

1. Pendahuluan

1.1 Latar Belakang

Pertumbuhan pengguna Twitter setiap tahun semakin meningkat dimana pada kuartal keempat tahun 2021 pengguna Twitter bertambah 26% dari tahun sebelumnya dengan rata-rata 192 juta pengguna. Peningkatan tersebut mengakibatkan adanya lonjakan dalam penggunaan tweet dalam Twitter. Dengan meningkatnya penggunaan sosial media saat ini, tentu akan menimbulkan dampak positif dan negative. Salah satu dampak negatif dari media sosial adalah banyaknya hoax yang tersebar di media sosial, karena manusia menjadi lebih mudah untuk berinteraksi, beropini, dan bertukar informasi satu sama lain. Berdasarkan data dari Kementerian Komunikasi dan Informatika, dari sejak tanggal 30 Agustus 2021 terdapat 4.163 unggahan hoax seputar Covid-19 meliputi Facebook mencapai 3.523, Twitter 554, YouTube 49, Instagram 35 dan TikTok 2. Temuan Hoax seputar Covid-19 tersebut meningkat dari 29 Juli 2021, sebanyak 1.814 isu dengan sebaran sebanyak 4.142 unggahan, sebaran Hoax seputar Covid-19 tersebut meliputi Facebook mencapai 3.502 unggahan, Twitter 554 unggahan, Instagram 35, YouTube 49 dan TikTok 2 unggahan [1].

Hoax merupakan berita atau informasi yang berisi data yang belum diketahui kebenarannya atau yang sebenarnya tidak terjadi [8]. Sehingga akan timbul banyak persepsi pro dan kontra untuk masyarakat luas yang menggunakan media sosial mengenai suatu berita, yang dapat berdampak ke dalam kehidupan sehari-hari.

Berdasarkan latar belakang meningkatnya kasus penyebaran Hoax terutama di Indonesia mengenai berita hoax virus Covid-19 di Indonesia, maka perlu adanya penelitian terkait hal tersebut. Penelitian yang telah ada misalnya penelitian yang dilakukan Rajdev et. al yang menggunakan Functional Tree, Decision Tree dengan Naive Bayes pada leaf-nya, serta Random Forest [6]. Berdasarkan penelitian tersebut, Functional Trees menghasilkan akurasi tertinggi, yaitu 91.71%. Penelitian lainnya mengenai hoax analyzer pada berita Bahasa Indonesia menggunakan deep learning oleh Nayoga et. al [7]. Pengujian dilakukan dengan deep learning, seperti Long Short-Term Memory (LSTM), Bidirectional LSTM (BI-LSTM), Gated Recurrent Unit (GRU), Bidirectional GRU (BI-GRU), dan 1Dimensional Convolutional Neural Network (1D-CNN) serta dua classifiers, yaitu Support Vector Machine (SVM) and Naive Bayes. Dari hasil penelitian, didapatkan bahwa deep learning lebih baik dibanding supervised text classification. Penelitian lainnya mengenai A Study on Fake News Detection Using Naive Bayes, SVM, Neural Networks and LSTM oleh Reddy et. al [11]. Pengujian dilakukan dengan menggunakan Model Naive Bayes, Support Vector Machine, Neural Network dan LSTM. Penelitian menghasilkan akurasi tertinggi, yaitu 94,27% untuk LSTM.

Beberapa penelitian tersebut memiliki hasil yang berbeda dalam melakukan deteksi hoax pada media sosial Twitter. Banyak faktor yang mempengaruhi hasil tersebut seperti perbedaan model dan jumlah dataset yang digunakan. Sehingga bisa dilakukan penelitian ulang untuk mendeteksi hoax twitter mengenai virus Covid19. Dengan memilih tingkat akurasi tertinggi dari penelitian sebelumnya Model LSTM [11], dan membandingkan dengan penelitian lainnya bahwa hasil penggabungan LSTM-CNN menghasilkan akurasi yang lebih baik dibandingkan CNN-LSTM, CNN, LSTM [10]. Kontribusi dari penelitian ini yaitu, membandingkan hasil yang akan didapatkan menggunakan metode LSTM-CNN dan Word2Vec pada media sosial twitter mengenai tweet virus Covid-19.

1.2 Perumusan Masalah

Berdasarkan Berdasarkan latar belakang yang sudah dijabarkan sebelumnya, didapatkan rumusan masalah sebagai berikut:

1. Bagaimana implementasi model LSTM-CNN dan Feature Extraction Word2Vec dalam Deteksi Hoax Tweet virus Covid-19 di Indonesia pada media sosial Twitter?
2. Bagaimana hasil tingkat akurasi dari penggunaan algoritma LSTM-CNN dalam Deteksi Hoax Tweet virus Covid-19 di Indonesia pada media sosial Twitter?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah melakukan deteksi Hoax dengan analisis mengimplementasikan metode LSTM-CNN.

1.4 Topik dan Batasannya

Berikut batasan-batasan yang dimuat pada penelitian ini pada hal-hal berikut:

1. Kumpulan Dataset Twitter mengenai berita virus COVID-19 pada rentang tahun 2020-2022.
2. Kumpulan Dataset Twitter yang akan dibagi menjadi dua bagian yaitu Hoax dan non Hoax yang merupakan hasil crawling.
3. Menggunakan Analisa Tweet Bahasa Indonesia untuk penelitian ini.

1.5 Organisasi Penulisan

Bagian penelitian ini terdiri dari lima bagian, yaitu pendahuluan, studi terkait, sistem yang dibangun, evaluasi, dan kesimpulan. Pada bagian pendahuluan berisi tentang penjelasan dan uraian dari permasalahan yang ada pada penelitian tugas akhir. Bagian penelitian terkait berisi tentang penjelasan literatur atau studi yang mendukung penelitian tugas akhir. Bagian sistem yang dibangun berisi tentang penjelasan perancangan sistem yang mencakup tinjauan umum serta arsitektur sistem. Bagian evaluasi berisi tentang penjelasan hasil pengujian dan analisis berdasarkan rumusan masalah dan tujuan pada bagian sebelumnya. Bagian kesimpulan berisi tentang rangkuman dari keseluruhan hasil penelitian yang telah diuji serta saran untuk pengembangan selanjutnya..