



Big Five Personality Assessment Using KNN method with RoBERTa

Athirah Rifdha Aryani¹, Erwin Budi Setiawan²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹athirahrifdha@students.telkomuniversity.ac.id, ²erwinbudisetiawan@telkomuniversity.ac.id

Abstract

Personality is the general way a person responds to and interacts with others. Personality is also often defined as the quality that distinguishes individuals. Social media was created to help people communicate remotely and easily. These personalities fall into five categories known as the Big Five personality traits, namely Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN). The use of K-Nearest Neighbour (KNN) is a method of classifying objects based on the training data closest to them. To overcome the data imbalance during training data, we use K-Means SMOTE (Synthetic Minority Oversampling Technique). Other features such as LIWC (Linguistic Inquiry Word Count), Information Gain, Robustly Optimized BERT Approach (RoBERTa), and hyperparameter tuning can improve the performance of the systems we build. The focus of this study is to present an analysis of Twitter user behavior that can be used to predict the personality of the Big Five Personality using the KNN method. The Important aspect to consider when using this method, namely accuracy in classifying the Big Five Personalities. The experimental results show that the accuracy of the KNN method is 72.09%, which is 95.28% gain above the specified baseline.

Keywords: Big Five Personality, K-Nearest Neighbours (KNN), RoBERTa, LIWC, Information Gain