

1. LATAR BELAKANG

Sejak pandemi, hampir semua kegiatan dilakukan secara daring yang meningkatkan penggunaan teknologi dan internet. Adanya teknologi ini semakin memudahkan manusia dalam melakukan segala aktivitas, salah satunya kemudahan dalam mendapatkan informasi. Sayangnya, hanya sebagian informasi atau berita yang terbukti kebenarannya. Banyak pihak yang menyalahgunakan hal ini untuk menyebarkan berita atau informasi bohong yang disebut juga hoax. Kesalahpahaman informasi dapat memicu banyak masalah.

Sebuah artikel berita [1] menjelaskan Kementerian Komunikasi dan Informatika (Kominfo) menyebut telah ditemukan 9.546 hoaks yang tersebar di berbagai platform media sosial di internet. Data ini dihitung sejak Agustus 2018 hingga awal 2022. Dijelaskan pula dalam artikel yang sama [1] bahwa masyarakat Indonesia mengambil sumber informasi dari media sosial dengan persentase 73% berdasarkan survei yang dilakukan oleh Kominfo. Di dalam artikel [2] Kementerian Komunikasi dan Informatika menemukan sekitar 800.000 situs di Indonesia yang terindikasi sebagai penyebar berita bohong. Individu tertentu menggunakan Internet untuk menyebarkan konten negatif yang menyebabkan kecemasan. Berdasarkan data dan penjelasan di atas dapat disimpulkan bahwa masalah penyebaran berita hoax sangatlah parah dan berbahaya. Oleh karena itu, penelitian terhadap berita bohong atau hoax harus dilakukan. Untuk mencegah masalah atau persepsi yang salah.

Penelitian tentang deteksi hoax telah dilakukan dengan menggunakan berbagai metode. Penelitian [3] menggunakan 3.300 berita Indonesia, 1.800 berita non hoax, dan 1.500 berita hoax sebagai data training. Akurasi sebesar 89,27% diperoleh dengan menggunakan metode Support Vector Machine untuk 10-fold dan akurasi sebesar 89,48% menggunakan metode Naïve Bayes untuk 5-fold [3]. Fitur perluasan dengan Word2Vec juga dapat digunakan untuk deteksi hoax, seperti yang dilakukan dalam penelitian [4]. Penelitian [4] menggunakan fitur perluasan Word2Vec dengan algoritma Logistic Regression, Support Vector Machine, Random Forest, dan sistem tanpa metode fitur perluasan Word2Vec. Penggunaan fitur ekspansi Word2Vec pada algoritma Random Forest meningkatkan akurasi sebesar 1,46% dengan nilai akurasi sebesar 89,53%. Berbeda dengan penelitian [4], penelitian [5] menggunakan fitur pembobotan TF-IDF dengan menggunakan algoritma Naïve Bayes Multinomial untuk mendapatkan akurasi sebesar 72,06%.

Dalam penelitian [6] mengenai pengertian dan analisis ciri-ciri fake news terkait sentimen untuk mendeteksi fake news. Riset tersebut diverifikasi dengan membandingkan beberapa metode pendeteksian berita palsu berbasis non-sentimen. Metode classifier yang digunakan adalah LOGIT, SVM-Linear, Decision Tree, Random Forest, XG-Boost, dan LSTM_HAN. Dengan hasil akurasi terbaik sebesar 86% menggunakan metode SVM untuk pendeteksian rumor hanya teks yang sadar sentimen. Oleh karena itu, penelitian ini akan menggunakan SVM sebagai model untuk menganalisis sentimen terhadap reply tweet.

Dalam penelitian [7], deteksi hoaks didasarkan pada analisis sentimen dengan algoritma Naïve Bayes dan analisis berbasis leksikon. Pada penelitian [7], metode yang digunakan untuk mendeteksi hoax adalah mengecek ulang komentar postingan di halaman bisnis Facebook. Dimana semua komentar pada setiap postingan akan dianalisa sentimennya, kemudian akan dilanjutkan ke proses pendeteksian hoax. Proses pendeteksian hoax adalah proses pengecekan ulang terhadap komentar postingan berdasarkan persentase poin komentar setiap postingan.

Berdasarkan penjelasan sebelumnya, pada penelitian ini penulis akan mendeteksi hoax pada tweet yang di-tweet dari Twitter berdasarkan analisis sentimen dengan algoritma Support Vector Machine. Berbeda dengan penelitian [7], pada penelitian ini proses pendeteksian hoax akan difokuskan pada pengecekan tweet postingan, bukan pada balasan/komentar tweet. Pada penelitian ini terdapat dua tugas yang berbeda, tugas pertama adalah analisis sentimen dari reply tweet, dan tugas kedua adalah pendeteksian hoax dari post tweet. Proses analisis sentimen untuk tweet reply merupakan proses untuk menentukan postingan tweet mana yang berpotensi menjadi hoax. Dari hasil tersebut, proses pendeteksian hoax pada post tweet menggunakan Bag of Words

untuk pembobotan kata untuk menentukan label “hoax” atau “non-hoax”. Penelitian ini berkontribusi dalam mendeteksi berita hoax di Twitter dengan dua tipe data labeling yang berbeda untuk proses analisis sentimen dan deteksi hoax menggunakan Support Vector Machine.

Struktur laporan adalah sebagai berikut. Bagian 2 merupakan penjelasan penelitian terkait dan metode yang digunakan. Bagian 3 merupakan penjelasan dari diagram sistem yang dibangun untuk penelitian. Bagian 4 merupakan penjelasan hasil penelitian dan evaluasi. Bagian 5 merupakan kesimpulan dari hasil penelitian ini.