

Abstract

Daily use of the internet can be seen increasing from year to year. The activities carried out are also varied, and one of them is giving comments on a post. This comment has a unique role, which will represent someone's thoughts from the posts they comment on. The content of each comment varies, but there will be problems when the comments are abusive. Commenting in abusive language can give a bad impression both to the reader of the comment and to the creator of the post. Because of this, many studies have made the detection of offensive language using a variety of methods, from machine learning to deep learning. But in the Indonesian commentary, it is still little or difficult to find detection of offensive language using deep learning methods. So that in this study, a deep learning method was developed to detect abusive language, namely the Bidirectional Encoder Representational from Transformers (BERT). The model used is a selfdesigned BERT and a multilingual BERT pre-train model to become baseline. The system will receive input in the form of comment text which will then issue a label to classify the comment text, whether it includes Offensive, Normal, or Non Offensive. The results of the Scratch model trained with the Indonesian language dataset got a Macro Average F1 Score of 50% compared to the BERT Multilingual of 54%.

Keywords: abusive language, BERT, comment, detection, news