1. INTRODUCTION

Stock investment has become popular among the public, and people often choose it because it can generate a high profit. Companies select stock for their financial need. Stock is an investment instrument investors favour because of its attractive return [1]. Stock is also added to by the number of investors, which keeps increasing yearly. People now realize the importance of investment, accompanied by increasingly advanced digital technology. The data obtained from the Central Securities Depository of Indonesia (KSEI) explained that, until August 2022, there were already 9.54 million stock investors. This number is very high because, in December 2021, the recorded number of investors stood at only 7.48 million. It means that stock market investors have increased by 27.38% [2]. The Efficient Market Hypothesis (EMH) theory defines that market information reflects stock price [3].

One of the most prominent social media platforms for sentiment analysis is Twitter. On Twitter, we can find specific keywords about a topic to determine whether the tweet is positive or negative [4].

Sentiment analysis is obtaining internet and social media data using text analytics. Sentiment analysis seeks to ascertain a person's perspective on social media, whether positive, negative, or neutral, with attention to the discussed topic [5]. Pradana et al.'s (2020) earlier investigation discovered a connection between the tone of a tweet and the movement of a company's stock price. A tweet from someone that contains a complaint tends to affect the stock price, as a tweet with many complaints will lead to a lower stock price, whereas positive tweets will lead to a higher stock price [6]. This sentiment is advantageous for organizations, corporations, and governments, but text mining and sentiment analysis are needed[7]. One example is the alleged Bank Central Asia's (BCA) tax objection in 2014, which led to negative perceptions of Bank Central Asia [8]. Sentiment analysis is performed on social media news and issues to determine how they influence changes in stock prices.

In this research, the researcher observes the impact of tweets on the stock price movement of Bank Central Asia (BBCA) using a hybrid method of Bidirectional Long Short-Term Memory (Bi-LSTM) and Convolutional Neural Network (CNN) with feature expansion Word2Vec. The researcher uses this hybrid method because it can enhance prediction accuracy using the method used [9]. This research uses the feature expansion Word2Vec, which converts words to vectors [10]. The advantage of using Word2Vec is that it can process small or large amounts of data [11]. The researcher expects this research to provide good accuracy results from the hybrid method and obtain sentiment information related to BBCA stock price movements to help investors make decisions.

Wang Yue and Lei Li [12] introduce this combination method of CNN and Bi-LSTM with additional Word2Vec feature expansion. The result obtained from this method is 91.48%, which is way bigger than other methods used in the same research. The accuracy result obtained from the only LSTM method is 79.83%, and with CNN, the accuracy obtained is 81.25% with Bi-LSTM, 85.69%, and with CNN-LSTM, 87.44%. Researchers conclude that using a combination method of CNN and Bi-LSTM with Word2Vec feature expansion is easier and more accurate. It also raises the accuracy of sentiment analysis for short text.

Kai Zhou and Fei Long [13] did a sentiment analysis of Chinese products using CNN as feature extraction in text and Word2Vec as feature expansion. They obtained high results after the metric evaluation with F1-Score, Precision, and Recall. Overall accuracy obtained from the hybrid method of CNN and Bi-LSTM dominates positive and negative metrics over other methods used in the research. These methods are LSTM, CNN, Bi-LSTM, and hybrid CNN-LSTM.

Tam et al. [14] also researched using different combined approaches, including CNN, LSTM, Bi-LSTM, and ConvBi-LSTM, for sentiment classification with two datasets and using GloVe and Word2Vec as feature expansions. The result of this research shows that using Word2Vec feature expansion on the tweets dataset achieves higher accuracy in the ConvBi-LSTM method, which has an accuracy of 93.76%, a CNN result of 91.89%, an LSTM result of 90.94%, a Bi-LSTM result of 91.52%, and a CNN-LSTM method with more than 86% accuracy.

Gandhi et al. [15] did the sentiment analysis for a dataset named IMDB, which was obtained from Kaggle.com and uses feature expansion Word2Vec. This research indicates that the LSTM method's accuracy is 88.02% while the CNN method is 87.72%.

Dedi et al.[16] did the sentiment classification for a dataset crawl from www.Finance.detik.com. This research uses CNN and LSTM for the method and Word2Vec as feature expansion. The accuracy result of the LSTM method is 51%, hybrid LSTM-CNN is 53%, and CNN-LSTM is 62%.

Based on the five journals above, the conclusion is that using CNN and Bi-LSTM methods can provide high accuracy and be applied to big data. This research refers to the journals above and uses CNN

combined with Bi-LSTM methods and Word2Vec as feature expansion. The researcher does not use the four scenarios directly explained in section 3, which are not used in the five journals above. The amount of the dataset used in this research is higher than in 3 previous research by Wang Yue and Lei Li, Kai Zhou, and Fei Long, and Dedi et al., whereas the amount of the dataset is lower than in two previous research by Gandhi et al. and Tam et al. Based on the different scenarios for this model, the researcher predicts that the overall accuracy in this hybrid method will reach more than 90%.