DISC Analysis of User Twitter with TF-IDF and TF-CHI-SQUARE Weighting Feature using C4.5 Algorithm

Mohammad Abbiyu[1], Erwin Budi Setiawan[2]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[4]Divisi Digital Service PT Telekomunikasi Indonesia
[1]biyubiu@students.telkomuniversity.ac.id, [2]erwinbudisetiawan@telkomuniversity.ac.id,

Abstract
Twitter is one of the development of microblogging-based social media that is currently widely used worldwide. With the development of today's technology, defining one's personality through social media like Twitter is a thing that is being surpassed. The problem is how to classify the text that is on social media Twitter into classes that will be created to get good value for performance. The research aims to build a system to classify the personality of Twitter users using the algorithm classification method C 4.5 (Decission Tree C 4.5) as well as the weighted methods of TF-IDF and TF Chi-Square. Still at least research on the use of DISC personality models as aids in assessing personality, use of two weighted methods and the addition of features based on social behaviour such as follower, following, retweet and others can Differentiator from previous research. From the experiment results gained the best average accuracy of 43.60% on the comparison of the training data and test data of 90:10 and a linguistic approach using TF-CHI-Square with the best 50 word selection.

Keywords: Twitter, DISC, Algorithm C 4.5, TF-IDF, TF-CHI-Square.