Abstract

Curiosity to predict someone personality through Twitter social media is no longer a new thing anymore. Every tweet that typed can provide a piece of information about someone personality. The problem is how to classify a text on Twitter social media into classes that will be created with good performance value. In this Final Task research, writer builds a system that classifies someone personality on Twitter social media using decision tree C4.5 classification method with TF-RF and TF-CHI^2 weighting method. Another thing that distinguishes this research is the weighting of each word using the weighting method TF-RF and TF-CHI^2 with the addition of new features for approaches based on user social behavior such as the number of characters in tweets, average characters in each tweet, average the word on the tweet, the media URL that reads how many users upload photos or videos, punctuation marks that counts the number of question marks (?) and exclamation marks (!) on each tweet, counts the uppercase letters and the emojis on tweets that can be detected as 2,552 characters different, and the last is trying to integrate the personality approach based on social behavior with a linguistic approach. From the results of the experiment on the 90% training data ratio and 10% test data (90:10) and the combination of personality features based on social behavior with a linguistic approach with TF-RF weighting obtained the accuracy results of 65.72%.

Keywords: Decision Tree C4.5, Term Weighting, TF-RF, TF-CHI^2, Big Five