BAB I
PENDAHULUAN

Bab ini membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang digunakan dalam dokumen penelitian yang berjudul Evaluasi Komparatif Antara Model BERT-BiLSTM dan BERT-BiGRU untuk Analisis Sentimen pada Dataset *Movie Review*.

1.1 Latar Belakang

Perkembangan teknologi informasi telah meningkatkan kemudahan individu dalam mengekspresikan opini, memberikan umpan balik, serta membagikan perasaan melalui berbagai *platform* daring, khususnya *platform* ulasan yang memungkinkan pengguna menyampaikan opini dan evaluasi secara tekstual. Ulasan film berperan sebagai sarana yang krusial dalam mengevaluasi kinerja sebuah film. Meskipun pemberian penilaian numerik atau bintang menyediakan evaluasi kuantitatif terhadap keberhasilan atau kegagalan suatu film, kumpulan ulasan film memungkinkan eksplorasi kualitatif terhadap berbagai aspek film seperti alur cerita, akting, sinematografi, dan kualitas produksi. Dengan demikian, ulasan film dapat memberikan gambaran mengenai kelebihan dan kekurangan sebuah film serta tingkat kepuasan audiens secara lebih komprehensif.

Menurut Bordoloi dan Biswas (2023), analisis sentimen, yang juga dikenal sebagai *opinion mining*, merupakan bidang dalam *Natural Language Processing* (NLP) yang berfokus pada pemrosesan komputasional terhadap opini, sentimen, dan subjektivitas yang diekspresikan dalam teks. Analisis sentimen berkembang dari kebutuhan untuk menganalisis teks evaluatif secara otomatis, khususnya pada data ulasan daring yang mengandung penilaian, sikap, dan persepsi pengguna terhadap suatu entitas, seperti produk, layanan, atau film. Dalam konteks ulasan film, analisis sentimen bertujuan untuk mengidentifikasi kecenderungan evaluasi audiens terhadap suatu film berdasarkan opini yang mereka tuliskan, yang sulit dilakukan secara manual akibat volume data yang besar.

Ketidakmampuan menganalisis sentimen ulasan film secara akurat berdampak signifikan terhadap berbagai pemangku kepentingan. Bagi produser dan distributor film, kesalahan interpretasi sentimen audiens dapat mengakibatkan keputusan strategis yang kurang tepat dalam pemasaran dan distribusi film, sehingga berpotensi menyebabkan

kerugian finansial (Asur & Huberman, 2010). Sementara itu, bagi konsumen, keberadaan sistem rekomendasi yang akurat berdasarkan sentimen dapat membantu konsumen dalam mengambil keputusan secara lebih cepat dan terinformasi, sehingga ketiadaannya mengakibatkan pemborosan waktu dan biaya (Zhang & Chen, 2020).

Analisis sentimen sering diformulasikan sebagai permasalahan klasifikasi polaritas, khususnya *binary classification*, yaitu pengelompokan opini ke dalam dua kategori sentimen yang saling berlawanan, positif dan negatif. Pendekatan ini relevan pada ulasan film karena dapat merepresentasikan evaluasi audiens secara langsung. *Binary classification* pada ulasan film telah direalisasikan menggunakan berbagai arsitektur *deep learning*, dengan *Movie Review* (MR) dari Pang dan Lee (2005), sebagai salah satu dataset *benchmark* yang banyak digunakan dalam penelitian analisis sentimen ulasan film. Atandoh dkk. (2021) mengembangkan model GloVe-CNN-BiLSTM pada dataset MR dan mencapai akurasi tertinggi sebesar 84,4%, namun penggunaan GloVe sebagai *embedding* statis membuat model belum mampu memahami konteks kata secara optimal.

Perkembangan model berbasis *transformer* kemudian memberikan peluang untuk mengatasi keterbatasan representasi kata statis. BERT (*Bidirectional Encoder Representations from Transformers*) merupakan model bahasa *pretrained* yang mampu menghasilkan representasi kontekstual melalui pemodelan dua arah pada level token, sehingga makna kata dapat dipahami berdasarkan konteks kalimatnya (Devlin dkk., 2018). Keunggulan ini terlihat pada penelitian Shen dan Liu (2021), di mana BERT-BiLSTM memperoleh akurasi 86,83% pada dataset ulasan film SST, lebih tinggi dibandingkan Word2Vec-BiLSTM sebesar 82,59%. Temuan tersebut menunjukkan bahwa penggunaan BERT sebagai representasi kata kontekstual mampu meningkatkan performa klasifikasi dibandingkan *embedding* statis seperti Word2Vec.

BERT juga banyak dikombinasikan dengan lapisan sekuensial untuk meningkatkan kemampuan model dalam menangkap pola urutan pada teks selain digunakan sebagai model bahasa secara mandiri. Pada domain ulasan film, Wang dkk. (2021) menunjukkan bahwa model RoBERTa-BiLSTM-TCN mencapai akurasi 89,7% pada dataset MR dan meningkat 6,5% dibandingkan CNN-BiLSTM. Di sisi lain, hasil penelitian Talaat (2023) menunjukkan bahwa kombinasi RoBERTa dengan BiGRU dapat memperoleh akurasi 86% pada dataset Airlines, sedikit lebih tinggi dibandingkan RoBERTa-BiLSTM sebesar 85,66%. Temuan

tersebut menunjukkan bahwa BiGRU juga berpotensi menjadi alternatif lapisan sekuensial yang kompetitif dalam model *hybrid* berbasis *transformer.*

Perbandingan antara BiLSTM dan BiGRU juga ditunjukkan oleh Zhou dan Bian (2019) pada ulasan film Douban, di mana BiGRU memperoleh akurasi 88,56%, sedikit lebih tinggi dibandingkan BiLSTM sebesar 88,23%. Namun, penelitian tersebut belum menggunakan representasi kontekstual BERT. Dengan demikian, meskipun penelitian terdahulu menunjukkan bahwa BERT, BiLSTM, dan BiGRU memiliki potensi dalam analisis sentimen, variasi arsitektur model, pendekatan pemrosesan, dataset, dan konfigurasi eksperimen menyebabkan hasil penelitian tersebut belum memberikan dasar perbandingan langsung yang setara antara BERT-BiLSTM dan BERT-BiGRU pada dataset ulasan film yang sama.

Kondisi tersebut menunjukkan perlunya kajian yang secara khusus membandingkan BERT-BiLSTM dan BERT-BiGRU dalam kerangka *fine-tuned* BERT pada dataset ulasan film yang sama. Pemilihan kedua arsitektur tersebut didasarkan pada perbedaan karakteristik BiLSTM dan BiGRU dalam memproses representasi kontekstual BERT. BiLSTM memiliki mekanisme *gating* yang lebih kompleks dengan *cell state* untuk mempertahankan informasi pada sekuens panjang, sedangkan BiGRU memiliki struktur yang lebih sederhana melalui *update gate* dan *reset gate*, tetapi tetap mampu menangkap dependensi (Chung dkk., 2014; Yu dkk., 2019). Perbedaan karakteristik tersebut berpotensi menghasilkan performa klasifikasi yang berbeda, baik dari sisi akurasi maupun *F1-score.*

Performa model *deep learning* tidak hanya ditentukan oleh arsitektur yang digunakan, tetapi juga dipengaruhi oleh konfigurasi *hyperparameter*. Pemilihan *hyperparameter* berpengaruh terhadap kemampuan model dalam mencapai konvergensi yang optimal. Oleh karena itu, penelitian ini membandingkan BERT-BiLSTM dan BERT-BiGRU secara empiris pada dataset MR dengan menerapkan *hyperparameter tuning*, guna mengidentifikasi arsitektur yang lebih efektif serta kombinasi *hyperparameter* yang menghasilkan performa optimal pada masing-masing model.

Keberhasilan solusi yang diusulkan diharapkan memberikan manfaat bagi berbagai pihak. Secara praktis, model yang akurat dapat membantu industri film dalam mengambil keputusan strategis berbasis data sentimen audiens yang lebih reliabel, meningkatkan efektivitas kampanye pemasaran, dan mengoptimalkan strategi rilis film sebagaimana telah

ditunjukkan oleh Asur dan Huberman (2010). Bagi *platform* ulasan dan sistem rekomendasi, model ini dapat meningkatkan kualitas layanan, sejalan dengan temuan Zhang dan Chen (2020) tentang peran sistem rekomendasi dalam efisiensi pengambilan keputusan konsumen.

1.2 Rumusan Masalah

Permasalahan utama dalam penelitian ini adalah belum diketahuinya efektivitas model BERT-BiLSTM dan BERT-BiGRU dalam melakukan analisis sentimen ulasan film pada dataset yang sama. Permasalahan tersebut dirinci ke dalam beberapa pertanyaan penelitian sebagai berikut:

1. Bagaimanakah perbandingan performa antara model BERT-BiLSTM dan BERT-BiGRU dalam melakukan analisis sentimen menggunakan pendekatan *binary classification* ulasan film?
2. Bagaimana pengaruh *hyperparameter tuning* terhadap nilai akurasi dan *F1-score* kedua model tersebut?
3. Manakah kombinasi *hyperparameter* yang menghasilkan performa paling optimal ditinjau dari metrik evaluasi?

1.3 Tujuan dan Manfaat

Beberapa tujuan ditetapkan dalam penelitian ini berdasarkan rumusan masalah yang telah diuraikan, yaitu:

1. Mengetahui perbandingan performa antara model BERT-BiLSTM dan BERT-BiGRU pada tugas analisis sentimen.
2. Mengetahui pengaruh variasi *hyperparameter tuning* terhadap metrik akurasi dan *F1-score* model.
3. Mendapatkan kombinasi *hyperparameter* terbaik yang memberikan performa optimal pada masing-masing model.

Manfaat yang diharapkan dari penelitian ini adalah memberikan kontribusi pada pengembangan literatur mengenai penggabungan model berbasis *transformer*, yaitu BERT, dengan arsitektur RNN, seperti BiLSTM maupun BiGRU dalam pemrosesan bahasa alami. Secara praktis, penelitian ini diharapkan dapat membantu industri film dalam mengambil

keputusan strategis berbasis analisis sentimen audiens, serta meningkatkan kualitas sistem rekomendasi film berbasis teks ulasan.

1.4 Ruang Lingkup

Penelitian dalam skripsi ini memerlukan adanya ruang lingkup agar penelitian yang dilakukan lebih terarah dan mencapai sasaran yang diharapkan. Ruang lingkup dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah *Movie Review Dataset* (*Sentence Polarity Dataset* v1.0) yang diperkenalkan oleh Pang dan Lee (2005). Dataset ini merupakan dataset ulasan film pada tingkat kalimat (*sentence-level*) yang terdiri dari 10.662 kalimat, dengan distribusi kelas yang seimbang, yaitu 5.331 sentimen positif dan 5.331 sentimen negatif.
2. Penelitian ini berfokus pada analisis sentimen dengan pendekatan *binary classification* untuk membedakan polaritas sentimen positif dan negatif.
3. Arsitektur dasar yang digunakan adalah model *pretrained* BERT (*Bidirectional Encoder Representations from Transformers*) varian *bert-base-uncased*. Model BERT akan dikombinasikan dengan dua varian arsitektur *Bidirectional Recurrent Neural Network* (RNN), yaitu *Bidirectional Long Short-Term Memory* (BiLSTM) dan *Bidirectional Gated Recurrent Unit* (BiGRU).
4. Performa model dievaluasi berdasarkan nilai *Accuracy, Precision, Recall*, dan *F1-score*.

1.5 Sistematika Penulisan

Gambaran mengenai pembahasan penyusunan laporan skripsi ini secara urut dan jelas, maka dibentuk sistematika sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah, rumusan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan dari penelitian dengan judul Evaluasi Komparatif Antara Model BERT-BiLSTM dan BERT-BiGRU untuk Analisis Sentimen pada Dataset *Movie Review*.

BAB II TINJAUAN PUSTAKA

Bab ini berisi tinjauan pustaka dan dasar teori yang digunakan dalam skripsi berjudul Evaluasi Komparatif Antara Model BERT-BiLSTM dan BERT-

BiGRU untuk Analisis Sentimen pada Dataset *Movie Review*. Bab ini terdiri atas beberapa subbab, antara lain *State of The Art*, *Natural Language Processing* (NLP), analisis sentimen, prapemrosesan teks, arsitektur *Transformer*, model BERT beserta mekanisme *pretraining* dan *fine-tuning*-nya, *word embedding* BERT, *Recurrent Neural Network* (RNN) beserta variannya yaitu BiLSTM dan BiGRU, model hibrida BERT-BiLSTM dan BERT-BiGRU, *classification layer*, *Stratified K-Fold Cross Validation*, *hyperparameter tuning*, *Cross Entropy Loss*, serta metrik evaluasi.

BAB III METODOLOGI PENELITIAN

Bab ini berisi rancangan penyelesaian mengenai penelitian dan tahap-tahap penelitian yang dilakukan dalam skripsi Evaluasi Komparatif Antara Model BERT-BiLSTM dan BERT-BiGRU untuk Analisis Sentimen pada Dataset *Movie Review*. Gambaran umum tahapan penelitian ini adalah pengumpulan data, eksplorasi dan pemahaman data, pembagian data, prapemrosesan data, pelatihan dan validasi model menggunakan *Stratified* 5-*Fold Cross Validation* dengan variasi *hyperparameter* pada arsitektur BERT-BiLSTM dan BERT-BiGRU, seleksi konfigurasi *hyperparameter* terbaik, *final training*, serta pengujian dan evaluasi model.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi hasil dan analisis yang dilakukan sesuai metodologi yang digunakan dalam skripsi Evaluasi Komparatif Antara Model BERT-BiLSTM dan BERT-BiGRU untuk Analisis Sentimen pada Dataset *Movie Review*. Bab ini membahas tentang lingkungan dan perangkat penelitian, skenario eksperimen, hasil pelatihan dan validasi menggunakan 5-*Fold Cross Validation* pada kedua model, hasil *final training*, hasil pengujian model pada data uji, serta perbandingan performa antara model BERT-BiLSTM dan BERT-BiGRU.

BAB V PENUTUP

Bab ini berisi kesimpulan dari bab-bab yang telah dijelaskan sebelumnya dan saran untuk penelitian lebih lanjut.