

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Penelitian

Aplikasi pemrosesan bahasa alami dikembangkan menggunakan kecerdasan buatan, yang memungkinkan komputer memiliki kemampuan pemrosesan bahasa alami seperti manusia [1], salah satunya adalah ringkasan teks otomatis. Dengan perkembangan era informasi, kebutuhan untuk pemrosesan dan ringkasan konten semakin meningkat. Oleh karena itu, diperlukan alat ringkasan dokumen yang dapat secara otomatis merangkum informasi dari berbagai sumber. Alat ringkasan dokumen bertujuan untuk menghasilkan ringkasan teks yang relevan dan sesuai untuk dokumen atau set dokumen tertentu [2]. Terdapat dua jenis ringkasan dokumen: abstraktif dan ekstraktif [2]. Metode ringkasan ekstraktif fokus pada identifikasi dan pengumpulan kalimat-kalimat kunci dari teks sumber untuk membentuk ringkasan. Pendekatan ini memiliki keuntungan dari implementasi yang lebih sederhana dan pemanfaatan langsung teks dari dokumen asli [3]. Sementara itu, ringkasan abstraktif menghasilkan kalimat-kalimat baru, sering kali memerlukan model kompleks untuk memparafrasa dan meringkas makna teks asli, yang dapat menghasilkan ringkasan yang lebih fasih dan kurang redundan [4]. Ringkasan ekstraktif secara langsung mengambil kalimat-kalimat bermakna yang telah dipilih dari dokumen asli [5], [6]. Faktanya, ringkasan yang ditulis oleh manusia umumnya bersifat abstraktif. Ringkasan abstraktif memungkinkan komputer untuk menghasilkan ringkasan teks dengan membuat set kalimat baru yang mewakili informasi yang terdapat dalam sumber dengan bentuk penyampaian yang berbeda dari teks asli [2], [7], [8].

Setiap bahasa memiliki keunikan dan karakteristiknya sendiri, termasuk Bahasa Indonesia. Alat peringkasan teks bahasa Indonesia telah mengalami kemajuan signifikan melalui penerapan model transformer, mengatasi tantangan pemrosesan bahasa yang digunakan oleh hampir 200 juta orang namun masih kurang terwakili dalam penelitian NLP [9]. Adanya penelitian tentang pengembangan data set bahasa Indonesia sebagai tolak ukur evaluasi dalam pengembangan alat ringkasan teks otomatis membuka peluang untuk penelitian

lebih lanjut. IndoSum adalah data set yang digunakan sebagai tolak ukur baru dalam ringkasan teks bahasa Indonesia [10]. Kemudian terdapat pula Liputan6 kumpulan data Indonesia skala besar untuk ringkasan teks [1]. Liputan6 dikumpulkan dari data portal berita bahasa Indonesia Liputan6.com selama sepuluh tahun, sehingga terdapat 215.827 dokumen data ringkasan teks [1]. Dalam penelitiannya, juga dikembangkan model ringkasan dokumen tunggal berbasis BERT menggunakan metode ekstraktif dan abstraktif. MultilingualBERT dan IndoBERT digunakan sebagai model yang telah dilatih dalam model ini. Penggunaan IndoBERT menghasilkan nilai evaluasi yang cukup baik untuk penggunaan data set Liputan6 [1]. Selain itu, terdapat juga data set komprehensif, Indonesia Language Evaluation Montage (IndoLEM), yang mencakup tujuh tugas NLP dan delapan sub-data set [10]. Model ringkasan teks ekstraktif lainnya untuk bahasa Indonesia, seperti fine-tuning Sentence Transformers (SBERT), telah menunjukkan peningkatan kinerja dalam menghasilkan ringkasan dokumen atau cuplikan, khususnya menggunakan dokumen tesis Indonesia untuk membangun data set baru untuk tugas ini [11].

Berdasarkan perkembangan penelitian tentang ringkasan teks otomatis untuk bahasa Indonesia, ditemukan bahwa ringkasan multi-dokumen abstraktif dalam penelitian bahasa Indonesia belum dilakukan menggunakan model berbasis Transformer, dan ini adalah peluang besar untuk dieksplorasi. Hal ini disebabkan oleh keterbatasan data set yang menyediakan data ringkasan multi-dokumen [2], [12], [13]. Dalam kasus selain bahasa Indonesia, terdapat penelitian tentang fine-tuning menggunakan BERT Sentence Embedding Model pada data set ringkasan multi-dokumen ekstraktif [9]. Selain itu, terdapat penelitian yang melakukan fine-tuning model Transformer yang telah dilatih untuk ringkasan dokumen tunggal pada data set multi-dokumen dengan menyesuaikan struktur *encoder* dan *decoder* [3], [5]. Penelitian lainnya juga mengadaptasi model Transformer yang dilatih untuk ringkasan dokumen tunggal untuk melakukan tugas ringkasan multi-dokumen menggunakan Decoding Controller [4]. Ketiga penelitian ini menggunakan model yang telah dilatih yang di-fine-tune menggunakan data multi-dokumen. Dari penelitian-penelitian ini, muncul bahwa model ringkasan dokumen tunggal memiliki potensi untuk melakukan tugas ringkasan multi-dokumen. Oleh karena itu, studi ini berkontribusi dengan menggunakan model yang dilatih dengan

data set ringkasan dokumen tunggal dalam tugas ringkasan multi-dokumen bahasa Indonesia dengan model Transformer. Dengan itu, penelitian ini mengambil judul “Penerapan Model Bert2bert Untuk Alat Peringkasan Abstraktif Multi Dokumen Otomatis”

## 1.2 Perumusan Masalah

Dari latar belakang yang telah dipaparkan di atas, maka dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana penerapan model IndoBERT pada proses *encode-decode* untuk peringkasan teks abstraktif multi dokumen?
2. Bagaimana kualitas hasil ringkasan dari penerapan model IndoBERT pada *encoder* dan *decoder* untuk peringkasan teks abstraktif multi-dokumen?

## 1.3 Tujuan Penelitian

Dari rumusan masalah di atas, tujuan penelitian ini adalah sebagai berikut:

1. Menganalisis kinerja model IndoBERT yang diterapkan pada *encoder* dan *decoder* dalam meringkas multi-dokumen.
2. Mengevaluasi kualitas hasil ringkasan dari penerapan model IndoBERT pada *encoder* dan *decoder* untuk peringkasan teks abstraktif multi-dokumen.

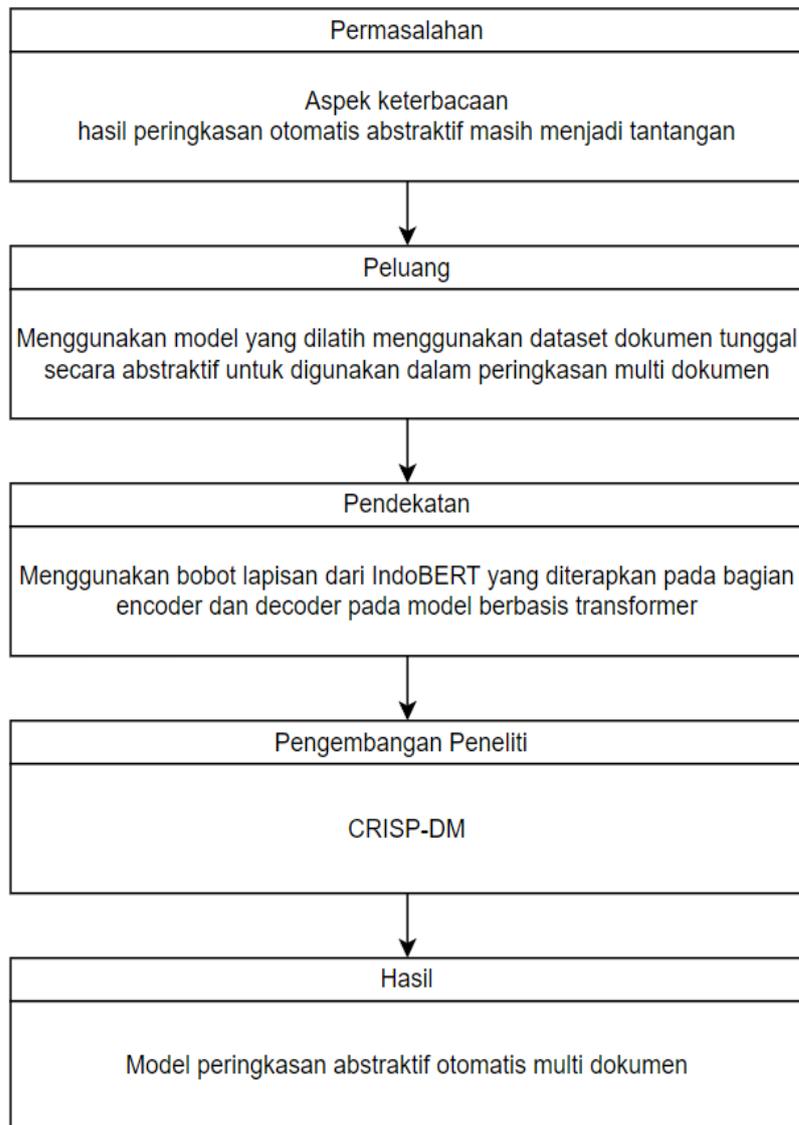
## 1.4 Batasan Masalah Penelitian

Batasan masalah yang diterapkan dalam penelitian ini adalah sebagai berikut:

1. Menerapkan model pretrained IndoBERT dalam mengembangkan model
2. Menggunakan dataset Liputan6 [1].
3. Evaluasi hasil ringkasan menggunakan ROUGE-1, ROUGE-2, ROUGE-N, Dwiyanto Djoko Pranowo, FKGL, dan FGI.
4. Objek penelitian dokumen berbahasa Indonesia.

## 1.5 Kerangka Pemikiran Penelitian

Tabel 1.1 Kerangka Pemikiran



## 1.6 Sistematika Penulisan

Tugas akhir ini disusun dengan sistematika penulisan untuk memberikan gambaran umum. Adapun sistematika penulisannya adalah sebagai berikut:

### **BAB I: Pendahuluan**

Bab I ini memaparkan latar belakang penelitian, perumusan masalah, tujuan penelitian, batasan masalah, kerangka pemikiran, dan sistematika penulisan. Bagian ini bertujuan untuk memberikan konteks dan alasan dilakukannya penelitian serta bagaimana penelitian ini diorganisasikan.

## **BAB II: Kajian Literatur**

Bab II mengulas penelitian-penelitian terdahulu yang relevan dan teori-teori yang digunakan untuk memecahkan masalah penelitian. Bagian ini membantu dalam memahami dasar teoretis dan penelitian sebelumnya yang menjadi landasan bagi penelitian yang dilakukan.

## **BAB III: Metodologi Penelitian**

Bab III menjelaskan metode yang digunakan dalam penelitian, termasuk langkah-langkah dan teknik yang diterapkan. Penjelasan yang detail mengenai prosedur penelitian memastikan bahwa penelitian dapat di replikasi dan hasilnya dapat dipercaya.

## **BAB IV: Hasil dan Pembahasan**

Bab IV memaparkan hasil penelitian serta pembahasan yang mendalam mengenai proses dan temuan yang dicapai. Analisis data dan interpretasi hasil juga disertakan untuk memberikan pemahaman yang komprehensif.

## **BAB V: Diskusi**

Bab V berfokus pada analisis dan interpretasi hasil penelitian. Penulis menjelaskan implikasi hasil, membandingkannya dengan penelitian sebelumnya, mengidentifikasi keterbatasan penelitian, dan memberikan rekomendasi untuk penelitian selanjutnya.

## **BAB VI: Simpulan dan Saran**

Bab VI menyajikan kesimpulan utama dari penelitian yang menjawab rumusan masalah, serta memberikan saran-saran untuk penelitian di masa depan yang bertujuan untuk mengembangkan lebih lanjut topik penelitian ini.