

PART-OF-SPEECH TAGGING PADA BAHASA JAWA MENGUNAKAN MODEL PRE-TRAINED BIDIRECTIONAL ENCODER REPRESENTATION from TRANSFORMERS

Nama Mahasiswa	: Muhammad Alvin Ajry
NIM	: 215430017
Pembimbing I	: Ahmad Izzuddin, S.T., M.Kom
Pembimbing II	: Nuzul Hikmah, S.T., M.Kom

ABSTRAK

Part-of-Speech Tagging (POS tagging) merupakan proses penentuan kelas kata dalam suatu teks yang penting dalam pemrosesan bahasa alami (*Natural Language Processing*). Pada bahasa Jawa, *POS tagging* masih merupakan tantangan karena keterbatasan sumber daya linguistik dan kompleksitas bahasa tersebut. Dengan perkembangan teknologi *deep learning*, metode *fine-tuning* BERT (*Bidirectional Encoder Representations from Transformers*) telah diterapkan untuk melakukan *POS tagging* dalam bahasa Jawa, yang merupakan bahasa dengan sumber daya terbatas. Model *javanese-bert-small* dilatih menggunakan dataset UD_Javanese-CSUI, dan dievaluasi menggunakan metrik *precision*, *recall*, *F1-score*, dan *accuracy*. Hasil penelitian menunjukkan bahwa model mencapai performa mumpuni dengan akurasi tercapai 88,87%, serta menunjukkan kestabilan selama pelatihan tanpa *overfitting* signifikan. Temuan ini menunjukkan bahwa pendekatan berbasis BERT efektif untuk menangani ambiguitas kelas kata dalam bahasa Jawa dan dapat menjadi pijakan untuk pengembangan lebih lanjut dalam sistem NLP untuk bahasa daerah.

Kata kunci : bahasa Jawa, *deep learning*, BERT, *Part-of-Speech tagging*

***PART-OF-SPEECH TAGGING IN JAVANESE
USING PRE-TRAINED BIDIRECTIONAL ENCODER
REPRESENTATION from TRANSFORMERS MODEL***

<i>Student Name</i>	: Muhammad Alvin Ajry
<i>NIM</i>	: 215430017
<i>Advisor I</i>	: Ahmad Izzuddin, S.T., M.Kom
<i>Advisor II</i>	: Nuzul Hikmah, S.T., M.Kom

ABSTRACT

Part-of-Speech Tagging (POS tagging) is the process of determining word classes in a text that is important in natural language processing. In Javanese, POS tagging is still a challenge due to limited linguistic resources and the complexity of the language. With the development of deep learning technology, the BERT (Bidirectional Encoder Representations from Transformers) fine-tuning method has been applied to POS tagging in Javanese, which is a language with limited resources. The javanese-bert-small model was trained using the UD_Javanese-CSUI dataset, and evaluated using precision, recall, F1-score, and accuracy metrics. The results showed that the model achieved good performance with an accuracy of 88,87%, and showed stability during training without significant overfitting. These findings indicate that the BERT-based approach is effective in handling word class ambiguity in Javanese and can be a stepping stone for further development in NLP systems for regional languages.

Keywords : BERT, deep learning, Javanese language, Part-of-Speech tagging