1. INTRODUCTION

Personality is a difference owned by each individual in thinking, feeling, and behaving. According to Gordon Allport, Personality is "something" in the individual concerned [1]. Personality has various theories used to determine personalities, such as the MBTI (Myers Briggs Type Indicator) and Big Five Personality. Personality can be seen in how individuals speak, share opinions, and communicate, which, of course, can be seen in the individual's daily life. Personality is also one of the determinants of the type of work that is more suitable because it involves taste and comfort [2].

People are now sharing social activities through online platforms called social media. Based on data taken in February 2022 by DataReportal[3], Twitter is one of the social media that is often used by Indonesians, with a percentage of 58.3%. Twitter users can share their social activities through tweets on Twitter. Therefore the Twitter platform can be used as a platform to detect personality, primarily through the tweets of users.

Personality detection certainly has various labels as output to classify each individual. The personality theory that will be used as a reference is the big five personality theory. Currently, psychologists believe that the description of the structure of traits is owned by the Big Five Personality[4]. Big five personality has five dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism[2]. The Openness dimension is an individual personality that has high curiosity, the Conscientiousness dimension is a very alert personality, the Extraversion dimension is an individual who is easy to socialize with, the Agreeableness dimension is an individual who tends to avoid conflict, and the Neuroticism dimension is an individual who can control emotions[2],[5]. This case requires an algorithm that can produce output from labels, namely the supervised learning algorithm.

Research on personality detection has been carried out using different supervised learning algorithm models in research conducted by Rendy Putra Pratama[5]using the Random Forest model. Random Forest is a model that can produce the best accuracy with unbalanced data. This is because Random Forest uses a Decision Tree as the basis for the classification that is built so that it can detect correlations between data. This research resulted in an accuracy rate of 69.23%. The research conducted by Nanda Yonda Hutama [6] used the Multinomial Naïve Bayes and Decision tree models. Naïve Bayes is a model that has an independent principle. This principle shows that naïve Bayes considers the data to be processed with no correlation between attributes, so balanced data is needed to maximize detection. This research resulted in an accuracy rate of 55% for Naïve Bayes and 33% for the Decision Tree. Therefore this study reveals that Naïve Bayes is a stable model for detection.

Another study regarding the Big Five personality detection was conducted by S. V. Therik [7] using the C4 model. This study obtained an accuracy of 62.02%. This accuracy was obtained because it used weighted data using TF-IDF and LIWC, with an increase of 17.24% compared to the baseline. Besides that, S. V. Therik implemented the SMOTE implementation, thereby increasing the accuracy value to 76.92%, an increase of 32.1% from the baseline. Finally, in the research by Y. A. I. Nanda [8]. This research predicts Personality Disorder using Naïve Bayes. This study validates it with the Confusion Matrix. The accuracy obtained is 88.2%. The results of these studies produce a good performance.

Random Forest and Multinomial Naïve Bayes have differences in performance. In the case of text detection, multinomial naïve Bayes has more stable performance. After all, it utilizes probability theory [6]. Random Forest has better accuracy because it has more features than multinomial naïve bayes.

Based on the background, this study compares the random forest model and the multinomial naive Bayes model to determine which algorithm is more suitable for personality detection cases. Personality theory focuses on the big five personality theory, and the data used is tweet data from Twitter to detect personality based on the Big Five Personality rules. Then the program will classify data and create models using random forests and multinomial naive bayes. Tests were carried out as many as three tests in the form to determining the best algorithm in this study in the form of an algorithm that has high accuracy and precision.

This study has a problem limitation in the form of the dataset used is tweets from Twitter that use Indonesian. This data is retrieved using the Crawling data method with the Twitter API. The system cannot handle typos, and personality theory focuses on the big five personality theories.