

1. Pendahuluan

Latar Belakang

Dalam beberapa tahun terakhir, pengguna media sosial mengalami peningkatan yang cukup tinggi. Pada bulan Januari 2022 pengguna media sosial di Indonesia mengalami peningkatan sebesar 12,35% dari tahun sebelumnya, yaitu mencapai 191 juta orang [1]. Dengan peningkatan yang cukup tinggi setiap tahunnya, maka semakin banyak masyarakat yang cenderung mencari dan mengkonsumsi informasi melalui media sosial [2]. Alasan masyarakat mendapatkan informasi melalui media sosial, karena cukup mudah dan lebih cepat mendapatkan informasi dibandingkan dengan mendapatkan informasi melalui media tradisional seperti koran atau majalah [3]. Meskipun banyak kelebihan yang diberikan oleh media sosial, namun kualitas informasi di media sosial lebih rendah dibandingkan dengan media berita tradisional karena banyak informasi hoax yang tersebar, hal ini disebabkan karena mudahnya akses dan juga kurangnya kontrol terhadap Internet [3]. Menurut data Kementerian Komunikasi dan Informatika Republik Indonesia (KEMKOMINFO), penyebaran hoax di media sosial dalam tiga tahun terakhir terjadi sebanyak 9.546 kasus. Informasi hoax sengaja ditulis untuk menyesatkan pembaca [4] dan dilakukan oleh orang-orang yang tidak bertanggung jawab dengan berbagai tujuan, seperti tujuan suatu kelompok atau individu yang dapat menimbulkan dampak negatif yang serius bagi masyarakat [2]. Dengan banyaknya kerugian yang dirasakan oleh informasi hoax, memunculkan banyak penelitian untuk mendeteksi informasi hoax di media sosial, khususnya informasi yang banyak tersebar di Twitter [5].

Terdapat beberapa penelitian sebelumnya yang menggunakan berbagai model dengan menggunakan machine learning [6]–[8] dan juga menggunakan deep learning [9]–[11] untuk mendeteksi hoax. Deep learning sangat baik digunakan untuk melakukan beberapa tugas klasifikasi teks, terutama dalam mendeteksi hoax [12]. Salah satu model yang banyak digunakan adalah Recurrent Neural Network (RNN). RNN dikembangkan untuk mengatasi data yang datang secara berurutan, namun RNN memiliki keterbatasan dalam menangkap informasi jangka panjang. LSTM merupakan salah satu model dari RNN yang dikembangkan atau dimodifikasi untuk mengatasi keterbatasan RNN dalam menangkap informasi jangka panjang dengan cara mengingat informasi jangka panjang [13]. Dengan kata lain keterbatasan pada RNN, tidak menjadi masalah pada LSTM, sehingga LSTM lebih efisien dalam proses, prediksi, dan klasifikasi data. Pada penelitian [13] untuk mendeteksi hoax berbahasa Indonesia menggunakan metode Long Short-Term Memory (LSTM), hasil dari penelitian tersebut didapatkan nilai rata-rata precision, recall, f1 score sebesar 0,819; 0,809; dan 0,807 dari beberapa kali percobaan. Pada penelitian lain [9] untuk mengidentifikasi hoax pada berita Indonesia dengan menggunakan beberapa model deep learning, model LSTM menghasilkan nilai akurasi yang tinggi yaitu 95,6% tanpa menggunakan dropout, setelah menggunakan dropout nilai akurasinya sedikit meningkat menjadi 95,6%.

Studi terbaru menunjukkan bahwa model *pre-trained* yang dilatih pada korpus yang besar dapat berhasil menyelesaikan berbagai tugas dengan *transfer learning*. Salah satu *pre-trained* model yang dapat memberikan hasil terbaik untuk tugas tertentu adalah arsitektur BERT [14]. IndoBERT merupakan model berbasis *transformer* dengan menggunakan BERT [14]. Saat ini ada dua macam IndoBERT yang dilatih pada korpus yang berbeda, yang dikembangkan oleh IndoNLU [15] dan IndoLEM [16]. Pada penelitian [14] yang dilakukan untuk mendeteksi hoax mengenai Covid-19 menggunakan berbagai model BERT, model IndoBERT memberikan hasil akurasi 97,67.

LSTM dan IndoBERT merupakan model yang banyak digunakan dalam bidang Natural Language Processing (NLP). LSTM memiliki memori yang dapat mengingat informasi dalam jangka waktu Panjang, dan IndoBERT dapat melakukan paralel *input* dan *output* dalam jumlah besar, berdasarkan kemampuan kedua model tersebut dalam memproses suatu informasi sehingga kedua model ini dapat memberikan hasil akurasi yang baik dalam melakukan tugas klasifikasi teks terutama dalam mendeteksi hoax. Sehingga pada penelitian ini menggunakan LSTM dan IndoBERT dari IndoNLU dalam membuat program untuk mendeteksi informasi hoax yang menghasilkan nilai akurasi dan juga membandingkan hasil akurasi yang didapatkan dari kedua metode tersebut.

Organisasi tulisan pada jurnal tugas akhir ini setelah bagian pendahuluan adalah studi terkait, yaitu mengenai studi literatur dan juga teori yang mendukung dalam penelitian ini. Bagian ketiga berisikan penjelasan perancangan sistem yang telah dibuat. Bagian keempat, evaluasi, yaitu berisi hasil pengujian dan analisis dari hasil pengujian yang dilakukan. Bagian kelima, berisikan kesimpulan yang menjelaskan rangkuman dari seluruh rangkaian penelitian yang dilakukan.

Perumusan Masalah

Berdasarkan latar belakang pada bagian sub bab sebelumnya, rumusan masalah pada Tugas Akhir ini adalah bagaimana menentukan parameter LSTM dan IndoBERT yang optimal dalam mendeteksi hoax, dan bagaimana kinerja dari model LSTM dan IndoBERT dalam melakukan pendeteksian terhadap hoax.

Tujuan

Tujuan yang ingin dicapai dalam Tugas Akhir ini adalah dapat merancang serta mengimplementasikan suatu sistem untuk melakukan proses pendeteksian hoax terhadap suatu tweet pada Twitter, dengan menentukan

parameter yang optimal dari model LSTM dan IndoBERT, dan juga pada penelitian ini akan melihat hasil kinerja yang diberikan dari penggunaan model LSTM dan IndoBERT serta membandingkan hasil yang didapatkan, dalam mendeteksi tweet di Twitter termasuk ke dalam hoax atau tidak.