Abstract

Twitter is one of the most popular social media that can facilitate users to write their thoughts into a tweet. Because tweet tend to be personal, information can be taken like emotion, sentiment, and personality. Personality is important to know because it can be useful for understanding user's potential, increasing accuracy for recommender system, and help mental health diagnosis. Currently, there are many previous studies that predict someone's personality using English data with many feature, but for bahasa Indonesia has not been well studied. So in this study a personality prediction system will be built and will classified into Big Five model, because it is considered to be the most appropriate model to determine the user's personality on social media. The feature used social feature, sentiment detection, and emotion detection by counting weight that will be calculated using Term Frequency Inverse Document Frequency (TF-IDF). The classification method used is Naive Bayes because it is fast in large data calculations, space efficiency, and implementations tend to be easier to build. Experimental results show an accuracy of 45,00% for all features combined. Average emotion features alone produce an accuracy 38.12%, sentiment features produce an accuracy 38.95%, and social features with an accuracy 23.15%.

Keywords: big five, emotion detection, sentiment analysis, tf-idf, naive bayes.