

BAB I

PENDAHULUAN

1.1. Latar Belakang

Di Indonesia, jumlah orang yang menggunakan internet secara konsisten meningkat setiap tahunnya. Jumlah pengguna internet telah meningkat dari 175 juta menjadi 202 juta, dengan peningkatan sebesar 11% atau sebanyak 27 juta pengguna sejak tahun 2021. Dari 175 juta pengguna tersebut, 160 juta di antaranya adalah pengguna aktif media sosial. Sementara itu, terdapat 272 juta orang yang tidak aktif menggunakan internet, di antaranya yakni balita dan orang dewasa [1].

Klasifikasi berita adalah proses mengkategorikan berita ke dalam kategori tertentu berdasarkan kontennya. Dalam beberapa tahun terakhir, teknik *deep learning* seperti LSTM (*Long Short-Term Memory*) dan BERT (*Bidirectional Encoder Representation from Transformers*) telah digunakan untuk meningkatkan akurasi klasifikasi berita. LSTM adalah jenis *Recurrent Neural Network* (RNN) yang dapat memproses data sekuensial seperti teks. Dalam klasifikasi berita, LSTM dapat digunakan untuk memodelkan hubungan antar kata dalam sebuah kalimat atau dokumen. LSTM dapat mempelajari pola-pola yang kompleks dalam data teks dan dapat digunakan untuk klasifikasi berita dengan akurasi yang tinggi [2]. IndoBERT adalah versi BERT yang telah dilatih pada *dataset* bahasa Indonesia. Klasifikasi teks, analisis sentimen, dan *question answering* merupakan contoh tugas dalam NLP (*Natural Language Processing*) yang dapat ditangani oleh IndoBERT. Dalam klasifikasi berita, IndoBERT dapat digunakan untuk mempelajari representasi kata-kata dan kalimat dalam bahasa Indonesia dan dapat meningkatkan akurasi klasifikasi berita [3].

Berdasarkan penjelasan sebelumnya, klasifikasi berita di Indonesia sangat penting agar dapat memudahkan para pembaca dalam mencari topik berita. Pada penelitian ini, akan dilakukan klasifikasi berita menggunakan metode IndoBERT dan LSTM yang diharapkan dapat memudahkan para pembaca dalam mencari topik

berita.

1.2. Rumusan Masalah

Perumusan masalah berdasarkan latar belakang di atas adalah sebagai berikut:

1. Bagaimana performansi metode IndoBERT dan LSTM dalam mengklasifikasikan berita di Indonesia?
2. Apa saja hal yang dapat memengaruhi dalam mengklasifikasikan berita di Indonesia?
3. Bagaimana pengaruh *precision* dan *recall* dalam mengklasifikasikan berita di Indonesia?

1.3. Tujuan

Tujuan yang ingin dicapai dalam penelitian ini antara lain:

1. Mengetahui tahapan dalam mengklasifikasikan berita di Indonesia menggunakan metode IndoBERT.
2. Memeroleh hasil klasifikasi berita di Indonesia menggunakan metode IndoBERT.

1.4. Rencana Kegiatan

Tahapan-tahapan yang dilakukan dalam penyelesaian penelitian tugas akhir ini adalah sebagai berikut:

1. Studi Literatur

Kegiatan ini bertujuan untuk mengumpulkan, memahami konsep, dan teori dasar dari beberapa literatur terkait klasifikasi berita, klasifikasi teks, IndoBert, serta hal-hal lain tentang pengolahan data.

2. *Pre-processing* Data

Kegiatan ini meliputi pengumpulan data untuk penelitian yang sesuai dengan tujuan dan tidak melebihi masalah yang ditentukan.

3. Pemrosesan Data

Kegiatan mengolah data menggunakan metode yang telah ditentukan.

4. Evaluasi dan Analisis Data

Kegiatan yang bertujuan untuk menganalisis hasil performansi yang didapatkan selama proses pemrosesan data hingga evaluasi berdasarkan parameter dan tujuan yang telah ditentukan.

5. Pembuatan Laporan

Kegiatan membuat laporan akhir mengenai hasil penelitian yang telah dilakukan melalui tahap pemrosesan data dan didapatkan hasil analisis data.

1.5. Jadwal Kegiatan

Adapun perencanaan jadwal kegiatan penelitian yang akan dilaksanakan dan dapat dilihat pada tabel berikut :

No	Kegiatan	Bulan ke-					
		1	2	3	4	5	6
1	Studi Literatur						
2	<i>Pre-processing</i> data						
3	Pemrosesan Data						
4	Evaluasi dan Analisis Data						
5	Pembuatan Laporan						