ABSTRACT

Legal judgment documents from the Supreme Court of Indonesia typically exhibit long textual structures composed of various essential sections, thus requiring an effective method to handle lengthy texts and categorize types of criminal sentences. This study proposes a classification approach based on fine-tuning the Hierarchical BERT model, combined with a BiLSTM-Attention Gated Mechanism architecture. Previous approaches using Word2Vec word embeddings and various deep learning methods were considered suboptimal, as Word2Vec fails to capture contextual relationships between words in a sentence. In this study, experiments were conducted using three pooling techniques—CLS pooling, mean pooling, and max pooling—for feature representation from BERT outputs. The results show that CLS pooling yields the best performance across both models. The fine-tuned Hierarchical BERT model achieved an accuracy of 80.25%, while the combination of Hierarchical BERT with BiLSTM-Attention Gated Mechanism reached the highest accuracy of 82.57%. This performance marks a significant improvement compared to previous approaches, such as Word2Vec with BiLSTM and Attention, which achieved 65.81%, and studies using Hierarchical BERT with mean pooling, which obtained 79.88%. The improvement in accuracy is influenced not only by the model architecture but also by the fine-tuning strategies tailored to the characteristics of Indonesian legal documents. Although the addition of BiLSTM and the Attention Gated Mechanism does not produce a drastic accuracy gain, this method still contributes to a better understanding of contextual and inter-segment relationships within long documents.

Keywords : BERT, BiLSTM, hyperparameters, legal documents, sentence length prediction,