ABSTRACT

In 2020, the Covid-19 virus caused a pandemic that made most people more active on social media, such as Twitter. Twitter has a tweet feature allows its users to send short messages about how they feel and think at that moment. Based on someone's tweet, we know their mindset, and it allows us to know the personality of that person. One model of personality is the Big Five personality. Big Five divides personality into five classes: openness, conscientiousness, extraversion, agreeableness, and neuroticism. Several ways can be done to determine personality, such as taking a psychological test. However, it can take a long time and total concentration. Therefore, this study conducted a Big Five personality detection on Twitter users using the Gradient Boosted Decision Tree (GBDT) method. This study aims to obtain a high accuracy value by weighting it through the TF-IDF method and using sentiment and emotion features. This study utilized an Indonesian dataset that was collected through Twitter API. This study consists of two scenario tests, with the first scenario test being carried out with an imbalanced dataset and the second scenario test being carried out by applying the oversampling technique with SMOTE method to handle the imbalanced dataset. By applying SMOTE method, this study obtained a high accuracy with a value of 60.36%.

Keywords: Twitter; Tweet; Personality; Big Five; Gradient Boosted Decision Tree;