

BAB I PENDAHULUAN

1.1 Latar Belakang

Perkembangan era digitalisasi saat ini telah menghasilkan data tekstual dalam jumlah besar yang membuat kesulitan dalam mendapatkan informasi yang penting dengan cepat dan akurat dikarenakan informasi yang berlebihan. Dengan besarnya informasi tekstual, kebutuhan untuk meng-ekstrak informasi yang berharga menjadi penting bagi pembaca agar mendapat pengalaman membaca yang lebih baik dan lebih mudah mencerna informasi yang diberikan. Untungnya, proses ini telah dilakukan secara otomatis selama beberapa dekade oleh metode pencarian informasi..

Peringkasan Teks adalah prosedur yang menghasilkan teks kental dari dokumen sumber. Tujuan dari peringkasan teks adalah untuk mengetahui konsep utama dan menghasilkan teks yang lebih pendek dibandingkan dengan dokumen sumber. Salah satu masalah yang paling menantang di bidang NLP adalah untuk menemukan bagian-bagian penting yang ada dalam dokumen input[1]. Dengan berkembang pesatnya data di internet membuat informasi menjadi berlimpah dan hal tersebut membuat pembaca kesulitan mendapatkan informasi yang relevan. Ketika kita mencari suatu informasi tertentu, banyak data yang akan diambil dan tidak mungkin bagi seseorang untuk membaca semua dengan seksama. Hal ini membuat pertumbuhan pesat pada area Automatic Text Summary (ATS).

Automatic Text Summary (ATS) bertujuan untuk mendapatkan versi dokumen yang dipersingkat dengan memfilter informasi terpenting dari satu atau beberapa sumber teks[2]. Sistem ATS dirancang dengan menerapkan salah satu pendekatan peringkasan teks: ekstraktif, abstrak, atau hibrid. Pendekatan ekstraktif memilih kalimat yang paling penting dari teks masukan dan menggunakannya untuk menghasilkan ringkasan. Pendekatan abstraksi merepresentasikan teks masukan dalam bentuk peralihan kemudian menghasilkan ringkasan dengan kata-kata dan kalimat yang berbeda dari kalimat teks aslinya. Pendekatan hybrid

menggabungkan kedua pendekatan ekstraktif dan abstraktif[3]. Dokumen yang lebih besar atau kumpulan dokumen, sebagian besar akan di ringkas menggunakan pendekatan ekstraktif. Teknik ringkasan ekstraktif terdiri dari memilih kalimat penting, paragraf, dll, dari naskah asli dan menggabungkannya ke dalam bentuk yang lebih pendek. Pentingnya kalimat sangat didasarkan pada statistik dan linguistik fitur kalimat[3]. Pada tahun 2014, menunjukkan bahwa vektor kontinu berbasis saraf model menjanjikan untuk peringkasan teks[4]. Ini menandai dimulainya penyebaran penggunaan model peringkasan teks berbasis jaringan saraf, karena kinerjanya yang unggul dibandingkan dengan teknik tradisional.

Recurrent Neural Network (RNN) digunakan dalam arsitektur Deep Neural Network untuk encoder dan juga decoder pada ekstraktif text summarization[5]. Namun, karena masalah "long-distance dependence"[6], model RNN bisa kehilangan beberapa informasi setelah sejumlah langkah tertentu, sehingga menghasilkan peringkasan teks yang kurang akurat. Long Short Term Memory (LSTM) menjadi perbaikan dari arsitektur Recurring Neural Network (RNN) yang menggunakan mekanisme gating yang terdiri dari input gate, forget gate, dan output gate[7]. Gate ini membantu menentukan apakah data dalam state sebelumnya harus dipertahankan atau dilupakan dalam state saat ini. Oleh karena itu, mekanisme gating membantu LSTM mengatasi masalah "long-distance dependence" dan masalah vanishing gradient pada RNN tradisional[8]. Kemampuan LSTM untuk mengekstrak informasi teks tingkat lanjut ini mempunyai peran penting dalam klasifikasi teks.

Bi-directional LSTM merupakan model RNN yang biasa digunakan pada Natural Language Processing (NLP). Model Bi-LSTM ini digunakan untuk mendapatkan representasi dua arah yang digabungkan menjadi ekspresi baru melalui convolutional neural network. Setiap ekspresi kata ditambahkan dengan sendirinya ke vector teks kiri dan vector teks kanan untuk indikasi. Struktur looping yang digunakan merupakan transformasi non-linier dari kata sebelumnya dan teks pada sisi kiri. Pendekatan ini mempertahankan informasi kontekstual dan urutan kata yang lebih luas dengan lebih baik[9]. Namun, riset tentang penggunaan Bi-LSTM

untuk ekstraksi dokumen masing kurang, sehingga kami mengusulkan untuk melakukan Pendekatan ekstraktif pada summarize dokumen menggunakan Bi-LSTM.

1.2 Rumusan Masalah

Berdasarkan permasalahan pada latar belakang tersebut, permasalahan dalam penelitian ini dirumukan sebagai berikut:

1. Bagaimana hasil dari ringkasan teks yang dilakukan menggunakan pendekatan ekstraktif dengan model BLSTM ?
2. Berapa nilai akurasi dari hasil peringkasan teks oleh model BLSTM dengan pendekatan ekstraktif?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini meliputi beberapa hal sebagai berikut:

1. *Dataset* yang digunakan dalam penelitian ini merupakan *dataset* berita online Liputan6 yang dikumpulkan dari tahun 2000 sampai tahun 2010.
2. *Dataset* yang digunakan berjumlah 10.000 *records*.
3. Penelitian ini menggunakan model BLSTM dalam melakukan peringkasan teks dengan pendekatan ekstraktif.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitian ini adalah sebagai berikut:

1. Mengetahui hasil peringkasan teks oleh model BLSTM secara ekstraktif pada dataset berita online berbahasa Indonesia.
2. Mengetahui nilai akurasi model BLSTM dalam melakukan peringkasan teks secara ekstraktif.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Secara praktis, hasil penelitian ini dapat menjadi rekomendasi dalam

melakukan *text summarize* secara ekstraktif dan menghasilkan ringkasan teks yang menyerupai ringkasan manusia dengan cepat sehingga dapat mempersingkat waktu untuk mendapatkan pokok bahasan dalam membaca sebuah teks.

2. Secara teoritis, hasil penelitian ini dapat menjadi referensi bagi penelitian selanjutnya mengenai model peringkasan teks berbahasa Indonesia.

1.6 Sistematika Penulisan

Adapun sistematika penulisan yang digunakan pada penelitian ini adalah sebagai berikut:

BAB I : PENDAHULUAN

Bab ini berisi penjelasan umum mengenai penelitian yang akan dilakukan, meliputi latar belakang, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, metode penelitian, dan sistematika penulisan.

BAB II : TINJAUAN PUSTAKA

Bab ini berisi tinjauan pustaka terkait penelitian yang akan dilakukan serta menguraikan secara terperinci teori-teori yang berhubungan dengan pokok permasalahan yang dijadikan sebagai acuan dan landasan dalam penelitian ini.

BAB III : METODE PENELITIAN

Pada bab ini akan menjelaskan mengenai perancangan dan cara kerja model yang digunakan dalam proses peringkasan text berita Bahasa Indonesia.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini berisi pembahasan dan pemaparan hasil dari penelitian yang dilakukan.

BAB V : PENUTUP

Bab ini akan menjelaskan kesimpulan yang didapat dari hasil analisis, perancangan serta saran untuk pengembangan lebih baik.