

Abstract

Personality is a unique set of motivations, feelings, and behaviors humans possess. Personality detection on social media is a research topic commonly conducted in computer science. Personality models often used for personality detection research are the Big Five Indicator (BFI) and Myers-Briggs Type Indicator (MBTI) models. Unlike the BFI, which classifies personalities based on an individual's traits, the MBTI model classifies personalities based on the type of the individual. So, MBTI performs better in several scenarios than the Big Five model. Many studies use machine learning to detect personality on social media, such as Logistic Regression, Naïve Bayes, and Support Vector Machine. With the recent popularity of Deep Learning, we can use language models such as DistilBERT to classify personality on social media. Because of DistilBERT's ability to process large sentences and the ability for parallelization thanks to the transformer architecture. Therefore, the proposed research will detect MBTI personality on Reddit using DistilBERT. The evaluation shows that removing stopwords on the data preprocessing stage can reduce the model's performance, and with class imbalance handling, DistilBERT performs worse than without class imbalance handling. Also, as a comparison, DistilBERT outperforms other machine learning classifiers such as Naïve Bayes, SVM, and Logistic Regression in accuracy, precision, recall, and f1-score.

Keywords : *Personality Detection, Reddit, DistilBERT.*

