1. Introduction

Social media has an impact on changes in human social life. Using internet-connected devices, users can communicate, interact, and share content on social media platforms. According to DataReportal's data in Digital 2022: Indonesia [1], Indonesia has 191.4 million active social media users out of a population of 277.7 million. As one of the most popular networking sites, Twitter allows users to post messages to the internet through text, pictures, and videos, known as tweets. In addition, social media user activities can reveal information about their personalities. Personality is a unique combination of behaviors, emotions, and thought patterns that influence cognition, motivation, behavior, and life choices [2], [3]. Twitter is considered an accurate application for identifying personality because Twitter users tend not to worry about the words used in tweets [4].

To identify a person's personality, research can be done using the Five-Factor Model approach, also known as The Big Five. This research method is widely used and has been used successfully in several previous studies [4]–[6]. This theory divides personality into five dimensions known as OCEAN, which stands for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism [2], [7]. Personality prediction research can also be done using the Myer-Briggs Type Indicator (MBTI) approach [8], [9] or DISC [10], [11] which classifies personality into dominance, inducement, submission, and compliance.

The flow of personality prediction research generally involves the processes of data extraction, preprocessing, feature extraction, and classification [12]. Pratama and Maharani [6] conducted a study in which personality prediction was performed using a Random Forest classification with a feature extraction process consisting of emotional, sentiment, and social analysis based on statistical data from user Twitter accounts. The highest accuracy is 69.23%, obtained by combining features and comparing test data 20:80 [6]. Classification with Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naïve Bayes is also used in personality prediction research by Maharani and Effendy [4]. Research [4] has found openness to be the simplest trait to measure, while neuroticism is considered quite difficult. In predicting personality traits, the SVM approach produces a good performance, with a score of 59.45% [4]. Larger sample size and more features will yield better results [4]. Aditi and Suja [13] compared the Multinomial Naïve Bayes (MNB), Latent Dirichlet Allocation (LDA), and AdaBoost methods to predict personality using an English dataset. Research [13] used a multi-label classification and found that MNB had the highest accuracy, while AdaBoost and LDA had nearly the same accuracy on all labels except openness.

In ensemble learning, the boosting algorithm iteratively trains the weak classifier and then adds it to the strong classifier by minimizing bias or variance due to training records [14]. Adaptive Boost, or AdaBoost, is one of the boosting algorithms that can improve accuracy by combining several weak learners and correcting previous weak classifier errors [15]. However, this method is noise-sensitive, which means that if there is more noise data, the AdaBoost algorithm will spend more time and be less efficient [16]. AdaBoost is considered to be adaptive because it assigns weights to the base model based on model accuracy and changes the weights of the training data based on prediction accuracy [14]. Thus, the background of this study uses AdaBoost as a method of identifying personality.

Due to limited resources, personality research is conducted using the Big Five approach. The Indonesian dataset is derived from user tweets and used as sentiment, emotion, and POS-tagging features. Twitter user account statistics are also used as a social feature consisting of the number of tweets, following, followers, and favorites. This paper is structured as follows. The research methods will discuss in section 2, the results will discuss in section 3, and conclusions and future work will discuss in section 4.