

Abstract

Twitter is one of the most popular social media used to interact online. Through Twitter, a person's personality can be determined based on that person's thoughts, feelings, and behavior patterns. A person has five leading personalities: openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. This study will make five personality predictions using the Naïve Bayes method – Support Vector Machine, Synthetic Minority Over Sampling Technique (SMOTE), Linguistic Inquiry Word Count (LIWC), and Bidirectional Encoder from Transformers Representations (BERT). A questionnaire was distributed to people who used Twitter to collect and become a dataset in this research. The dataset obtained will be processed into SMOTE to balance the data. Linguistic Inquiry Word Count is used as a linguistic feature, and BERT will be used as a semantic approach. The Naïve Bayes method is used to perform the weighting, and the Support Vector Machine is used to classify Big Five Personalities. To help improve accuracy, the Optuna Hyperparameter Tuning method will be added to the Naïve Bayes Support Vector Machine model. This study has an accuracy of 87.82% from the results of combining SMOTE, BERT, LIWC, and Tuning, where the accuracy increases from the baseline.