Abstract

Personality can be defined as the overall way in which someone reacts and interacts with the environment and with other individuals. Research related to someone personality has been carried out by experts for certain purposes. Research on someone personality in the use of social media has also begun to be developed. One of the social media used to observe and examine someone personality is Twitter. With many active users on Twitter, each individual must have different characteristics in using their Twitter account. In this thesis research, the authors built a personality classification system for Twitter users. Research carried out using two approaches, namely the linguistic approach and features of Twitter itself. The data used are data from 143 Twitter users with a total of 351,197 tweets with a comparison ratio of training data and test data 70%:30%. Using weighting Term Frequency Inverse Document Frequency (TF - IDF) and Logistic Regression as classification algorithms, the accuracy produced by the system built in this final project reaches 69% for the social behavior approach and 76,20% for the linguistic approach and social behavior approach.
Keywords: twitter, weighting, TF – IDF, logistic regression