

IMPROVEMENT OF ACCURACY IN THE TWITTER USER MBTI PERSONALITY PREDICTION USING DATA AUGMENTATION

Abstract

The personality of an individual needs to be known to help people in considering things, one of them is career recruitment. In general, personality can be known through interviews, observations, and questionnaire surveys. However, the conventional method is judged to be impractical in terms of time and material because it takes a long time and has considerable costs to process data. After all, the use of conventional methods can also cause bias because it involves a third person in data processing. The research tries to provide a solution by building a system that can predict the personality of a person based on the analysis of data and information from social media Twitter. The data and information will be processed so that the personality prediction is obtained. The personality classification theory used is the Myers-Briggs Type Indicator (MBTI) theory. The research also tries to implement data augmentation techniques to improve the performance of text mining tasks that have a slight dataset. The best results are obtained by the Random Forest method using the Term Frequency-Inverse Document Frequency (TF-IDF) weighted and the features available on Twitter. The use of augmentation techniques can increase accuracy by up to 20% from initial accuracy. So, the use of data augmentation techniques can be used to improve the performance of MBTI personality prediction models.

Keywords: MBTI, random forest, twitter, data augmentation