INTISARI

Penyebaran berita hoaks melalui media daring di Indonesia kian mengkhawatirkan, mendorong perlunya sistem deteksi otomatis yang andal. Penelitian ini mengusulkan pendekatan optimasi deteksi berita hoaks berbahasa Indonesia dengan memanfaatkan Global Vectors for Word Representation (GloVe) sebagai metode word embedding dalam arsitektur Long Short-Term Memory (LSTM). Dataset terdiri dari 6.258 berita (3.107 hoaks dan 3.151 valid) yang diperoleh dari Kaggle dan Kompas. Untuk mengamati pengaruh penyeimbangan data terhadap performa model, diterapkan teknik Synthetic Minority Oversampling Technique (SMOTE), meskipun dataset awal relatif seimbang. Dua model LSTM dibangun, yaitu model dengan data asli dan model dengan data yang telah melalui SMOTE. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil evaluasi menunjukkan bahwa model LSTM dengan GloVe tanpa SMOTE menghasilkan performa terbaik dengan akurasi 94,73%, presisi 93,06%, recall 96,92%, dan F1-score 94,95%. Sementara model dengan SMOTE menunjukkan performa yang stabil namun sedikit lebih rendah di semua metrik evaluasi. Hasil penelitian ini mengindikasikan bahwa integrasi GloVe dalam arsitektur LSTM efektif untuk deteksi berita hoaks, dan penyeimbangan data dengan SMOTE pada dataset yang hampir seimbang tidak selalu meningkatkan performa model.

Kata kunci: klasifikasi teks, LSTM, GloVe, berita hoaks, SMOTE.

ABSTRACT

The spread of hoax news through online media in Indonesia is increasingly alarming, necessitating a reliable automated detection system. This study proposes an optimized approach for detecting Indonesian-language hoax news by utilizing Global Vectors for Word Representation (GloVe) as a word embedding method within a Long Short-Term Memory (LSTM) architecture. The dataset consists of 6,258 news articles (3,107 hoaxes and 3,151 valid), collected from Kaggle and Kompas. To observe the impact of data balancing on model performance, the Synthetic Minority Oversampling Technique (SMOTE) was applied, although the original dataset was relatively balanced. Two LSTM models were developed: one using the original data and the other using SMOTE-balanced data. The models were evaluated using accuracy, precision, recall, and F1-score metrics. Evaluation results show that the LSTM model with GloVe without SMOTE achieved the best performance, with an accuracy of 94.73%, precision of 93.06%, recall of 96.92%, and F1-score of 94.95%. Meanwhile, the model with SMOTE showed stable but slightly lower performance across all evaluation metrics. These findings indicate that the integration of GloVe is effective for hoax news detection, and applying SMOTE to an already balanced dataset does not necessarily improve model performance.

Keywords: text classification, LSTM, GloVe, hoax news, SMOTE.