ICIC Express Letters ICIC International ©2023 ISSN 1881-803X
Volume **17**, Number **2**, February **2023** pp. 171–179

UTILIZING BERT AND CNN-LSTM IN STOCK PRICE PREDICTION USING DATA SENTIMENT ANALYSIS AND TECHNICAL ANALYSIS OF STOCK AND COMMODITY

Bambang Sulistio[1,*] and Derwin Suhartono[2]

[1]Computer Science Department, BINUS Graduate Program – Master of Computer Science
[2]Computer Science Department, School of Computer Science
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
dsuhartono@binus.edu
[*]Corresponding author: bambang.sulistio@binus.ac.id

Received February 2022; accepted May 2022

Abstract. *The COVID-19 pandemic undoubtedly has affected people's lifestyles and stock investment activities. The government's policies to deal with the pandemic have an impact on increasing the number of investors in the stock market. Apart from profits, there are also risks associated with investing in stocks. To reduce the risk required analysis for stock price predictions. The data often used are stock data, commodity prices, and social media. The application of deep learning and natural language processing can help investors to process data. This paper proposes Convolutional Neural Network and Long Short-Term Memory (CNN-LSTM) for technical analysis predicting stock prices using stock and commodity price data and urges BERT for sentiment analysis using social media data. The CNN-LSTM method has the best performance compared to the other four methods. The results showed that the performance of this method was the best, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) were the smallest, and R Square (R^2) was the largest. The BERT method has the best classification performance using 5-epochs, Weight Macro Avg, Weighted Avg, Accuracy, and the highest F1-Score. CNN-LSTM and BERT are more appropriate to predict stock prices and give investors suggestions to make stock investment decisions based on technical analysis and sentiment analysis.*

Keywords: Bidirectional encoder representations from transformers, Convolutional neural network, Stock prediction, Long short-term memory

1. **Introduction.** Since the emergence of COVID-19, many newspapers have reported that the disease will cause a setback for the economies of the affected countries, as the spread and severity of the disease causes a high mortality rate, making the response of policymakers and individual behavior difficult to predict [1]. The effective policies taken by the government in dealing with the COVID-19 pandemic have had a massive impact on investment by increasing trust and increasing the number of market participants to invest in the stock market [2]. There are advantages such as dividends and capital gains, but there is also a risk of loss such as capital loss and liquidation risk. This risk creates depression and stress, which is caused by profit pressure, social pressure, workload, and decision making process [3].

To reduce the risk of errors, it is necessary to have a program that can analyze and predict stock prices using data often used, namely technical analysis and sentiment analysis. Technical analysis focuses on stock prices, and trading volume is used to predict stock prices [4]. Sentiment analysis focuses on investors' reactions to financial news and daily events related to issuers [5].

DOI: 10.24507/icicel.17.02.171

It is necessary to apply computer science in processing technical data and stock sentiment analysis. The Recurrent Neural Networks (RNN) Long Short-Term Memory (LSTM) algorithm can be used to study patterns in data, because it stores information about data patterns in its network architecture [6] and proves to give a solution to overcome the vanishing gradient in RNN when processing long sequential data. In other research LSTM is also proven applicable to predicting the stock market in the future [6]. There is also a CNN-LSTM hybrid application that uses a method that combines the advantages of Convolutional Neural Network (CNN) to extract data features, and LSTM to find the dependence between data on time series [4] in which the result can provide stock price forecasting with the highest prediction accuracy. Related to this study, the CNN-LSTM method is used to examine stock price predictions using technical data combined with commodity price data produced by the issuer company, because researchers find that there is a correlation between stock index prices, energy, gold, and food commodity prices, which are used to estimate future volatility in the stock market [7].

Apply Natural Language Processing (NLP) model to explore online text information for stock prediction [8]. Bidirectional Encoder Representations from Transformers (BERT) language representation NLP model is successful in many tasks such as answering questions, and text classification [9], where this model can handle a series of NLP tasks with a significant increase in accuracy values. This study examines BERT for modeling Twitter social media data for sentiment classification for stock price movement analysis.

One of the pillars on the rise of the Composite Stock Price Index (IHSG) in early 2022 is mining sector shares [11]. Technical data and social media news related to mining are used to predict stock prices of issuers in the mining sector: PT. Aneka Tambang (ANTM) represents gold, Vale Indonesia Tbk (INCO) represents Nickel and PT. Medco Energy (MEDC) represents the oil commodity. Data used include daily technical (closing price and volume) on the Indonesian stock exchange trading, daily data on the price of gold, oil, and nickel commodities in international markets, and sentiment data on stocks on Twitter social media related to ANTM, INCO, and MEDC shares.

2. **Proposed Method.** This study uses the stages of the process used in technical analysis and sentiment analysis which is illustrated in Figure 1, which consists of Initialization data, Pre-processing, Prediction Model, and Evaluation.

2.1. **Technical analysis process.** Based on the research method above, the following details will explain the stages of stock price prediction using technical data analysis. Initialization Data: collecting data on Indonesian stock exchange and commodity prices from the finance.yahoo.com website. The stock and commodity data taken are transaction data for the last 15 years in IDR currency units. Then we download the data using the csv file format. The stock data feature that is used is the closing price of shares and volume, as well as commodity data using the closing price. The csv file is read and combined into a dataset with a 2D array data frame type [10]. Pre-processing: the process of converting data into a NumPy type, forming the input data into a tensor [11]. The normalization process uses MinMaxScaler to calculate all features at the same scale and speed up the computing process. The dataset is split into three: training, evaluation, and test datasets. The training dataset is used in the training process, the evaluation dataset is used in the optimization process to compare the prediction error of one model with another model and the test dataset is used to predict data and evaluate the performance of the model in this experiment. Then the sliding window dataset consists of input and output in the form of a 3D array dataset that has the same window size. Prediction Model: carry out the training process using a sliding window dataset, starting with calculating the CNN layer. The data is successively passed through the convolution layer and the pooling layer. This process functions to extract input data features and produce output data, and

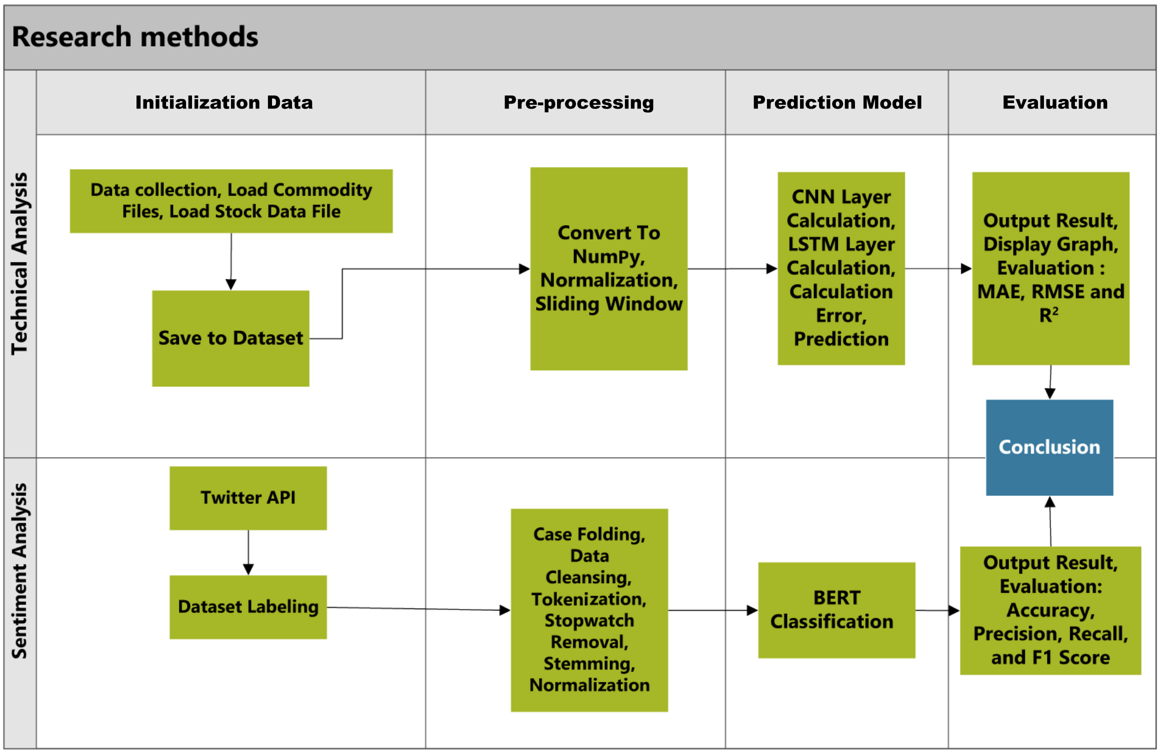

Figure 1. Flow process: Technical analysis and sentiment analysis process

then it is used as input data in the calculation of the LSTM layer that creates an output value, which is used as a complete connection layer input to get an approximate output value. Evaluation: To find the best method for predicting the value of this stock, the authors perform a technical analysis by comparing the error rates of the models: Gated Recurrent Unit (GRU), Multilayer Perceptron (MLP), CNN, LSTM and CNN-LSTM. The process of calculating the prediction error rate of the various models above uses the formula: MAE, RMSE and R^2

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{2}$$

$$\mathrm{R}^2 = 1 - \frac{\left(\sum_{i=1}^{n} (y_i - \hat{y}_i)^2\right)/n}{\left(\sum_{i=1}^{n} (\bar{y}_i - \hat{y}_i)^2\right)/n} \tag{3}$$

The closer the MAE and RMSE values to 0, the higher the forecasting accuracy. The closer R^2 to 1, the better the level of match value.

2.2. **Sentiment analysis process.** Based on the research method, the authors will explain the stages of stock sentiment analysis. Initialization Data: collecting news datasets from Twitter social media, using API (Application Programming Interface) to filter ANTM, MEDC, and INCO news. The news dataset is carried out by a sentence splitting process to separate the data in the text into sentence, then stored in the form of a csv file. Furthermore, the comments sentence is labeled with a dataset with a positive sentiment value of 2, a neutral value of 1, and a negative value of 0 [12], so that the model can see and understand how comments have negative, neutral, and positive sentiments [13]. Pre-processing is the process of preparing unstructured data into more structured data by performing several stages, namely Case folding: making all letters lowercase; Data cleaning: cleaning two or more repeated characters, links, usernames, hashtag numbers, symbols, extra spaces, punctuation marks, and numbers; Tokenization: breaking sentences into chunks of words, punctuation marks, and other meaningful expressions according to the terms of the language used; Stopwords removal: removing words without meaning using the Indonesian stopwords library provided by NLTK and added with the Tala dictionary; Stemming: changing words that have affixes into root forms by removing affixes such as prefixes, suffixes, and confixes; Normalization: the process of changing non-standard words into standard words according to spelling using the Alay2 Dictionary [14]. BERT Classification: Performing tokenizer aimed at adjusting the sentence dataset so that it can be accepted by BERT. The sentence dataset inputs through the encoder by self-attention and produces output through the feed-forward network which is then continued to the next encoder. After passing through all the encoders, each token per position gives an output in the form of a vector with a hidden size of 768 on BERTBASE which has 12 layers [15]. To achieve good results, BERT output uses a single neural network that functions as a classifier [9]. The classifier layer generates logits in the form of predictions of the probability of the comments to be classified. The layer used is a fully connected neural network with the softmax function to convert these logits into probabilities using the formula:

$$softmax(z)_i = \frac{e^{zi}}{\sum_{j=1}^{K} e^{zj}} \tag{4}$$

Evaluation: to see the value of the accuracy of the model which is the result of sentiment analysis of the comments in the BERT training dataset, a confusion matrix is used. Furthermore, the values of accuracy, precision, recall, and F1-score are calculated. Accuracy aims to show the percentage of input that the neural network predicts correctly. So in evaluating the effectiveness of BERT for the classification of sentiment analysis, methods are used, namely Accuracy, Precision, Recall, and F1 Score [13], with the formula:

$$Accuracy = \frac{TP + TNt + TN}{TP + FP + TNt + FNt + TN + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + TF} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$\frac{1}{F1} = \frac{1}{2}\left(\frac{1}{precision} + \frac{1}{recall}\right) \tag{8}$$

To find the best results of sentiment analysis, an evaluation of the number of epochs was carried out using the BERT method.

3. **Numerical Example.** This research produces two evaluation results on technical analysis and sentiment analysis as follows.

3.1. **Results of sentiment analysis.** Comments related to ANTM, INCO, and MEDC stocks were taken using the Twitter API totaling 4120 comments dataset. The dataset is then split and labeled. Furthermore, the dataset was carried out by case folding, data cleaning, tokenization, stopwords removal, stemming, normalization, and dataset split (training, validation, and testing). Next, setting the hyper parameters as follows: Dropout layer of 0.1 [9], Learning rate: 0.00002, and Batch size: 16 [16]. Then the dataset was studied using BERT by comparing 5 epochs [17], 10 epochs [18], and 16 epochs [19]. Then perform sentiment analysis predictions using the model formed using dataset testing. The prediction results are then evaluated and written in Table 1, Table 2, and Table 3

Table 1. Accuracy with 5 epochs

	Precision	**Recall**	**F1-score**	**Support**
Negative	0.76	0.38	0.51	66
Neutral	0.82	0.89	0.86	588
Positive	0.55	0.49	0.52	170
Accuracy			0.77	824
Macro avg	0.71	0.59	0.63	824
Weighted avg	0.76	0.77	0.76	824

Table 2. Accuracy with 10 epochs

	Precision	**Recall**	**F1-score**	**Support**
Negative	0.63	0.47	0.54	66
Neutral	0.81	0.79	0.84	588
Positive	0.55	0.46	0.50	170
Accuracy			0.75	824
Macro avg	0.66	0.60	0.63	824
Weighted avg	0.74	0.75	0.74	824

Table 3. Accuracy with 16 epochs

	Precision	**Recall**	**F1-score**	**Support**
Negative	0.57	0.48	0.52	66
Neutral	0.81	0.88	0.84	588
Positive	0.54	0.43	0.48	170
Accuracy			0.75	824
Macro avg	0.64	0.60	0.62	824
Weighted avg	0.74	0.75	0.74	824

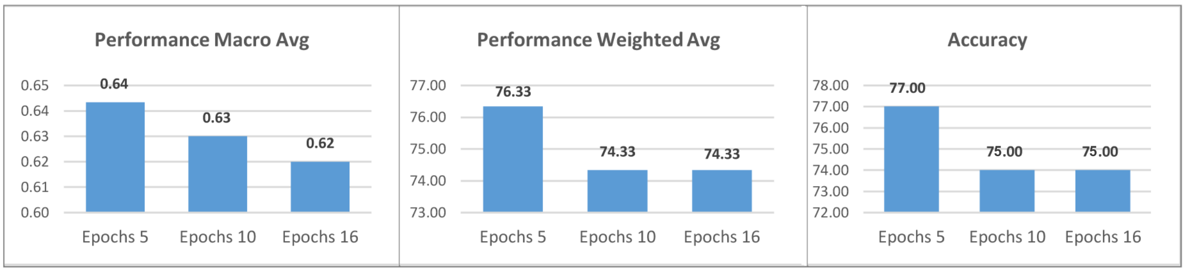

Figure 2. Performance macro avg, weighted avg, and accuracy graphic

respectively showing the accuracy of the test results in each experiment using 5 epochs, 10 epochs, and 16 epochs.

The results of the calculation of the averaged precision, recall, and F1-score values are seen in the macro avg performance graph in Figure 2 that the experiment using 5 epochs has the highest value of 64%, higher than 10 epochs which is worth 63%, 1% difference, and higher than 16 epochs which are worth 62%, a difference of 2%. The results of the weighted avg calculation for precision, recall, and F1-score values that the averaged can be seen in the weighted avg performance graph in Figure 2, an experiment with 5 epochs has the highest value of 76.33%, higher than 10 epochs and 16 epochs which are worth 74.33%, 2% difference. The results of the accuracy test can be seen in the accuracy graph in Figure 2 that the F1-score 5 epoch value is 77%, higher than 10 epochs and 16 epochs which are worth 75% with a difference of 2%.

3.2. **Results of technical analysis.** To collect historical data on stock from Indonesia stock exchange and commodity prices, the data is available on the finance.yahoo.com website. Daily data interval for the past 15 years of trading days (3925 data) in IDR using the csv file format stored in the variable dataset. It consists of features: closing price of shares, volume, and closing price of commodities. The dataset is then converted to NumPy type, normalized, split the dataset (training, evaluation, and test), and form a sliding window. Then the training process is carried out using the architecture as shown in Figure 3 which illustrates the CNN-LSTM model used, consisting of 2 convolutional layers, 2 padding layers, 2 LSTM layers, 3 dense layers, and 2 dropout layers. The activation function used in the layer is the real function. The dense layer consists of 1 neuron because the output required is only 1 value, namely the stock price prediction value. This dense layer uses a linear activation function, this function is used because it has the highest prediction accuracy value [20].

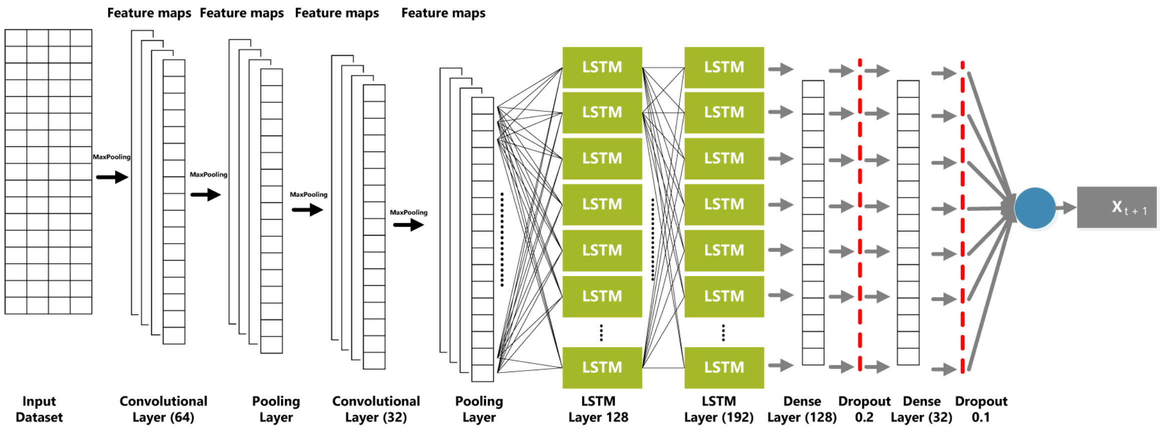

Figure 3. CNN-LSTM hyper parameters

The training dataset is processed using MLP, GRU, CNN, LSTM, and CNN-LSTM. The training model generated by each of these algorithmic processes is used to predict the test dataset, and the actual values are compared with the predicted ones. The graph in Figure 4 is the result of predictions using CNN-LSTM which compares the predicted and real values of the shares used (ANTM, INCO, and MEDC).

Evaluation of the predicted value of the stock price and the actual value uses: MAE, RMSE, and R^2. Table 4 contains information on the calculation results of the error rate prediction of ANTM, INCO, and MEDC stock prices.

The graph in Figure 5 is a comparison of the average MAE value. RMSE and R^2 are obtained from the prediction results using the MLP, GRU, CNN, LSTM and CNN-LSTM models.

From the values in Table 4 and Figure 5, the CNN-LSTM method has the smallest MAE and RMSE values, and the maximum R^2 is close to 1. When compared to CNN-LSTM and LSTM, it is found that CNN-LSTM has lower MAE and RMSE values than LSTM, MAE decreased from 0.904 to 0.798 by 11.72%, RMSE decreased from 1.281 to 1.100 by 14.12%, R^2 increased from 0.920 to 0.956 by 3.9%. This shows that the LSTM forecasting performance can be improved effectively by extracting data features through CNN.

4. **Conclusions.** To reduce investment risk during the COVID-19 pandemic, analysis is needed to predict stock prices based on technical analysis and sentiment analysis. After doing the research, we conclude that the prediction results of technical analysis using CNN-LSTM are better than the MLP, GRU, CNN, and LSTM methods. With values: MAE was successfully reduced by 0.798 by 11.72%, RMSE decreased by 1.100 by 14.12%,

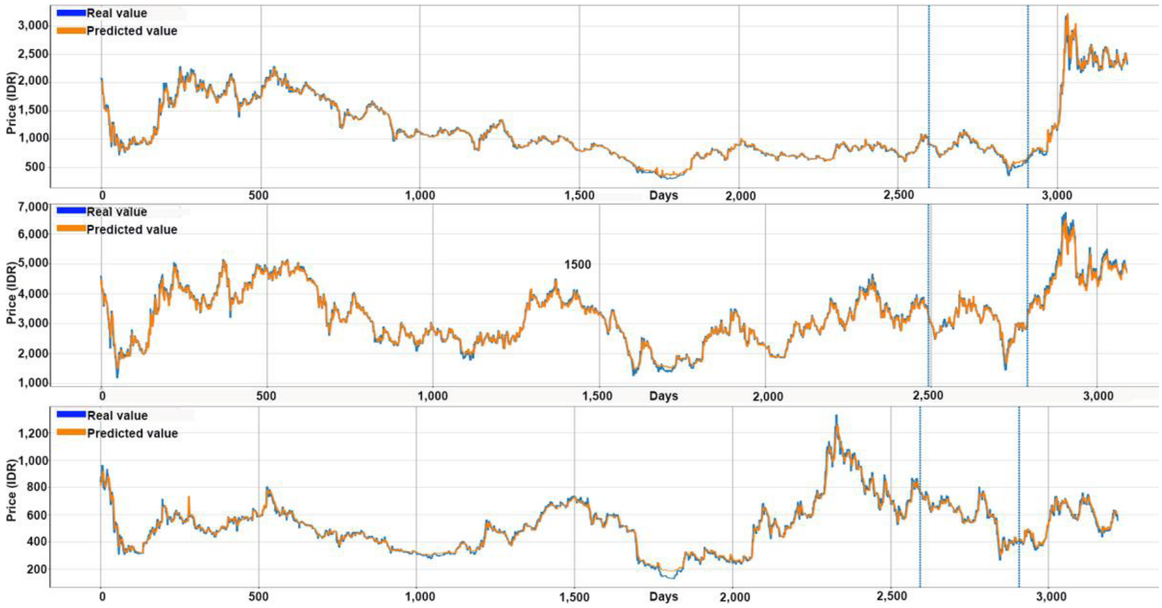

FIGURE 4. ANTM, INCO, and MEDC predicted and real values for CNN-LSTM

TABLE 4. MAE, RMSE and R^2 result

Method	Stock	MAE	AVG MAE	RMSE	AVG RMSE	R^2	AVG R^2
MLP	ANTM	2.380	2.500	3.497	3.405	0.779	0.585
	INCO	4.529		5.980		0.386	
	MEDC	0.590		0.739		0.589	
GRU	ANTM	1.524	1.386	2.377	1.945	0.898	0.861
	INCO	2.266		2.998		0.845	
	MEDC	0.368		0.461		0.839	
CNN	ANTM	2.133	2.523	3.157	3.390	0.820	0.594
	INCO	4.893		6.331		0.312	
	MEDC	0.542		0.682		0.649	
LSTM	ANTM	1.140	0.904	1.660	1.281	0.950	0.920
	INCO	1.229		1.759		0.946	
	MEDC	0.344		0.424		0.864	
CNN-LSTM	ANTM	0.745	0.798	1.091	1.100	0.978	0.956
	INCO	1.460		1.973		0.933	
	MEDC	0.188		0.235		0.958	

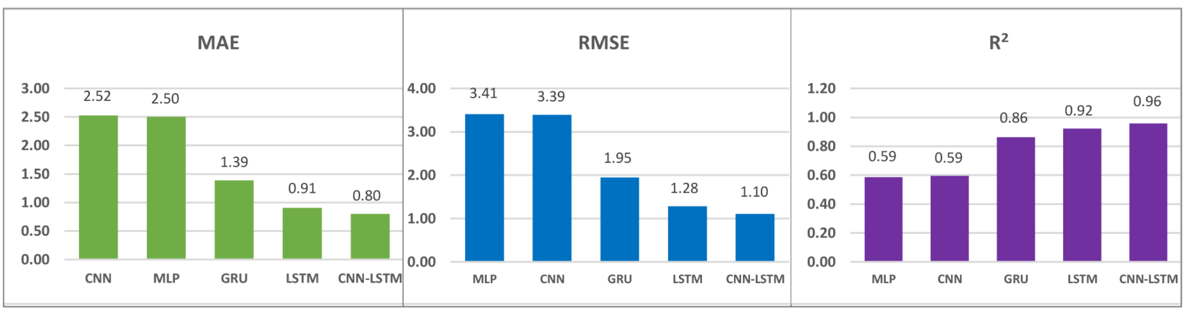

FIGURE 5. Comparison of the average value of MAE, RMSE and R^2

and R^2 increased by 0.956 by 3.9%. So the combination of CNN and LSTM is proven reliable to reduce the error value and increase the efficiency of stock prediction using technical analysis. Sentiment analysis results using 5 epochs BERT is better than using 10 epochs and 16 epochs. The value obtained is performance macro avg that 5 epochs have the highest value of 64%, higher than 10 epochs which is 63% with a 1% difference, and higher than 16 epochs which are 62% with a 2% difference. Performance weighted avg got 5 epochs having the highest value of 76.33%, higher than 10 epochs and 16 epochs which worth 74.33% with a difference of 2%. Performance accuracy obtained an F1-score of 5 epochs of 77%, higher than 10 epochs and 16 epochs worth 75% with a difference of 2%. So BERT using 5 epochs is proven to have the best macro avg, weighted avg, and F1-score performances. The researchers suggest that further research should experiment with a combination of the CNN-LSTM and BERT methods, using technical data analysis and sentiment analysis.

Acknowledgment. The authors would like to thank Bina Nusantara University for their support so that this research can be carried out.

REFERENCES

[1] R. Baldwin and B. W. di Mauro, *Economics in the Time of COVID-19*, CEPR Press, London, 2020.
[2] K. Khan, H. Zhao, H. Zhang, H. Yang, M. H. Shah and A. Jahanger, The impact of COVID-19 pandemic on stock markets: An empirical analysis of world major stock indices, *Journal of Asian Finance, Economics and Business*, vol.7, no.7, pp.463-474, 2020.
[3] T. Oberlechner and A. Nimgade, Work stress and performance among financial traders, *Stress and Health*, vol.21, no.5, pp.285-293, 2005.
[4] W. Lu, J. Li, Y. Li, A. Sun and J. Wang, A CNN-LSTM-based model to forecast stock prices, *Complexity*, vol.2020, 6622927, 2020.
[5] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia and D. C. Anastasiu, Stock price prediction using news sentiment analysis, *2019 IEEE 5th International Conference on Big Data Computing Service and Applications (BigDataService)*, Newark, CA, USA, 2019.
[6] M. Roondiwala, H. Patel and S. Varma, Predicting stock prices using LSTM, *International Journal of Science and Research (IJSR)*, vol.6, no.4, pp.1754-1756, 2017.
[7] W. Mensi, M. Beljid, A. Boubaker and S. Managi, Correlations and volatility spillovers across commodity and stock markets: Linking energies, food, and gold, *Economic Modelling*, vol.32, pp.15-22, 2013.
[8] X. Ding, Y. Zhang, T. Liu and J. Duan, Knowledge-driven event embedding for stock prediction, *Proc. of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pp.2133-2142, 2016.
[9] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, *Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, MN, USA, pp.4171-4186, 2019.
[10] K. A. Althelaya, E.-S. M. El-Alfy and S. Mohammed, Evaluation of bidirectional LSTM for short- and long-term stock market prediction, *2018 9th International Conference on Information and Communication Systems (ICICS)*, 2018.
[11] F. Chollet, *Deep Learning with Python*, Simon and Schuster, New York, 2021.
[12] Y. Goldberg, *Neural Network Methods for Natural Language Processing*, Morgan and Claypool Publishers, CA, 2017.
[13] M. Li, W. Li, F. Fang, X. Jia and G. Rui, Applying BERT to analyze investor sentiment in stock market, *Neural Computing and Applications*, vol.33, pp.4663-4676, 2021.
[14] N. A. Salsabila, Y. A. Winatmoko, A. A. Septiandri and A. Jamal, Colloquial Indonesian lexicon, *2018 International Conference on Asian Language Processing (IALP)*, 2018.
[15] H. Xu, B. Liu, L. Shu and P. Yu, BERT post-training for review reading comprehension and aspect-based sentiment analysis, *Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, MN, USA, pp.2324-2335, 2019.
[16] C. Sun, X. Qiu, Y. Xu and X. Huang, How to fine-tune BERT for text classification?, *China National Conference on Chinese Computational Linguistics*, Kunming, China, 2019.

[17] I. Annamoradnejad, M. Fazli and J. Habibi, Predicting subjective features from questions on QA websites using BERT, *2020 6th International Conference on Web Research (ICWR)*, 2020.
[18] Y. Song, J. Wang, Z. Liang, Z. Liu and T. Jiang, Utilizing BERT intermediate layers for aspect based sentiment analysis and natural language inference, *arXiv.org*, arXiv: 2002.04815, 2020.
[19] M. R. Yanuar and S. Shiramatsu, Aspect extraction for tourist spot review in Indonesian language using BERT, *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2020.
[20] M. Rana, M. M. Uddin and M. M. Hoque, Effects of activation functions and optimizers on stock price prediction using LSTM recurrent networks, *Proc. of the 2019 3rd International Conference on Computer Science and Artificial Intelligence*, pp.354-358, 2019.