INTRODUCTION

In today's digital era, social media has become the main source of information for many people around the world. With advances in technology and the widespread availability of the internet, people can easily access various social media platforms to get the latest news. Often, the news presented on social media no longer goes through a strict editing process so that the truth of the information conveyed cannot be ensured [1]. Everyone has access to information that is shared without restriction, so anyone can disseminate news that is ambiguous or even contains hoax material, sometimes known as "fake news" [2]. Fake news is frequently created to appear legitimate and convincing; this is done to present false information that could mislead the reader's perception or comprehension [3].

The spread of hoaxes can cause significant losses, such as decreased public trust in the media, social instability, and disruption in the democratic process [4]. In the context of social media and online information-sharing platforms, the spread of hoaxes has become easier and more widespread. Anyone with internet access can quickly spread false information without first verifying its truth. Therefore, it is important to develop methods that can effectively identify and classify hoax news.

Several studies have been conducted to classify fake news using various methods. One of them is research carried out using the BERT (Bidirectional Encoder Representations from Transformers) method, which achieved an accuracy value of 67% [5]. Other studies using CNN and LSTM methods achieved a CNN method accuracy value of 88% and an LSTM method accuracy value of 84% [6]. The conducted study contrasts various approaches to categorizing fake news [7]. This research compares several model classification approaches, including SVMs, logistic regression, Bi-LSTMs, and CNNs, to categorize the same dataset in this research. Based on the research findings, it can be seen that the CNNs model is the highest test results on the dataset provide an accuracy of approximately 0.260 and 0.270 on the validation and test sets, respectively.

In this research, the dataset that will be used is obtained from the same LIAR Dataset as in previous research [7], but uses different model and data processing. LIAR Dataset is a collection of data obtained from the Politifact website API. Politifact itself is a website that labels news statements on social media as true or false. The model that will be used in this research is the BERT model. BERT has revolutionized natural language processing (NLP) in recent years because of its exceptional performance [8]. With preprocessing the data, training the model on real news, and using more validation data to validating the model are part of this strategy. This research is expected to obtain better performance scores in classifying fake news.