

ABSTRAK

Penyebaran berita *hoax* telah menjadi ancaman serius dalam era informasi digital saat ini. Penyebaran berita *hoax* atau berita palsu dapat memiliki dampak yang merugikan dan berbahaya, baik bagi individu maupun masyarakat secara keseluruhan. Penyebaran Berita palsu dapat mempengaruhi persepsi dan pemahaman orang terhadap suatu peristiwa atau isu. Tujuan penelitian ini adalah untuk menerapkan sebuah model yang efektif dalam mengenali dan membedakan teks berita *hoax* dari teks berita *non-hoax*. metode TF-IDF digunakan untuk menghasilkan *Vector Space Model* dari teks berita, dengan mempertimbangkan frekuensi kemunculan kata dan bobotnya dalam korpus keseluruhan. Kemudian, arsitektur LSTM dipakai sebagai model klasifikasi yang mampu memahami konteks jangka panjang dalam teks. Tahapan penelitian meliputi pengumpulan data berita *hoax* dan *non-hoax* dari Kaggle (*fake.csv* dan *true.csv*), preprocessing teks (membersihkan dan mengubah teks menjadi representasi vektor), pelatihan model LSTM, dan evaluasi kinerja model menggunakan confusion matrix. Setelah pengujian, kombinasi parameter LSTM dengan akurasi tertinggi sebesar 99,65% dicapai pada skenario uji coba 14 (80:20, neuron hidden layer 16, epoch 100, kompilasi model dengan *binary crossentropy* sebagai fungsi *loss*, adam sebagai *optimizer*, dan metrik akurasi untuk evaluasi) dengan penerapan preprocessing yakni *tokenizing*, *stopword removal*, dan *stemming*. Hasil ini menunjukkan bahwa model klasifikasi yang dikembangkan sangat akurat dalam mendeteksi berita *hoax*

Kata kunci: klasifikasi teks, berita *hoax*, TF-IDF, LSTM, deteksi berita *hoax*

ABSTRACT

The spread of hoax news has become a serious threat in today's digital information era. The dissemination of hoax news or false news can have harmful and dangerous impacts, both for individuals and society as a whole. The spread of fake news can influence people's perceptions and understanding of an event or issue. The aim of this research is to implement an effective model to recognize and distinguish hoax news texts from non-hoax news texts. The TF-IDF method is used to generate a Vector Space Model from the news texts, considering the frequency of word occurrences and their weights in the overall corpus. Then, the LSTM architecture is employed as a classification model capable of understanding long-term context in the text. The research stages include collecting hoax and non-hoax news data from Kaggle (fake.csv and true.csv), preprocessing the text (cleaning and converting the text into vector representation), training the LSTM model, and evaluating the model's performance using a confusion matrix. After testing, the combination of LSTM parameters with the highest accuracy of 99.65% was achieved in trial scenario 14 (80:20 split, hidden layer neurons 16, 100 epochs, compiling the model with binary crossentropy as the loss function, Adam as the optimizer, and accuracy as the evaluation metric) with preprocessing applied, including tokenizing, stopwords removal, and stemming. These results show that the developed classification model is highly accurate in detecting hoax news.

Key word: text classification, hoax news, TF-IDF, LSTM, hoax news detection