

ABSTRAK

Penyebaran berita palsu dalam berbagai bahasa telah menjadi tantangan yang signifikan dalam memastikan keaslian informasi di era digital. Penelitian ini menyelidiki pendekatan pembelajaran hibrida untuk mendeteksi berita palsu lintas bahasa, dengan fokus pada bahasa Indonesia sebagai bahasa sumber dan mengevaluasi kinerja di seluruh bahasa target yaitu bahasa Melayu, Jerman, Spanyol, dan Inggris. Model deep learning (DL) hibrida, yang menggabungkan arsitektur CNN dan LSTM, dipilih karena kekuatannya yang saling melengkapi: CNN secara efektif menangkap pola lokal dan ketergantungan kata, sementara LSTM unggul dalam pemodelan informasi sekuensial dan ketergantungan jangka panjang dalam teks. Kombinasi ini memungkinkan model hibrida untuk memanfaatkan fitur spasial dan temporal, sehingga sangat cocok untuk tugas-tugas multibahasa dan lintas-bahasa di mana struktur teks sangat bervariasi. Untuk menyelaraskan representasi lintas bahasa, penelitian ini menggunakan penyematan MUSE, yang menyelaraskan vektor kata dalam ruang multibahasa bersama. Hasil eksperimen menunjukkan bahwa penggunaan embedding MUSE secara signifikan meningkatkan kinerja deteksi, terutama dengan memungkinkan model untuk menggeneralisasi dengan lebih baik di seluruh set data yang beragam secara linguistik. Temuan ini mengungkapkan bahwa kesamaan linguistik antara bahasa sumber dan bahasa target meningkatkan kinerja deteksi, dengan CNN yang unggul dalam pasangan bahasa yang mirip, seperti bahasa Indonesia-Melayu, mencapai F1-score 82%. Sementara LSTM berkinerja baik dalam bahasa yang berbeda secara struktural, terutama dalam pasangan bahasa Indonesia-Jerman yang mencapai 97% dan juga pada LSTM-CNN, mencapai 95%. Selain itu, penelitian ini menyoroti bahwa memilih bahasa selain bahasa Inggris sebagai bahasa sumber dapat memberikan hasil yang menjanjikan, terutama dalam skenario dengan sumber daya yang terbatas. Penelitian ini menggarisbawahi potensi model hibrida dan penyematan lintas bahasa dalam memajukan deteksi berita palsu lintas bahasa, sekaligus mengidentifikasi tantangan dan area untuk perbaikan dengan bahasa yang menunjukkan perbedaan struktural yang lebih besar.

Kata kunci: Lintas Bahasa, Deteksi Berita Palsu, Pembelajaran Hibrida, Penyematan MUSE, Kesalahan Informasi Digital