I. INTRODUCTION

With the development of technology, accessing information through social media has become easier, as the use of social media continues to grow. Social media platforms enable users to share stories, express opinions, and interact with one another. One of the most popular social media platforms is X, formerly known as Twitter. X users can share tweets to reveal many things about themselves. Publicly uploaded tweets can serve as a basis for determining the characteristics of these users [1]. Personality is a characteristic that distinguishes a person in terms of behavior, reactions to certain environments, and thought processes [2]. There are several methods to describe a personality, one of which is the Big Five Personality, which will be used in this study. Big Five Personality has five personality dimensions, including openness, conscientiousness, extroversion, agreeableness, and neuroticism. The Big Five Personality is significantly related to social user behaviour. For example, highly neurotic individuals use more negative words in their tweets [3]. This research uses the Indonesian text information publicly uploaded by user X to classify the user's personality using a machine learning algorithm called Logistic Regression. The algorithm can be used to predict whether a user has certain personality traits based on their text data by modeling the relationship between text features and these traits [4]. Logistic Regression algorithm is the simplest classification algorithm that can be used for binary and multiclass classification [5]. The Logistic Regression model with the TF-IDF Vectorizer feature achieved the highest accuracy of 97% compared to other algorithms used in the research conducted by Shah et al. [6]. On the classification of X news data, Logistic Regression produces an accuracy of 91%, outperforming SVM, Random Forest, Gradient Boosting, and Naïve Bayes [7]. Both studies [6] [7] support the idea that Logistic Regression is an effective and good method for classification. Therefore, in this study, the Logistic Regression method will be used to classify the personality of X users based on the Big Five Personality. Although several studies have used logistic regression to perform personality classification, most of these studies have not explored the various conditions that can affect the performance of the model. This research will focus on combining the Logistic Regression method with several scenarios, including data preprocessing, data split ratio, data balancing techniques, and feature extraction. These scenarios are designed to find the best combination that affects the accuracy of personality classification.