

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi yang sangat pesat telah mengubah secara signifikan cara masyarakat mengakses dan mengolah informasi, terutama dalam bentuk berita *online* (Adek et al., 2022). Di era digital ini, kita dihadapkan pada arus informasi yang terus bertambah dengan *volume* yang masif. Khususnya pada berita. Kondisi ini memunculkan tantangan baru bagi pembaca untuk memilah informasi yang benar-benar relevan dan penting dari sejumlah besar konten berita yang tersedia (Idhafi et al., 2023; Maghfiroh et al., 2021). Fenomena ini sangat relevan di Indonesia, di mana banyak masyarakat cenderung malas membaca teks panjang dan lebih memilih informasi yang cepat dan instan (Tahmidaten & Krismanto, 2020). Oleh karena itu, Peringkasan teks otomatis menjadi solusi untuk memenuhi kebutuhan ini, memberikan ringkasan yang padat namun tetap informatif.

Peringkasan teks yang sesuai dan relevan dengan teks aslinya merupakan salah satu tantangan dalam pengolahan bahasa alami (NLP). Karena dasar tersebut penelitian ini berfokus pada perbandingan efektivitas algoritma BART dan *TextRank* dalam peringkasan teks berita *online* secara otomatis. Algoritma BART, yang merupakan model pembelajaran mendalam dengan pendekatan *encoder-decoder*, telah terbukti efektif dalam berbagai tugas pemrosesan bahasa, termasuk peringkasan, dengan menghasilkan ringkasan yang lebih koheren dan informatif dibandingkan metode tradisional. Di sisi lain, *TextRank*, yang merupakan algoritma berbasis graf, telah digunakan secara luas untuk peringkasan ekstraktif dan menunjukkan hasil yang kompetitif dalam berbagai peringkasan (Gulati et al., 2023).

Penelitian mengenai peringkasan teks otomatis telah berkembang, dengan berbagai pendekatan yang diusulkan untuk meningkatkan kualitas ringkasan. Misalnya, penelitian oleh Nisrina Arintia Maghfiroh, Galih Wasis Wicaksono, dan

Christian Sri Kusuma Aditya, dengan judul “Peringkasan Berita *Online* Corona Virus dengan Metode *Lexical Chain* dan *Word Sense Disambiguation*” pada tahun 2021 menunjukkan bahwa peringkasan dengan pendekatan ekstraktif cocok untuk peringkasan yang membahas topik yang cukup spesifik dan dapat meningkatkan kinerja peringkasan berita (Maghfiroh et al., 2021). Penelitian lainnya yang menggunakan algoritma *TextRank* dilakukan oleh Lizza Nur Fadhila dan Kadek DwiNuryana dengan judul “Teks Ringkas Otomatis pada Portal Berita CNN Indonesia Menggunakan Algoritma *Textrank*” pada tahun 2024 menunjukkan bahwa peringkasan teks otomatis pada portal berita CNN Indonesia menggunakan algoritma *Textrank* cocok untuk digunakan untuk peringkasan teks dengan pendekatan yang sederhana (Fadhila et al., 2024).

Selanjutnya sebuah studi tentang algoritma BART berbasis abstraktif yang dilakukan oleh Gaduh Hartawan, Dian Sa'adillah Maylawati, dan Wisnu Uriawan dengan judul “*Bidirectional and Auto-Regressive Transformer (BART) for Indonesian Abstractive Text Summarization*” pada tahun 2024 menunjukkan kemampuan model dalam menghasilkan ringkasan abstraktif berkualitas tinggi untuk teks berbahasa Indonesia. Penelitian ini menggunakan dataset Liputan6 dan mengevaluasi kinerja model dengan metrik ROUGE. Hasilnya menunjukkan bahwa BART mencapai skor ROUGE-1, ROUGE-2, dan ROUGE-L masing-masing sebesar 37.19, 14.03, and 33.85. Skor ini menunjukkan bahwa BART dapat menghasilkan ringkasan dengan keterbacaan dan koherensi yang baik. Jika dibandingkan dengan model lain seperti mBERT, IndoBERT, BERT2BERT, dan Pegasus, BART menunjukkan kinerja yang kompetitif, yang mengkonfirmasi efektivitasnya untuk tugas peringkasan teks berbahasa Indonesia. Hal ini membuktikan efektivitas pra-pelatihan dengan memanipulasi dokumen dan merekonstruksi teks asli (Hartawan et al., 2024).

Berdasarkan penelitian sebelumnya, penelitian membandingkan antara BART dan *TextRank* menarik untuk dilakukan karena keduanya merepresentasikan dua pendekatan yang sangat berbeda dalam peringkasan teks, *TextRank* menggunakan pendekatan ekstraktif berbasis graf yang sederhana namun efektif, sedangkan BART adalah model berbasis *transformer* yang kompleks dan canggih. Evaluasi

keduanya dilakukan menggunakan evaluasi kuantitatif melalui metrix ROUGE dan evaluasi kualitatif melalui survei untuk menilai kualitas ringkasan yang dihasilkan. Dataset yang digunakan mencakup artikel berita dari portal berita Indonesia yang representatif. Evaluasi yang dilakukan dengan kualitatif dan kuantitatif didasari berdasarkan ringkasan dengan kemiripan teks asli sebelum diringkas dan efisiensi antara kedua algoritma tersebut.

Dengan latar belakang tersebut, penelitian saya ini bertujuan untuk membandingkan efektivitas algoritma BART dan *TextRank* dalam peringkasan teks berita *online*. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem peringkasan teks otomatis yang lebih efektif dan efisien, serta memberikan wawasan baru bagi penelitian di bidang ini.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka yang menjadi rumusan masalah dalam penelitian ini adalah :

1. Bagaimana implementasi algoritma BART dan *TextRank* dalam melakukan peringkasan teks ?
2. Bagaimana efektivitas algoritma BART dan *TextRank* dalam menghasilkan ringkasan teks berita *online* ?
3. Rekomendasi algoritma mana yang lebih cocok untuk digunakan dalam peringkasan teks berita *online* berdasarkan hasil penelitian ini ?

1.3 Tujuan Penelitian

Adapun tujuan penelitian ini adalah sebagai berikut :

1. Untuk menjelaskan implementasi algoritma BART dan *TextRank* dalam menghasilkan ringkasan teks berita *online*.
2. Untuk mengevaluasi efektivitas kedua algoritma dalam menghasilkan ringkasan yang berkualitas.
3. Untuk memberikan rekomendasi mengenai algoritma yang lebih efektif digunakan dalam peringkasan teks berita *online* berdasarkan hasil penelitian.

1.4 Manfaat Penulisan

Manfaat yang didapatkan dari penelitian ini antara lain :

1. Memberikan wawasan baru mengenai implementasi dan efektivitas algoritma BART dan *TextRank* dalam konteks peringkasan teks berita *online*.
2. Menjadi referensi bagi peneliti di bidang pemrosesan bahasa alami (NLP) dalam memilih algoritma yang tepat untuk peringkasan teks.
3. Meningkatkan pemahaman tentang kualitas ringkasan teks, sehingga dapat digunakan untuk pengembangan sistem peringkasan teks yang lebih baik.

1.5 Ruang lingkup dan Batasan Masalah

Penelitian ini ditentukan pada ruang lingkup tertentu antara lain :

1. Penelitian ini hanya akan membandingkan algoritma BART dan *TextRank* dalam konteks peringkasan teks berita *online* berbahasa Indonesia.
2. Data yang digunakan untuk pengujian dalam penelitian ini akan diambil dari berita *online* yang diterbitkan dalam bahasa Indonesia, bersumber dari Liputan 6 dan CNN Indonesia, sehingga lebih optimal dan terfokus.
3. Berita *online* akan dibatasi pada artikel berita tidak lebih dari 0 sampai 400 kata, untuk memastikan bahwa kedua algoritma dapat diterapkan secara optimal pada *website*.
4. Efektivitas ringkasan akan diukur berdasarkan tiga aspek utama, yaitu relevansi informasi, kemudahan dipahami, dan kelengkapan informasi penting berita dengan melakukan evaluasi kualitatif dan kuantitatif.
5. *Output* dari penelitian ini berupa hasil perbandingan 2 algoritma dan aplikasi berbasis *website* dengan menggunakan bahasa python yang dapat digunakan untuk melakukan peringkasan dengan 2 algoritma BART dan *TextRank*.