



Available online at <http://scik.org>

Commun. Math. Biol. Neurosci. 2023, 2023:102

<https://doi.org/10.28919/cmbn/8168>

ISSN: 2052-2541

# STOCK PRICE PREDICTION ON INDONESIA STOCK MARKET WITH THE INFLUENCE OF EXTERNAL FACTORS USING RECURRENT NEURAL NETWORK WITH ATTENTION MECHANISM

HADRIAN\*, GEDE PUTRA KUSUMA

Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia, 11480

Copyright © 2023 the author(s). This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract:** In recent years, more businesses and individuals have started to rely on data as their decision-making factor. Past data has been very useful in making the next decision about business activities in many business sectors, such as in investing activities. Many people invest in stocks without capable knowledge to analyze a proper performance of the stocks as there are a lot of factors affecting the value of stocks. To predict the value of the stock close value, deep learning especially recurrent neural network is applicable to predict the stock price. Previous research only involves the stock price data and vanilla recurrent neural network. This research predicts the stock price with the influence of external factors using recurrent neural network with attention mechanism. A block each for LSTM and attention mechanism model is created and can be repeated several times as needed. The AALI stock price data chosen to represent one of the biggest sectors in Indonesia, with 10 external factors related to the stocks. The result shows that the additional method of using external data with feature selection and attention mechanism helps to improve the model performance in predicting the stock prices. With the right combination through tuning the model, a single block

---

\*Corresponding author

E-mail address: [hadrian001@binus.ac.id](mailto:hadrian001@binus.ac.id)

Received August 09, 2023

each for LSTM and attention mechanism giving the best performance of the model with average MAPE of 9.438 for validation and 17.593 for testing. Future research will be able to improve the model accuracy by involving other stock price data and more external data to get more relevant supporting data.

**Keywords:** stock prediction; recurrent neural network; deep learning; attention mechanism; external factors.

**2020 AMS Subject Classification:** 68T07.

## 1. INTRODUCTION

Nowadays, investment has been one of the new alternatives in preparing future savings. In Indonesia, a lot of people start to invest recklessly with minimum knowledge [1], [2]. In terms of investment, it is important to know what to be analyzed to predict the next value of the stock. Neural networks can be utilized as a tool to help investors predict the next price of the stock and help them to make investment decisions. There are many business activities, including investing in different business sectors using machine learning techniques, especially neural networks, to be data-driven by making predictions and decisions using available data [3]. There are two kinds of data that can be analyzed in stock prediction, fundamental data and technical data [4]. Main data of technical analysis involves the daily stock prices of the company. Closing price, open price, highest price, and lowest price, are the common data used in analyzing the stock price of a company. On the other hand, fundamental analysis analyzes the qualitative data of the company such as financial reports and news regarding the company. Due to these data, fundamental analysis method is less preferable to predict stock prices as it is hard to find the correlated data and the data analyzed is subjective to each investor [5]. Therefore, technical analysis is preferable to predict the price of the stock.

Predicting the value of stock is not limited to the use of the company's data, but also external data. Previous research in predicting stock prices indicates good accuracy when involving external data as the supporting data. In some related works to data prediction, external data such as commodity prices and exchange rates are being used as they affect the price of a company's stock. Some previous works also use news data for sentiment analysis to represent fundamental analysis

in price prediction [6]. But still, sentiment analysis is not preferred as it cannot provide enough data related to the business and subjective.

To find the best algorithm for stock prediction with the effect of the external data from the technical approach, this research proposed a hybrid algorithm which is Long Short-Term Memory Network (LSTM) combined with a feature selection process and attention mechanism. This hybrid method might be a chance to increase the accuracy and speed of prediction. By comparing the proposed algorithm with the solely working algorithm, the outcome of the best algorithm combination is desired in terms of its performance. The model was compared one to another through MAPE as performance metrics.

## **2. PRELIMINARIES**

Stock prediction using a machine learning system has been done using various methods. The most common algorithm used by researchers to predict the stock, which is a time-series data, is the long short-term memory (LSTM) algorithm [7],[8]. But researchers keep improving methods discovered to improve the performance of the algorithm by creating a hybrid algorithm with LSTM as the base model. Thus, a number of these hybrid algorithms have been showing good results compared to the common approach, such as Bayesian-LSTM [9], PSO-LSTM [10], K-means clustering LSTM [11], LSTM-GRU [12], and CNN-LSTM [13].

Researchers not only combine LSTM with other algorithms or modify the algorithms itself [14], [15]. They also add sentiment analysis to increase the accuracy of the predicted results [16]. Another researcher used the same process in 2021 but use BERT to process the sentiment analysis [17]. But the news data can be treated as subjective data as it is based on what people say about the stock. Moreover, the data collection for sentiment analysis is hard to get as the news data should match the time of the stock price data.

Based on all the related work reviewed above, LSTM has a very good performance in terms of predicting the value of stock value. However, data used in stock prediction is not just limited to technical data of the stock price itself. Commodity prices such as exchange rates, gold prices, as

well as crude oil, are a potential factor affecting the price of stock prices in the market. But using all the commodity prices to predict the stock price increases the computing cost and becomes inefficient. To make the prediction process effective and accurate, feature selection is involved in the process to choose which commodity is significantly correlated.

In this research, LSTM with attention mechanism used to increase the accuracy of the stock price prediction. Before the data is processed by the LSTM algorithm, feature pre-processing is conducted. The filter method is chosen for the feature selection process and compares the Pearson correlation coefficient between each commodity price with the stock price data.

By the proposed input data and model, this research takes the opportunity to improve the accuracy of the stock value prediction using the hybrid LSTM algorithm. The dataset used in the research is taken from the historical data of a company stock value in Indonesia Stock Market and the common commodity prices.

### **3. THEORY AND METHODS**

**3.1. Feature Selection.** Working with data sometimes can be overwhelming due to the dimension of the data. As a result, it creates some difficulties in the prediction process as there are a lot of features that need to be used as predicting features. Moreover, it might also drag down the performance of the predicting model. To resolve the problem, feature selection can be done. It is one of the dimensionality reduction techniques to choose the only relevant features and remove the irrelevant one [18]. By reducing the dimension or the feature of the dataset, it can increase the efficiency and reduce the computational cost of the predicting process. But differ from dimensionality reduction, feature selection removes a feature from the dataset while dimensionality reduction reconstructs the whole dataset into a new input. Feature selection is also a technique to prevent the overfitting of prediction due to the irrelevant data input. Feature selection is divided into two methods, supervised and unsupervised methods. In the unsupervised method, the feature selection ignores the target variable and removes all irrelevant variables commonly by correlation. On the other hand, the supervised method recognizes the target variable

and removes all the variables irrelevant to the target. There are two common methods used for feature selection, namely filter method and wrapper method [19]. However, the filter method handles the feature selection process in this research.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \quad (1)$$

Filter method is one of the feature selection methods which assess the data and select the features that fulfill the criteria needed for the process without the involvement of learning algorithms [20]. To choose the relevant features, the feature importance level is determined by the criterion applied where each data type uses a different correlation coefficient. In this research, the features and target used are both continuous. Hence, the Pearson's Correlation Coefficient is used to assess the correlation between the feature and the target. The highest correlation is taken as the relevant features and the other is removed. The correlation coefficient value varies from -1 to +1, calculated using Equation (1), where  $cov$  is the covariance of both features and target and  $\sigma$  for the standard deviation of both features and target.

**3.2. Recurrent Neural Network.** Recurrent neural network is one kind of neural network packed with a short-term memory used to give feedback inputs for the recurrent process [21]. This short-term memory makes the algorithm able to process sequential data as the algorithm analyzes the patterns from each repetition to solve the problem. Figure 1 indicates the simple flow of a recurrent neural network.

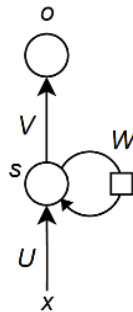


Figure 1 Recurrent Neural Network Flow

The recurrent neural network starts with the input  $x$  at iteration  $t$  into  $s$ , which is the memory of the neural network. The variable  $s$  is a hidden state which usually uses the activation function of

$\tanh$ . The variable  $U$ ,  $V$ , and  $W$  are the parameters that used in the repetition of the network and passed to the next neural network unit. These parameters are configured the same for each layer of the parameter and reduce the complexity of the computation.

There are a lot of recurrent neural networks models that have been developed to solve certain cases as each model performed differently based on the dataset used and the output desired. In processing time-series data, Long-Short Term Memory (LSTM) is the commonly used model and therefore is chosen for this research.

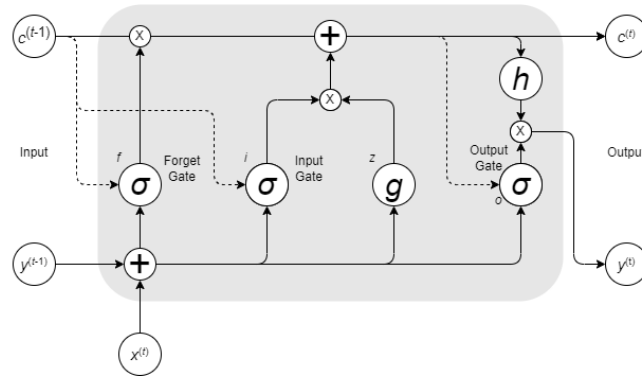


Figure 2 Long-Short Term Memory Block

Long-short term memory (LSTM) is categorized as one kind of recurrent neural network introduced in 1997 by Hochreiter and Schmidhuber [22]. This neural network is developed to solve time series related problems. The neural network itself is constructed from three gate structures where each of them is called input gate, forget gate, and output gate [23]. The original structure proposed by Hochreiter and Schmidhuber was composed of just two gates which is input and output gate. The forget gate is a new solution suggested by Ger et al in 2000 [24]. The function of this gate is to give the opportunity for each block to reset its' state back. Figure 2 is an example of an LSTM block where  $x^{(t)}$  is the input signal,  $y^{(t)}$  is the output signal,  $\sigma$  is the gate,  $g$  and  $h$  are the input activation function,  $+$  is to sum all inputs,  $\times$  is to multiply all inputs, and  $c^{(t)}$  is the cell state which determines what information should be kept in the network.

$$z^{(t)} = g(W_z x^{(t)} + R_z y^{(t-1)} + b_z) \quad (2)$$

The workflow of an LSTM block is simply to pass the signals and inputs through the configured gates in the block. The input of the block is configured of the current input used,  $x^{(t)}$ , and the

processed output from the previous LSTM block,  $y^{(t-1)}$ . Equation (2) is the configured input equation to the LSTM block where  $W_z$  is the weight imposed for  $x^{(t)}$ ,  $R_z$  is the weight for  $y^{(t-1)}$ ,  $b_z$  is the bias weight vector, and  $i$  represents the tanh activation function.

$$i^{(t)} = \sigma(W_i x^{(t)} + R_i y^{(t-1)} + p_i \times c^{(t-1)} + b_i) \quad (3)$$

The input signal then travels to the input gate. The input gate filters the input signals according to the cell state,  $c^{(t-1)}$  from the previous iteration. The input signal also processed through the forget gate to remove information based on the previous cell state,  $c^{(t-1)}$ . The process undergoes using Equation (3) where  $W_i$  is the weight imposed for  $x^{(t)}$ ,  $R_i$  is the weight for  $y^{(t-1)}$ ,  $p_i$  is the rate for  $c^{(t-1)}$ ,  $b_i$  is the bias weight vector, and  $\sigma$  represents the sigmoid activation function.

$$f^{(t)} = \sigma(W_f x^{(t)} + R_f y^{(t-1)} + p_f \times c^{(t-1)} + b_f) \quad (4)$$

The input signal also processed through the forget gate to remove information based on the previous cell state,  $c^{(t-1)}$ . The process is done with the Equation (4) where  $W_f$  is the weight imposed for  $x^{(t)}$ ,  $R_f$  is the weight for  $y^{(t-1)}$ ,  $p_f$  is the rate for  $c^{(t-1)}$ ,  $b_f$  is the bias weight vector, and  $\sigma$  represents the sigmoid activation function.

$$c^{(t)} = z^{(t)} \times i^{(t)} + c^{(t-1)} \times f^{(t)} \quad (5)$$

The next step is to calculate the cell for the block to be used in the next block. The cell,  $c^{(t)}$ , is constructed by the block input, input gate, and the forget gate using the Equation (5) where  $z^{(t)}$  is the block input,  $i^{(t)}$  is the value of input gate,  $f^{(t)}$  is the value of forget gate, and  $c^{(t-1)}$  is the cell value of previous iteration.

$$o^{(t)} = \sigma(W_o x^{(t)} + R_o y^{(t-1)} + p_o \times c^{(t-1)} + b_o) \quad (6)$$

The output gate is calculated using the current input,  $x^{(t)}$ , the last output of LSTM block,  $y^{(t-1)}$ , and the cell value of the previous block,  $c^{(t-1)}$ . The calculation is done using the Equation (6) where  $W_o$  is the weight imposed for  $x^{(t)}$ ,  $R_o$  is the weight for  $y^{(t-1)}$ ,  $p_o$  is the rate for  $c^{(t-1)}$ ,  $b_o$  is the bias weight vector, and  $\sigma$  represents the sigmoid activation function.

$$y^{(t)} = \sigma(W_i x^{(t)} + R_i y^{(t-1)} + p_i \times c^{(t-1)} + b_i) \quad (7)$$

The last product of the block is the block output, which is passed to the next block,  $y^{(t)}$ . Block output is the product of the cell and the output gate value, calculated using the Equation (7) where

$c^{(t)}$  is the cell of the block,  $o^{(t)}$  is the output gate value and  $g$  represent the tanh activation function.

The process of individual LSTM blocks is integrated with one another to be able to provide a good model to train the dataset and predict an accurate result. In terms of the original LSTM, it has been proven that the neural network has performed great regarding the past related works discussed. But the performance can still be improved by collaborating LSTM with other methods.

**3.3. Attention Mechanism.** Attention mechanism (AM) is a deep learning model created to solve the problem on the fixed length encoding vector which is unable to access the input with longer sequence [25]. The mechanism is taken from the human brain where humans only look and give attention to the relevant only. To apply this mechanism attention method, a neural machine translation is applied where it calculates the correlation level between the input and the output of translation model. The attention mechanism is proposed to handle Natural Language Processing (NLP) problems as it involves a very long input data. But the mechanism itself is applicable to other problems other than NLP problems to help the neural network focusing on the relevant data only and remove the redundant. There are two common attention mechanisms available to be applied such as Dot-Product Attention and Multi-Head Attention [26]. However, in this opportunity, Multi-Head Attention is chosen to give a better performance.

$$MultiHead(Q, K, V) = [head_1, \dots, head_h] W_0 \quad (8)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (9)$$

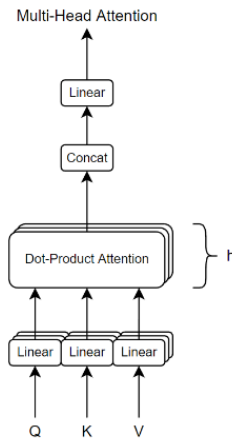


Figure 3. Multi-Head Attention Framework



Multi-Head Attention mechanism offers simultaneous attention calculation to make the process efficient [27]. Each of the input goes through separate Head, where each Head performs attention function, and combined (concatenated) at the end of the parallel computation. At last, the attention function is applied once again to the output of the concatenated values to produce the final output. Equation (8), Equation (9), and Figure 3 visualize the calculation and framework of the mechanism.

#### **4. PROPOSED METHOD**

Stock price prediction is very beneficial for investors to determine when do they need to buy the shares and when to sell. The prediction process of stock prices has been involving external data as that data affects the price of the stock on a particular day. There are lots of data involved such as commodity price, exchange rate, index rate, and sentiment analysis [28]. However, the use of sentiment analysis can cause problems in terms of data synchronization and tends to be subjective. In relation to external data, there is a lot of data involved and a method to choose the most relevant data is needed to reduce the computational cost and increase the accuracy of stock prediction. The application of deep learning in the search of the optimum model for stock price prediction has been through various models, including solely working models and hybrid models. RNN, especially LSTM, have been assessed to perform better than others as the algorithm have the advantage in processing time series data type, which is the data type for stock prices. There are some researches involved optimization algorithms to help optimize the hyperparameters in the neural network. Some also utilize clustering algorithms to cluster the input data to be used in the main processing model.

Most of the related works regarding stock price prediction only use the core technical data of the stock price from the company. This means that it only assesses and analyzes the pattern of the data and predicts based on the pattern. But the data of these works are limited and subjective. This research aims to develop a stock prediction model with attention mechanism and feature selection process for Indonesian stock in agricultural sector involving external supporting data.

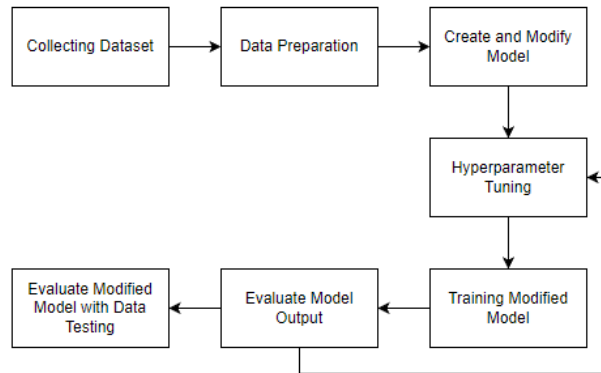


Figure 4. Model Development Process

This research proposed a set of processes to analyze the performance of the model in predicting the stock price. The flow of the model development process is illustrated in Figure 4 and discussed in detail in the following sub-sections.

**4.1. Dataset Collection.** The stock data used in this research is AALI in IDR. The commodity external data includes ZG gold price, Newcastle coal price, and CIF Rotterdam crude oil price. All the mentioned data is in USD. For composite index data, including IHSG, JKSRI, and Kompas100, is in IDR. The exchange rate external data used is the rate from US Dollar (USD), Chinese Yuan (CNY), and Euro (EUR), to Indonesian Rupiah (IDR). All data is available for the desired timeframe, which is from January 1, 2013, until December 30, 2022, with a total of 2487 data for each dataset.

**4.2. Dataset Preparation.** The collected individual datasets are combined into a single data frame that contains all the columns from each table. The date column represents the index of the data frame. Before being used, all date data was transformed into the same format. A column was created for the target variable which is the close price of the next day. This condition simulates the use of today's data to predict tomorrow's closing price. The table is created through manual selection from the data available. All the selected features are joined into one table. Table 1 summarized the joined table for the model input.

**Table 1.** Input Data Structure

Column Name	Data Type	Description
Date	Object	Specific date of stock price
Target	Float	Close price of the next date
Open	Float	Open price of the date
High	Float	Highest price of the date
Low	Float	Lowest price of the date
Close	Float	Close price of the date
Volume	Integer	Volume of stock traded on the date
g_Close	Float	Close gold price of the date
c_Close	Float	Close coal price of the date
o_Close	Float	Close crude oil price of the date
p_Close	Float	Close crude palm oil price of the date
i_Close	Float	Close IHSG price of the date
k_Close	Float	Close Kompas100 price of the date
j_Close	Float	Close JKSRI price of the date
u_Close	Float	Close USD rate to IDR of the date
c_Close	Float	Close CNY rate to IDR of the date
e_Close	Float	Close EUR rate to IDR of the date

**4.3. Feature Selection Method.** The feature selection process applied in this research is the filter method. The filter method is chosen due to better computational time and flexibility than wrapper method [29]. The features available in the table are assessed for its correlation towards the target. The target and features assessed can be seen in Table 1 from the previous section. The feature selection method is done by comparing the Pearson's Correlation Coefficient between each feature and the target. Equation (10) is used to evaluate the correlation.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \quad (8)$$

The feature is selected if the coefficient value is above the threshold given. This research used the threshold value of 0.6. The feature selected from this process becomes the input of the proposed model.

**4.4. Proposed Model.** The model proposed combines the process of multi-head attention mechanism as well as the LSTM layer. In relation to the combination, two main blocks are present in the architecture of the proposed model. One block handles the multi-head attention model, and the other block handles the LSTM layer. Both blocks are repeatable according to the desired value. The architecture of the model is shown in Figure 5.

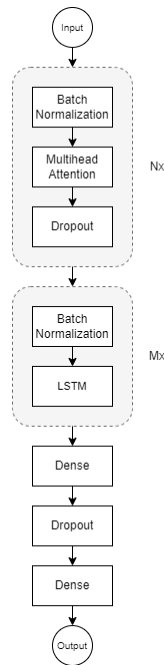


Figure 5. Proposed Model Architecture

The process starts with the input into the first block. The input data is first being normalized to prevent explosive gradient descent as well as there are different units in the dataset. The normalized data are then transferred to the multi-head attention module. The first package can be repeated a  $Nx$  times according to the desired number. The dimensions needed are predicted and created through the linear combination of the attention outputs and passed to the next block.

**Table 2.** Model Layers Configuration

Block	Layer	Layer Configuration	Output Shape
1	BatchNormalization_0	momentum = 0.09	(30, 6)
	MultiheadAttention	num_heads = 3 key_dim = 6	(30, 6)
	Dropout_0	rate = 0.2	(30, 6)
2	BatchNormalization_1	momentum = 0.09	(30, 6)
	LSTM	units = 15 activation = 'tanh'	(30, 15)
3	Dense_0	units = 8 activation = 'relu'	(8)
	Dropout_1	rate = 0.2	(8)
	Dense_1	units = 1 activation = 'relu'	(, 1)

The second block starts with batch normalization to ensure the data is normalized. Then the data is passed to the next module, the LSTM module. This type of recurrent network can process time series data which is the type of all input data. The second block can be repeated  $Mx$  times according to the number of iteration desired. The output is then passed to dense layers to extract the features of all the neurons and then to the drop layer to avoid the chance of overfitting occurred in the model. The layers are configured as shown in Table 2.

**4.5. Incremental Training.** Incremental training is a method in machine learning where a model is trained using updated data to keep the model updated with new knowledge about the data [30]. The method is beneficial in time-series prediction as the data changes along the period. While using the method, the model keeps its settings based on the previous data and improves in the next process through the new updated data. The new updated data fed into the model updates the parameter values in the model, adapts, and predicts more accurately based on the new data. Figure 6 visualizes the application of incremental training to a model.

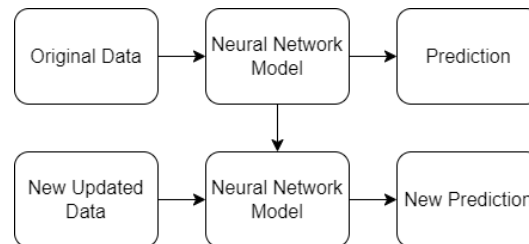


Figure 6. Incremental Training Implementation

**4.6. Hyperparameter Tuning.** The tuning of the hyperparameters involved in the prediction process will be handled by the Keras Tuner. The tuning will be optimized by the Bayesian Optimization algorithm and will run ten trials where each trial will be executed three times. The configuration of the hyperparameters will be summarized in Table 3.

Hyperparameter	Variation
Learning Rate	[0.01, 0.001, 0.0001]
Number of LSTM Block (Mx)	[1, 2, 3]
Number of Attention Block (Nx)	[1, 3, 5]
Optimizer	[SGD, Adam]

**4.7. Evaluation.** The dataset should be divided into ten equal segments to train the model and test the model when ready. The experimental scenario follows the incremental training scenario. The models are trained with the first segment, validated with the second segment, and tested with the third segment. Then the trained model continues to be trained with the second segment data, validated with the third segment data, and tested with the fourth segment data. The sequence keeps on going, until the tenth segment data is tested. In addition, each data holds 242 records.

**Table 4.** Model Configuration

Model
LSTM
LSTM + External
LSTM + External + FS (0.6)
LSTM + AM
LSTM + AM + External
LSTM + AM + External + FS (0.6)

For all the time where the model is trained and tested, the result of MAPE is averaged for the validation and testing MAPE. These results are then compared between each configuration tested listed in Table 4.

## 5. RESULT AND DISCUSSION

All the model configurations are optimized by Keras Tuner using Bayesian Optimization with 10 trials and each trial is executed 3 times with 500 epochs. The Keras Tuner optimizes the learning

## STOCK PRICE PREDICTION USING RECURRENT NEURAL NETWORK

rate between 10<sup>-2</sup>, 10<sup>-3</sup>, and 10<sup>-4</sup>, as well as the optimizer of the model between SGD or Adam. The validation and testing average MAPE are summarized in Table 5. The models other than the proposed model are included and discussed to see if there are improvements in the predicting performance when applying certain methods to the vanilla LSTM. When comparing the vanilla LSTM model and the LSTM model involving external data, the average MAPE increases as it shows that there are irrelevant external data which causes the error in the prediction increases. When feature selection is added to the LSTM model involving the external data, the average MAPE improves compared to the vanilla LSTM model. This shows that involving external data with feature selection improves the predicting performance of the model.

**Table 5.** Average MAPE Result for Each Configuration

Model	Hyperparameters	Average Validation MAPE	Average Testing MAPE
1 LSTM	Learning Rate: [0.01, 0.0001, 0.01, 0.01, 0.01, 0.0001, 0.0001, 0.01] Optimizer: Adam	14.508	34.842
1 LSTM + External	Learning Rate: [0.01, 0.01, 0.0001, 0.001, 0.0001, 0.01, 0.001, 0.01] Optimizer: Adam	27.422	56.808
1 LSTM + External + FS (0.6)	Learning Rate: [0.01, 0.01, 0.001, 0.001, 0.01, 0.01, 0.01, 0.01] Optimizer: [Adam, Adam, Adam, SGD, Adam, Adam, Adam, Adam]	10.810	22.644
2 LSTM + 1 AM	Learning Rate: 0.01 Optimizer: Adam	11.398	18.869
<b>3 LSTM + 1 AM + External</b>	<b>Learning Rate: 0.01</b> <b>Optimizer: Adam</b>	<b>11.777</b>	<b>17.290</b>
1 LSTM + 1 AM + External + FS (0.6)	Learning Rate: 0.01 Optimizer: Adam	9.438	17.593

The result shows that the model with the configuration of 1 LSTM block and 1 AM block with feature selection of the external data obtained the best average MAPE both for validation and

testing. Every combination of the LSTM + AM model has a better average testing MAPE compared to the vanilla LSTM model. With the increase of the number of AM blocks used, the average validation MAPE also slightly increases, and some ends up worse than the vanilla LSTM. This is where the block number should be tuned to get an optimum result of the prediction.

Involving external data in the LSTM + AM model increases the average MAPE which is not improving the performance. However, when feature selection is applied, the performance improves for certain combinations. The results show that adding external data with feature selection improves the performance compared to only using the stock price data and adding attention mechanism to the model also improves the performance even further with condition to tune the number of LSTM and AM blocks used. In this case, the best model performing is the model with 1 LSTM block and 1 AM block, involving external data and feature selection. The best model result is bolded in Table 5.

## **6. CONCLUSION AND FUTURE WORKS**

The aim of the research is to predict the stock price in Indonesia stock market with the influence of external factors, in this case 10 commodity prices, using recurrent neural network with attention mechanism. The proposed model is compared to the vanilla LSTM. Incremental training is also implemented to all the model analyzed to increase its performance. The result shows that using external data with feature selection increases the performance of the prediction compared to the vanilla LSTM model. The addition of attention mechanism also improves the predicting performance even further but with fine tuning on the number of LSTM and AM blocks used in the model. The output of the research shows that the model with 1 LSTM block and 1 AM block involving external data with feature selection have performed better than the vanilla LSTM model, giving an average MAPE score of 9.438 for validation and 17.593 for testing. Thus, the proposed model can be used to predict the next day's closing price of the stock market.

However, there are some problems that occurred during the process of the research. The duration of the model computation time is still too long. The data used in the research is also



limited to Indonesia stock market and 10 external data, which might affect the performance of the model. Another limitation of the research is the hardware used is not powerful enough to compute more configurations which might improve the performance and the result of the model. Further research can be done by implementing the proposed model to various stock price data and involving more external data that might be related to the stock price data to increase the prediction accuracy. Better hardware can also be utilized to tune more hyperparameter configurations.

### **ACKNOWLEDGEMENTS**

The authors would like to thank all parties involved and contributed to the process of this research.

### **CONFLICT OF INTERESTS**

The authors declare that there is no conflict of interests.

### **REFERENCES**

- [1] S. Wrenn, Top financial literacy education gaps across generations. Investopedia, (2022). Retrieved June 24, 2022. from <https://www.investopedia.com/financial-literacy-survey-5223919>.
- [2] P. Lauria, Generation Z: Stepping into Financial Independence. Investopedia, (2022). Retrieved June 24, 2022. from <https://www.investopedia.com/generation-z-stepping-into-financial-independence-5224362>.
- [3] P. Taylor, Data-driven decision-making in organizations worldwide as of 2020, by sector, Statista, (2020). <https://www.statista.com/statistics/1235436/worldwide-data-driven-decision-making-organizations-by-sector>.
- [4] D. Spahija, S. Xhaferi, Fundamental and technical analysis of the stock price, *Int. Sci. J. Monte*. 1 (2018), 5-11. <https://doi.org/10.33807/monte.1.201904160>.
- [5] P. Christie, R. Isidore, Fundamental analysis versus technical analysis-a comparative review, *Int. J. Recent Sci. Res.* 9(2018), 23009-23013.
- [6] S. Mohan, S. Mullapudi, S. Sammeta, et al. Stock price prediction using news sentiment analysis, in: 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService), IEEE,

- Newark, CA, USA, 2019: pp. 205–208. <https://doi.org/10.1109/BigDataService.2019.00035>.
- [7] A. Sravani, C. Anusha, N.V.S. Shankar, A comparative analysis of machine learning algorithms in stock prediction, in: Proceedings of the International Conference on Industrial Engineering and Operations Management Rome, Italy, August 2-5, 2021.
- [8] J. Dhilipan, D.B. Shanmugam, I. Quraishi, Stock prediction using machine learning, *Int. J. Linguist. Comput. Appl.* 8 (2021), 33-36. <https://doi.org/10.30726/ijlca/v8.i4.2021.84008>.
- [9] B. Huang, Q. Ding, G. Sun, et al. Stock Prediction based on Bayesian-LSTM, in: Proceedings of the 2018 10th International Conference on Machine Learning and Computing, ACM, Macau China, 2018: pp. 128-133. <https://doi.org/10.1145/3195106.3195170>.
- [10] H. Liu, Z. Long, An improved deep learning model for predicting stock market price time series, *Digit. Signal Process.* 102 (2020), 102741. <https://doi.org/10.1016/j.dsp.2020.102741>.
- [11] Y. Chen, J. Wu, Z. Wu, China's commercial bank stock price prediction using a novel K-means-LSTM hybrid approach, *Expert Syst. Appl.* 202 (2022), 117370. <https://doi.org/10.1016/j.eswa.2022.117370>.
- [12] Y. Li, Y. Pan, A novel ensemble deep learning model for stock prediction based on stock prices and news, *Int. J. Data Sci. Anal.* 13 (2022), 139-149. <https://doi.org/10.1007/s41060-021-00279-9>.
- [13] Z. Shi, Y. Hu, G. Mo, et al. Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction, preprint, (2022). <https://doi.org/10.48550/ARXIV.2204.02623>.
- [14] A. Doshi, A. Issa, P. Sachdeva, et al. Deep stock predictions, preprint, (2020). <https://doi.org/10.48550/ARXIV.2006.04992>.
- [15] X. Zhang, C. Li, K.-L. Chen, et al. Stock prediction with stacked-LSTM neural networks, in: 2021 IEEE 21st International Conference on Software Quality, Reliability and Security Companion (QRS-C), IEEE, Hainan, China, 2021: pp. 1119–1125. <https://doi.org/10.1109/QRS-C55045.2021.00166>.
- [16] Lakshya, Prateek, D. Sethia, Stock price prediction using news sentiment analysis, in: 2022 2nd International Conference on Intelligent Technologies (CONIT), IEEE, Hubli, India, 2022: pp. 1–6. <https://doi.org/10.1109/CONIT55038.2022.9847747>.
- [17] C.R. Ko, H.T. Chang, LSTM-based sentiment analysis for stock price forecast, *PeerJ Computer Sci.* 7 (2021), e408. <https://doi.org/10.7717/peerj-cs.408>.

- [18] C. Wang, D. Han, Q. Liu, et al. A deep learning approach for credit scoring of peer-to-peer lending using attention mechanism LSTM, *IEEE Access*. 7 (2019), 2161-2168. <https://doi.org/10.1109/access.2018.2887138>.
- [19] Y.B. Wah, N. Ibrahim, H.A. Hamid, et al. Feature selection methods: case of filter and wrapper approaches for maximising classification accuracy, *Pertanika J. Sci. Technol.* 26 (2018), 329-340.
- [20] N. Sánchez-Marroño, A. Alonso-Betanzos, M. Tombilla-Sanromán, Filter methods for feature selection - a comparative study, in: H. Yin, P. Tino, E. Corchado, W. Byrne, X. Yao (Eds.), *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007: pp. 178-187. [https://doi.org/10.1007/978-3-540-77226-2\\_19](https://doi.org/10.1007/978-3-540-77226-2_19).
- [21] C.M. Wang, Network of recurrent neural networks, preprint, (2017). <https://doi.org/10.48550/ARXIV.1710.03414>.
- [22] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (1997), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [23] S.S. Abubaker, S.R. Farid, Stock market prediction using LSTM, *Int. J. Res. Appl. Sci. Eng.* 10 (2022), 3178-3184. <https://doi.org/10.22214/ijraset.2022.42039>.
- [24] F.A. Gers, J. Schmidhuber, F. Cummins, Learning to forget: continual prediction with LSTM, *Neural Comput.* 12 (2000), 2451-2471. <https://doi.org/10.1162/089976600300015015>.
- [25] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, preprint, (2014). <https://doi.org/10.48550/ARXIV.1409.0473>.
- [26] A. Vaswani, N. Shazeer, N. Parmar, et al. Attention is all you need, preprint, (2017). <https://doi.org/10.48550/ARXIV.1706.03762>.
- [27] E. Voita, D. Talbot, F. Moiseev, et al. Analyzing multi-head self-attention: specialized heads do the heavy lifting, the rest can be pruned, in: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Florence, Italy, 2019: pp. 5797-5808. <https://doi.org/10.18653/v1/P19-1580>.
- [28] W. Gu, L. Zhang, H. Xi, et al. Stock prediction based on news text analysis, *J. Adv. Comput. Intell. Intell. Inform.* 25 (2021), 581-591. <https://doi.org/10.20965/jaciii.2021.p0581>.
- [29] P. Narkhede, R. Walambe, S. Poddar, et al. Incremental learning of LSTM framework for sensor fusion in attitude

estimation, *PeerJ Computer Sci.* 7 (2021), e662. <https://doi.org/10.7717/peerj-cs.662>.

- [30] N.A. Nnamoko, F.N. Arshad, D. England, et al. Evaluation of filter and wrapper methods for feature selection in supervised machine learning, in: the 15th annual postgraduate symposium on the convergence of telecommunication, networking and broadcasting, Liverpool, (2014).