

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Bahasa menjadi sarana utama bagi manusia untuk menyampaikan pikiran dan perasaannya, baik secara verbal melalui tulisan atau ucapan, maupun non-verbal melalui gerakan tubuh (Sarasati & Nurvia, 2021; Wibawa & Pramatha, 2023). Di era digital, media sosial menjadi sarana dalam mengekspresikan emosi, didukung oleh lebih dari 221 juta pengguna internet di Indonesia pada 2024 dan lebih dari 5 miliar pengguna media sosial global (APJII, 2024; Simon, 2024). Dalam aspek psikologi, analisis emosi di media sosial penting untuk memahami kecerdasan emosional dan mengenali kondisi emosional individu serta mengidentifikasi gangguan seperti depresi dan kecemasan (Shankar & Tewari, 2021). Melalui analisis emosi berbasis teks di media sosial, deteksi dini dan pemantauan kondisi emosional masyarakat dapat dilakukan secara lebih efektif.

Analisis emosi berbasis teks ini menghadapi tantangan dengan adanya fenomena bilingualisme dan penggunaan emoji yang meluas di media sosial (Rohayati, 2023). Media sosial diisi oleh pengguna dari latar belakang linguistik yang beragam, di mana fenomena bilingualisme seperti alih kode dan campur kode menjadi umum (Hayati & Anindhita, 2024).

Bahasa Indonesia sendiri dituturkan oleh lebih dari 200 juta orang, sementara Bahasa Inggris berfungsi sebagai *lingua franca* global (Koto et al., 2020; Macías Lóor et al., 2024). Kombinasi keduanya menciptakan peluang dalam menganalisis emosi berbasis teks di media sosial. Penggunaan Bahasa Indonesia dan Bahasa

Inggris dalam penelitian ini didasarkan pada peran keduanya yang mewakili komunikasi digital di tingkat nasional maupun global.

Komunikasi digital turut diperkaya dengan penggunaan emoji. Emoji berfungsi sebagai penanda emosi tambahan dan dapat meningkatkan akurasi dalam analisis sentimen jika ditangani dengan tepat (Shalahuddin & Fajrianti, 2024; Yoo & Rayz, 2021). Tantangan dalam integrasi emoji muncul dari perbedaan interpretasi serta ketidaksesuaian makna antara emoji dan teks (Shoeb & de Melo, 2020). Emoji yang mendukung teks dapat mempercepat pemrosesan emosi, sedangkan yang bertentangan justru menghambatnya. Penentuan emosi tetap bergantung pada teks sebagai komponen utama (Dai et al., 2024). Agar integrasi emoji dalam analisis emosi berjalan optimal, dibutuhkan pendekatan yang cermat.

Penelitian sebelumnya menunjukkan bahwa analisis emosi berbasis *XLNet*, dengan *fine-tuning* pada *dropout* dan *learning rate* mencapai akurasi 95,56%, tetapi masih terbatas dalam keberagaman dataset dan pemrosesan emoji (Aripin et al., 2023). *XLNet* diterapkan pada data *bilingual* judul berita menghasilkan akurasi 93,3%, tetapi performa model cenderung lebih akurat pada Bahasa Inggris. Penelitian tersebut juga menyoroti pentingnya eksplorasi terhadap *hyperparameter*, khususnya dalam konteks multibahasa (Wiciaputra et al., 2021).

Kedua penelitian tersebut menunjukkan potensi model *XLNet*, tetapi belum ada penelitian yang secara bersamaan mengeksplorasi pengaruh *hyperparameter*, keberadaan emoji, dan data *bilingual* dalam analisis emosi berbasis *XLNet*. Ketiadaan pendekatan terintegrasi ini menciptakan kesenjangan penting, karena tanpa mempertimbangkan ketiga aspek tersebut,

model berisiko kehilangan konteks emosional yang kompleks dalam komunikasi digital multibahasa.

Penelitian ini bertujuan untuk mengatasi keterbatasan dari penelitian sebelumnya seperti, eksplorasi *hyperparameter*, penggunaan data *bilingual*, dan pengelolaan emoji dengan mengimplementasikan *XLNet* untuk analisis emosi berdasarkan teks dua bahasa, Indonesia dan Inggris pada postingan media sosial. Model ini dipilih karena kemampuannya menangani lebih dari 100 bahasa (Conneau et al., 2020).

Penelitian ini berfokus pada eksplorasi terhadap *hyperparameter*, khususnya jumlah *epoch*, untuk mengevaluasi pengaruhnya terhadap performa model. Konfigurasi awal menggunakan *learning rate*  $2e-5$  dan *epoch* 10, serta mengeksplorasi nilai lebih rendah seperti 3 dan 5 (Afaq & Rao, 2020; Marthen & Novita, 2024). Teknik *early stopping* digunakan untuk mencegah *overfitting* dan menghentikan pelatihan secara otomatis jika akurasi validasi tidak meningkat selama beberapa iterasi (Hussein & Shareef, 2024). Penelitian ini juga mencakup analisis pada data teks yang mengandung emoji dan yang tidak mengandung emoji sehingga dapat menghasilkan klasifikasi emosi yang lebih informatif. Dengan menggunakan pendekatan ini, penelitian ini diharapkan dapat memberikan solusi yang adaptif untuk kebutuhan analisis emosi berbasis teks di era digital.

## 1.2 Rumusan Masalah

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

1. Bagaimana penerapan model *XLM-RoBERTa* dalam klasifikasi emosi berbasis teks *bilingual* serta pengaruh representasi emoji terhadap akurasi?
2. Bagaimana penerapan konfigurasi *hyperparameter* pada model *XLM-RoBERTa* mempengaruhi performa klasifikasi emosi teks?

### 1.3 Tujuan Penelitian

Dari uraian masalah yang sudah dirumuskan, didapat tujuan penelitian sebagai berikut:

1. Menganalisis penerapan model *XLM-RoBERTa* dalam klasifikasi emosi berbasis teks *bilingual* serta pengaruh keberadaan emoji terhadap akurasi dan performa model.
2. Menganalisis pengaruh penerapan konfigurasi *hyperparameter* pada model *XLM-RoBERTa* terhadap performa klasifikasi emosi teks.

### 1.4 Manfaat Penelitian

Adapun beberapa manfaat yang didapatkan dari penelitian ini, sebagai berikut:

1. Penelitian ini berkontribusi dalam meningkatkan pemahaman tentang penerapan model *XLM-RoBERTa* untuk klasifikasi emosi dalam teks *bilingual* serta mengoptimalkan performanya, termasuk analisis terhadap pengaruh emoji, guna mendukung sistem analisis emosi yang lebih akurat.
2. Bagi pengembangan model *transformer*, penelitian ini memberikan wawasan mengenai optimasi konfigurasi *hyperparameter* pada *XLM-RoBERTa* untuk klasifikasi emosi, sehingga dapat dijadikan referensi untuk penelitian lebih lanjut.

## 1.5 Batasan Masalah

Berikut merupakan batasan masalah yang terdapat pada penelitian ini:

1. Dataset yang digunakan berasal dari dataset publik yaitu Kaggle dan Github yang mencakup data postingan di media sosial.
2. Dataset tersebut terdiri dari dua bahasa, yaitu bahasa Indonesia dan bahasa Inggris, yang merepresentasikan analisis emosi dalam teks *bilingual*.
3. Penelitian berfokus pada empat kategori emosi, yaitu *joy* (kebahagiaan), *sadness* (kesedihan), *anger* (kemarahan), dan *fear* (ketakutan).
4. Eksplorasi *hyperparameter* dalam penelitian ini hanya terbatas pada variasi jumlah *epoch*.
5. Implementasi model hanya terbatas pada eksperimen berbasis perangkat lunak tanpa mencakup penerapannya dalam aplikasi atau sistem nyata.