

Abstract

Social media has become everyone's most used communication and interaction tool in recent years. Direct interaction between humans is decreasing because people tend to communicate indirectly through smartphones. Thus, it is quite difficult to recognize someone's personality. The Big Five method is a theory that can be used to determine a person's personality by means of OCEAN analysis, namely Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. In this study, the Convolutional Neural Network (CNN) method was used to detect Big Five personalities by utilizing data generated by Twitter social media users through the Twitter Application Programming Interface (API). The data is processed through preprocessing by removing unnecessary components so that the data becomes efficient for CNN modeling. The data is then weighted based on the occurrence of each word through TF-IDF. Tests were carried out with several comparisons of training data and test data, namely 80:20, 70:30 and 90:10. The test results were obtained with the highest F1-score of 86% using a comparison of training data and test data of 90:10 by carrying out the oversampling technique, namely the technique of taking minority data so that it has a larger proportion than the initial proportion.

Keywords: Personality, Big Five, Twitter, Convolutional Neural Network
