

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era digital ini, perkembangan pengetahuan informasi sudah semakin pesat, memungkinkan akses informasi yang lebih efektif dan efisien dibandingkan pada era sebelumnya [1]. Masyarakat tidak dapat dipisahkan dengan informasi sehingga perkembangan informasi berkembang dengan cepat tanpa disadari [2]. Bukan hanya mempermudah dalam proses penelitian atau pencarian informasi yang dilakukan, tetapi juga dapat membantu dalam meningkatkan performa dan akurasi informasi yang dihasilkan [3] [4]. Dengan meningkatnya perkembangan teknologi informasi, masyarakat semakin mudah dalam pencarian informasi yang dibutuhkan.

Seiring dengan berkembangnya kemajuan teknologi informasi, pencarian informasi semakin dimudahkan sehingga berdampak signifikan dalam berbagai bidang, seperti pada kesehatan, ekonomi, teknologi, terutama dalam bidang pendidikan [5]. Pada umumnya, informasi maupun pengetahuan yang sudah ada akan dikembangkan untuk mendapatkan informasi yang baru [6]. Semakin tahun, semakin banyak jurnal yang telah dipublikasi. Salah satu penerbit akademik “Elsevier”, pada tahun 2021 telah menerbitkan lebih dari enam ratus ribu artikel [7], dan terus bertambah hingga pada tahun 2024 tercatat telah mempublikasi artikel lebih dari tujuh ratus dua puluh ribu jurnal [8]. Setiap tahunnya jurnal yang dibuat semakin Panjang halaman dan konteks yang ada, dengan perbandingan 3.5 kali lebih Panjang dibandingkan jurnal pada 30 tahun lalu [9]. Tentu saja, dengan banyaknya informasi yang ada dapat memberikan berbagai manfaat sehingga mempermudah dalam mendapatkan informasi yang mudah didapatkan oleh semua orang, tetapi makin lama informasi menjadi berlebihan sehingga menyebabkan berbagai masalah [10].

Informasi yang terlalu banyak dapat menyebabkan dampak buruk dalam masyarakat [11]. Mahasiswa seringkali mencari informasi yang ada melalui *website* untuk keperluan kuliah, mencari solusi dari permasalahan,

hingga pembelajaran online. Dalam pencarian informasi tersebut, seringkali mahasiswa merasakan kebingungan dan kesulitan mendapatkan informasi yang relevan dengan yang dibutuhkan. Hal ini disebabkan banyaknya *website* yang dibuka dan waktu yang dihabiskan terlalu lama sehingga mengganggu pencarian solusi yang dibutuhkan. Informasi yang berlebihan menyebabkan berbagai masalah, seperti kecemasan dan kebingungan dalam pengambilan keputusan, menghambat waktu dan menambah tenaga ekstra dalam pencarian informasi yang relevan. Dengan melihat terlalu banyak informasi dapat menyebabkan kebingungan dan ketidakpuasan terhadap informasi yang dibutuhkan atau diinginkan. penelitian yang dilakukan pada 46 mahasiswa di universitas Taiwan yang melakukan pengecekan *eye tracking* saat melakukan pencarian informasi yang *landslide*. Hasil yang didapatkan terdapat tiga kategori dimana terdapat mahasiswa yang merasa kebingungan dan tidak mendapatkan informasi yang relevan, kategori kedua adalah mahasiswa yang mendapatkan informasi yang kurang relevan, tetapi melakukan pengecekan kembali terhadap situs yang relevan sehingga membutuhkan waktu yang cukup lama dalam mendapatkan informasi yang tepat, kategori ketiga adalah mahasiswa yang dapat mencari informasi yang relevan dan cepat dengan mengabaikan situs yang berisi informasi yang tidak relevan. Hal tersebut menunjukkan bahwa 2 dari tiga kategori mengalami dampak dari informasi yang berlebihan sehingga menyebabkan kebingungan terhadap informasi yang relevan [12]. Disekitar kawasan penelitian pun masih sering terjadi kesulitan pencarian informasi karena adanya kelebihan data.

Terdapat survei yang dilakukan untuk mengetahui bagaimana dampak kelebihan data yang terjadi di kalangan mahasiswa. Survei dibuat melalui Google Form dan disebarluaskan melalui media sosial. Hasil survei didapatkan dengan 80 responden yang terdiri dari 77 responden pernah atau sedang menjadi mahasiswa dan 3 orang tidak pernah menjadi mahasiswa. Didapatkan hasil bahwa 33.8% (27) responden memilih bahwa kelebihan informasi sangat mempengaruhi terhadap membuat Keputusan atau menyelesaikan tugas dengan efisien, 41.3% (33) responden memilih cukup mempengaruhi, 20% (16)

responden memilih sedikit mempengaruhi, dan hanya 5% (4) responden yang memilih tidak mempengaruhi. Hal ini menunjukkan bahwa responden merasakan bahwa kelebihan informasi memberikan dampak yang mempengaruhi terhadap membuat keputusan dan menyelesaikan tugas dengan efisien. Terdapat juga pengambilan bagaimana kondisi responden saat menghadapi kelebihan data. terdapat beberapa pernyataan responden yang menyatakan bahwa kesulitan mencari informasi yang telah pada saat mencari melalui internet. Terdapat banyak sekali situs tetapi setiap informasi yang diberikan seringkali berbeda-beda pada satu konteks yang sama, sehingga diperlukan pencarian lebih lanjut terhadap informasi yang kredibel dan dipercaya menyebabkan menghabiskan tenaga dan waktu ekstra. hasil survei juga menunjukkan bahwa kelebihan informasi mempengaruhi kualitas belajar atau produktivitas akademik seperti memperlambat waktu penyerapan informasi dan menguras tenaga untuk mencari informasi. lebih dari setengah responden menjawab cukup atau sangat mempengaruhi responden dalam pencarian informasi. Hal ini terjadi oleh berbagai situasi, seperti terlalu banyaknya sumber informasi yang tidak kredibel sehingga menyebabkan kebingungan terhadap informasi seperti apa yang harus dicari, bagaimana menghadapi informasi yang terlalu banyak, apakah informasi yang didapatkan merupakan informasi yang relevan [13]. Diperlukan sebuah metode penanganan data yang tidak dapat ditangani oleh manusia dimungkinkan dengan menggunakan peringkasan teks otomatis. Dengan melakukan peringkasan teks, banyak orang akan mendapatkan informasi yang relevan yang dipadatkan menjadi sebuah ringkasan [14].

Peringkasan teks adalah sebuah proses untuk menghasilkan versi ringkas dari dokumen yang lebih panjang [15]. Peringkasan teks mempunyai tugas untuk mereduksi dokumen atau artikel yang panjang menjadi ringkasan yang ringkas dan koheren, dengan tetap mempertahankan makna inti dan informasi penting. Teknik ini sangat penting dalam berbagai aplikasi, termasuk pencarian informasi, analisis sentimen, dan manajemen pengetahuan [16]. Dengan adanya peringkasan teks memberikan dampak mempermudah mendapatkan informasi dari berbagai sumber secara cepat, tanpa harus membaca keseluruhan informasi

yang ada sehingga menghemat waktu dan tenaga yang diperlukan. Dengan cepatnya informasi yang didapatkan, penelitian dapat memberikan hasil yang lebih akurat, dan peneliti dapat mengambil keputusan yang tepat sesuai dengan kebutuhan informasi yang sesuai [17]. Dengan berkembangnya informasi yang ada, diperlukan sebuah metode peringkasan teks yang mengambil intisari dari dokumen yang panjang secara efisien dari keseluruhan informasi, tidak hanya mengambil gambaran umum suatu informasi saja. Untuk meningkatkan akurasi pada peringkasan teks, diperlukan sebuah metode yang dapat meningkatkan akurasi dan memenuhi kebutuhan peringkasan teks dengan menggunakan ATS. ATS (*Automatic Text Summarization*) merupakan sebuah teknik peringkasan teks untuk mendapatkan informasi secara ringkas dan informatif dengan menggunakan komputasi LLM [18].

Dengan menggunakan LLM (*Large Language Models*), teknik ringkasan teks semakin berkembang melalui pemanfaatan arsitektur berbasis transformer yang mampu memahami konteks secara mendalam dan menghasilkan ringkasan yang lebih akurat dan relevan [19] [20]. Arsitektur Transformer telah merevolusi bidang Pemrosesan NLP (*Natural Language Processing*) dengan kemampuannya untuk menangkap ketergantungan jangka panjang dalam teks. LLMs telah menjadi landasan dalam berbagai aplikasi NLP, termasuk otomatisasi pengambilan informasi, generasi konten, dan penyederhanaan analisis dokumen dalam skala besar [21]. Model-model ini dilatih menggunakan sejumlah besar data teks, yang memungkinkan untuk mempelajari pola bahasa yang kompleks dan menghasilkan teks yang koheren serta relevan. Selain itu, model LLM juga telah terbukti memberikan hasil yang dapat diandalkan dalam hal pemrosesan, sehingga LLM terbukti dapat membuat peringkasan teks yang lebih lengkap dan tepat pada perbandingan dengan peringkasan teks yang dibuat oleh manusia [22]. Terdapat model Mistral 7B yang merupakan model LLM dengan 7 miliar parameter yang dibuat oleh tim Mistral AI dengan mode yang dapat memberikan kinerja yang tinggi. model Mistral 7B mencapai skor evaluasi yang lebih tinggi dibandingkan pendahulunya, Mistral 13B. Model Mistral 7B seringkali digunakan oleh

berbagai penelitian yang menggunakan LLM untuk penelitian pengolahan bahasa, dan memberikan hasil yang baik dalam penelitian yang dilakukan [23]. Dalam penggunaan sebuah LLM, diperlukan optimalisasi model melalui proses *tuning* atau penyesuaian model dasar (*base model*) agar mendapatkan hasil yang lebih baik dalam melakukan tugas atau domain tertentu. Pada fungsi peringkasan teks, diperlukan sebuah optimasi model sehingga dapat menciptakan ringkasan yang sesuai dengan kebutuhan informasi yang diringkaskan. Penyediaan informasi spesifik yang disediakan pada peringkasan yang dibuat dapat membantu model memberikan informasi yang tepat dan sesuai dengan kebutuhan informasi dari peringkasan yang dibuat. Maka dari itu, diperlukan sebuah optimasi menggunakan *instruction tuning* untuk menyesuaikan hasil *output* peringkasan teks yang dibuat oleh model. *Instruction tuning* digunakan untuk model LLM sehingga hasil dapat menyesuaikan dengan instruksi yang diberikan [25].

Tujuan dari penelitian yang dilakukan adalah menciptakan sebuah model yang dapat meringkas teks artikel sesuai dengan kebutuhan penelitian, dan dengan nilai akurasi yang tinggi. Penelitian ini bertujuan mengembangkan model peringkasan teks berbasis LLM Mistral 7B melalui *instruction tuning*, sehingga *output* yang dihasilkan dapat disesuaikan dengan instruksi pengguna. Untuk mengevaluasi kualitasnya, hasil model akan dibandingkan dengan penelitian sebelumnya menggunakan teknik evaluasi yang sama. Dengan demikian, dapat diketahui sejauh mana performa Mistral 7B dalam hal akurasi dan keunggulan dibandingkan model LLM lain. berdasarkan pendapat masyarakat terhadap peringkasan yang dibuat. Diharapkan dengan adanya penelitian ini dapat membantu pembaca untuk mengetahui lebih baik dalam LLM Mistral 7B dan peringkasan teks

1.2 Rumusan Masalah

Berdasarkan penjelasan diatas, maka dari itu penelitian ini memiliki rumusan masalah, yaitu:

1. Bagaimana hasil implementasi *instruction tuning* terhadap model Mistral 7B dalam pembuatan model peringkasan teks?
2. Bagaimana performa model peringkasan teks Mistral 7B dengan menggunakan perbandingan penelitian terdahulu?
3. Bagaimana pendapat mahasiswa terhadap model peringkasa teks Mistral 7B dan bagaimana hasil perbandingannya dibandingkan dengan penelitian terdahulu?

1.3 Batasan Masalah

Terdapat beberapa Batasan masalah pada penelitian yang dilakukan, yaitu:

1. Dataset yang digunakan untuk penelitian merupakan data laporan berita yang berisikan laporan berita dan *highlight* berita telah ditulis oleh reporter.
2. Model Mistral 7B yang digunakan memiliki jumlah parameter sebanyak 7 miliar parameter.
3. Hasil akhir model peringkasan teks dengan menggunakan Mistral 7B diintegrasikan melalui sebuah situs peringkasan teks.
4. Bahasa yang dapat digunakan untuk peringkasan teks adalah bahasa Indonesia dan Inggris.
5. Ukuran data pada teks yang akan diringkas maksimal mencapai 10 MB.
6. Perangkat komputer yang digunakan untuk model LLM memiliki spesifikasi RAM 8 GB dan GPU 1080 RTX.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Terdapat tujuan penelitian yang telah ditetapkan, yaitu:

1. Mengimplementasi model Mistral 7B yang sudah di-*instruction tuning* untuk membantu meringkas teks.

2. Mengetahui hasil evaluasi dari ringkasan text yang telah dibuat dengan model Mistral 7B.
3. Mengetahui pendapat masyarakat mengenai hasil dari ringkasan yang dibuat oleh model Mistral 7B

1.4.2 Manfaat Penelitian

Terdapat manfaat penelitian yang diharapkan, yaitu:

1. Sebagai referensi dan pengetahuan untuk penelitian-penelitian selanjutnya terhadap pemahaman lebih lanjut mengenai pembuatan ringkasan teks maupun mengenai model Mistral 7B.
2. Hasil dari penelitian ini dapat digunakan untuk melakukan peringkasan teks untuk informasi, jurnal, artikel, maupun sumber pengetahuan lainnya yang bersifat tulisan atau kata.

