

ABSTRACT

The Supreme Court plays a crucial role in Indonesia's judicial system, with the number of cases increasing every year. This has led to a rapid growth in the number of court ruling documents, which contain important legal information, including case categories and sentencing levels. However, legal documents are generally very lengthy, have a complex structure, and contain a large amount of interrelated information, making standard BERT-based natural language processing models limited in capturing relationships between different sections of the document. Conventional BERT models have a token length limitation, causing a loss of context when applied directly to long legal documents. Additionally, standard approaches such as neural networks cannot optimally identify critical information in legal texts, such as key sentences that influence a judge's decision. This study aims to develop a classification model for sentencing levels into the categories mild, moderate, heavy, and very heavy using the Fine-tuning Hierarchical BERT + Attention approach, designed to handle long and complex legal documents more effectively. The proposed method consists of three main stages: (1) document segmentation using a chunking technique with overlapping words, (2) progressive learning with an *attention* mechanism to capture essential legal information, and (3) *Hierarchical* feature integration with positional *embeddings* to preserve contextual relationships across document sections. Additionally, fine-tuning is performed using Grid Search for hyperparameter optimization and freeze layers to prevent overfitting. Experimental results show that the Fine-tuning *Hierarchical BERT + Attention* model achieves 82% accuracy, an improvement over the standard BERT fine-tuning model, which only achieved 81% accuracy. Similar improvements are observed in precision, recall, and F1-score, indicating the model's effectiveness in handling Indonesian legal documents. However, strategies to address overfitting still require refinement, such as further regulation of frozen layers or the implementation of a more optimal dropout technique.

Keywords: Attention Mechanism, Fine-tuning, Hierarchical BERT, Legal Document Classification, Supreme Court, Natural Language Processing