ABSTRACT

MODEL PERFORMANCE OF NAMED ENTITY RECOGNITION USING THE BLSTM-HMM METHOD ON INDONESIAN-LANGUAGE TWITTER DOCUMENTS

By

Adi Wijaya

19102099

Twitter is a popular *microblogging* platform in Indonesia, providing text data in Indonesian language that can be utilized for *Natural Language Processing* (*NLP*) applications. *Named Entity Recognition* (*NER*) is a crucial subtask in *Information Extraction* (*IE*) aiming to identify named entities such as persons, locations, and organizations within text. Limitations in Indonesian *NER* research on the Twitter platform have prompted the urgency to develop optimized *NER* models. NER plays a significant role in *IE* and *NLP* in general, enabling automatic extraction of useful information from Twitter text. The objective of this research is to determine the *F1 score* performance value in the *NER* model using the *BLSTM-HMM* method, with *FastText* and *POS tag* features. Also, to ascertain whether the application of these methods can enhance the *F1 score* performance of the *NER* model. This study implements the *Bidirectional Long Short-Term Memory* (*BLSTM*) – *Hidden Markov Model* (*HMM*) method with *FastText* features and *POS tags* to enhance the performance of *NER* models on Indonesian-language Twitter document data. The developed model achieved an F1 score of 89.23%. The combination of the *NER* model with *BLSTM-HMM* and the addition of *FastText* and *POS tag* features can be an interesting alternative to enhance entity recognition performance. In this study, this combination successfully yielded positive results.

Keyword: ***BLSTM, FastText, HMM, Named Entity Recognition, POS tag***