Bab 1

I. INTRODUCTION

Along with technological advances, social media is often the main access for people to get all information quickly. The information provided is complete and diverse, so the growth of social media users in Indonesia yearly has increased significantly [1]. One social media that most Indonesians visit is the X platform [2]. On the X platform, users usually write opinions about a topic, which will later get other people's responses [3].

Regional head elections are an interesting topic that allows people to have opinions on each other [4]. There are many positive and negative issues in the campaign period leading up to the regional head elections. Differences in opinion often create a reciprocal response between users [5]. The diverse responses from the community motivate researchers to analyze these responses [6]. Sentiment analysis involves the study of analyzing responses [7]. The sentiment analysis system divides words into groups of positive opinions, negative opinions, and neutral opinions [8]. Sentiment analysis research is often carried out using various algorithms. Sentiment analysis is suitable for identifying public opinion, such as in regional head elections [9].

Accuracy is often not maximized against data with ambiguity, undetected words, and other similar things [9]. Therefore, a good combination of algorithms is needed. Based on research [10], a hybrid trial of Support Vector Machine and Convolutional Neural Network using TF-IDF achieved an increased accuracy of 2.01% over the baseline. Based on other research, to optimize the accuracy value to be stable even with variants of models, features, or datasets, it is necessary to apply the Word2Vec expansion feature [9]. Research [11] also shows increased accuracy when adding semantic features, specifically IndoBERT, to capture the meaning of Indonesian words in a document.

To improve previous research, this study tests with the main contribution of testing a hybrid classification combination with all complex features. To our knowledge, research has yet to be done on that test. This research tests the hybrid classification between CNN and SVM. The method is assisted by Word2Vec feature expansion, TF-IDF feature extraction, and IndoBERT Semantic Features in solving the dataset obtained through the crawling process on platform X with several keywords relevant to “Pilkada 2024”. Deep learning methods such as CNN are one part of machine learning whose work process can reach the deepest layers and recognize text patterns even with many data. In this research, TF-IDF helps determine the statistical weight between essential words in a document by looking at the frequency of word presence. Especially in sentiment analysis research, TF- IDF is useful for finding uncommon words to distinguish between sentiments and generating vectors to indicate the significance of words in the text. In the next stage, the TF-IDF matrix and the Word2Vec model trained based on the corpus created will capture the semantics of words in the document, which is very helpful in finding words with many synonyms with relevant context. To maximize the research, the integrated Semantic Features IndoBERT is also used to train Indonesians to enrich words not detected in TF-IDF and consider all possible sequences that Word2Vec cannot. Semantic Features IndoBERT is very influential in sentiment analysis because it can understand document sentences by looking at the relationship and order between Indonesian words. The combination of features will produce a matrix that will be further processed in hybrid CNN and SVM. The combination of features is also believed to include statistical, semantic, and contextual completeness, which will play an essential role in this research to find the classification model and feature optimization that has the most influence on the highest accuracy in analyzing sentiment on the X platform community in the 2024 regional head election.