

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan internet dan dunia digital telah mengubah cara pandang masyarakat dalam mengakses dan memperoleh informasi. Berdasarkan survei Asosiasi Penyelenggara Jasa Internet (APJII), jumlah pengguna internet di Indonesia telah mencapai sekitar 229 juta orang pada tahun 2025 [1]. Tingginya jumlah pengguna internet tersebut sangat memudahkan masyarakat untuk mengakses berita dan informasi melalui perangkat digital. Berbagai portal berita saat ini memberikan kemudahan bagi masyarakat dalam memperoleh informasi secara cepat dan luas melalui platform digital. Portal berita yang tersedia mampu menyajikan berbagai topik informasi mulai dari bidang sosial, politik, ekonomi, dan lain sebagainya secara *real time*.

Kecepatan dalam memperoleh informasi memungkinkan untuk mengambil keputusan secara cepat. Namun, di sisi lain, pengguna seringkali dihadapkan oleh masalah, dimana banyaknya berita yang ditampilkan serta banyaknya artikel yang panjang serta bertele-tele membuat masyarakat merasa membaca berita secara *online* adalah sesuatu kegiatan yang menghabiskan waktu. Masalah ini tidak hanya mengurangi efisiensi dalam mendapatkan informasi penting, tetapi juga dapat menyebabkan masyarakat merasa kewalahan dan akhirnya mengabaikan informasi yang sebenarnya penting [2].

Berdasarkan permasalahan tersebut, diperlukan sistem yang mampu membantu masyarakat dalam memahami inti informasi secara praktis dan efisien. Oleh karena itu, ringkasan (*summary*) adalah cara yang cukup efektif dalam menyajikan suatu teks yang panjang menjadi lebih singkat [3]. Secara umum, peringkasan teks otomatis terbagi menjadi dua pendekatan utama yaitu peringkasan *abstractive* dan *extractive*. Pendekatan *abstractive* menghasilkan ringkasan baru dengan parafrase semantik, sedangkan pendekatan *extractive* memilih kalimat penting langsung dari teks sumber [4]. Hasil ringkasan dari *abstractive summarization* menghasilkan ringkasan dengan membangun terlebih dahulu representasi semantik internal yang lalu digunakan kembali sebagai sumber untuk menghasilkan ringkasan dengan harapan mendekati standar ringkasan buatan manusia [5].

Meskipun berbagai teknologi peringkasan teks telah berkembang pesat seperti fitur ringkasan berbasis *AI* yang telah dikembangkan oleh *Google* serta aplikasi ekstensi

tambahan lainnya, masih terdapat sejumlah keterbatasan dalam menghasilkan terjemahan yang benar benar akurat dalam segala situasi terutama dalam konteks Bahasa Indonesia dan berita lokal [6]. Selain itu, sebagian besar *tools* bersifat umum (*general purpose*) dan tidak dirancang secara khusus untuk *domain* berita sehingga hasil ringkasan berita yang dihasilkan menjadi kurang akurat dalam menangkap inti informasi.

Dalam konteks metode, pendekatan *extractive* dinilai kurang mampu menghasilkan ringkasan yang alami dan kurang kohesif. Oleh karena itu, pendekatan *abstractive* menjadi lebih sesuai untuk kasus peringkasan berita karena mampu menghasilkan ringkasan yang lebih natural. Selain itu, penelitian ini juga memfokuskan pada berita lokal wilayah Medan yang bersumber dari portal seperti Tribun Medan, Waspada, dan Detik Sumut. Hingga saat ini, sistem peringkasan teks otomatis yang secara khusus di optimalkan untuk berita lokal Indonesia khususnya di kota Medan masih sangat terbatas. Hal ini menunjukkan adanya kebutuhan untuk mengembangkan sistem yang mampu memahami dan merangkum berita lokal secara lebih akurat dan relevan.

Selain implementasi sistem, evaluasi terhadap kualitas ringkasan yang dihasilkan juga menjadi aspek penting dalam penelitian ini. Metode evaluasi yang digunakan dalam penelitian ini adalah *ROUGE* (*Recall-Oriented Understudy for Gisting Evaluation*) dan *BERTScore*. Nilai *ROUGE* digunakan untuk mengevaluasi kualitas sebuah ringkasan yang dihasilkan oleh model, baik itu dibandingkan dengan ringkasan referensi yang telah disediakan maupun dibandingkan dengan ringkasan manusia [7]. Sementara itu, *BERTScore* digunakan untuk mengukur kesamaan semantik antara hasil ringkasan referensi dan ringkasan dari hasil model dengan memanfaatkan representasi *embedding* dari model BERT. Penggunaan kedua metrik tersebut memberikan evaluasi yang lebih komprehensif, karena tidak hanya mempertimbangkan kesamaan kata maupun frasa, tetapi juga kesamaan makna yang menjadi karakteristik utama dari metode *Abstractive Summarization*. Dengan menggunakan *ROUGE* dan *BERTScore*, performa model T5 Base Indonesian Cased dapat diukur secara lebih objektif.

Berdasarkan penjelasan yang telah diuraikan, penelitian ini bertujuan untuk menerapkan metode *abstractive summarization* menggunakan model T5 Base Indonesian Summarization Cased dalam sistem peringkasan berita lokal Medan berbasis web serta mengevaluasi efektivitas model menggunakan metrik *ROUGE* dan *BERTScore*. Penelitian ini diharapkan dapat membantu masyarakat, khususnya masyarakat Kota Medan, dalam mengakses informasi secara lebih cepat, efektif, dan efisien.

## 1.2 Rumusan Masalah

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

1. Bagaimana merancang dan membangun sistem peringkasan teks otomatis berbasis metode *abstractive summarization* untuk berita lokal berbahasa Indonesia berbasis *website*?
2. Bagaimana performa model T5 Base Indonesian Summarization Cased dalam menghasilkan ringkasan berita berdasarkan evaluasi menggunakan ROUGE Score dan BERTScore?

## 1.3 Tujuan

Sesuai dengan rumusan masalah yang telah diuraikan, penelitian ini memiliki tujuan yang dapat dirincikan sebagai berikut :

1. Merancang dan membangun sistem peringkasan teks otomatis berbasis *abstractive summarization* untuk berita lokal berbahasa Indonesia berbasis *website*.
2. Mengevaluasi performa model T5 Base Indonesian Summarization Cased dalam menghasilkan ringkasan berita menggunakan metrik ROUGE dan BERTScore
3. Menganalisis kualitas ringkasan yang dihasilkan dalam mempertahankan informasi penting dan makna utama dari teks berita sumber.

## 1.4 Manfaat

Setelah tujuan penelitian tercapai, penelitian ini diharapkan dapat memberikan sejumlah manfaat positif baik kepada :

1. Bagi Masyarakat Kota Medan
  - a) Membantu masyarakat dalam memahami isi berita secara lebih cepat dan efisien melalui sistem berbasis *website*.
  - b) Mempermudah masyarakat dalam memperoleh inti informasi dari suatu artikel berita tanpa harus membaca keseluruhan artikel.
  - c) Menghemat waktu masyarakat dalam mengakses dan memahami informasi yang tersedia dalam beberapa portal berita.

2. Bagi Pengembang Ilmu Pengetahuan (Akademis)
  - a) Menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan pengembangan sistem peringkasan teks otomatis.
  - b) Memberikan kontribusi dalam penerapan metode *abstractive summarization* pada pengolahan teks berita berbasis digital.
3. Bagi Pengembang Sistem
  - a) Menjadi dasar atau acuan dalam pengembangan sistem peringkasan teks otomatis pada berbagai platform digital.
  - b) Mendorong pengembangan teknologi yang dapat membantu pengguna dalam mengolah dan memahami informasi secara lebih efisien.

### 1.5 Ruang Lingkup

Untuk menjaga agar penelitian ini tetap terfokus dan terarah, maka penelitian ini dibatasi pada ruang lingkup sebagai berikut :

1. Sumber Data  
Portal berita yang digunakan sebagai sumber agregasi dibatasi pada tiga portal berita yaitu, Tribun-Medan, WaspadaOnline, dan detikSumut.
2. Format Data  
Data yang diproses terbatas pada artikel berita dalam format teks berbahasa Indonesia. Penelitian ini tidak mencakup pemrosesan data dalam bentuk *audio*, *video*, maupun gambar.
3. Metode Peringkasan  
Penelitian ini menggunakan metode *abstractive summarization* tanpa melakukan perbandingan dengan metode *extractive summarization*. Hal ini dikarenakan penelitian difokuskan pada pengembangan sistem yang mampu menghasilkan ringkasan yang lebih ringkas, natural, dan menyerupai cara manusia untuk merangkum teks.
4. Model yang digunakan  
Model *abstractive* yang digunakan adalah T5 Base Indonesian Summarization Cased sebagai model dasar (*base model*). Meskipun model tersebut telah di *fine-tune* untuk peringkasan berita, penelitian ini juga melakukan *fine tuning* lanjutan (*further fine tuning*) menggunakan dataset yang didapatkan melalui portal berita Detik Sumut, Waspada Online dan Detik Sumut yang didapatkan melalui tahap

*scrapping*. Proses ini dilakukan agar model dapat mempelajari karakteristik penulisan berita pada dataset penelitian hingga mampu menghasilkan ringkasan yang lebih alami dan sesuai.

#### 5. Rentang Data Pengujian

Pengujian fungsional aplikasi serta evaluasi kualitas ringkasan dilakukan menggunakan artikel berita yang diterbitkan dalam tahun 2021 – 2026 guna menjaga konsistensi dan relevansi pengujian.

#### 6. Platform *Website*

Purwarupa (*prototype*) *website* “MedanBrief” yang dikembangkan dalam penelitian ini berbasis web dan dapat diakses melalui perangkat *laptop* maupun komputer (PC).

