

1. Pendahuluan

1.1 Latar Belakang

Artikel berita merupakan kebutuhan bagi banyak orang, di mana biasanya digunakan sebagai bahan bacaan atau referensi untuk mendapatkan suatu informasi terbaru yang sedang beredar. Membaca sebuah artikel berita juga merupakan kebiasaan rutin untuk mayoritas orang. Salah satu bentuk atau cara mendapatkan informasi dengan cepat pada artikel berita adalah dengan cara membaca ringkasannya. Peringkasan artikel berita juga memiliki tujuan untuk meminimalkan informasi yang berlebihan [1]. Dalam dunia *Natural Language Preprocessing* terdapat hal yang menarik untuk mendapatkan informasi dari sebuah teks panjang. Hal tersebut dapat digunakan untuk peringkasan artikel berita untuk mendapatkan informasi yang ringkas dari sebuah teks [2]. Pendekatan peringkasan teks dapat dibagi menjadi dua kelompok: peringkasan ekstraktif dan peringkasan abstraktif. Dalam pendekatan ekstraksi, untuk mendapatkan kalimat kunci dan objek tanpa merubah objek itu sendiri adalah dengan metode ekstraksi. Metode ini dilakukan dengan menggunakan frasa kunci atau ekstraksi kalimat ad-hoc dan menjaga kalimat tetap utuh [4]. Sedangkan pada pendekatan abstraktif melibatkan parafrasa kalimat dengan memperhatikan konteks setelah memahami bahasa yang digunakan [5]. Selain itu peringkasan abstraktif membuat teks lebih sederhana dengan hanya menyajikan informasi penting dengan metode abstraksi yang menitikberatkan pada pemahaman konteks yang utama dari sumber dokumen dan membuat ringkasan yang singkat dan jelas dalam bahasa yang mudah dipahami [3]. Peringkasan teks abstraksi dianggap lebih rumit dibandingkan dengan peringkasan teks ekstraktif karena melibatkan pembuatan kalimat baru dengan berbagai metode pengulangan atau penggunaan kata-kata baru [6]. Teks yang diringkas menggunakan pendekatan abstraksi diharapkan dalam bentuk yang koheren sehingga dapat berkontribusi pada keterbacaan dan akurasi tata bahasanya [7].

Salah satu metode *state-of-the-art* peringkasan abstraktif menggunakan model *fine-tuned generation* adalah dengan menggunakan metode berbasis *transformer* (pada sistem *Summaformer*) [8]. *Fine-tuned generation* merupakan proses menyesuaikan model bahasa yang sudah dilatih sebelumnya (*pre-trained*) untuk tugas atau domain tertentu dengan dataset yang lebih kecil dan sistem *Summaformer* menduduki peringkat pertama pada tantangan peringkasan [8]. Selain itu dalam penelitian lain untuk tugas peringkasan abstraktif menggunakan *fine-tuned* model T5 memiliki keefektifan yang lebih bagus daripada BART [9].

Selain itu untuk meningkatkan akurasi model peringkasan abstraktif perlu dilakukan augmentasi data karena biasanya sangat terbatas. Hal ini dapat menyebabkan model tidak mampu menangani variasi dalam bahasa dan konteks, sehingga menyebabkan akurasi yang rendah. Salah satu penelitian menunjukkan bahwa penggunaan metode augmentasi data dapat memperbaiki hasil ringkasan menjadi lebih baik [13].

Pada Tugas Akhir ini mengeksplorasi metode peringkasan artikel berita yang dapat menghasilkan ringkasan yang lebih baik dan efektif untuk mempermudah proses mendapatkan informasi dengan dilakukan percobaan augmentasi data berdasarkan klasifikasi topik. Hal ini dimotivasi untuk meningkatkan performansi model dengan menambahkan dataset sesuai dengan topik klasifikasi yang ditentukan. Selain itu juga untuk membuat model yang dibangun agar lebih memahami artikel berita pada kategori yang telah ditentukan. Dataset yang digunakan pada Tugas Akhir ini adalah XLSum dan Multi News yang terdiri atas kategori, yaitu *government politic*, *health*, *economic*, *infrastructure development*, *crime*, dan *technology industry*. Pada setiap kategori memiliki kata kunci yang ditentukan untuk melakukan pemilahan data.

1.2 Topik dan Batasan

Dataset yang digunakan berasal dari dua penelitian terkait peringkasan artikel berita yaitu XLSum dataset dari BBC News dan Multi News dengan menggunakan kategori augmentasi: *government politic*, *health*, *economic*, *infrastructure development*, *crime*, dan *technology industry*.

1.3 Tujuan

Tujuan Tugas Akhir ini adalah untuk memodelkan peringkasan artikel berita dengan pendekatan abstraktif dengan pengaplikasian augmentasi data. Pendekatan peringkasan yang digunakan adalah metode abstraktif dengan memanfaatkan *pre-trained model* berbasis transformers.

1.4 Organisasi Tulisan

Pada jurnal ini berisi bagian abstrak, pendahuluan, studi terkait, sistem yang dibangun, evaluasi, kesimpulan.