1. Introduction

Social media has become a platform where people can freely express their opinions on various topic, including politic. In democratic systems, it is expected that citizens stay updated on current events, participate in political discussions, and take an active role in the processes that shape governance and policy [1]. In Indonesia, the utilization of social media has experienced a high growth. In early 2024, approximately 139 million individuals in the nation were active social media users. Data Reportal shows that Indonesia has 139 million social media users in January 2024 representing 49.9% of the total population of Indonesia. There is also a presentation of the number of active users of X or in Indonesia as much as 24.69 million and X's advertising planning tool recorded an increase of 693 thousand in the span of early 2023 to early 2024. Sentiments refer to an individual's feelings, emotions, opinions, or an event [2]. Comments across social media platforms can be converted into data for sentiment analysis. Sentiment analysis can play a critical role for strategic planning, decision-making, and content filtering. By using advanced data analysis techniques, including Natural Language Processing (NLP) [3], researchers can analyze the opinions, emotions, and reactions of individuals as expressed in online environments.

Despite significant progress in sentiment analysis, challenges remain. Particularly in understanding complex linguistic phenomena such as sarcasm, sentiment shifts and the use of slang. In Indonesian context, regional dialects, mixed languages used per tweets, and cultural expressions add complexity. For instance, like "gua", "lo" and many other examples to complicate interpretation. Lexicon-based methods, such as SentiWordNet [4], TextBlob [5], and VADER [6], have proven effective for basic sentiment analysis tasks, However, these methods struggle with nuanced language and require substantial manual labelling efforts, limiting their scalability and accuracy in handling diverse datasets. This highlights the need for advanced machine learning approaches.

This study utilizes IndoBERT, a transformer-based model specifically designed for Indonesian text, to address sentiment analysis challenges. IndoBERT was developed by (Koto et al., 2020). Unlike the original BERT (Devlin et al., 2019), IndoBERT was trained on Indonesian linguistic datasets, making it far more effective to capture the nuances of the language. Built using Hugging Face framework [9] this model is a powerful tool for analyzing sentiments in Indonesian text. To enhance its performance, this study also incorporates advanced text preprocessing techniques, such as stemming, and normalization, using the Sastrawi library [10]. These steps ensure that text is cleaned and standardized for better analysis. By combining IndoBERT's language-specific training with careful preprocessing, this research aims to deliver a more accurate and reliable approach to sentiment analysis for Indonesian text, addressing common challenges that arise in working with this language. This study improves the understanding of slang, dialects, and mixed-language usage, this study provides practical benefits for businesses, policymakers, and researchers who rely on accurate sentiment analysis for strategic decision-making. Furthermore, the development of a voting-based labelling system enhances classification accuracy, reducing biases, inherent in individual methods. These contributions are not only relevant to NLP [11] research but also demonstrate the practical application of IndoBERT in understanding public sentiment.