Personality Classification Based on Twitter "X" Posts With Big Five Theory Using Naive Bayes Method

Karen Novita Rahmadhani
Informatics Engineering Study Program
Departement of Information Technology

ABSTRACT

Twitter is a widely used social media platform for sharing opinions, feelings, and daily activities. The textual data generated by Twitter users can serve as valuable information to understand individual personality traits, especially for prospective employees considered for recruitment. This study utilizes 500 recent tweets to identify the five dimensions of the Big Five personality traits by applying the Naive Bayes machine learning method along with TF-IDF (Term Frequency-Inverse Document Frequency) feature extraction techniques. The data undergoes preprocessing stages including text cleaning, tokenization, stopword removal, and stemming to create an optimal numerical representation for model training. The model is evaluated on test data and achieves an accuracy rate of 61%. The prediction distribution in the training data reveals proportions of Extraversion at 22.5%, Neuroticism at 21.5%, Agreeableness at 19%, Conscientiousness at 19%, and Openness at 18%. The model implementation is realized in an interactive web application capable of displaying real-time prediction results, supported by key visualizations such as Word Clouds and comprehensive evaluation reports. This research aims to assist companies in the employee selection process by providing insights into personality traits through social media text analysis.

***Keywords:** Big Five, Naive Bayes, TF-IDF, Twitter, personality prediction, prospective employees.*