

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dengan munculnya internet mengakibatkan jumlah dokumen yang tersedia dalam bentuk daring meningkat drastis. Dokumen-dokumen tersebut diantaranya adalah berita, situs web, blog, buku, jurnal ilmiah dan media sosial. Dampak dari meningkatnya jumlah dokumen daring adalah untuk mendapatkan informasi yang relevan bagi pengguna internet menjadi semakin sulit karena jumlah informasi yang ada sangat masif [1]. Meringkas teks menjadi salah satu cara yang efektif untuk mendapatkan informasi penting dalam dokumen. Namun terdapat keterbatasan waktu dan tenaga untuk meringkas seluruh teks secara manual. Oleh karena itu ringkasan teks otomatis menjadi topik penelitian yang populer saat ini [2].

Tujuan utama dari sistem ringkasan teks otomatis adalah untuk menghasilkan teks ringkas yang tetap mempertahankan informasi utama dan mengurangi teks repetitif. Ringkasan teks otomatis dapat membantu pembaca untuk mendapatkan informasi utama tanpa harus membaca seluruh isi dokumen. Ringkasan teks otomatis juga menghemat waktu dan tenaga pembaca.

Ringkasan teks otomatis dapat diklasifikasikan menjadi dua yaitu ringkasan teks dokumen tunggal dan dokumen banyak. Ringkasan dokumen tunggal menggunakan satu dokumen teks sebagai masukan, sementara ringkasan dokumen banyak menggunakan banyak dokumen sebagai masukan untuk menghasilkan ringkasan. Pada umumnya sistem ringkasan teks otomatis dibuat menggunakan tiga pendekatan: ekstraktif, abstraktif, atau campuran. Ringkasan ekstraktif akan memilih kalimat terpenting dalam dokumen dan menggunakannya sebagai hasil ringkasan. Ringkasan abstraktif dapat menghasilkan ringkasan berisi informasi utama dengan menggunakan kata dan kalimat yang tidak ada di dokumen asli sebelumnya. Sementara ringkasan campuran (*hybrid*) adalah perpaduan antara ringkasan ekstraktif dan ringkasan abstraktif [3].

Salah satu metode yang dapat diterapkan untuk mendapat ringkasan teks

abstraktif adalah dengan *sequence-to-sequence* (Seq2Seq). Seq2Seq terdiri atas komponen *encoder* untuk menerima masukan dan komponen *decoder* yang menghasilkan keluaran. Algoritma yang pernah digunakan untuk membangun model Seq2Seq antara lain Convolutional Neural Network (CNN) [4], Long Short-Term Memory (LSTM) [5], Gated Recurrent Units (GRUs) [6]. Namun ketiga algoritma tersebut masih memiliki kendala dalam hal performa dan hasil. LSTM dan GRU akan memproses data secara urut satu per satu yang mengakibatkan kecepatannya menurun. LSTM juga kesulitan untuk memproses dokumen dengan teks panjang [7]. LSTM tidak mampu menangkap konteks informasi secara luas saat memproses teks panjang. Transformer dapat digunakan sebagai solusi berbagai kendala dari algoritma sebelumnya. Transformer memproses data secara paralel dan memanfaatkan mekanisme *self-attention* untuk proses *encode-decode*, sehingga menghasilkan performa yang lebih baik dari algoritma sebelumnya [8]. Transformer telah banyak digunakan dalam berbagai tugas *Natural Language processing* (NLP) termasuk ringkasan teks otomatis. Berbagai model transformer untuk ringkasan teks otomatis diantaranya Bidirectional and *Auto-Regressive Transformers* (BART) [9], Pegasus [10], Prophetnet [11], Transformer-XL [12], dan *Text-to-Text Transfer Transformer* (T5) [13].

Penelitian ini bertujuan untuk mengisi kesenjangan penelitian dengan menerapkan metode abstrak Sequence to Sequence berbasis algoritma T5 pada ringkasan teks otomatis dalam bahasa Indonesia. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ringkasan teks otomatis yang dapat digunakan dalam dokumen bahasa Indonesia terutama dengan pendekatan abstraktif.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah dikemukakan, maka bisa dirumuskan beberapa rumusan masalah berikut ini :

- a. Bagaimana penerapan ringkasan teks otomatis menggunakan Seq2Seq berbasis algoritma *Text to Text Transfer Transformer* (T5)?
- b. Bagaimana tingkat akurasi ringkasan teks yang dihasilkan

menggunakan Seq2Seq berbasis algoritma *Text to Text Transfer Transformer (T5)*?

1.3 Batasan Masalah

Beberapa Batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut :

- a. Ringkasan teks otomatis menggunakan Seq2Seq berbasis algoritma *Text to Text Transfer Transformer (T5)*.
- b. Penelitian ini menggunakan 50000 dataset IndoSUM yang berisi teks berita bahasa Indonesia
- c. Penelitian ini menggunakan pengujian ROUGE Score untuk mengukur tingkat akurasi

1.4 Tujuan Penelitian

Adapun tujuan peneliti dalam melakukan penelitian adalah sebagai berikut:

- a. Memanfaatkan metode Seq2Seq berbasis algoritma *Text to Text Transfer Transformer (T5)* untuk digunakan dalam meringkas teks bahasa Indonesia secara otomatis.
- b. Menunjukkan bahwa metode Seq2Seq berbasis algoritma *Text to Text Transfer Transformer (T5)* dapat melakukan peringkasan abstraktif teks bahasa Indonesia secara otomatis.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini antara lain :

- a. Memudahkan pembaca untuk mendapat informasi utama dari teks panjang.
- b. Dapat mengetahui nilai akurasi yang didapat dari penggunaan metode Seq2Seq berbasis algoritma *Text to Text Transfer Transformer (T5)* dalam peringkasan teks otomatis.
- c. Penelitian ini diharapkan dapat menjadi referensi bagi penulis atau pengembang untuk penelitian di masa yang akan datang.

- d. Menambah wawasan serta ilmu pengetahuan dalam pengembangan sistem yang lebih lanjut.

1.6 Sistematika Penulisan

Sesuai dengan petunjuk penulisan laporan skripsi yang berlaku di Universitas Amikom Yogyakarta, sistematika penulisan laporan ini adalah sebagai berikut yang berisi paparan garis besar isi masing - masing bab.

BAB I PENDAHULUAN

Bab pendahuluan berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab tinjauan Pustaka berisi teori atau konsep yang digunakan sebagai landasan berpikir dalam penelitian ini. Tinjauan pustakan diperoleh dari buku, jurnal, dan beberapa penelitian sebelumnya yang terkait dengan topik *Automatic Text Summarization*.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan alur tahap penelitian, alat berupa algoritma, serta perangkat dan bahan berupa data yang akan digunakan dalam penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab hasil dan pembahasan berisi hasil penelitian secara menyeluruh serta pembahasan pada penelitian yang dikaji.

BAB V KESIMPULAN DAN SARAN

Bab kesimpulan dan saran berisi mengenai kesimpulan yang didapat dari penelitian di bab-bab sebelumnya, serta saran yang diharapkan dapat berguna untuk pengembangan penelitian di masa mendatang