

BAB I

PENDAHULUAN

1.1 Latar Belakang

Interaksi sehari-hari antarmanusia menunjukkan bahwa emosi menjadi peran penting dalam komunikasi dan pengambilan keputusan. Kemampuan untuk mengenali emosi dan meresponsnya dengan tepat dapat meningkatkan interaksi sosial dan profesional (Kamal, 2025). Emosi adalah reaksi seseorang terhadap suatu situasi dan dapat diekspresikan dalam bentuk kebahagiaan, kesedihan, atau rasa takut (Aripin dkk., 2023). Di era digital, media sosial telah menjadi salah satu *platform* yang menghasilkan data teks dalam jumlah sangat besar setiap hari (Almalki, 2025). Melalui media sosial, pengguna dapat saling berkomunikasi, berbagi informasi, dan bahkan menyampaikan pendapat mereka melalui teks, gambar, dan video (Ishak, 2023). Kekayaan data ini membuka peluang yang sangat besar bagi pengembangan sistem berbasis teks untuk analisis emosi.

Klasifikasi emosi berbasis teks multibahasa merupakan suatu tantangan karena setiap bahasa memiliki struktur dan cara tersendiri dalam mengekspresikan emosi. Perbedaan-perbedaan ini menyulitkan model untuk secara konsisten mengenali dan menggeneralisasi *cross-lingual* emosi (Raharko & Yamasari, 2024). Topik ini relevan dengan beberapa tujuan *Sustainable Development Goals* (SDGs), termasuk tujuan 3 (Kehidupan Sehat dan Sejahtera), tujuan 4 (Pendidikan Berkualitas), dan tujuan 5 (Kesetaraan Gender). Selain itu, penelitian ini selaras dengan inisiatif “Asta Cita” pemerintah (tujuan 4) untuk memperkuat sumber daya manusia, *sains*,

teknologi, pendidikan dan kesetaraan gender, serta berkontribusi pada peningkatan kualitas pembelajaran dan komunikasi digital yang inklusif (Tlogopayung, 2024).

Perkembangan metode *deep learning* berbasis Transformer, seperti *Bidirectional Encoder Representations from Transformers* (BERT), *Robustly Optimized BERT Pre-training Approach* (RoBERTa), *Multilingual-BERT* (mBERT), dan *cross-lingual language model robustly optimized bidirectional encoder representations from transformers* (XLM-RoBERTa) telah secara signifikan meningkatkan kinerja pada tugas-tugas *Natural Language Processing* (NLP) serta representasi kontekstual teks yang efektif di berbagai tugas multibahasa, termasuk klasifikasi emosi (Almalki, 2025) (Abdul Aziz dkk., 2022). Namun, tugas ini tetap memiliki tingkat kompleksitas yang tinggi, karena teks informal sering kali mengandung makna implisit, bahasa kiasan, dan konteks yang ambigu (Alqarni dkk., 2025).

Beberapa penelitian sebelumnya menunjukkan bahwa klasifikasi emosi multibahasa menggunakan XLM-RoBERTa-Large mencapai skor *macro-F1* hingga 0,84. Namun kinerjanya terus menurun pada bahasa-bahasa dengan sumber daya terbatas (*low-resource languages*) (Hosseinzadeh dkk., 2025). XLM-RoBERTa, seperti model tradisional seperti Naïve Bayes, SVM, dan TF-IDF, diterapkan pada teks bahasa Indonesia untuk mengklasifikasikan kategori emosi utama dan emosi gabungan dengan akurasi hingga 95,56%. Namun, penelitian lebih lanjut diperlukan terkait variasi yang lebih kompleks dalam pola semantik (Aripin dkk., 2023). Studi lain menunjukkan bahwa algoritma mBERT digunakan untuk pengenalan emosi pada teks-teks bahasa Hindi dengan akurasi sebesar 91,84%.

Namun, karena jumlah kelas emosi masih terbatas, pengembangan klasifikasi emosi secara *cross-lingual* dengan kategori emosi yang lebih banyak diperlukan (T. Kumar dkk., 2023).

Model mBERT dan XLM-RoBERTa dapat meningkatkan kinerja klasifikasi teks *multilingual* (Aripin dkk., 2023) (A.Akshaya dkk., 2025). Penelitian sebelumnya menunjukkan bahwa klasifikasi emosi pada teks *multilingual* telah mengalami kemajuan signifikan dengan menggunakan XLM-RoBERTa (Magistra Apti Paramarta dkk., 2025). Namun, upaya penelitian ini terus menghadapi tantangan, seperti penurunan kinerja pada bahasa dengan data terbatas dan kesulitan menangkap emosi yang lebih halus *cross-lingual*.

Kompleksitas multibahasa dalam klasifikasi emosi lintas bahasa muncul akibat variasi pola semantik dan sintaksis di antara berbagai bahasa (Aripin dkk., 2023). Model seperti mBERT dan XLM-RoBERTa sering kali cenderung ke arah orientasi semantik yang lebih kuat, yang menyebabkan salah tafsir emosi dalam bahasa target karena perbedaan budaya, konsep emosi yang bervariasi, dan struktur sintaksis yang berbeda (Gundam, Marri, dan Mamidi, 2025). Perubahan sintaksis dalam urutan kata, seperti dari “Ibu makan ikan” menjadi “Ikan makan ibu,” mengubah logika struktural dan menyulitkan model untuk memetakan hubungan yang dihasilkan antara kata-kata. Secara semantik, kompleksitas muncul ketika kata yang sama memiliki makna emosional yang berbeda dalam berbagai konteks budaya. Tanpa *fine-tuning*, model hanya mengenali makna asli dan gagal menangkap makna emosional, sehingga menghasilkan kinerja yang kurang optimal dalam klasifikasi emosi dalam bahasa target (Zhang dkk., 2024).

Penelitian ini mengkaji kemampuan model berbasis BERT, khususnya mBERT dan XLM-RoBERTa, dalam menangani kompleksitas kumpulan data multibahasa. Kompleksitas tersebut mencakup perbedaan struktur linguistik, gaya penulisan, dan ekspresi emosi antarbahasa, yang berpotensi mempengaruhi kinerja model. Penelitian ini juga membandingkan kinerja model multibahasa sebelum dan sesudah penyempurnaan. Penelitian ini diharapkan dapat memberikan wawasan mengenai keefektifan model berbasis BERT dalam klasifikasi emosi multibahasa.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, penelitian ini dapat dirumuskan masalah berikut ini:

1. Bagaimana menganalisis kompleksitas multibahasa yang memiliki variasi pola semantik pada klasifikasi emosi *cross-lingual*?
2. Bagaimana mengevaluasi performa model multibahasa sebelum dan setelah dilakukan *fine-tuning* menggunakan model mBERT dan XLM-RoBERTa dalam memperoleh klasifikasi emosi yang optimal pada beberapa bahasa secara *cross-lingual*?

1.3 Tujuan Penelitian

Sesuai pada uraian masalah yang telah dirumuskan, didapat tujuan penelitian ini sebagai berikut:

1. Menganalisis model dalam kompleksitas teks multibahasa yang memiliki variasi pola semantik dalam klasifikasi emosi *cross-lingual* dengan memanfaatkan model bahasa multibahasa berbasis BERT.

2. Mengevaluasi perbedaan performa model multibahasa sebelum dan setelah dilakukan *fine-tuning* menggunakan model mBERT dan XLM-RoBERTa dalam memperoleh klasifikasi emosi yang optimal pada beberapa bahasa secara *cross-lingual*.

1.4 Manfaat Penelitian

Adapun manfaat yang diharapkan pada penelitian ini, yaitu:

1. Memberikan kontribusi pada pengembangan teknologi klasifikasi emosi berbasis teks multibahasa.
2. Menambah pemahaman tentang kemampuan model *Transformer*, khususnya mBERT dan XLM-RoBERTa dalam menangani perbedaan struktur bahasa dan emosional pada antar bahasa.
3. Memberikan wawasan bagi pengembangan solusi berbasis AI yang mendukung interaksi sosial, layanan publik, dan *platform* digital multibahasa.

1.5 Batasan Masalah

1. Penelitian ini hanya fokus pada pengembangan dan evaluasi model klasifikasi emosi, tanpa membangun sistem atau aplikasi.
2. Dataset yang digunakan dalam penelitian ini hanya mencakup empat bahasa, yaitu bahasa Inggris, bahasa Arab, Spanyol, dan Indonesia.
3. Evaluasi dilakukan pada dataset multibahasa dengan cakupan enam label emosi utama yaitu *anger*, *disgust*, *fear*, *joy*, *sadness*, dan *surprise*.
4. Penelitian ini melakukan klasifikasi pada *single-label*, setiap data hanya dapat memiliki satu label emosi.