

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Data dan informasi yang tersedia di internet akan terus menerus mengalami peningkatan setiap harinya. Hal tersebut timbul karena adanya perkembangan internet yang semakin pesat yang dapat menyebabkan munculnya beberapa portal penyedia berita *online*. Portal berita *online* merupakan situs yang memuat informasi mengenai fenomena yang sedang terjadi di lingkungan masyarakat, informasi yang tersedia pada situs tersebut berupa artikel berita yang digunakan untuk menampung data yang melimpah di internet sehingga dapat menghasilkan informasi yang bermanfaat (Kencana et al., 2022).

Berdasarkan data yang tersedia di Kementerian Komunikasi dan Informatika jumlah penyedia portal berita *online* di Indonesia mencapai 43 ribu, tetapi yang sudah terverifikasi oleh Dewan Pers jumlahnya tidak mencapai 100 portal berita (Kominfo, 2018).

Kehadiran portal penyedia berita *online* dapat membantu masyarakat untuk memperoleh berbagai informasi dimanapun dan kapanpun tanpa mengenal jarak dan waktu, sehingga membuat jumlah peminat media *online* lebih banyak dibanding media lain seperti koran atau surat kabar (Kurniawan et al., 2012).

Dengan adanya portal penyedia berita *online*, artikel berita yang tersedia mengenai berbagai topik atau peristiwa tertentu akan semakin melimpah serta memiliki jumlah yang sangat banyak yang dapat mengakibatkan pembaca merasa kesulitan dan kebingungan untuk memahami garis besar dari suatu peristiwa. Selain itu, informasi yang tersedia pada artikel berita sering bertele-tele dan menampilkan informasi yang tidak penting sehingga membutuhkan waktu yang cukup lama untuk membaca keseluruhan isi berita. Permasalahan tersebut dapat diatasi dan diminimalisir dengan menggunakan peringkasan teks secara otomatis.

Peringkasan teks otomatis merupakan peringkasan pada sekumpulan kalimat yang tersedia, apakah suatu dokumen atau artikel berita memiliki kalimat-kalimat penting yang dapat disajikan secara ringkas tanpa menampilkan semua isi

dari artikel berita dengan tetap memperhatikan konteks topik yang sebenarnya agar tetap menyajikan informasi dengan makna yang sama (Andriani et al., 2019). Peringkasan berita secara otomatis termasuk kedalam bagian dari *Automatic Text Summarization* di bidang *Natural Language Processing* (NLP) yang dapat menghasilkan ringkasan dari kumpulan beberapa kata atau kalimat yang ada pada dokumen asli (Timalsina et al., 2022).

Dalam melakukan peringkasan teks terdapat dua pendekatan yang dapat digunakan yaitu pendekatan ekstraktif dan pendekatan abstraktif. Pendekatan ekstraktif menghasilkan ringkasan berupa ekstraksi kalimat-kalimat penting yang paling relevan dengan sebuah dokumen asli, sedangkan pendekatan abstraktif menghasilkan ringkasan berupa kalimat baru yang berbeda atau hasil parafrase dari dokumen asli (Verma et al., 2019).

Seiring dengan perkembangan teknologi, *deep learning* menjadi semakin populer dalam berbagai aplikasi pemrosesan bahasa alami termasuk peringkasan teks, *deep learning* merupakan teknik yang menggunakan *neural networks* dan terdiri dari beberapa *layers* atau *hidden layers* yang dapat membuat komputer memproses data dalam jumlah yang besar (Gupta & Gupta, 2018). Teknik ini mampu menangkap pola-pola kompleks dalam data dan menghasilkan keluaran yang lebih akurat dibandingkan dengan metode tradisional. *Deep learning* dapat menghasilkan akurasi yang lebih tinggi apabila jumlah data pelatihannya banyak, tetapi sebaliknya jika data pelatihannya sedikit akan sulit untuk menganalisis hasilnya.

Deep learning untuk peringkasan teks memiliki beberapa model diantaranya model *Recurrent Neural Networks* (RNN), *Long Short Term Memory* (LSTM) dan *Transformer*. RNN adalah model yang digunakan untuk memproses data secara berurutan dan beroperasi secara berulang sehingga menghasilkan *output* yang diperoleh dari inputan saat ini dan sebelumnya (Das et al., 2023). Sedangkan LSTM merupakan model dalam pemrosesan bahasa alami yang dapat membantu mengurangi permasalahan *vanishing gradient* yang sering terjadi pada RNN. *Vanishing gradient* terjadi ketika gradien dari fungsi *loss* menjadi sangat kecil selama proses *backpropagation*, sehingga jaringan sulit belajar dari data dan tidak mampu menangkap dependensi jangka panjang atau dengan kata lain, jaringan

kesulitan untuk mengingat dan memproses informasi yang telah lama terlewat, sehingga tidak dapat menggunakan informasi tersebut untuk membuat prediksi yang lebih akurat (Das et al., 2023).

Dalam beberapa tahun terakhir, peran LSTM mulai tergantikan oleh metode baru yang lebih efektif untuk peringkasan teks, yaitu *Transformer*. *Transformer* adalah model yang dapat memproses data secara paralel, sehingga lebih efisien dalam menangani data sekuensial dibandingkan dengan arsitektur RNN tradisional yang memproses data secara berurutan (Bisong, 2019).

Penelitian mengenai peringkasan berita otomatis yang menggunakan *deep learning* sudah pernah dilakukan sebelumnya. Diantaranya sistem peringkasan berita menggunakan model SHEG dengan hasil akurasi skor Rouge-1 mencapai skor 28.2 (Singh et al., 2021). Pendekatan lain menggunakan teknik RNN-LSTM, dengan hasil evaluasi Rouge-1 mencapai skor 29.10 yang dilakukan terhadap 13 ribu kata artikel berita (Sheela & Janet, 2021).

Untuk artikel berita bahasa Indonesia penelitian yang pernah dilakukan yaitu penelitian menggunakan teknik Bi-LSTM dengan skor 43.53 untuk Rouge-1 (Yuliska & Syaliman, 2022). Penelitian lain menggunakan algoritma TextTeaser terhadap 778 artikel berita dengan hasil F-Score diantara range 0.7-0.79 dengan persentase mencapai 25.30% (Gunawan et al., 2017). Selain itu, penelitian dengan metode *Transformer* berbasis BERT dilakukan pada artikel berita Liputan6.com yang menggunakan model mBERT dengan hasil evaluasi Rouge-1 mencapai 39.81 dan IndoBERT mencapai skor 41.08 (Koto et al., 2020).

Berdasarkan penelitian yang telah dilakukan sebelumnya, skripsi ini akan mengembangkan sebuah sistem peringkasan berita bahasa Indonesia secara otomatis menggunakan teknik *deep learning* yang mengimplementasikan metode *fine-tuning* pada *Transformer*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, dapat dirumuskan beberapa masalah dari penelitian ini diantaranya sebagai berikut.

- 1) Bagaimana cara mengembangkan sebuah sistem peringkasan teks otomatis yang dapat meringkas artikel berita menggunakan teknik *fine-tuning* dari metode *Transformer*?

- 2) Bagaimana evaluasi hasil ringkasan berita secara otomatis menggunakan metode *Transformer*?

1.3 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dari penelitian mengenai peringkasan berita secara otomatis diantaranya sebagai berikut.

- 1) Mengembangkan sebuah sistem peringkasan teks otomatis yang dapat meringkas artikel berita menggunakan teknik *fine-tuning* dari metode *Transformer*.
- 2) Mengevaluasi hasil ringkasan berita secara otomatis menggunakan metode *Transformer*.

1.4 Batasan Masalah

Dalam penelitian ini, batasan masalah dari penelitian ini adalah sebagai berikut:

- 1) Berita yang dapat diringkas merupakan berita bahasa Indonesia yang termasuk kedalam artikel berita yang tersedia pada portal penyedia berita *online*.
- 2) Panjang artikel berita dibatasi hanya sebanyak 512 kata.
- 3) Eksperimen yang dilakukan merupakan hasil *fine-tuning* dari *pre-trained* model T5 dan Pegasus pada *Transformer*
- 4) *Fine-tuning* pada model T5 dilakukan dengan 10 *epoch*, dan model Pegasus dengan 7 *epoch*. Karena keterbatasan *resource* GPU, jumlah *epoch* dibatasi sehingga eksperimen tidak dapat dilanjutkan untuk mengevaluasi potensi *overfitting* pada jumlah *epoch* yang lebih banyak.

1.5 Manfaat Penelitian

- 1) Penelitian ini diharapkan dapat menjadi salah satu referensi untuk peneliti berikutnya agar dapat mengembangkan sistem peringkasan berita secara otomatis yang lebih luas lagi dimasa yang akan datang.
- 2) Penelitian ini diharapkan dapat membantu pembaca memahami dan memperoleh informasi dari artikel berita dengan lebih cepat.

1.6 Sistematika Penulisan

Dalam penyusunan skripsi, rincian sistematika penulisan yang dibuat oleh penulis terdiri dari lima bab yaitu Bab I Pendahuluan, Bab II Kajian Pustaka, Bab III Metodologi Penelitian, Bab IV Hasil dan Pembahasan, Bab V Penutup. Adapun uraian penjelasan dari setiap bab dipaparkan sebagai berikut:

- 1) Bab I Pendahuluan berisi latar belakang penelitian, rumusan masalah, tujuan penelitian yang ingin dicapai, batasan masalah yang akan dikaji, manfaat yang didapat dari penelitian serta sistematika penulisan dari skripsi.
- 2) Bab II Kajian Pustaka berisi studi literatur dari konsep dan dasar teori penelitian yang akan dikaji berupa pendapat para ahli dan studi literatur mengenai penelitian sebelum-sebelumnya yang berkaitan dengan penelitian yang akan dikaji.
- 3) Bab III Metodologi Penelitian berisi paparan metode penelitian yang digunakan meliputi desain penelitian, serta alat dan bahan penelitian.
- 4) Bab IV Hasil dan Pembahasan berisi penjabaran dari tahapan yang dilakukan dan penjelasan mengenai hasil eksperimen yang telah dilakukan sesuai dengan skenario eksperimen yang dijelaskan dalam desain penelitian.
- 5) Bab V Penutup berisi kesimpulan dan saran dari hasil penelitian yang telah dilakukan.