

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Perkembangan teknologi informasi dan komunikasi memudahkan masyarakat mengakses serta menyebarkan informasi melalui media sosial, portal berita, dan aplikasi chatting. Kemajuan ini mempercepat arus informasi namun juga memicu meningkatnya penyebaran berita hoaks [1]. Hoaks merupakan informasi yang salah atau menyesatkan yang disebar dengan tujuan untuk memengaruhi persepsi publik atau menimbulkan kebingungan [2]. Berdasarkan data Kementerian Komunikasi dan Informatika (Kominfo), terdapat 12.547 isu hoaks di Indonesia selama 2018-2023 dengan berbagai topik seperti politik, kesehatan, ekonomi, dan isu sosial [3].

Peredaran berita hoaks yang meluas menimbulkan tantangan dalam menjaga kualitas informasi di ruang digital. Banyak masyarakat kesulitan membedakan berita benar dan palsu karena kemiripan gaya bahasa serta sumber yang tampak kredibel [4]. Kondisi ini diperparah oleh penyebaran cepat melalui media sosial seperti Twitter, Facebook, dan WhatsApp, yang memungkinkan satu informasi menjangkau ribuan orang dalam waktu singkat [1], [5]. Oleh karena itu, diperlukan upaya untuk menganalisis dan mengklasifikasikan berita hoaks secara efektif agar penyebarannya berita hoaks dapat diminimalkan.

Dalam beberapa tahun terakhir, *deep learning* semakin banyak digunakan untuk klasifikasi teks, termasuk deteksi berita hoaks. Model *Recurrent Neural Network* (RNN) seperti *Long Short-Term Memory* (LSTM) dan *Gated Recurrent Unit* (GRU) mampu memahami hubungan konteks antar kata dengan baik [6]. Versi dua arah dari keduanya, yaitu *Bidirectional LSTM* (BiLSTM) dan *Bidirectional GRU* (BiGRU), dapat menangkap informasi dari kedua arah teks, baik dari awal ke akhir maupun dari akhir ke awal, sehingga dapat memahami makna secara lebih menyeluruh [7].

Penelitian ini bertujuan untuk menganalisis dan membandingkan performa

model BiLSTM dan BiGRU dalam klasifikasi berita hoaks berbahasa Indonesia. Perbandingan dilakukan dengan memanfaatkan dua jenis *word embedding*, yaitu FastText dan GloVe [8], untuk melihat pengaruh representasi kata terhadap kemampuan model dalam memahami konteks teks. Selain itu, penelitian ini juga menerapkan *hyperparameter tuning* untuk memperoleh konfigurasi model yang paling optimal, lalu *k-fold cross validation* untuk memastikan evaluasi performa model yang lebih stabil dan tidak bergantung pada satu pembagian data tertentu.. Dengan membandingkan kedua model, penelitian diharapkan dapat mengetahui model mana yang lebih efektif dalam mengenali berita hoaks dari hasil evaluasi seperti akurasi, presisi, *recall*, *F1-score* dan *confusion matrix*.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana performa model BiLSTM dan BiGRU dalam klasifikasi berita hoaks berdasarkan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*?
2. Bagaimana perbedaan performa fitur ekstraksi *word embedding* FastText dan GloVe dalam klasifikasi berita hoaks ?
3. Bagaimana penerapan dan pengaruh *hyperparameter tuning* terhadap peningkatan performa model BiLSTM dan BiGRU?

## 1.3 Batasan Masalah

Berdasarkan latar belakang diatas, maka batasan masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Penelitian ini hanya berfokus pada analisis performa dua model deep learning, yaitu *Bidirectional Long Short-Term Memory* (BiLSTM) dan *Bidirectional Gated Recurrent Unit* (BiGRU) dalam melakukan klasifikasi berita hoaks berbahasa Indonesia.
2. Penelitian ini hanya menggunakan teks berita berbahasa Indonesia, tanpa melibatkan berita dalam bahasa asing.

3. Dataset yang digunakan dalam penelitian ini berasal dari Kaggle dengan sumber data yang mencakup TurnBackHoax.id, AntaraNews, Kompas, dan Detik. Dataset tersebut disajikan dalam format CSV dengan jumlah data yang terdiri atas 12.744 dari TurnBackHoax.id, 4.200 dari AntaraNews, 3.500 dari Kompas, dan 3.500 dari Detik.
4. Penelitian ini hanya menggunakan dua jenis *word embedding*, yaitu FastText dan GloVe, sebagai representasi kata yang diuji pada model BiLSTM dan BiGRU.
5. Proses klasifikasi dalam penelitian ini difokuskan pada dua label, yaitu Berita Non-Hoaks (0) dan Berita Hoaks (1).
6. Penelitian ini hanya memfokuskan pada judul dan narasi teks berita tanpa mempertimbangkan elemen non-teks seperti gambar, video, atau metadata lainnya.
7. Evaluasi kinerja model dibatasi pada metrik klasifikasi standar seperti akurasi, presisi, *recall*, dan *F1-score*, serta visualisasi hasil evaluasi menggunakan *confusion matrix*.

#### 1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitian ini adalah sebagai berikut:

1. Menerapkan model BiLSTM dan BiGRU dalam proses klasifikasi berita hoaks menggunakan dataset berbahasa Indonesia.
2. Menganalisis dan membandingkan performa kedua model berdasarkan metrik evaluasi akurasi, presisi, *recall*, *F1-score* dan *confusion matrix*.
3. Mengetahui model yang lebih efektif dalam mengklasifikasikan berita hoaks berdasarkan hasil pengujian dan analisis performa model.

#### 1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoritis maupun praktis sebagai berikut:

1. Manfaat teoritis dari penelitian ini adalah dapat menjadi referensi bagi

pengembangan ilmu pengetahuan di bidang *Natural Language Processing* (NLP), khususnya dalam penerapan model *deep learning* seperti BiLSTM dan BiGRU untuk klasifikasi teks berbahasa Indonesia. Hasil penelitian ini juga diharapkan dapat menambah wawasan mengenai perbandingan performa model sejenis dalam mendeteksi berita hoaks sehingga dapat dijadikan dasar bagi penelitian-penelitian selanjutnya.

2. Manfaat praktis dari penelitian ini diharapkan dapat memberikan kontribusi bagi pengembang sistem deteksi berita hoaks, lembaga pemeriksa fakta, serta platform media digital dalam memilih model *deep learning* yang paling efektif untuk klasifikasi berita hoaks berbahasa Indonesia.

## 1.6 Sistematika Penelitian

Skripsi ini disusun dalam lima bab utama dengan sistematika sebagai berikut:

### BAB I PENDAHULUAN

Bab ini berisi penjelasan mengenai latar belakang yang melandasi penelitian, rumusan masalah yang ingin diselesaikan, batasan masalah untuk memperjelas ruang lingkup penelitian, tujuan yang ingin dicapai, manfaat penelitian baik secara teoritis maupun praktis, serta sistematika penelitian skripsi ini.

### BAB II TINJAUAN PUSTAKA

Bab ini berisi kajian literatur dan penelitian terdahulu yang relevan, serta dasar teori yang mendukung penelitian.

### BAB III METODE PENELITIAN

Bab ini menjelaskan secara rinci metode yang digunakan dalam penelitian, termasuk objek penelitian, alur penelitian dalam bentuk flowchart, data dan alat yang digunakan, serta proses yang dilakukan dalam pelaksanaan penelitian.

### BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil penelitian yang meliputi proses pengumpulan data, tahapan preprocessing teks, serta exploratory data analysis untuk memahami

karakteristik dataset berita hoaks berbahasa Indonesia. Selanjutnya dibahas proses word embedding dan pelatihan model BiLSTM dan BiGRU, termasuk pengaturan hyperparameter dan penerapan k-fold cross validation. Evaluasi kinerja model dilakukan menggunakan classification report dan confusion matrix, yang kemudian dilanjutkan dengan analisis perbandingan performa model BiLSTM dan BiGRU berdasarkan metrik evaluasi yang digunakan. Berdasarkan hasil analisis tersebut, model dengan performa terbaik selanjutnya diimplementasikan ke dalam aplikasi web pada tahap deployment untuk menguji kinerja prediksi terhadap data baru.

## BAB V PENUTUP

Bab ini berisi kesimpulan yang diperoleh dari hasil penelitian serta saran yang dapat dijadikan bahan pertimbangan untuk penelitian selanjutnya agar dapat menghasilkan pengembangan yang lebih baik di masa mendatang.

