Abstract

Nowadays many examples are commonly used in the development of deep learning problems, such as Transformers. This research uses one of transformers architecture namely IndoBERT, IndoBERT itself consists of BERT namely Bidirectional Encoder Representations from Transformers which is commonly used for deep learning problems. IndoBERT method is implemented to detect the use of abusive sentences in Indonesian text. The dataset used in this study has an imbalance in the amount of data in each class so that there will be additional data to find out the effect of increasing the amount of data on the performance of the architecture results. The stages of work in this research started from dataset, data pre-processing, modeling with IndoBERT method to detect abusive sentences, training and testing. Testing was carried out on the architecture of KNN, SVM, Naive Bayes, BERT Multilingual Base and BERT Base and then compared to IndoBERT. The test results showed that in addition to BERT Multilingual Base, BERT Base and IndoBERT can only predict against the majority class so that the addition of dataset usage is carried out. The test results showed IndoBERT could be better at classifying abusive sentences in Indonesian texts. In the BERT Base model, successfully generating an F1 Score for all classes of 0.6842 IndoBERT can already produce an F1 Score for all classes better.

Keywords : Abusive Sentences, IndoBERT, F1 Score, SVM, Naive Bayes, BERT