PROJECT REPORT

APPLICATION OF DEEP LEARNING FOR CLASSIFICATION OF INDONESIAN LANGUAGE ONLINE NEWS

Nikodemus Karuna Wijaya
22.K1.0011

Faculty of Computer Science
Soegijapranata Catholic University
2026

ABSTRACT

Rapidly increasing online news distribution in Indonesia increases the risk of misinformation, including in the form of clickbait titles and hoax news. This conditions negatively affect the formation of public opinion, reduces the credibility of information, and is a severe challenge for digital literacy. According to the literature review, transformer-based models such as M-BERT and IndoBERT perform well on text classification tasks, but still suffer from data imbalance handling limitations, fine-tuning optimization, and comparative analysis between multilingual and monolingual models. This paper aims to complement those limitations by building a deep learning-based clickbait and hoax news classification system in Indonesian. The proposed method applies fine-tuning to M-BERT and IndoBERT models and uses Focal Loss to make the model more sensitive to minority classes. This dataset is taken from several open sources containing the titles and contents of clickbait and hoax news in Indonesian. Some preprocessing stages are case folding, removing URLs, and removing symbols. Those steps are intended to maintain the quality of text representation according to the BERT tokenizer characteristics. This study tested six experimental configurations that combined two transformer models with three loss functions, namely, Cross Entropy, Weighted Cross Entropy, and Focal Loss. The evaluation of the performance of these deep learning models uses accuracy, precision, recall, and F1-score. Experimental results using IndoBERT along with Focal Loss provided the best performance, with an F1-score of 0.91 in clickbait classification and 0.89 in hoax news classification, proving that the monolingual transformer model optimized with Focal Loss can be an alternative for automatic misinformation detection in Indonesian-language online news.

***Keywords**: clickbait detection, hoax classification, IndoBERT, M-BERT, Focal Loss*