

1. Pendahuluan

Latar Belakang

Berita palsu atau hoax adalah suatu informasi yang mengandung kepalsuan yang disengaja. Hoax dipercaya memiliki tujuan untuk menyesatkan pembaca dalam mempercayai informasi yang tidak benar. Hal tersebut dapat merugikan orang lain hingga membuat permasalahan. Dengan pengembangan teknologi saat ini, informasi dengan mudah didapatkan dan dibagikan oleh pengguna menggunakan berbagai media sosial. Menurut penelitian [1] platform media sosial yang populer sebagai sumber terjadinya hoax adalah Twitter, karena jumlah penggunanya yang banyak [2]. Dengan kebebasan berbagi informasi di Twitter, pengguna dapat menemukan ber berbagai informasi yang mengandung hal positif maupun negatif. Hal ini dapat berdampak buruk bagi masyarakat yang tidak kebal terhadap berita palsu (hoax) [3][4], karena pengguna dapat mengunggah tweetnya yang mengandung informasi palsu keseluruhan pengguna yang menggunakan platform tersebut [5]. Saat ini, berita palsu sering disalahgunakan sebagai alat untuk menggiring opini masyarakat tanpa mengetahui kebenarannya. Kasus hoax yang sering ditemukan sekarang pada media sosial adalah politik [6]. Dengan menggunakan media sosial, pihak yang bersangkutan menggunakan media sosial untuk menyebarkan hoax dengan tujuan untuk menjatuhkan pihak lawannya [6]. Tanpa disadari hal tersebut dapat memicu keresahan pada masyarakat karena hal tersebut sering dijadikan senjata oleh pihak lain untuk melakukan penipuan, pencemaran nama, *cyber bullying*, dan lainnya [5].

Penelitian mengenai deteksi hoax pada media sosial telah dilakukan dengan menggunakan metode *deep learning*. Penelitian [7], Implementasi *deep learning* yang digunakan adalah *Convolutional Neural Network* (CNN) dan *Long Short Term Memory* (LSTM), hasil akurasi yang didapatkan pada metode CNN adalah 0.88 sedangkan metode LSTM memperoleh hasil akurasi sebesar 0.84. hal ini membuktikan bahwa CNN memiliki tingkat akurasi yang tinggi dibandingkan LSTM dalam mendeteksi berita palsu. Kedua metode tersebut dibandingkan kembali dengan jumlah 2 data untuk berita fakta dan 2 untuk berita palsu, melalui perbandingan tersebut menunjukkan bahwa model CNN memprediksi data tersebut tepat sedangkan model LSTM terdapat 1 data yang terprediksi tidak tepat. Maka dari itu penelitian ini mengungkapkan bahwa CNN merupakan metode yang memiliki hasil prediksi terbaik. Penelitian [8], implementasi *deep learning* dilakukan dengan menggunakan metode LSTM, LSTM Drop, dan hybrid LSTM dengan CNN. Pada penelitian ini, digunakan dataset sebesar 5.800 tweets. Hasil pada penelitian ini menunjukkan bahwa performansi dari metode LSTM vanilla mengungguli performansi dari metode LSTM Drop dan model hybrid LSTM-CNN, yaitu akurasi sebesar 82.29, presisi sebesar 44.35, recall sebesar 40.50, and F-M sebesar 40.59. Penelitian [9] menganalisis deteksi berita hoax pada dataset berita hoax *benchmark* dengan menggunakan beberapa metode *deep learning* dengan mengkombinasikannya dengan *word embedding* GloVe, Word2Vec, dan fastText. *Word embedding* tersebut merupakan *word embedding pre-trained*, dengan arti *word embedding* tersebut sudah dilatih dengan menggunakan korpus besar untuk menghasilkan kosakata yang cukup [9]. Untuk menentukan hasil performansi dalam mendeteksi hoax, setiap *word embedding* dikombinasikan dengan menggunakan metode *deep learning* seperti CNN, Bi-LSTM, dan RestNet. Pada penelitian [9] mencoba menggunakan 4 dataset yang berbeda dengan hasil akhir didapatkan berupa Bi-LSTM mengungguli hasil CNN dan RestNet pada setiap dataset.

Berdasarkan bahasan penelitian terdahulu, membuktikan bahwa pendekatan *deep learning* sangat cocok dalam sistem deteksi hoax. Dengan menggunakan *Deep Neural Network* (DNN) dapat membuat hasil prediksi menjadi lebih baik. Namun masih ada penelitian yang belum menggunakan kombinasi ekstraksi fitur *Term Frequency-inverse Term Frequency* (TF-IDF) dengan ekspansi fitur *Global Vektor* (GloVe). Pada tugas akhir ini, berfokus pada deteksi hoax dengan mengimplementasikan model *deep learning* *Convolutional Neural Network* (CNN) dan *Recurrent Neural Network* (RNN). Penelitian ini memiliki batasan masalah berupa data yang digunakan merupakan cuitan dari media sosial Twitter dalam Bahasa Indonesia. Dataset yang digunakan sejumlah 25.325 yang diolah lebih lanjut dengan menggunakan fitur ekspansi berupa *word embedding global vector* (GloVe) dan fitur ekstraksi menggunakan *Term Frequency - Inverse Document Frequency* (TF-IDF). Topik yang diangkat untuk dijadikan dataset untuk tugas akhir ini adalah peristiwa yang terjadi pada tahun 2022, yaitu “Ferdinand Sambo” dan “Tragedi Kanjuruhan”. Tujuan tugas akhir ini adalah menghasilkan akurasi yang optimum pada sistem dengan menggunakan model CNN, RNN, dan hybrid CNN-RNN.

Susunan pada tugas akhir ini yaitu berisi pendahuluan, kemudian pada bagian kedua yaitu studi terkait yang membahas tentang studi literatur pada penelitian yang terkait pada tugas akhir ini. Bagian ketiga terdiri dari pembangunan sistem, penjelasan Fitur Ekstraksi *Term Frequency – Inverse Document Frequency* (TF-IDF), penjelasan Fitur Ekspansi *word embedding Global Vektor* (GloVe), penjelasan metode *Convolutional Neural Network* (CNN), penjelasan mengenai metode *Recurrent Neural Network* (RNN), dan penjelasan *Confusion Matrix*. Pada bagian keempat adalah terdiri dari penjelasan mengenai hasil akurasi pada penelitian beserta evaluasi penelitian. Pada bagian kelima adalah kesimpulan yang berisikan rangkuman dari penelitian yang telah dilakukan beserta saran.