari-hari, serta konteks kalimat yang umum ditemukan dalam komentar media sosial.

Selanjutnya, lapisan LSTM digunakan untuk memodelkan ketergantungan sekuensial antar token dalam sebuah komentar, sehingga informasi yang berkaitan dengan urutan kata dan dinamika ekspresi sentimen dapat dipelajari secara lebih efektif. Kombinasi IndoBERT dan LSTM diharapkan mampu mengintegrasikan keunggulan pemahaman konteks semantik dan pemodelan informasi temporal secara bersamaan. Dengan demikian, model *IndoBERT-LSTM* diharapkan dapat mengidentifikasi sentimen positif, negatif, maupun netral secara lebih akurat dan stabil dibandingkan pendekatan tradisional, khususnya dalam menganalisis opini publik digital yang bersifat kompleks dan emosional seperti komentar YouTube terkait isu WNI di Jepang.

1.2 Rumusan Masalah

1. Bagaimana cara melakukan analisis sentimen terhadap komentar *YouTube* berbahasa Indonesia yang membahas isu WNI di Jepang?
2. Bagaimana kinerja model *IndoBERT-LSTM* dalam menganalisis sentimen komentar *YouTube* ?

1.3 Batasan Masalah

1. Data yang digunakan adalah komentar *YouTube* berbahasa Indonesia mengenai isu WNI di Jepang.
2. Model yang digunakan terbatas pada *IndoBERT, LSTM*.
3. Klasifikasi sentimen dibatasi menjadi tiga kategori: positif, negatif, dan netral.
4. Evaluasi model menggunakan metrik *accuracy, precision, recall*, dan *F1-score*.

1.4 Tujuan Penelitian

1. Mengimplementasikan model *hybrid IndoBERT-LSTM* untuk analisis sentimen komentar *YouTube* berbahasa Indonesia.
2. Mengembangkan model *hybrid IndoBERT-LSTM* untuk melakukan analisis sentimen komentar *YouTube* serta mengevaluasi performa model menggunakan metrik *akurasi*, *precision*, *recall*, dan *F1-Score*.

1.5 Manfaat Penelitian

1.5.1 Manfaat Akademis

1. Memberikan kontribusi terhadap pengembangan bidang *Natural Language Processing* (NLP) untuk bahasa Indonesia melalui penerapan model hibrida *IndoBERT-LSTM*.
2. Menyediakan dasar bagi penelitian lanjutan dalam optimasi model hibrida berbasis *deep learning* untuk pemrosesan teks informal di media sosial.

1.5.2 Manfaat Praktis

1. Menyediakan pendekatan yang dapat diadaptasi untuk pengembangan sistem analisis sentimen otomatis berbasis bahasa Indonesia.
2. Mendemonstrasikan penerapan model *IndoBERT-LSTM* sebagai solusi dalam menangani permasalahan konteks semantik dan urutan temporal pada teks berbahasa alami.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian ini disusun untuk memberikan gambaran umum mengenai isi laporan dari awal hingga akhir. Adapun sistematika penulisan laporan adalah sebagai berikut:

1. Bab 1 PENDAHULUAN

Bab ini berisi latar belakang penelitian, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, serta sistematika penulisan.

2. Bab 2 LANDASAN TEORI

Bab ini membahas kajian pustaka dan landasan teori yang mendukung penelitian, termasuk penelitian terdahulu yang relevan.

3. Bab 3 METODOLOGI PENELITIAN

Bab ini menjelaskan metode penelitian yang digunakan, meliputi pendekatan, sumber data, teknik pengumpulan data, serta metode analisis.

4. Bab 4 HASIL DAN DISKUSI

Bab ini berisi hasil penelitian dan pembahasan yang diperoleh berdasarkan analisis data.

5. Bab 5 KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang diperoleh dari hasil penelitian serta saran untuk penelitian selanjutnya.

