

BAB I

PENDAHULUAN

1.1 Latar Belakang

Era digital ini merupakan era yang dapat kita rasakan dampaknya secara langsung. Pertukaran arus informasi di era ini sangat cepat. Informasi datang dalam bentuk media daring maupun media luring. Penyajian informasi yang beragam dalam bentuk teks, audio, maupun video dapat kita temukan di berbagai platform dunia maya. Informasi dalam bentuk digital merupakan hal yang lumrah atau wajar saat ini. [1] Indonesia sendiri memiliki 73,7% pengguna internet dari total populasi yang ada. Jumlah ini terus meningkat dari tahun ke tahun. Dengan rata-rata waktu yang digunakan dalam menggunakan internet sebanyak 8 jam 52 menit ini, 1 jam 38 menit pengguna internet di Indonesia digunakan untuk membaca press media atau berita baik daring maupun dalam bentuk fisik. Pertumbuhan media *online* yang sangat massif ini menjadikan kemudahan akses bagi masyarakat Indonesia dalam mengakses informasi di mana saja dan kapan saja.

Peralihan dari kebiasaan masyarakat Indonesia dalam mendapatkan informasi secara cepat ini berdampak pada data atau ukuran dokumen elektronik yang dapat diakses secara daring. Menurut [2] detik.com, kompas.com, tribunnews.com, cnbcindonesia.com, suara.com, liputan6.com, dan cnnindonesia.com merupakan website atau portal berita yang sering dikunjungi selama tahun 2021. Melalui media daring, masyarakat lebih cepat mendapatkan akses mengenai topik terbaru atau topik yang sedang hangat dibicarakan tanpa memerlukan biaya yang lebih. Dengan kemajuan teknologi dan tingkat kesibukan masyarakat yang beragam dalam membaca berita terkini, penting bagi media menyajikan berita yang singkat, padat, namun memiliki informasi bermakna lengkap dari berita yang diterbitkan. Menurut data Badan Pusat Statistik (BPS) [3] 70,72% penduduk dengan usia produktif mendominasi jumlah penduduk di Indonesia. Data ini terus bertambah sejak tahun 1971 sampai dengan sekarang. Dengan menyajikan berita yang memiliki ringkasan teks, [4] inti dari bacaan atau informasi yang ada dapat ditemukan dan dimaknai lebih cepat oleh pembaca.

Memahami kebutuhan informasi tersebut, mendorong peneliti dalam bidang *Natural Language Processing* (NLP) melakukan penelitian untuk meringkas teks bacaan yang panjang dengan tujuan membuat dokumen yang lebih singkat tanpa menghilangkan ide atau pesan dari dokumen berita asli. Peringkasan teks ini dapat dilakukan dengan dua pendekatan yaitu secara ekstraktif dan abstraktif. [5], [6] Peringkasan teks secara ekstraktif dilakukan dengan mengekstrak kata kunci dari teks asli berdasarkan frekuensi teks tersebut digunakan. Sedangkan, [7], [8] Peringkasan teks secara abstraktif didapatkan dari proses peringkasan teks seperti melakukan parafrasa. Pendekatan secara abstraktif menghasilkan frasa atau kata baru yang sebelumnya tidak ada di teks asli tanpa merubah ide atau gagasan pokok dari teks asli sebelum dilakukan peringkasan. [9] Pendekatan abstraktif pada peringkasan teks memiliki keunggulan yang lebih baik daripada pendekatan ekstraktif. Metode NLP yang populer dalam penelitian mengenai peringkasan teks adalah *Recurrent Neural Network* (RNN). Namun, dalam implementasinya, [10] RNN sering kali menghasilkan ringkasan yang berisi frasa berulang atau bahkan kalimat berulang yang menyebabkan hasil ringkasan dari teks bacaan tersebut kurang memiliki arti seperti pada teks aslinya. Hal ini disebabkan karena RNN [11] tidak memiliki kemampuan dalam menyimpan memori untuk menyeleksi terlebih dahulu sehingga kata yang dihasilkan dari proses meringkas teks asli ke dalam konteks teks yang telah diringkas. Untuk menutupi kekurangan dari RNN tersebut, model *Long Short Term Memory* (LSTM) digunakan sebagai alternatif untuk melakukan peringkasan teks karena LSTM memiliki kemampuan dalam menyimpan memori dan mekanisme *attention* untuk menghasilkan kata yang berhubungan dengan konteks yang ada. Namun, dalam penelitian [12] LSTM dinilai tidak dapat mendeteksi sebagian kata pada kalimat teks asli dikarenakan LSTM melakukan proses pembelajaran konteks kalimat hanya melalui satu arah. Sehingga, informasi yang diperlukan atau penting dari teks asli tidak terdeteksi dan tidak tersimpan di dalam memori.

Dibandingkan dengan LSTM, penggunaan *Bidirectional Long Short-Term Memory* (BLSTM) dinilai mampu mengatasi masalah yang ada pada LSTM. BLSTM memiliki proses pembelajaran yang lebih kompleks karena berisi dua

lapisan paralel yang merambat dari dua arah. Sehingga, BLSTM memungkinkan peneliti untuk memperoleh inti dari informasi yang lebih kompleks daripada yang dilakukan oleh LSTM. Pada penelitian [13] BLSTM digunakan sebagai *encoder* dalam melakukan ekstraksi ringkasan dan menunjukkan peningkatan untuk menangani encoding pada teks panjang untuk generate ringkasan secara abstraktif. Berdasarkan hal tersebut, peneliti melakukan penelitian menggunakan model *Bidirectional Long Short-Term Memory* (BLSTM) dalam melakukan peringkasan teks melalui pendekatan secara abstraktif.

1.2 Rumusan Masalah

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

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

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 abstraktif.

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 abstraktif pada dataset berita online berbahasa Indonesia.
2. Mengetahui nilai akurasi model BLSTM dalam melakukan peringkasan teks secara abstraktif.

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 abstraktif 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.

