

I. INTRODUCTION

The rapid development of digital technologies has transformed the ways in which people communicate and express themselves, particularly through social media platforms. Services such as X (formerly Twitter), Facebook, and Instagram serve as prolific sources of user-generated data, producing massive volumes of information on a daily basis. This data has significant potential for various analyses, including personality prediction. In psychology, the Big Five personality model—which includes five main dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism—has become a standard in measuring a person’s personality characteristics [1].

Using text data from social media to predict personality is a growing field of study. Previous research has shown that text-based approaches can provide deep insights into a person’s personality [2]. In this context, natural language processing (NLP) methods have become key tools for exploring social media text data. A study by Rohmah and Maharani [3] investigated personality detection using the IndoBERT method on Twitter (now known as X). Their results demonstrated that IndoBERT could classify personality with a fairly high level of accuracy, making it one of the most promising approaches in Indonesian-language NLP.

In contrast, Lee and Sibaroni [4] conducted a comparative analysis of the IndoBERTweet and Support Vector Machine (SVM) models in performing sentiment analysis on public discourse related to the construction of a racing circuit in Indonesia. They found that IndoBERTweet excelled in capturing linguistic context over SVM, demonstrating the potential of transformer-based models for text analysis tasks in the Indonesian language.

Additionally, XLNet is another widely adopted model in various studies. Kumar et al. [5] used a fine-tuned XLNet to detect fake news and found that it outperformed conventional models in terms of accuracy for text analysis tasks. Despite the proven effectiveness of BERT and XLNet in numerous NLP applications, there remains a lack of comprehensive research that directly compares their performance in the context of Big Five personality prediction using social media data. A study by Danyal et al. [6] highlights the importance of conducting comparative analyses of transformer model performance to better understand the strengths and weaknesses of each model for specific applications.

This study aims to address the gap in current research by comparing the performance of BERT and XLNet in predicting Big Five personality traits from social media texts on platform X. Unlike previous studies that focus primarily on classification accuracy, this research also takes into account the computational efficiency of each model during both training and inference. The results are expected to advance personality prediction techniques and provide deeper insights into the trade-offs between performance and resource consumption of transformer-based models in Indonesian NLP applications.