

ABSTRACT

The spread of fake news in multiple languages has become a significant challenge in ensuring information authenticity in the digital era. This research investigates a hybrid learning approach for cross-lingual fake news detection, focusing on Indonesian as the source language and evaluating the performance across the target languages of Malay, German, Spanish, and English. The hybrid deep learning (DL) model, which combines CNN and LSTM architectures, was chosen due to its complementary strengths: CNN effectively captures local patterns and word dependencies, while LSTM excels in modeling sequential information and long-term dependencies in text. This combination allows the hybrid model to take advantage of both spatial and temporal features, making it well-suited for multilingual and cross-lingual tasks where text structures vary widely. To harmonize the representation across languages, this study uses MUSE embedding, which aligns word vectors in a shared multilingual space. Experimental results show that the use of MUSE embedding significantly improves detection performance, mainly by allowing the model to generalize better across linguistically diverse data sets. The findings reveal that linguistic similarity between the source and target languages improves detection performance, with CNNs excelling in similar language pairs, such as Indonesian-Malay achieving F1-score 82%. While LSTMs perform well in structurally different languages, especially in the Indonesian-German language pair achieved 97% and also in LSTM-CNN, achieved 95%. Additionally, this study highlights that choosing a language other than English as the source language can provide promising results, especially in scenarios with limited resources. This research underscores the potential of hybrid models and cross-lingual embedding in advancing cross-lingual fake news detection while identifying challenges and areas for improvement with languages that exhibit greater structural differences.

Keywords: Cross-Lingual, Fake News Detection, Hybrid Learning, MUSE Embeddings, Digital Misinformation