ABSTRACT

The management of legal data is crucial to supporting an efficient legal information system. Still, the complexity of legal language, the diversity of document structures, and the large volume of data pose significant challenges in the automatic classification process. This study aims to optimize the DistilBERT model through a fine-tuning approach using a multi-task learning scheme to predict two labels simultaneously: the status of the regulation (valid/invalid) and the type/form of the regulation. The research stages include data collection, preprocessing, model training, and model evaluation. The model achieved high performance on both classification tasks, with an accuracy of 96%, precision of 94%, recall of 96%, and an F1-score of 94% for regulation status classification, as well as perfect results of 100% on all evaluation metrics for regulation type/form classification, demonstrating the model's accuracy and reliability in comprehensively understanding and classifying legal documents. These findings confirm that this optimized model is highly reliable in classifying the status and type of regulations.

Keywords: Classification, DistilBERT, Fine-tuning, Natural Language Processing, Optimization.