



PROJECT REPORT
HYPERPARAMETER ANALYSIS FOR FINE-TUNING
BERT-BASE-CASED IN CURRICULUM VITAE
INFORMATION EXTRACTION

ALBERT CHRISTIAN WIBOWO
22.K1.0008

Faculty of Computer Science
Soegijapranata Catholic University
2025

ABSTRACT (ABSTRACT TITLE)

This research focuses on analyzing hyperparameter tuning in the fine-tuning process of a BERT-based model for information extraction from Curriculum Vitae (CV) datasets using Natural Language Processing (NLP), specifically Named Entity Recognition (NER). The main objective of this study is to investigate the influence of several hyperparameters, including batch size, learning rate, number of training epochs, and maximum sequence length, on the performance of the BERT model in extracting information from unstructured CV text.

The model is developed using a Kaggle dataset consisting of 5,029 annotated CV samples. The research methodology includes dataset preparation, data preprocessing, tokenization, model fine-tuning, hyperparameter experimentation, and model evaluation. Multiple experiments are conducted by varying the hyperparameter configurations to observe their impact on the model's learning behavior and extraction performance.

Experimental results shows that different hyperparameter settings could affect the performance of the model. The study identifies an optimal combination of hyperparameters that improves the model's ability to extract relevant information from CV documents dataset. Furthermore, the results demonstrate that hyperparameter tuning can provide better performance compared to the default BERT configuration, as reflected in improved evaluation metrics such as precision, recall, and F1-score.

Keyword: Natural Language Processing, Named Entity Recogniton, Information extraction, BERT-based Model, Hyperparameter Tuning