I. Introduction

Social media has become a common thing in our daily lives. Especially after the pandemic in 2020, many people have become active on social media to interact and socialize with others [1]. One of the most frequently used social media is Twitter [2]. Twitter has a feature called a tweet, where users can express themselves about what they are feeling and share information that the user wants to convey [3]. Based on these tweets, we can get information and an overview of the feelings and mindset of the user who wrote the tweet. Twitter is one of the social media that makes it easy for many users to see the latest or trending information through the tweet feature. In addition, tweets can make it easier for researchers to get datasets because tweets can be categorized as big data [4]. Several studies have used tweets from Twitter users as datasets, such as research conducted by Balakrishnan et al. [5], [6], which detects cyberbullying by associating the personality of the user to create a pattern used to detect cyberbullying.

Personality is an important thing that is owned by everyone [7]. Personality influences the mindset, nature, and feelings of a person. However, everyone has a different personality, so personality can make someone unique. One well- known method in the world of psychology that can divide a person's personality into five classes is called the Big Five personality. The five classes are openness, conscientiousness, extraversion, agreeableness, and neuroticism [8]. The Big Five personality is one of the most researched personality structure measurements and is considered good by many experts [8].

Personality is one of the critical factors that allows one to know oneself better. Knowing individual personality allows the person to know their weaknesses or deficiencies that can be improved. You can discover your personality in various ways, such as by taking a psychological test. However, working on a psychological test takes a long time and requires total concentration. So, suppose someone is not focused on doing the work. In that case, that person can fill in answers that are not following their personality, which results in personality test results, not in accordance with their actual personality. Today's technology has developed far. So

it is no stranger if many things in everyday life are done digitally.

Previously, there has been research carried out to detect personality on Twitter users using various methods, such as research by Shantika Valerin Therik, who used the Decision Tree C4.5 method [9] in 2021. This research was conducted with a total of 549,151 tweets from 287 Twitter users. After that, TF-IDF and LIWC data were added and applied to the SMOTE method, which was tested by applying the hyperparameter tuning technique using Grid Search with social behavior as a baseline. This study achieved an accuracy of 62.06%, with an increase in accuracy of 17.24% from the baseline. After using the SMOTE method, the accuracy value increased to 76.92% resulting in an accuracy value of 32.1% from the baseline.

In 2022 there was also research by Alvinda Amalia Rizkita, who uses three different methods, which are Support Vector Machine (SVM), Naïve Bayes (NB), and Naïve Bayes-Support Vector Machine (NBSVM) [10]. This research is pretty similar to research by Shantika Valerin Therik [9] who used TF-IDF and LIWC methods with SMOTE method to oversample the data and used hyperparameter tuning technique with Grid Search and social behavior. The accuracy results obtained from this study were 82.41% using the Support Vector Machine algorithm, 71.42% accuracy using the Naïve Bayes algorithm, and 53.86% accuracy using the Naïve Bayes-Support Vector Machine algorithm.

In 2019, there was also a study by Roji Ellandi, who used k-Nearest Neighbor (kNN) method [11]. This research succeeded in getting data from 331,439 tweets from 137 accounts. The results obtained using the k Nearest Neighbor method get the best performance with a value of k = 9, which is 60.97%.

In this research, the method used is the Gradient Boosted Decision Tree (GBDT). GBDT is a DT-based supervised learning technique with a boosting approach for predicting a value [12]. The boosting algorithm will identify weak attributes through a looping process that begins by giving each observation the same weight to find the features that cause predictions to be less accurate. After being successfully identified, these features will be weighted higher in the next iteration [13]. Finally, the results will be combined using voting for classification or mean for regression to produce the final model. GBDT is one of the most popular

algorithmic methods in recent years, with many awards in machine learning and data mining competitions [14]. which uses the GBDT method to expand features with fastText on topic classification in 2022. This research obtains the highest accuracy results, namely 91.39%. The same year, Dhuhita Trias Maulidia [15], conducted feature expansion research with Word2vec for topic classification using the GBDT method. The highest accuracy result obtained in this study was 85.44%. Based on these studies, the GBDT method can potentially obtain high- accuracy results. Therefore, this study uses the GBDT method.