

BAB I

PENDAHULUAN

1.1. Latar Belakang

Berita palsu mengacu pada penyebaran informasi palsu atau disinformasi yang disajikan seolah-olah merupakan berita faktual dengan tujuan membingungkan, memanipulasi dan memengaruhi opini publik serta bukan berasal dari pendapat tokoh publik kredibel [1]. Setiap tahunnya penyebaran berita palsu meningkat secara masif dimana 2018 hingga Mei 2023, sebanyak 11.642 konten telah diidentifikasi oleh KOMINFO sebagai hoaks. Dari jumlah tersebut sebanyak 1.373 diantaranya adalah konten hoaks kategori politik [2]. Terdapat sebuah penelitian yang membandingkan model *deep learning* CNN dan LSTM dengan menambahkan Word2Vec untuk *word embedding* pada masalah pendeteksian berita palsu. Hasilnya, model CNN memiliki akurasi yang lebih baik dibandingkan dengan model LSTM. Meskipun demikian, model yang dihasilkan *overfit*, terdapat pula saran perbaikan pada tahapan *pre-processing* data dan representasi makna atau *word embedding* khususnya untuk data berbahasa Indonesia [3]. Junita Amalia et al (2022) juga melakukan penelitian yang sama menggunakan dataset berbahasa Indonesia dengan menggabungkan Bi-LSTM dan Word2Vec. Performa model yang dihasilkan memiliki akurasi baik dan bergantung pada parameter *windows size*, *embedding size* dan unit Bi-LSTM yang digunakan. Namun dalam pengujian dataset judul, hasil akurasi tidak sebaik narasi karena cenderung mengarah bias bila hanya mendeteksi kalimat yang singkat [4].

Untuk mengatasi masalah representasi makna kata yang lebih baik dari Word2Vec, Ryan Adipradana et al (2021) membandingkan model LSTM, GRU, Bi-LSTM dan Bi-GRU menggunakan FastText dan GloVe. Dari hasil penelitian ini, model Bi-GRU-FastText memiliki akurasi terbaik. Penelitian ini menyatakan bahwa di seluruh model yang diteliti, kombinasi dengan FastText *embedding* menghasilkan akurasi yang lebih baik dibandingkan dengan GloVe. Kesimpulannya adalah FastText mampu menangkap vektor kata yang lebih kompleks daripada GloVe. Tetapi masih terdapat saran perbaikan untuk menggunakan berita Indonesia asli tanpa terjemahan sebelumnya [5].

Angeline Karen et al (2023) pun meneliti berita palsu dengan membandingkan Bi-LSTM dan IndoBERT untuk mendeteksi berita palsu. Hasilnya adalah Bi-LSTM memiliki akurasi lebih baik dibandingkan IndoBERT. Penelitian ini menyatakan IndoBERT unggul di bagian pemahaman relasi dua kalimat pendek dan perlu dilakukan utilisasi *fine-tuning* untuk hasil yang lebih baik [6]. Penelitian terbaru dilakukan oleh Rifky Maulana et al (2023)

dengan menggunakan LSTM memanfaatkan IndoBERT. Model ini berhasil membuktikan bahwa IndoBERT dapat bekerja dengan baik jika digabungkan dengan LSTM, memanfaatkan keunggulan IndoBERT dalam pemahaman relasi kalimat pendek. Namun, dalam penerapannya, data yang digunakan masih terlalu sedikit dan kurang bervariasi [7]. Tantangan yang dihadapi dari penelitian-penelitian sebelumnya sebagian besar diakibatkan bentuk teks yang heterogen, artinya tidak menentu panjangnya setiap kalimat dalam artikel berita dan potensi bias dalam pendeteksian pada judul berita.

Berdasarkan pemaparan di atas, maka akan dilakukan penelitian deteksi berita palsu dengan menggunakan IndoBERT-Bi-GRU, di mana akan dilakukan perbandingan juga dengan model IndoBERT-Bi-GRU bila dikombinasikan dengan FastText dan GloVe sebagai *word embedding*. Masing-masing model akan ditambahkan mekanisme atensi untuk memberikan hasil yang lebih halus dan peningkatan performa dari model *deep learning*. Model ini diharapkan dapat memberikan hasil yang lebih baik dari sebelumnya [8], [9].

Dari permasalahan dan kajian literatur yang dijabarkan, maka dilakukanlah penelitian ini dengan judul “**Deteksi Berita Palsu Politik Indonesia Menggunakan IndoBERT-BiGRU-Attention**”.

1.1. Rumusan Masalah

Berdasarkan latar belakang, pendeteksian berita palsu perlu dilakukan untuk menekan peningkatan pengguna internet percaya terhadap berita palsu. Maka permasalahan yang dihadapi penelitian ini adalah sebagai berikut.

1. Bagaimana mengidentifikasi dan mengurangi potensi bias dalam pendeteksian berita palsu pada judul berita menggunakan model IndoBERT-Bi-GRU-Attention?
2. Bagaimana meningkatkan akurasi model deteksi hoaks IndoBERT-Bi-GRU-Attention dalam menghadapi heterogenitas panjang kalimat dalam artikel berita?
3. Apakah dengan menambahkan model *word embedding* dapat meningkatkan metrik evaluasi?

1.2. Tujuan

Tujuan yang ingin dicapai dalam penelitian tesis ini adalah sebagai berikut.

1. Mengembangkan dan menerapkan metode untuk mengidentifikasi dan mengurangi potensi bias dalam pendeteksian berita palsu pada judul berita menggunakan model IndoBERT-Bi-GRU-Attention.

1. Merancang strategi atau teknik *preprocessing* teks yang dapat meningkatkan akurasi model IndoBERT-Bi-GRU-Attention dalam menghadapi heterogenitas panjang kalimat dalam artikel berita.
2. Mengevaluasi pengaruh penambahan model *word embedding* terhadap peningkatan metrik evaluasi dalam deteksi hoaks menggunakan model IndoBERT-Bi-GRU-Attention.

1.1. Manfaat Penelitian

Penelitian tesis diharapkan dapat memberikan manfaat:

1. Model dapat membantu memberikan hasil deteksi yang lebih akurat dalam pendeteksian berita palsu.
2. Model yang diusulkan dapat menjadi pertimbangan untuk menambah sarana ilmu pengetahuan yang dapat digunakan sebagai referensi tambahan dalam model pendeteksian berita palsu.

1.2. Ruang Lingkup

Batasan dan ruang lingkup yang diterapkan dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan untuk pendeteksian bersumber dari platform berita terpercaya seperti CNN, Tempo.co, Kompas dan TurnBackHoax.ID.
2. Dataset yang terpisah untuk pengujian judul berita dan narasi berita politik yang berasal dari 4 platform berita tersebut pada rentang 2020 hingga April 2023 dengan total 10875 data narasi dan 17278 data judul.
3. Penelitian akan menguji dataset yang berisikan judul saja dan narasi saja pada ketiga model.
4. Model IndoBERT yang digunakan adalah IndoBERT-lite-base-p1, maka *transformer* yang akan dijalankan adalah konfigurasi Albert.
5. Masing-masing model akan memiliki 4 lapisan *dense* yang bernilai berturut-turut 256, 128, 64 dan 1.