

1. Pendahuluan

Media sosial berkembang mengikuti perkembangan zaman. Keefisienan penggunaan media sosial ini menjadikan wadah penyebaran informasi yang berpengaruh. Media sosial dapat mengubah cara informasi disampaikan dan juga mengubah cara isi dari informasi tersebut dapat dicerna [1].

Hoax adalah informasi yang tidak sesuai dengan yang terjadi. Hoax dapat tersebar karena pembaca hoax tersebut merasa yakin bahwa penyebar hoax dianggap kredibel, sehingga pembaca hoax menyebarluaskan tanpa mengetahui fakta [2]. Tujuan penyebaran hoax salah satunya untuk merugikan salah satu pihak demi suatu keuntungan. Hoax juga dapat digambarkan seperti kebohongan yang dikemas dengan baik agar tidak terlihat seperti kebohongan. Dampak yang diberikan akibat penyebaran hoax adalah munculnya rasa benci, curiga, dan sentimen terhadap perbedaan pendapat. Penyebaran hoax yang semakin banyak disebabkan salah satu faktornya kurangnya pengawasan dari pihak yang berwenang.

Sistem identifikasi hoax pada sosial media yang dibangun pada penelitian kali ini adalah menerapkan klasifikasi dengan dua metode, yaitu *Recurrent Neural Network* (RNN) dan *Long Short-Term Memory* (LSTM). Sebelum pengklasifikasian akan dilakukan ekstraksi fitur dengan menggunakan pembobotan kata Term Frequency-Inverse Document Frequency (TF-IDF) dan juga menggunakan ekspansi fitur GloVe. Pada penelitian sebelumnya telah dilakukan percobaan dengan metode yang sama yaitu dengan LSTM-RNN yang menghasilkan akurasi hanya sebesar 91,48% [3]. Perbedaan pada penelitian kali ini adalah layer yang digunakan. Pada penelitian sebelumnya dua metode klasifikasi digunakan menjadi satu layer, tetapi yang digunakan pada penelitian kali ini adalah layer yang digunakan akan berbeda. Pada layer RNN akan digunakan layer SimpleRNN dan untuk LSTM akan digunakan layer LSTM pula. Sehingga hasil akhir yang didapat adalah untuk membandingkan metode mana dari dua klasifikasi tersebut yang menghasilkan akurasi paling tinggi.

Tujuan dari digunakannya ekstraksi fitur dan juga pembobotan kata dengan TF-IDF adalah untuk melihat pengaruh pada penggunaan kedua metode klasifikasi.

Adapun batasan yang digunakan pada penelitian kali ini adalah pengambilan dataset pada media sosial Twitter dengan jumlah sebesar 25.324 dengan menggunakan dua pelabelan, yaitu hoax dan non-hoax. Serta untuk penggunaan n-gram akan digunakan unigram, bigram, trigram, unigram bigram, unigram trigram, dan unigram bigram trigram.

Pada penulisan jurnal tugas akhir ini akan terdiri dari beberapa bagian, yaitu bagian 1 akan dipaparkan pendahuluan, bagian 2 akan memaparkan tentang studi terkait, dan pada bagian 3 akan dipaparkan mengenai pembangunan sistem disertai penjelasan pada setiap metode yang digunakan. Pada bagian 4 akan dipaparkan hasil dari pengujian dan juga analisis dari hasil pengujian. Dan pemaparan kesimpulan serta saran untuk penelitian selanjutnya akan ada pada bagian 5.