

BAB I PENDAHULUAN

1.1 Latar Belakang

Dalam beberapa dekade terakhir, dunia telah memasuki era digital di mana informasi menjadi sangat mudah diakses melalui internet. Tren global menunjukkan bahwa 59% pengguna internet lebih memilih untuk mengakses berita melalui platform digital dibandingkan media tradisional seperti televisi atau cetak (Newman dkk., t.t.). Namun, kemudahan ini diiringi tantangan berupa volume informasi yang sangat besar, yang sering kali membuat pembaca kewalahan dalam memilah berita yang relevan.

Dalam konteks berita Politik, yang memiliki peran penting dalam pembentukan opini publik dan pengambilan keputusan, minat masyarakat Indonesia terhadap jenis berita ini tergolong tinggi. *Survei Katadata Insight Center (KIC)* menunjukkan bahwa dari 59,8% responden yang tertarik dengan Politik, 80,1% di antaranya terdorong untuk mengikuti berita-berita Politik (Annur, 2021). Angka ini mencerminkan antusiasme yang tinggi terhadap berita Politik yang memberikan wawasan tentang peristiwa, kebijakan pemerintahan, serta isu Politik lainnya. Namun, tingginya minat terhadap berita Politik diiringi dengan tantangan berupa volume informasi yang sangat besar serta kompleksitas narasi yang sering kali sulit dipahami oleh pembaca awam. Kombinasi antara panjangnya artikel berita dan kebutuhan akan informasi yang relevan membuat pembaca sering kewalahan, terutama dalam memilah berita yang benar-benar informatif.

Penelitian dalam bidang *Natural Language Processing (NLP)* telah menawarkan solusi untuk masalah ini, terutama melalui tugas peringkasan teks otomatis. Peringkasan teks bertujuan merangkum artikel panjang menjadi versi yang lebih singkat tanpa kehilangan makna utama. Pendekatan ini terbagi menjadi metode *Ekstraktif* dan *Abstraktif*. Meskipun metode *Abstraktif* menawarkan fleksibilitas lebih tinggi, metode ini sering menghadapi tantangan berupa kesalahan generatif dan kompleksitas implementasi (Liu & Lapata, 2019). Sebaliknya, metode *Ekstraktif* lebih cocok untuk berita Politik karena mampu mempertahankan keakuratan informasi serta stabilitas hasil, terutama untuk teks yang panjang dan kompleks (Nenkova & McKeown, 2012).

Salah satu inovasi terkini dalam peringkasan teks otomatis adalah penggunaan model berbasis *Bidirectional Encoder Representations from Transformers (BERT)*. Model seperti *BERT-SUM* telah menunjukkan hasil yang sangat baik dalam peringkasan teks *Ekstraktif*, dengan performa yang signifikan pada dataset berita seperti *CNN/Daily Mail* dan *XSum* dalam bahasa Inggris (Liu & Lapata, 2019). Penelitian lebih lanjut menunjukkan bahwa *DISTILBERT*, sebagai versi ringan dari *BERT*, dapat menghasilkan kinerja yang mendekati *BERT* dengan efisiensi pemrosesan yang lebih tinggi, sehingga cocok untuk aplikasi yang membutuhkan respons cepat (Sanh dkk., 2019b).

Namun, sebagian besar penelitian peringkasan berbasis *BERT* masih terbatas pada teks berbahasa Inggris. Sebagai contoh, penelitian yang menggunakan model ekstraktif untuk dataset *CNN/Daily Mail* membuktikan bahwa metode ini efektif untuk berita (Nallapati dkk., 2016). Demikian pula, penelitian yang berfokus pada pemodelan struktur dokumen menunjukkan bahwa

pendekatan *ekstraktif* dapat menghasilkan ringkasan yang relevan untuk teks yang panjang. Penelitian-penelitian ini, meskipun berhasil, belum mempertimbangkan kekhasan bahasa Indonesia, seperti tata bahasa, istilah khas, dan konteks lokal (Koshorek dkk., 2018).

Selain itu, penelitian sebelumnya juga menunjukkan keterbatasan pada penggunaan dataset lokal. Sebagian besar penelitian menggunakan dataset bahasa Inggris, seperti *CNN/Daily Mail*, yang tidak mencerminkan struktur atau isi berita berbahasa Indonesia (Liu & Lapata, 2019). Penelitian dalam bahasa Indonesia masih minim, terutama untuk domain spesifik seperti berita Politik. Berita Politik memiliki karakteristik unik yang memerlukan penanganan khusus, seperti penggunaan istilah teknis dan kebutuhan untuk menjaga keakuratan informasi menekankan bahwa karakteristik ini memerlukan pendekatan khusus dalam analisis dan summarisasi untuk memastikan bahwa informasi yang disampaikan tetap akurat dan relevan (Boydston, 2013). Penelitian yang menerapkan *BERT* untuk teks bahasa Indonesia juga masih menghadapi keterbatasan pada dataset dan evaluasi dalam domain tertentu.

Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi performa model *BERT-SUM* dan *DISTILBERT* dalam menghasilkan peringkasan *ekstraktif* berita Politik berbahasa Indonesia. Dengan memanfaatkan teknologi NLP terkini, penelitian ini diharapkan dapat menjawab tantangan peringkasan teks pada bahasa Indonesia dan memberikan kontribusi nyata terhadap pengembangan teknologi lokal.

1.2 Identifikasi Masalah

Belum ada evaluasi yang mendalam mengenai performa model berbasis BERT, seperti *BERT-SUM* dan *DISTILBERT*, dalam menghasilkan ringkasan berita Politik berbahasa Indonesia. Penelitian ini penting untuk memahami sejauh mana model-model tersebut dapat menghasilkan ringkasan yang akurat dan relevan dengan konteks lokal.

1.3 Rumusan Masalah

Berdasarkan identifikasi masalah yang telah diuraikan, rumusan masalah dalam penelitian ini adalah bagaimana performa model berbasis *BERT*, khususnya *BERT-SUM* dan *DISTILBERT*, dalam menghasilkan ringkasan berita Politik berbahasa Indonesia?

1.4 Batasan Masalah

1. Penelitian ini hanya akan mengevaluasi dua model berbasis BERT, yaitu *BERT-SUM* untuk ekstraktif summarization, serta *DISTILBERT* yang merupakan versi lebih ringan dari BERT.
2. Evaluasi performa model dilakukan menggunakan metrik evaluasi standar untuk peringkasan teks yakni *ROUGE*.
3. Penelitian ini fokus pada peringkasan teks *ekstraktif* artikel berita Politik berbahasa Indonesia yang diambil dari dataset Politik dari *CNN Indonesia* (Januari-Juni 2024).
4. Penelitian ini juga akan menggunakan dua dataset tambahan, yaitu dataset Variasi berita dari *CNN Indonesia* (yang mencakup kategori berita selain

Politik seperti hiburan, olahraga, teknologi, dan lainnya) dan *Indosum*, untuk memperkaya pelatihan model dengan berbagai jenis konten berita.

1.5 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk menilai kemampuan model berbasis *BERT*, yaitu *BERT-SUM* dan *DISTILBERT*, dalam menghasilkan ringkasan relevan, serta mengukur kualitas ringkasan yang dihasilkan menggunakan metrik evaluasi standar seperti *ROUGE*.

1.6 Manfaat Penelitian

1. Menyediakan wawasan tentang efektivitas model berbasis *BERT*, yaitu *BERT-SUM* dan *DISTILBERT*, dalam tugas peringkasan .
2. Menyediakan acuan bagi penelitian selanjutnya dalam bidang *Natural Language Processing (NLP)* khususnya dalam peringkasan teks berbahasa Indonesia.

