

Bab I

Pendahuluan

1.1 Latar Belakang

Salah satu metode yang digunakan dalam pemrosesan bahasa alami untuk membuat ringkasan teks yang panjang, termasuk artikel berita, studi penelitian, catatan data, dan sebagainya, adalah peringkasan teks. Karena meningkatnya jumlah teks yang tersedia secara online, teknik ini menjadi semakin penting. Teknik ini dapat membantu orang-orang yang perlu memadatkan materi yang panjang untuk laporan singkat mereka, seperti artikel berita, makalah, dan bahkan dokumen kebijakan. Studi tentang analisis teks juga telah banyak dilakukan, tetapi aplikasinya dalam bahasa Indonesia masih sangat terbatas karena sebagian besar penelitian dilakukan untuk analisis teks dalam bahasa Inggris, yang merupakan bahasa internasional.

Pemrosesan bahasa Indonesia sendiri menghadapi banyak masalah, terutama peringkasan teks. Bahasa Indonesia yang fleksibel dapat menyebabkan masalah umum seperti perbedaan dalam konstruksi kalimat dan bahasa kontekstual, di mana makna kata berubah-ubah sesuai dengan konteks diskusi. Keterbatasan model bahasa adalah masalah lain yang cukup signifikan dalam pemrosesan bahasa. Keterbatasan data, terutama data bahasa Indonesia dalam bentuk korpus atau dataset besar, yang masih kurang dan sangat dibutuhkan untuk melatih model dan sistem, adalah masalah yang cukup signifikan.

Ada dua pendekatan untuk meringkas teks: pendekatan abstrak dan pendekatan ekstraktif. Menggunakan pendekatan ekstraktif berarti membiarkan kalimat-kalimat yang ada sebagai bagian utama teks tanpa perubahan apa pun. Sebaliknya, pendekatan abstrak memparafrasekan teks yang telah diringkas untuk memberikan ringkasan kosakata yang lebih luas, termasuk istilah-istilah yang tidak ada dalam teks aslinya. Sementara pendekatan abstraktif lebih sulit tetapi dapat menghasilkan ringkasan dengan kohesi kalimat yang tinggi dan lebih alami karena hasilnya lebih mirip dengan ringkasan buatan manusia, pendekatan ekstraktif biasanya lebih mudah tetapi sering kali menghasilkan ringkasan dengan kohesi antarkalimat yang rendah. Pada titik ini, pendekatan abstrak memiliki keunggulan ringkasan yang mirip dengan ringkasan buatan manusia, sedangkan pendekatan ekstraktif memiliki keunggulan

informasi penting karena keaslian teks.

Dengan menggunakan model bahasa Indonesia yang telah dilatih sebelumnya, pendekatan peringkasan teks hibrida yang menggabungkan metode ekstraktif dan abstrak diusulkan dalam penelitian ini untuk teks bahasa Indonesia. Penelitian lebih lanjut mengenai dampak peringkasan ekstraktif terhadap hasil dari model abstraktif juga diperlukan.

1.2 Rumusan Masalah

1. Penerapan model text summarization pada bahasa Indonesia menghasilkan kualitas rendah akibat tingginya variasi bahasa dan ketergantungan konteks bahasan
2. Pada metode penggabungan pendekatan ekstraktif dan abstraktif dalam summarization, masih belum diketahui apakah hasil ringkasan dari pendekatan ekstraktif memengaruhi hasil dari pendekatan abstraktif.

1.3 Tujuan

1. Membangun hybrid text summarization pada bahasa Indonesia dengan kualitas yang lebih tinggi
2. Mengetahui pengaruh hasil pendekatan hasil pendekatan ekstraktif pada hasil ringkasan pendekatan abstraktif dalam hybrid text summarization.

1.4 Batasan Masalah

Berikut adalah ruang lingkup yang ada pada penulisan tugas akhir ini:

1. Dataset yang digunakan adalah dataset yang terdiri dari teks bahasa Indonesia.
2. Model yang digunakan berfokus pada *pre-trained model* khusus untuk bahasa Indonesia.
3. Evaluasi model yang dilakukan dengan membandingkan hasil ringkasan yang dihasilkan dengan hasil ringkasan pada data referensi.

1.5 Sistematika Penulisan

Tugas Akhir ini disusun dengan sistematika penulisan sebagai berikut:

- **BAB I Pendahuluan.** Bab ini membahas mengenai latar belakang, rumusan masalah, dan tujuan pengerjaan Tugas Akhir ini.

- **Bab II Kajian Pustaka.** Bab ini membahas fakta dan teori yang berkaitan dengan perancangan sistem untuk mendirikan landasan berfikir. Dengan menggunakan fakta dan teori yang dikemukakan pada bab ini penulis menganalisis kebutuhan akan rancangan arsitektur sistem yang dibangun.
- **BAB III Metodologi dan Desain Sistem.** Bab ini menjelaskan metode penelitian, rancangan sistem dan metode pengujian yang dilakukan dalam penelitian.