

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sistem peringkasan teks otomatis (*automatic text summarization*) merupakan salah satu tugas penting dalam bidang *Natural Language Processing* (NLP) yang bertujuan menghasilkan ringkasan singkat dari dokumen panjang tanpa menghilangkan informasi utama [1]. Seiring meningkatnya jumlah dokumen digital, pengembangan metode peringkasan otomatis menjadi area penelitian yang terus berkembang untuk membantu proses ekstraksi informasi secara efisien.

Tingginya kebutuhan akan pengolahan informasi yang cepat menjadikan sistem peringkasan teks otomatis (*Automatic Text Summarization*) sebagai solusi yang paling efektif. Secara umum, sistem ini terbagi ke dalam dua paradigma utama, yaitu *extractive* dan *abstractive*. Pendekatan *abstractive* dinilai mampu menghasilkan ringkasan yang lebih koheren dan alami karena bekerja dengan cara memahami konteks lalu menyusun ulang kalimat baru (parafrase) layaknya manusia [2]. Hal ini berbeda dengan pendekatan *extractive* yang cenderung kaku karena hanya menyeleksi dan menyalin potongan kalimat yang dianggap penting dari dokumen asli tanpa mengubah struktur bahasanya [2] [3].

Sistem peringkasan otomatis modern kini telah mengadopsi teknologi *Transformer* sebagai fondasi utama untuk memahami kompleksitas bahasa manusia. Secara definisi, *Transformer* adalah model jaringan saraf yang dirancang untuk memproses data teks secara paralel, berbeda dengan pendekatan lama yang memproses kata demi kata [4]. Penggunaan *Transformer* menjadi sangat krusial dalam tugas peringkasan karena arsitektur ini mampu mengatasi masalah hilangnya informasi pada dokumen panjang melalui mekanisme perhatian diri yang efisien dan representasi kontekstual global [5]. Melalui representasi data yang lebih kaya, *Transformer* dapat memetakan intisari dari sebuah wacana dengan presisi tinggi, menjadikannya metode yang paling andal untuk menangani tugas-tugas pemahaman teks yang rumit.

Penerapan arsitektur *Transformer* untuk tugas peringkasan umumnya dikonfigurasi menggunakan kerangka kerja *encoder-decoder*. Bagian *encoder* bertugas memahami representasi semantik dari teks sumber, sedangkan *decoder*

berfungsi untuk menghasilkan teks ringkasan baru [6]. Variasi model dalam kerangka ini melahirkan dua pendekatan yang berbeda. Pendekatan pertama adalah BERT2BERT, yang memanfaatkan arsitektur BERT baik sebagai *encoder* maupun *decoder*. Model ini sangat kuat dalam aspek pemahaman konteks (*understanding-oriented*) karena sifat *bidirectional* yang dimilikinya [7]. Pendekatan kedua adalah BERT2GPT, yang menggabungkan *encoder* BERT dengan *decoder* GPT. Berbeda dengan BERT, GPT merupakan model *autoregressive* yang dirancang spesifik untuk tugas generatif (*generation-oriented*) [8]. Kendati demikian, kemampuan generatif yang tinggi ini memiliki *trade-off* tersendiri berupa risiko halusinasi (*hallucination*). Penelitian menunjukkan bahwa terdapat keseimbangan yang harus dicapai antara kreativitas model dalam menyusun kata dengan potensi penyimpangan informasi dari fakta aslinya [9].

Penentuan model dengan performa paling optimal dalam menghasilkan ringkasan berkualitas memerlukan serangkaian instrumen evaluasi yang terukur. Kualitas ringkasan dalam penelitian ini akan diukur menggunakan metrik evaluasi otomatis yang mencakup aspek kesamaan teks dan keterbacaan. Metrik utama yang digunakan adalah *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE) untuk mengukur tingkat tumpang tindih kata antara ringkasan sistem dan referensi manusia [10]. Selain itu, tingkat keterbacaan ringkasan akan diuji menggunakan *Flesch-Kincaid Grade Level* (FKGL) [11] dan *Gunning Fog Index* (GFI) [12] yang disesuaikan untuk teks umum, serta formula Dwiyanto Djoko Pranowo yang dirancang spesifik untuk mengukur tingkat keterbacaan teks berbahasa Indonesia [13].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengevaluasi dan membandingkan performa model BERT2BERT dan BERT2GPT. Fokus utama penelitian terletak pada analisis kemampuan kedua model tersebut dalam menghasilkan ringkasan abstraktif dari artikel berita yang akurat dan alami. Melalui pengujian menggunakan metrik evaluasi otomatis ROUGE dan instrumen keterbacaan yang telah ditetapkan, penelitian ini diharapkan dapat mengidentifikasi arsitektur terbaik yang mampu menyeimbangkan kualitas linguistik dengan akurasi konten, sekaligus memberikan

wawasan empiris mengenai dampak *trade-off* penggunaan *decoder* generatif GPT terhadap risiko halusinasi pada ringkasan berita Bahasa Indonesia.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang tersebut, diperlukan adanya kajian yang lebih mendalam untuk menganalisis penggunaan model generatif dalam *abstractive summarization* Bahasa Indonesia, khususnya melalui penggunaan arsitektur *encoder-decoder* yang berbeda serta evaluasi kualitas ringkasan yang dihasilkan. Oleh karena itu, penelitian ini dirumuskan ke dalam beberapa permasalahan sebagai berikut:

1. Bagaimana perbandingan kualitas konten ringkasan yang dihasilkan oleh model BERT2GPT dan BERT2BERT?
2. Bagaimana tingkat keterbacaan ringkasan abstraktif yang dihasilkan oleh model BERT2GPT dan BERT2BERT?
3. Bagaimana perbandingan tingkat konsistensi faktual (*faithfulness*) dari ringkasan yang dihasilkan oleh model BERT2GPT dan BERT2BERT?

1.3 Batasan Masalah

Untuk menjaga fokus penelitian serta menyesuaikan ruang lingkup kajian dengan keterbatasan sumber daya dan tujuan penelitian, maka penelitian ini diberikan beberapa batasan agar pembahasan yang dilakukan tetap terarah dan tidak melebar. Batasan-batasan tersebut adalah sebagai berikut:

1. Model BERT dan GPT yang digunakan sebagai *encoder* maupun *decoder* adalah model *pre-trained* yang telah tersedia.
2. Dataset yang digunakan adalah Liputan6 dan dibatasi pada dataset berita berbahasa Indonesia (https://github.com/fajri91/sum_liputan6).
3. Evaluasi kualitas konten ringkasan dilakukan menggunakan metrik evaluasi otomatis ROUGE dan *BERTScore*.
4. Evaluasi keterbacaan (*readability*) ringkasan dilakukan menggunakan metrik FKGL, GFI, dan Formula Dwiyanto Djoko Pranowo.
5. Pengukuran konsistensi faktual (*faithfulness*) dilakukan menggunakan pendekatan *LLM-as-a-Judge*.

1.4 Tujuan

Adapun tujuan dari penelitian ini sebagai berikut:

1. Menganalisis dan membandingkan kualitas konten antara ringkasan yang dihasilkan oleh model BERT2GPT dan BERT2BERT.
2. Mengevaluasi dan membandingkan tingkat keterbacaan dari ringkasan yang dihasilkan oleh model BERT2GPT dan model BERT2BERT.
3. Mengukur dan membandingkan tingkat konsistensi faktual (*faithfulness*) dari ringkasan yang dihasilkan oleh model BERT2GPT dan model BERT2BERT.

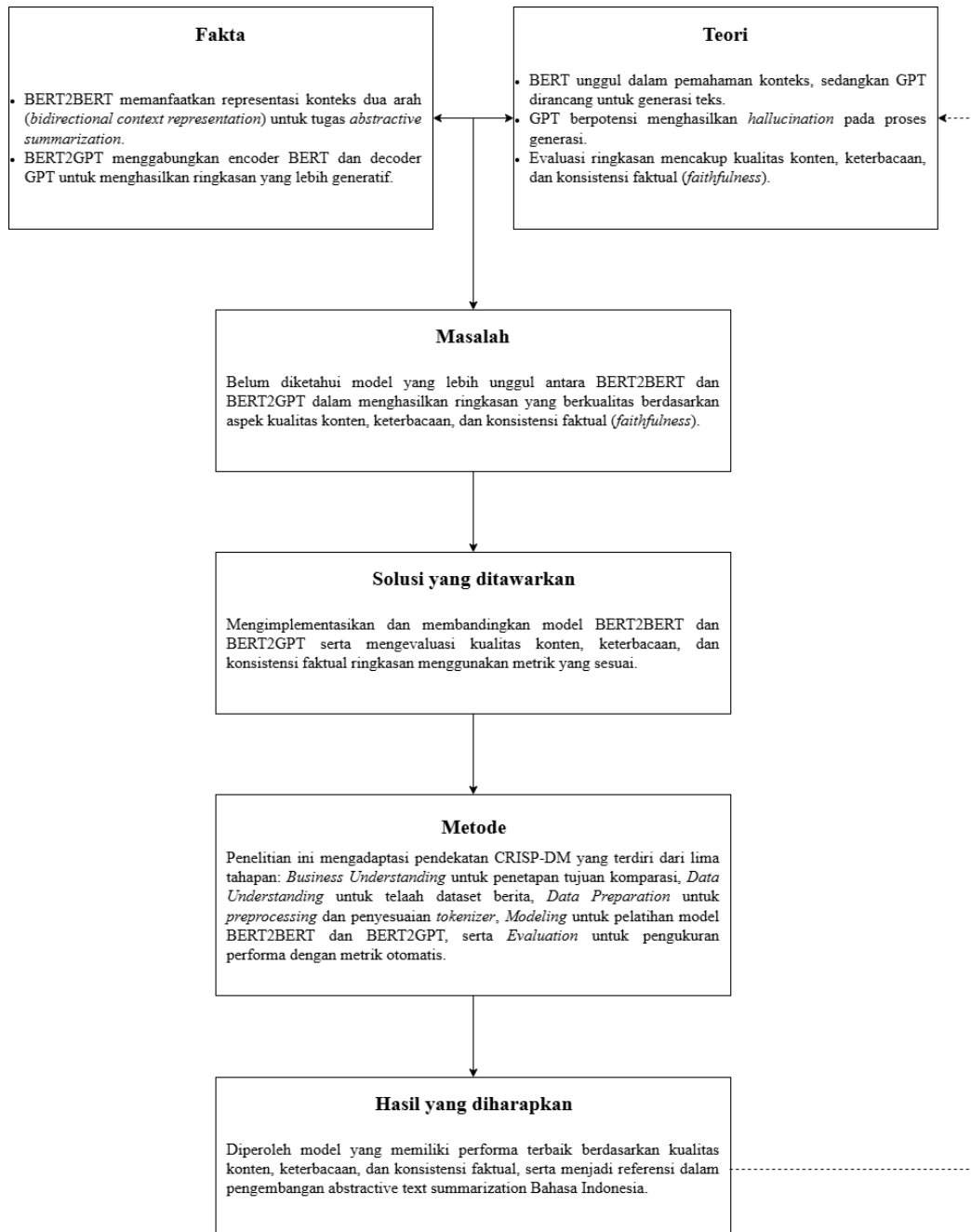
1.5 Manfaat

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan penelitian dan penerapan *abstractive summarization* Bahasa Indonesia. Adapun manfaat yang diharapkan sebagai berikut:

1. Menjadi bahan pertimbangan dalam pemilihan arsitektur model untuk pengembangan model peringkasan teks otomatis Bahasa Indonesia.
2. Menambah kajian ilmiah terkait perbandingan arsitektur *encoder-decoder*, khususnya antara model *decoder* generatif (BERT2GPT) dengan non-generatif (BERT2BERT) pada tugas *abstractive summarization*.

1.6 Kerangka Berpikir

Gambar 1.1 ini adalah kerangka berpikir yang menggambarkan alur penelitian, berdasarkan fakta, literatur, permasalahan, solusi, metodologi yang digunakan, dan hasil yang diharapkan.



Gambar 1.1. Kerangka Berpikir

Penelitian ini berangkat dari perkembangan metode *abstractive text summarization* berbasis *Transformer* yang mampu menghasilkan ringkasan secara lebih natural melalui pemahaman konteks dokumen. Berbagai penelitian menunjukkan bahwa model BERT2BERT memanfaatkan representasi konteks dua arah (*bidirectional contextual representation*) sehingga memiliki kemampuan yang baik dalam memahami hubungan semantik teks. Di sisi lain, BERT2GPT menggabungkan encoder BERT dengan decoder GPT untuk memanfaatkan kemampuan pemahaman konteks dan generasi bahasa pada tugas *abstractive text summarization*. Secara teoretis, BERT unggul dalam memahami konteks, sedangkan GPT dirancang sebagai model generatif yang mampu menghasilkan teks lebih natural, namun memiliki potensi menghasilkan *hallucination*, yaitu informasi yang tidak sepenuhnya didukung oleh teks sumber. Oleh karena itu, kualitas ringkasan perlu dievaluasi secara komprehensif berdasarkan aspek kualitas konten, keterbacaan, dan konsistensi faktual (*faithfulness*).

Berdasarkan kondisi tersebut, permasalahan dalam penelitian ini adalah belum diketahui model yang lebih unggul antara BERT2BERT dan BERT2GPT dalam menghasilkan ringkasan berbahasa Indonesia berdasarkan ketiga aspek evaluasi tersebut. Sebagai solusi, penelitian ini mengimplementasikan dan membandingkan kedua model pada tugas *abstractive text summarization*, kemudian mengevaluasi hasil ringkasan menggunakan metrik ROUGE dan BERTScore untuk kualitas konten, FKGL, GFI, dan Formula Dwiyanto Djoko Pranowo untuk keterbacaan, serta pendekatan *LLM-as-a-Judge* untuk mengukur konsistensi faktual (*faithfulness*). Proses penelitian dilaksanakan menggunakan pendekatan adaptasi dari metodologi CRISP-DM, yang meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, dan *Evaluation*. Melalui pendekatan tersebut, diharapkan diperoleh perbandingan performa yang komprehensif antara BERT2BERT dan BERT2GPT serta rekomendasi arsitektur yang lebih sesuai untuk pengembangan sistem *abstractive text summarization* berbahasa Indonesia.

1.7 Sistematika Penulisan

Bagian sistematika penulisan menjelaskan susunan laporan tugas akhir secara terstruktur dengan memberikan gambaran isi setiap bab serta hubungan antar bab dalam laporan penelitian. Adapun sistematika penulisan laporan tugas akhir ini adalah sebagai berikut.

BAB I PENDAHULUAN

Bab ini berisi gambaran umum mengenai penelitian yang dilakukan. Di dalamnya dijelaskan latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, kerangka pemikiran yang digunakan dalam penelitian, serta sistematika penulisan laporan tugas akhir.

BAB II TINJAUAN PUSTAKA

Bab ini memuat berbagai kajian teori, konsep, dan penelitian terdahulu yang relevan dengan topik penelitian. Teori dan konsep tersebut digunakan sebagai dasar atau landasan dalam melakukan analisis terhadap permasalahan yang dibahas dalam penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metode yang digunakan dalam penelitian, termasuk tahapan-tahapan penelitian, teknik pengumpulan data, serta metode atau pendekatan yang digunakan dalam proses analisis. Penjelasan disajikan secara sistematis agar alur penelitian dapat dipahami dengan jelas.

BAB IV HASIL DAN PEMAPARAN

Bab ini menyajikan hasil yang diperoleh dari proses penelitian yang telah dilakukan. Selain itu, bab ini juga memuat pembahasan mengenai hasil tersebut dengan mengaitkannya dengan teori atau konsep yang telah dijelaskan sebelumnya, sehingga dapat menjawab rumusan masalah yang telah ditetapkan.

BAB V SIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang diperoleh dari hasil penelitian serta pembahasan yang telah dilakukan. Selain itu, disampaikan pula beberapa saran yang dapat menjadi pertimbangan untuk penelitian selanjutnya maupun pengembangan lebih lanjut dari penelitian yang telah dilakukan.