ABSTRACT

Personality reflects an individual's patterns of thinking, emotions, and behavior, and one of the models used to analyze it is the Big Five Personality model, which consists of five main dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. This study aims to classify Big Five personality traits using data from social media X with the RoBERTa method. The main challenge in this research is the limited dataset, which includes only 381 users, and the imbalance in data distribution across labels, which affects prediction accuracy.

To address this issue, Random Oversampling and Easy Data Augmentation techniques were applied to balance and enrich the data. The model was tested with two types of preprocessing: Half Preprocessing and Full Preprocessing, as well as hyperparameter tuning using Random Search. The experimental results show that the combination of Data Augmentation and Random Oversampling provides the best performance, with an accuracy of 73.73% and an F1-score of 73.37% in the Full Preprocessing scenario. This research contributes to the development of text-based classification methods for personality analysis on small datasets.

Keywords: *Big Five Personality, RoBERTa, Random Oversampling, Data Augmentation, social media, text classification.*