

1. Pendahuluan

Menghasilkan dan menyebarkan informasi menjadi lebih mudah diakses dan lebih cepat. Platform media sosial menyediakan informasi yang bermanfaat untuk menginformasikan berita terbaru kepada masyarakat. Meskipun memiliki banyak manfaat, media sosial menghadirkan beberapa tantangan karena adanya informasi yang salah, rumor, pernyataan palsu, sindiran, dan berita palsu [1]. Indonesia merupakan salah satu negara yang sering dilanda masalah terkait berita palsu. Hal ini didukung oleh survei yang dilakukan oleh Tim AIS (Kementerian Komunikasi dan Informatika) pada bulan Agustus 2018 hingga 30 September 2019, ditemukan lebih dari 3.356 hoaks di Indonesia dalam kurun waktu satu periode. 3.356 hoaks tersebut terdiri dari 900 hoaks terkait isu politik dan 500 hoaks terkait pemilu. Terkait dengan permasalahan di atas, dibutuhkan sebuah metode yang dapat mendeteksi berita palsu. Metode yang diusulkan adalah machine learning dan hybrid deep learning, dimana model ini memiliki potensi yang besar untuk analisis sentimen dan klasifikasi teks [2], [3].

Beberapa penelitian telah mengembangkan sistem pendeteksi berita palsu dengan model deep learning dan machine learning. Abdullah All-Tanvir [4] menggunakan metode Long Short-Term Memory (LSTM) dan Support Vector Machine (SVM) untuk mendeteksi berita palsu di Twitter, sedangkan Xianyang [5] mengusulkan klasifikasi teks dengan menggunakan model Word2Vec sebagai embedding untuk kombinasi Convolutional Neural Network (CNN) dan LSTM. [6]-[8] Menggabungkan dua (atau lebih) metode dengan menggunakan masing-masing kelebihan dari kedua metode tersebut dan mengisi beberapa kelemahan dari masing-masing metode. Pembelajaran mesin tunggal atau pembelajaran mendalam relatif dapat diandalkan jika diterapkan pada kasus-kasus tertentu. Metode CNN membutuhkan lebih sedikit hiperparameter dan lebih sedikit kontrol daripada LSTM. Di sisi lain, proses LSTM membutuhkan waktu yang lebih lama tetapi berkinerja baik untuk kalimat yang lebih panjang [9]. SVM memiliki banyak kelebihan dibandingkan dengan pengklasifikasi lainnya, seperti kecepatan, toleransi terhadap data yang berisik, dan akurasi yang tinggi [10].

Disinformasi di media sosial menjadi tantangan tersendiri karena sumber data yang beragam, termasuk topik dan bahasa yang berbeda. Berdasarkan hal tersebut, model deep learning dan metode supervised learning, yaitu CNN, LSTM, dan SVM, memiliki potensi untuk mendeteksi berita palsu di Twitter. CNN dan LSTM dapat secara efektif mengurangi panjang urutan dengan mengekstraksi fitur-fitur penting dan menangkap pola dan ketergantungan temporal [5], sementara SVM dapat secara efektif memanfaatkan fitur-fitur ini dan mempelajari batas keputusan yang diskriminatif. Selain itu, penggunaan Word2Vec memberikan model hibrida dengan pemahaman semantik yang lebih baik, pengurangan dimensi yang efisien, generalisasi untuk kata-kata yang tidak terlihat, dan kemampuan untuk menangkap hubungan sintaksis dan semantik. Oleh karena itu, penelitian ini mengusulkan Hybrid Deep Supervised Learning yang dapat secara efektif menggabungkan dan mengintegrasikan informasi dari berbagai modalitas. Penelitian ini membandingkan varian individual deep learning dan hybrid deep learning dengan metode supervised learning menggunakan dataset berita politik Indonesia dan memvalidasi efisiensi klasifikasi.