Abstract

This research investigates the use of Support Vector Machine (SVM) and Naive Bayes models to classify the personality traits based on the social media posting patterns. This study integrates textual features obtained from the Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) methods, and along with the feature expansion using the Linguistic Inquiry and Word Count (LIWC) tool, to assess their influence on accuracy Classification Personality characteristics were mapped from social media posts using the Big Five Inventory (BFI-44). The research findings show that the SVM model in which uses the TF-IDF + LIWC feature set, provides the best performance, and achieve 76.60% of accuracy on the base model with a linear kernel. In comparison to the Naive Bayes model performed best with the same feature set, achieving 59.57% accuracy with a smoothing parameter of 1xE-2. Although the oversampling improved recall and precision, the undersampling was found to have a negative effect on model performance. These findings highlight the benefits of combining TF-IDF and LIWC features which improve model effectiveness, with SVM producing the best overall results in personality classification from social media data.