

## I. INTRODUCTION

The electoral procedure to choose the President and Vice President of Indonesia for the term 2024-2029 is set to occur during the 2024 Indonesian Presidential Election, slated for Wednesday, February 14, 2024. Based on this, it is anticipated that there will be a plethora of public opinions regarding the Indonesian presidential candidates for 2024 on social media platforms. The impact of social media, notably platforms like Twitter, is progressively becoming pivotal in molding public sentiment and impacting voter choices [1]. With the abundance of public opinions expressed on Twitter, valuable insights into the current sentiments toward the Indonesian presidential candidates for 2024 can be obtained [2]. Hence, the objective of this study is to collect and analyze tweets from Twitter users to gauge public sentiment regarding the 2024 Indonesian presidential candidates.

This sentiment analysis aims to understand the public's views on three presidential candidates who will compete in the 2024 presidential election: Prabowo Subianto, Anies Baswedan, and Ganjar Pranowo. In terms of positive, negative, or neutral sentiment, this analysis will be used to assess the candidates' image in the eyes of Twitter users, determining the extent to which they are accepted or rejected by the public. IndoBERT is able to capture the meaning of Indonesian text more effectively. Consequently, it is plausible to infer that the IndoBERT model could exhibit enhanced effectiveness in distinguishing between positive, negative, or neutral sentiments in Indonesian tweet data, as opposed to conventional sentiment analysis models like machine learning-based or lexicon-based models [3].

Bidirectional Encoder Representations from Transformers (BERT) is a pre-trained model that performs word embedding in Natural Language Processing, where each word is transformed into a set of numerical vectors using the Transformer architecture [4]. Meanwhile, IndoBERT is a variant of the pre-trained BERT model specifically developed using the Indonesian language corpus. Throughout the training phase of BERT, it undergoes training using Masked Language Modeling (MLM) and Next Sentence Prediction techniques. These methods enable BERT to acquire proficiency in understanding language and its contextual nuances [5]. Sentiment analysis using IndoBERT can demonstrate better results compared to regular BERT, which uses a multilingual corpus and other pre-trained models [6]. Subsequently, IndoBERT has demonstrated superior efficacy when compared to K-Nearest Neighbors (KNN), Support Vector Machines (SVM), naive Bayes, decision trees, and random forest models [7].

Based on the exposition above, this research will conduct sentiment analysis on potential candidates for the Indonesian Presidency in 2024 on Twitter using the IndoBERT model. The dataset used will consist of tweets from Twitter users, collected using the Tweet Harvest library based on Python, with three keywords: "Prabowo Subianto," "Anies Baswedan," and "Ganjar Pranowo." The performance evaluation of the model will be calculated using the confusion matrix method. With this research, it is hoped that new methodologies for sentiment analysis specific to the context of Indonesian presidential candidates in 2024 using IndoBERT can be developed and evaluated, demonstrating the superiority of this model.