

ABSTRAK

Permasalahan utama dalam penelitian ini adalah masih terbatasnya penelitian dan pengembangan teknologi pemrosesan bahasa alami (NLP) yang secara khusus mendukung bahasa daerah seperti bahasa Jawa. Meskipun bahasa Jawa memiliki jumlah penutur yang sangat besar, penggunaannya dalam konteks teknologi digital masih minim, termasuk dalam pengklasifikasian kelas kata. Keterbatasan ini menyebabkan sulitnya pengembangan sistem berbasis NLP seperti penerjemah otomatis atau chatbot dalam bahasa Jawa. Penelitian ini bertujuan untuk menerapkan algoritma *Long Short Term Memory* (LSTM) dalam melakukan pelabelan kelas kata *Part of Speech Tagging* pada teks berbahasa Jawa. Data yang digunakan berasal dari repositori *Universal Dependencies CSUI* yang berisi 1000 kalimat bahasa Jawa dengan 14.000 kata yang telah diberi anotasi. Untuk melatih dan mengevaluasi model, digunakan metode *K-Fold Cross Validation* sebanyak 15 *fold* agar hasil yang diperoleh lebih akurat dan merata. Dalam eksperimen ini, dua jenis representasi kata *word embedding* digunakan, yaitu *FastText* dan *BERT*. Hasil pengujian menunjukkan bahwa model LSTM yang menggunakan *FastText* memberikan hasil paling baik, dengan akurasi tertinggi mencapai 86,75%, jauh lebih unggul dibandingkan dengan penggunaan *BERT*. Hasil ini mengindikasikan bahwa *FastText* lebih mampu menangkap makna kata dalam konteks bahasa Jawa. Diharapkan, penelitian ini dapat mendukung pengembangan teknologi NLP untuk bahasa Jawa di masa mendatang.

Kata Kunci: Bahasa Jawa, *POS Tagging*, *Long Short Term Memory* (LSTM), *Word Embedding*, *FastText*, *BERT*, *Natural Language Processing* (NLP)

ABSTRACT

The main issue addressed in this study is the limited research and development of Natural Language Processing (NLP) technologies that specifically support regional languages such as Javanese. Despite having a large number of native speakers, Javanese remains underrepresented in digital technology applications, including word class classification. This limitation hinders the development of NLP-based systems such as automatic translators and chatbots for the Javanese language. This study aims to implement the Long Short-Term Memory (LSTM) algorithm for Part-of-Speech (POS) tagging on Javanese texts. The dataset used originates from the Universal Dependencies CSUI repository, which contains 1,000 annotated Javanese sentences totaling 14,000 words. To train and evaluate the model, 15-fold cross-validation was employed to ensure more accurate and reliable results. Two types of word embedding representations—FastText and BERT—were applied in the experiments. The results show that the LSTM model utilizing FastText achieved the best performance, reaching a maximum accuracy of 86.75%, significantly outperforming the BERT-based model. These findings indicate that FastText is more effective at capturing word meanings in the context of the Javanese language.

Keywords: Javanese Language, POS Tagging, Long Short-Term Memory (LSTM), Word Embedding, FastText, BERT, Natural Language Processing (NLP)

