ABSTRACT
HYPERPARAMETER OPTIMIZATION IN CONVOLUTIONAL NEURAL NETWORK (CNN) FOR CLASSIFYING SCRIPT TYPES

By
Firda Kamal
NIM : 20102125

The use of regional languages is declining, especially among the Post-gen Z generation. According to 2023 data from BPS, only 61.7% of Post-gen Z individuals can speak regional languages in their neighborhood or with relatives. This decline is due to a lack of understanding of regional scripts, which can be addressed through information technology by classifying the scripts. One suitable algorithm for this classification is CNN. CNN is an artificial neural network capable of recognizing predictive information such as images. However, CNN models are prone to overfitting, where the model performs well on training data but not on new data. To address this, CNN hyperparameters were optimized to improve accuracy and reduce the likelihood of overfitting. The hyperparameters used include filter kernels to focus the model on important data components, learning rate to control CNN performance, and dropout to randomly remove some neurons during training. This study tested 66 hyperparameter schemes and selected the best results based on testing accuracy. The data used included 4400 images of Javanese, Balinese, Sundanese, and Lontara scripts after augmentation, and 400 new images for testing. The best results were obtained from the scheme with a 5x5 kernel filter, a dropout of 0.6, and no learning rate adjustment, yielding a training accuracy of 0.9910 and a testing accuracy of 0.765, with a 55.69% reduction in overfitting compared to the model without hyperparameter optimization.

Keywords: CNN, hyperparameters, optimization, overfitting, scripts