

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital di Indonesia menyebabkan konsumsi informasi melalui internet meningkat secara signifikan. Berdasarkan laporan Asosiasi Penyelenggara Jasa Internet Indonesia (APJII), tingkat penetrasi internet nasional pada tahun 2025 mencapai 80,66% dari total populasi Indonesia (APJII, 2025). Tingginya arus informasi tersebut memunculkan fenomena *information overload*, yaitu kondisi ketika pengguna menerima terlalu banyak informasi dalam waktu singkat. Akibatnya, masyarakat membutuhkan penyajian informasi yang lebih ringkas, cepat, dan tetap mudah dipahami.

Salah satu solusi dalam bidang *Natural Language Processing* (NLP) adalah *Automatic Text Summarization* (ATS) atau penyimpulan teks otomatis. Metode ini mampu menghasilkan ringkasan dari satu maupun beberapa dokumen secara otomatis. Dalam perkembangannya, pendekatan abstraktif dinilai lebih unggul dibandingkan pendekatan ekstraktif karena mampu menghasilkan kalimat baru yang lebih alami dan koheren seperti cara manusia merangkum informasi (El-Kassas et al., 2021). Namun, pada penyimpulan multi-dokumen, sistem harus mampu menggabungkan informasi dari berbagai artikel, mengurangi redundansi, serta menjaga konsistensi makna agar ringkasan tetap relevan.

Arsitektur *Transformer* menjadi teknologi utama dalam pengembangan sistem penyimpulan modern. Salah satu model yang memiliki performa tinggi pada tugas generatif adalah *Bidirectional and Auto-regressive Transformer* (BART) karena mampu memahami konteks dokumen dan menghasilkan teks abstraktif yang koheren (Lewis et al., 2020). Di sisi lain, MiniLM dikembangkan untuk menghasilkan representasi bahasa yang ringan dan efisien melalui teknik *knowledge distillation*. Dalam penelitian ini, MiniLM diterapkan melalui pendekatan *pseudo-abstractive summarization* berbasis *sentence embedding*, yang memungkinkan model digunakan tanpa *fine-tuning*, sehingga dibandingkan dengan BART pada penyimpulan multi-dokumen berita Indonesia (Wang et al., 2020).

Penelitian ini bertujuan untuk menganalisis kinerja model BART dan MiniLM dalam menghasilkan ringkasan abstraktif berita Indonesia. Evaluasi dilakukan menggunakan metrik ROUGE dan BERTScore untuk mengukur kualitas leksikal dan kesamaan semantik ringkasan, serta didukung *Human Evaluation* untuk mengetahui kesesuaian hasil sistem dengan penilaian manusia (Zhang et al., 2020). Selain itu, penelitian ini juga mengimplementasikan prototipe sistem penyimpulan berita otomatis menggunakan Python guna membantu pengguna memperoleh informasi secara lebih cepat dan efisien.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana kinerja model BART dan MiniLM dalam menghasilkan ringkasan abstraktif pada penyimpulan multi-dokumen berita berbahasa Indonesia?
2. Bagaimana kualitas ringkasan yang dihasilkan berdasarkan evaluasi menggunakan metrik ROUGE, BERTScore, dan *Human Evaluation*?
3. Bagaimana implementasi prototipe sistem penyimpulan berita otomatis menggunakan BART dan MiniLM serta analisis *trade-off* kualitas dan efisiensi komputasi pada kedua model?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijabarkan, maka tujuan penelitian ini adalah sebagai berikut:

1. Menganalisis dan membandingkan kinerja model BART dan MiniLM dalam menghasilkan ringkasan abstraktif multi-dokumen berita berbahasa Indonesia.
2. Mengevaluasi kualitas ringkasan menggunakan metrik ROUGE, BERTScore, dan *Human Evaluation*.
3. Mengimplementasikan prototipe sistem penyimpulan berita otomatis menggunakan model BART dan MiniLM.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Menjadi referensi dalam pengembangan sistem penyimpulan teks otomatis berbasis *Transformer* untuk berita berbahasa Indonesia.
2. Memberikan informasi mengenai perbandingan performa model BART dan MiniLM dalam tugas penyimpulan abstraktif multi-dokumen.
3. Membantu pengembangan penelitian di bidang *Natural Language Processing* (NLP), khususnya pada teknologi penyimpulan teks otomatis.

1.5 Ruang Lingkup dan Batasan Masalah

Agar penelitian lebih terarah dan sesuai dengan tujuan yang telah ditetapkan, maka ruang lingkup dan batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian difokuskan pada penggunaan model BART dan MiniLM untuk penyimpulan abstraktif multi-dokumen berita berbahasa Indonesia.
2. Data penelitian diperoleh dari portal berita daring nasional, yaitu Kompas dan Detik.
3. Proses pelatihan dan pengujian model dilakukan menggunakan Python dan Google Colab.
4. Evaluasi kualitas ringkasan dilakukan menggunakan metrik ROUGE, BERTScore, dan *Human Evaluation*.
5. Sistem yang dibangun berupa prototipe penyimpulan berita otomatis menggunakan Python dan Google Colab, dengan masukan satu kata kunci dan keluaran lima artikel dari klaster yang paling relevan.