1. INTRODUCTION

In the digital era, X social media has emerged as a primary platform for conveying political information during elections, allowing voters to engage directly in political discussions and receive real-time information. However, the presence of social media also opens the door wide for the spread of hoaxes, which can undermine the integrity of elections. A hoax is false information deliberately created and spread as if it were true [1]. The hoaxes circulating on X social media are usually false or misleading information designed to influence public opinion during elections. Examples include false claims about certain candidates, manipulated election data, or fake news that stirs ethnic or religious tensions. Since X social media has a character limit per post, information is often presented in a brief format that can easily be misunderstood or intentionally misrepresented. Therefore, detecting hoaxes on social media, especially on the X platform, is crucial to maintaining the validity of political information and protecting the democratic process.

These hoaxes undermine the integrity of elections by creating confusion among voters, spreading distrust in the electoral process, and even potentially sparking social conflict. If left unchecked, the spread of hoaxes can disrupt the democratic process by leading voters to make decisions based on false or misleading information, ultimately affecting the election outcome. The spread of hoaxes can cause anxiety, discomfort, and even anger, especially if the information is provocative. This can incite hatred and lead to division within society [2]. Previous research has identified that the spread of hoaxes and disinformation during election periods tends to increase [3]. Therefore, measures are needed to address the spread of hoaxes in society, with an approach that can understand text within character limits.

One approach is to utilize machine learning to identify certain patterns that frequently appear in hoaxes. Machine learning can be applied through various methods, such as Support Vector Machine (SVM), Decision Tree, Self-Organizing Map, and Convolutional Neural Network (CNN) [4]. Previous research on hoax detection has been conducted by Nurhikam et al. [5] using the random forest algorithm, achieving an accuracy of 84%. Additionally, research by Fauzy et al. [6] used a Convolutional Neural Network (CNN) and successfully improved the accuracy of hoax detection in microblogging content. Therefore, this study chooses the Convolutional Neural Network (CNN) method as the main approach to detect fake news on the X social media platform.

Convolutional Neural Network (CNN) is an architecture originally designed to handle image data processing but has been successfully used in various text-processing tasks, including text classification [7]. CNN has several advantages in text classification, such as automatic feature extraction, the ability to handle big data, and more. Additionally, CNN can extract and understand complex patterns in text, which are often found in fake news content. This capability is the reason behind using the CNN method in this study.

Additionally, the author applies weighting methods using Term Frequency Relevance Frequency (TF-RF) and Term Frequency Inverse Document Frequency (TF-IDF) to enhance detection accuracy. TF-RF combines TF and RF to improve performance, focusing on the occurrence of a term within a document [8]. Conversely, TF-IDF is a weighting feature designed to represent the significance of a term within a document in a corpus. TF-IDF is often used to determine how important a term is in the context of search engines, text summarization, or text classification [9]. In similar studies, using TF-RF and TF-IDF weighting features has shown significant improvements in predicting potential hoax text content [9] [10].

The difference from previous research is that [11] found that CNN only achieved an accuracy of 60% using 804 FA-KES data as the dataset. Therefore, this study aims to build upon that research to achieve more optimal results. In [12], CNN only used TF-IDF weighting features, whereas, in this study, the author uses CNN with TF-RF and TF-IDF weighting features. Previous research on fake review detection [13] compared six classification models but did not provide a deep theoretical explanation of the methods, and all models produced low accuracy. Study [10] used the TF-RF and TF-IDF as weighting features with K-Nearest Neighbor classification, achieving an accuracy of 62%. In contrast, this study uses TF-RF and TF-IDF as weighting features and CNN for classification to compare the classification results with previous studies. Study [14] used Word2Vec for word vectorization, whereas this study uses Count Vectorizer to compare the results of text representation methods with previous research.

The author chose Convolutional Neural Network (CNN) because this method has been infrequently applied to the classification of hate speech using Indonesian language [15]. CNN excels in recognizing complex patterns, achieving high performance in tasks like classification, and is very effective in handling large and complex data. The effectiveness of this research method is evaluated using a confusion matrix.