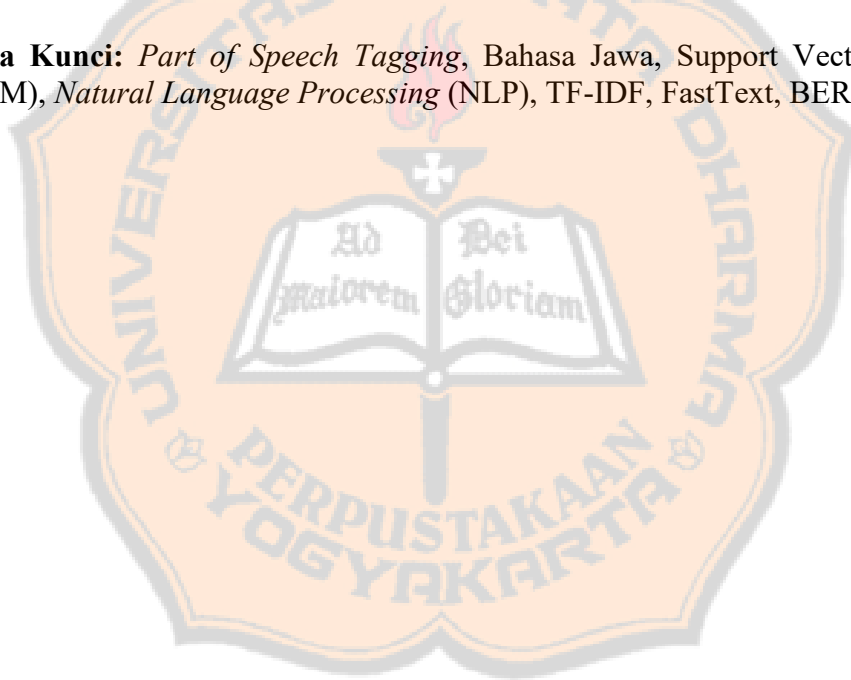


ABSTRAK

Penelitian ini mengkaji penerapan teknik Part Of Speech Tagging untuk Bahasa Jawa menggunakan metode Support Vector Machine (SVM). Bahasa Jawa, yang digunakan oleh mayoritas penduduk Pulau Jawa, memiliki potensi besar dalam bidang Natural Language Processing (NLP). Dalam penelitian ini, data Bahasa Jawa yang digunakan berasal dari Universal Dependencies (UD) Javanese-CSUI, yang mencakup 1000 kalimat dan 14.000 kata dengan anotasi manual. Fitur yang digunakan meliputi vektor TF-IDF kata, vektor gabungan TF-IDF dengan atribut lainnya, serta pre-trained embeddings seperti FastText dan BERT. Data yang telah diproses kemudian dibagi untuk klasifikasi menggunakan empat jenis kernel SVM: linear, RBF, polynomial, dan sigmoid. Hasil penelitian menunjukkan bahwa kombinasi fitur vektor gabungan TF-IDF dan atribut lainnya dengan kernel linear menghasilkan akurasi tertinggi sebesar 87,42%, dengan precision 87,40%, recall 87,42%, dan F1-score 87,28%. Hasil penelitian diharapkan dapat memberikan kontribusi signifikan dalam pengembangan teknologi NLP untuk Bahasa Jawa, serta menjadi dasar bagi penelitian lanjutan di masa depan.

Kata Kunci: *Part of Speech Tagging*, Bahasa Jawa, Support Vector Machine (SVM), *Natural Language Processing* (NLP), TF-IDF, FastText, BERT.



ABSTRACT

This study examines the application of Part Of Speech Tagging techniques for Javanese using the Support Vector Machine (SVM) method. Javanese, which is spoken by the majority of the population of Java, has great potential in the field of Natural Language Processing (NLP). In this study, the Javanese language data used was obtained from Universal Dependencies (UD) Javanese-CSUI, which includes 1,000 sentences and 14,000 words with manual annotations. The features used include TF-IDF word vectors, combined TF-IDF vectors with other attributes, and pre-trained embeddings such as FastText and BERT. The processed data was then divided for classification using four types of SVM kernels: linear, RBF, polynomial, and sigmoid. The results showed that the combination of TF-IDF vector features and other attributes with a linear kernel produced the highest accuracy of 87.42%, with precision of 87.40%, recall of 87.42%, and F1-score of 87.28%. The results of this study are expected to contribute significantly to the development of NLP technology for Javanese and serve as a basis for further research in the future.

Keywords : Part of Speech Tagging, Javanese, Support Vector Machine (SVM), Natural Language Processing (NLP), TF-IDF, FastText, BERT.

