

BAB I PENDAHULUAN

1.1 Latar Belakang

Berita palsu atau hoaks menyebar melalui internet[1]. Akses yang mudah terhadap internet menyebabkan peningkatan penyebaran berita palsu. Masyarakat Indonesia menerima berita maupun informasi dengan cepat melalui internet setiap harinya[2]. Hal tersebut terjadi karena keberadaan infrastruktur internet yang telah menjangkau lapisan masyarakat serta kepemilikan *smartphone*[3]. Berdasarkan survei yang dilaksanakan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pengguna internet di Indonesia pada tahun 2023 mencapai 215 juta pengguna[4]. Masyarakat yang dibanjiri informasi tanpa batas sulit untuk mengkritisi kebenaran suatu informasi yang didapat[5]. Berita palsu yang dikonsumsi masyarakat dapat memberikan dampak buruk seperti ujaran kebencian yang dapat menimbulkan kerusuhan[6]. Maka dari itu, dibutuhkan cara untuk mengurangi penyebaran berita palsu.

Klasifikasi menggunakan *deep learning* dapat digunakan untuk mengidentifikasi berita[7]. *Deep learning* adalah cabang dari *machine learning* yang menggunakan arsitektur jaringan saraf (*neural networks*) untuk mengekstraksi *features* (karakteristik data) yang abstrak[8]. *Deep learning* mampu mengidentifikasi data yang berbentuk teks berdasarkan hubungan semantik antar kata dalam teks[9]. Hal tersebut memungkinkan penggunaan *deep learning* untuk menganalisis pola-pola semantik yang muncul dalam data berita dan menunjukkan adanya berita palsu.

Salah satu arsitektur *deep learning* yang digunakan untuk klasifikasi berita palsu adalah arsitektur *recurrent* (berulang)[7], [10]–[14]. Arsitektur ini memiliki kemampuan untuk memproses data *sequential* secara berurutan dan mengambil informasi dari data yang diberikan[8]. Penelitian [10] membandingkan kinerja algoritma-algoritma yang menggunakan arsitektur *recurrent* yakni *Recurrent Neural Network* (RNN), *Long Short-Term Memory* (LSTM), dan *Gated Recurrent*

Unit (GRU) dalam melakukan deteksi hoaks pada berita lokal Indonesia. Selain itu, penelitian [14] juga membandingkan kinerja algoritma yang menggunakan arsitektur *recurrent* yaitu LSTM dengan algoritma yang menggunakan arsitektur *feedforward* yaitu *Convolutional Neural Network* (CNN) untuk menentukan berita palsu.

Arsitektur *recurrent* memiliki keterbatasan dalam pengolahan teks[15]. Keterbatasan yang utama adalah mengatasi dependensi antar kata dalam kalimat yang panjang dan pemahaman konteks yang kurang baik[16]. Keterbatasan lain arsitektur *recurrent* dalam hal pemrosesan adalah data *sequential* harus diproses secara berurutan sehingga memakan waktu yang lama[17]. Arsitektur *Transformer* dibuat untuk menangani masalah-masalah tersebut[15]. Algoritma yang didasarkan pada arsitektur *Transformer* yang dapat digunakan untuk tugas klasifikasi adalah *Bidirectional Representations from Transformers* (BERT)[18]. Pada dasarnya BERT merupakan pengembangan dari algoritma *Transformer* yang difokuskan untuk menghasilkan representasi kata yang kaya dan kontekstual[18]. Namun, penelitian [19] menemukan bahwa walaupun BERT sangat canggih, pada dataset kecil LSTM yang memiliki arsitektur paling sederhana dapat mengalahkan performa BERT. Hal tersebut karena pada dataset kecil BERT cenderung lebih *overfit* daripada LSTM sederhana. Penelitian ini dilakukan untuk menguji algoritma BERT dalam melakukan klasifikasi berita palsu berbahasa Indonesia serta mengetahui perbandingan kinerja BERT terhadap model *recurrent* pada penelitian [10] dan [14].

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, masalah-masalah yang ada dapat dirumuskan sebagai berikut.

1. Bagaimana hasil performa model BERT dalam melakukan klasifikasi berita palsu berbahasa Indonesia?
2. Bagaimana perbandingan model BERT terhadap model dengan arsitektur *recurrent* dalam melakukan klasifikasi berita palsu berbahasa Indonesia?

1.3 Batasan Masalah

Pelaksanaan penelitian ini dibatasi pada hal-hal sebagai berikut.

1. Data berita yang digunakan berasal dari penelitian [10] yang merepresentasikan dataset kecil dan penelitian [14] yang merepresentasikan dataset besar.
2. Model *pre-trained* yang digunakan adalah model yang telah dilatih dalam Bahasa Indonesia yakni IndoBERT_{BASE}.
3. Penelitian ini tidak menerapkan model dalam bentuk perangkat lunak.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitiannya adalah sebagai berikut.

1. Menghasilkan sebuah model BERT yang digunakan untuk melakukan klasifikasi berita palsu berbahasa Indonesia beserta evaluasinya.
2. Menghasilkan analisis perbandingan model BERT terhadap model dengan arsitektur *recurrent* dalam melakukan klasifikasi berita palsu berbahasa Indonesia.

1.5 Manfaat Penelitian

Adapun manfaat penelitian adalah sebagai berikut.

1. Model BERT dapat dijadikan dasar pengembangan sistem deteksi berita palsu di masa depan.
2. Hasil kinerja model BERT dan perbandingannya dengan model berarsitektur *recurrent* dapat dijadikan pijakan dan referensi pada penelitian-penelitian selanjutnya.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut.

1. BAB I PENDAHULUAN

Bab ini berisi tentang gambaran dari penelitian yang meliputi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian,

dan sistematika penulisan

2. BAB II TINJAUAN PUSTAKA

Bab ini berisi studi literatur yang mencakup penelitian-penelitian terdahulu dan dasar teori yang digunakan dalam penelitian ini.

3. BAB III METODE PENELITIAN

Bab ini berisi langkah-langkah dan metode yang digunakan untuk menyelesaikan penelitian meliputi alur penelitian serta alat dan bahan.

4. BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi penjelasan mengenai tahap-tahap yang telah dilakukan pada penelitian ini yang mencakup pengumpulan data, pra-pemrosesan, pelatihan, hingga evaluasi. Bab ini juga membahas hasil temuan yang didapat setelah proses evaluasi dan perbandingan model dilaksanakan.

5. BAB V PENUTUP

Bab ini berisi kesimpulan yang menjawab rumusan masalah. Kesimpulan diperoleh dari hasil dan pembahasan sebelumnya. Bab ini juga berisi saran yang bisa digunakan untuk mengembangkan penelitian ini.