Forecasting the Stock Price of PT Astra International Using the LSTM Method

Edwin Setiawan Nugraha1,2, Zalfani Alika1, Dadang Amir Hamzah3*
1,2,3Department of Actuarial Science, Faculty of Business, President University, Bekasi, Indonesia
1edwin.nugraha@president.ac.id, 2zalfani.alika@student.president.ac.id, 3dadang.hamzah@president.ac.id

Abstract
Stocks are one of the long-term investment options and represent ownership in a company that can be acquired through buying and selling. Investment carries both profit potential and risks that must be faced by investors when they provide their capital to companies. Accurate stock price forecasts are very important because they provide an estimate of risk. This research aims to forecast the stock price of PT Astra International Tbk (ASII.JK) using a Long Short-Term Memory (LSTM) method. Data set closing stock prices were taken from January 2, 2015, to December 30, 2020, with a total observation of 1506. This data set is divided into 80% for training and 20% for training. The forecasting results show that the best performance has MSE, MAE, and MAPE are 151.910, 23076.561, 118.128, and 2.3%, respectively. The model has a batch size of 4 and epochs of 50. This research recommends that other parties consider this method when they need to manage their investment risk in stocks.

Keywords: forecasting; stock price; recurrent neural network; long short-term memory

DOI: https://doi.org/10.29207/resti.v8i3.5699

1. Introduction
Investment is a commitment to a number of funds with the aim of making profits in the future. Investments have the purpose of getting good long-term financing and can be purchased through a property, gold, or sell/buy stocks. The sale/purchase of shares is one of the choices that have been made to explain this study. Shares are a sign of an individual or group's equity stake in a corporation or limited liability company[1]. Investors who provide their capital for the stock investment will obtain a stock return. Return is a form of compensation for the courage of investors to take their investment risk, and return will encourage investors to invest[2].

However, investors must deal with financial risks that can cause losses at any time. The higher the risk taken, the greater the return that will be obtained and vice versa. Primarily, investors check the financial statements to see the company's performance and to decide whether to buy/sell shares. If the financial statements are profitable, the demand to buy the stocks will increase, and the stock price will increase, too, and vice versa.

However, companies have targets that must be achieved in a predetermined time, and these targets are obtained to expand the ability and value of the company. Potential investors can invest in companies that have gone public and registered their shares with IDX or the Indonesia Stock Exchange. A stock exchange is a group of people who organize and provide systems and/or facilities to make offers to buy and sell securities from people who want to trade them together[3].

Stock prices fluctuate continuously within minutes or seconds without knowing whether to go up or down. The condition of increasing stock prices can be caused by a high demand from investors to own shares, which indicates that the company manages its finances well. Additionally, investors do not want to lose money while investing, so they must maintain their profits by predicting and deciding to buy or sell shares using various calculation options. Recurrent neural networks, or RNNs, are among the deep learning algorithms that are used to analyze time-series data.

RNN has a function of managing the data that does not have a long time period due to vanishing loss gradient problems. Vanishing gradient loss occurs when the model has a lot of layers, and during the
recently, the backpropagation process, the output goes to the hidden and then back to the input, causing the gradient value used to update weight to become smaller. In order to solve the problems, RNN has developed a method called Long Short Term Memory, or LSTM, because it is able to manage long-term cases, deal with vanishing gradients, and also store past information and forget data when there are irrelevant data. This happens because the LSTM has forgotten gates, input gates, output gates, cell states, and two activation functions, where each of them has its own function and duties.

For companies that have been registered on the IDX and are capable in their respective fields, the researchers chose PT Astra International Tbk as the object of this research. PT Astra International officially launched an IPO on 4 April 1990[4]. The company was able to survive the 1998 monetary crisis and the COVID-19 pandemic, which proved the company’s ability to manage well. A large number of subsidiaries can also prove how subsidiaries can help and expand other sectors so that they become efficient and effective, thus they can fulfill the needs of society in various fields.

The LSTM method has been widely used for forecasting in various fields in previous research. In[5], Aldi et al. implemented LSTM to analyze and forecast the price of the kilohertz coins. The results show that their forecasts achieve an average MSE of 93.5%. Similar study by Karakoyun and bdkiken[6], but using LSTM and ARIMA model. The MAPE of their forecasting is 11.86 by using the ARIMA model and 1.4% by using the LSTM method. Abadi applied the LSTM method to predict the number of COVID-19. The accuracy of their forecast represented MSE and MAE, 0.03 and 0.15, respectively.

Nurjaman et al.[7] applied LSTM to the predicted stock price of Pfizer Inc.. The forecast result demonstrates that the RSME value is 0.937 which shows that the forecast model is close to the variation in the observed value. The LSTM method was also used by Pratama et al. to predict the closing price of the apple stock [8]. In this work, the forecast performance was measured by RSME and MAPE. The result shows that both values are $8.43162 \text{ and } 4.514\%$. In 2023[9], Setiawan et al. forecast the stock price of PT. Unilever Indonesia using ARIMA and the ARIMA-Neural Network. The performance of the forecasting shows that the MAPE value using ARIMA is 0.86% while using hybrid ARIMA-NN, the MAPE value is 0.82%.

In 2019, the prediction of the PT stock price. Polychym was carried out by Putri and Aghisini using ARIMA [10]. They got the best model ARIMA(1,1,0) with MAPE 27.45 %. In[11], Aprilianti et al. predicted the stock price of PT. XL Axiata Tbk. using the maximum overlap discrete wavelet transform-ARIMA method. They got ARIMA(2,0,1), ARIMA(3,0,3) and ARIMA (5,0,7) for each wavelet and ARIMA (11,1,6) for scale coefficient. The forecast accuracy is presented with a MAPE value of 6.22%. In [12] Rezaldi and Sugiman forecast the stock price of PT. Telekomunikasi Indonesia using ARIMA. The analysis shows ARIMA (0,2,1) as the best model with an MSE value is 3.07. The prediction of the stock price of PT Garuda Indonesia using ARIMA is presented by Rasyida and Pratama [13]. Their analysis indicates that ARIMA(3,1,2) is the best model with an RMSE value of 38.03. In the other study, Wardani investigated and predicted the Sharia stock price using the LSTM method[14]. The result demonstrates that the MAPE of this forecast is 2.98%. Some of the studies mentioned above show excellent results from time-series forecasting with the LSTM approach. Therefore, LSTM methods are worth considering when making forecasts.

This research has the purpose of studying the Long-Short-Term Memory (LSTM) method, to forecast the stock price of PT Astra International Tbk using the LSTM method. Although the data set comes from the same company as the previous work[15]click or tap here to enter text., the data set has different intervals and methods.

2. Research Methods

The data used for this study aim to predict the daily stock prices of PT Astra International Tbk (ASILJK), where the data start from January 2, 2015, to December 30, 2020, totals 1506 data prices data contain date, open, high, low, close, Adj. Close and volume. Besides that, this study used one variable, which is the closing stock price.

The data source for this research is secondary data from the Yahoo Finance website to obtain historical data of PT Astra International Tbk (ASILJK) daily stock prices. The data type used quantitative data that was analysed using the Long Short Term Memory (LSTM) method.

In analyzing the data, the first step is data preprocessing. This function removes inconsistent data, missing values, or incomplete data. In carrying out data preprocessing, several stages are needed, namely cleaning datasets, splitting data, and normalization data. Cleaning dataset provides to delete data that has a missing value or null. This study has one missing value, so the number of datasets from 1506 changes to 1505. After that, continue to split data into 80:20 for training and testing data, respectively. The last step is the normalization of data, the process of transforming data within a certain range of values is the definition of data normalization[16]. Data transformation is to make it easier for the system to determine the final prediction results[17]. MinMax Scaling was used for this study with Equation 1.

\[ Y_t = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \] (1)
\[ Y_t \] is the normalization value at time \( t \), \( x \) is the actual value, \( X_{\text{min}} \) is the minimum value from data \( X \), and \( X_{\text{max}} \) is the maximum value from data \( X \).

The Training Dataset process is to get a model determined by the number of epochs, batch size, input units, hidden units, and output units. Besides that, there are other hyperparameters called epoch and batch size. Epoch is a hyperparameter that specifies how many times the entire training dataset will be processed by the learning algorithm, and batch size is a hyperparameter that specifies how many samples must be processed before the internal model parameters are updated[18]. The total of the batch is equal to the total of iterations to get one epoch. The presence of these two hyperparameters is intended to assist the computer in training big data.

Test the model with performance measurement using Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). To calculate the error variation in the forecast value, MAE and RMSE together. The best model has the smallest performance measurement value compared to other models.

The Flowchart process is to input the predetermined dataset and do the preprocessing. Then initialize the parameter to train the data and process it with the several best models using the LSTM method. Lastly, test the data and predict the stock price with four performance measurements to get the best result. The flowchart can be seen in Figure 1.

\[
\begin{align*}
&\text{Start} \\
&\text{Input Dataset} \\
&\text{Preprocessing Data} \\
&\text{Training Dataset} \\
&\text{Parameter Initialization} \quad \text{(Batch Size, Epoch, and Node for each layer)} \\
&\text{Testing Dataset} \\
&\text{Forecasting the Stock Price} \\
&\text{Result}
\end{align*}
\]

Figure 1. Flowchart of Forecasting the Stock Price using the LSTM Method

In dealing with the ability to manage information for the long term, RNNs are not able to handle it because they do not save previous information properly due to gradient loss problems[18]. Therefore, in 1997, Hochreiter & Schmidhuber created a special type of RNN named Long-Short Term Memory (LSTM) to solve the problem of long-term information storage. Long short-term memory is widely used in learning long-term dependencies and it is a development from recurrent neural networks. The development means it is the same as RNNs, there is an input layer, a hidden layer, and an output layer. However, the hidden layer of LSTM has three gates and two activation functions shown in Figure 2 and the description of the architecture of LSTM can be seen in Figure 3.

![Figure 2. Architecture of LSTM][20]

LSTM has several steps, the first step is the forget gate \((f_t)\) that determines which information needs to be deleted from the cell state. Forget gate has two inputs, \(h_{t-1}\) is the output of the LSTM process at time \( t - 1 \) and \( x_t \) is the input at the time step \( t \). Formula of forget gate can be seen in Equation 1.

\[
f_t = \sigma (W_f [h_{t-1}, x_t] + b_f)
\]

\(f_t\) is the forget gate, \(\sigma\) is the sigmoid activation function, \(W_f\) is the weight of forget gate, \(h_{t-1}\) is the output value before time \( t \), \(x_t\) is the input value at time \( t \), and \(b_f\) is the bias of forget gate.

After that, the input gate is to select the new information that will be kept in the cell state. The input gate has a sigmoid layer and a tanh layer. The sigmoid layer determines the value to be updated and the tanh layer creates a new vector to be added to the cell state. The formula of input gate can be seen in Equations 2 and 3.

\[
i_t = \sigma (W_i [h_{t-1}, x_t] + b_i)
\]

\[
i_t \quad \text{is input gate,} \quad \sigma \quad \text{is sigmoid activation function,} \quad W_i \quad \text{is weight of input gate,} \quad h_{t-1} \quad \text{is output value before at time} \quad t, \quad x_t \quad \text{is input value at time} \quad t, \quad b_i \quad \text{is bias of input gate,} \quad \tilde{C}_t \quad \text{is new value to be added in the cell state,} \quad tanh \quad \text{is tanh activation function,} \quad W_c \quad \text{is weight of cell state, and} \quad b_c \quad \text{is bias of cell state. After that, output forget gate} \quad (f_t) \quad \text{multiply with the value of the old cell state} \quad (C_{t-1}) \quad \text{and}
\]

![Figure 3. Description of Architecture LSTM][20]
sum up to the result of multiply \(i_t\) and \(\tilde{C}_t\) that can be seen in Equation 4,

\[
C_t = f_t * C_{t-1} + i_t * \tilde{C}_t
\]  
(4)

\(C_t\) is the new cell state, \(f_t\) is the forget gate, \(C_{t-1}\) is the cell state at time \(t - 1\), \(i_t\) is the input gate, and \(\tilde{C}_t\) is the new value to be added in the cell state.

Last step is output gate. The process is carried out by sigmoid that to determine portions of the cell state that will be output. After that, put the cell state through tanh, which has a result in a range -1 to 1. Output gate shows in Equations 5 and 6.

\[
o_t = \sigma (W_o [h_{t-1}, x_t] + b_o)
\]  
(5)

\[
h_t = o_t * \tanh(C_t)
\]  
(6)

\(o_t\) is output gate, \(\sigma\) is sigmoid activation function, \(W_o\) is weight of output gate, \(h_{t-1}\) is output value before at time \(t\), \(x_t\) is input value at time \(t\), \(b_o\) is bias of output gate, \(h_t\) is output value at time \(t\), \(\tanh\) is tanh activation function, and \(C_t\) is cell state. On the other side, the activation function is very necessary for deep learning to process a lot of data with high complexity, it is to determines whether a neuron will be activated or not. In LSTM, there are two activation functions that are used, sigmoid activation function and tanh activation function. Sigmoid activation function has an output from 0 to 1, whereas if the output is 0 the activation function will not be active and if the output is 1 the activation function is active. Tanh activation function can be called hyperbolic tangent, and it has a value between -1 to 1. The equation of the Sigmoid and the Tanh activation function can be seen in Equations 7 and 8.

\[
f(x) = \frac{1}{(1+e^{-x})}
\]  
(7)

\[
tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}
\]  
(8)

Performance measurement has four types that are used in this study, there are Root Mean Square Error (RMSE), Mean Square Error (MSE), Mean Average Error (MAE), and Mean Average Percentage Error (MAPE). These four performance measurements shows that on Equations 9, 10, 11, and 12, respectively.

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}
\]  
(9)

\[
MSE = \frac{1}{n} \sum_{i=1}^{n}(y_i - \hat{y}_i)^2
\]  
(10)

\[
MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n}
\]  
(11)

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100\%
\]  
(12)

\(y_i\) is the actual value, \(\hat{y}_i\) is the predicted value, and \(n\) is the number of observation.

3. Results and Discussions

The data is obtained from yahoo finance website with CSV (Comma Separated Values) format and using closing stock price of PT Astra International Tbk from January 2nd, 2015 to December 30th, 2020 with 1506 data [17]. Figure 4 shows daily stock price of PT Astra International Tbk from 2015 to 2020 where the horizontal line represents time and the vertical line represent close price.

The highest closing stock price is 9150 on April 21st, 2017 and the smallest closing stock price is 3280 on March 24th, 2020. In this period, it has average closing stock price is 7094.58. Table 1 indicates the date and closing stock price of PT Astra International Tbk.

![Closing Stock Price ASII.JK](image)

**Figure 4. Graph of Closing Stock Price 2015 – 2020**

<table>
<thead>
<tr>
<th>Date</th>
<th>Close Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/1/2015</td>
<td>7400</td>
</tr>
<tr>
<td>5/1/2015</td>
<td>7225</td>
</tr>
<tr>
<td>6/1/2015</td>
<td>7050</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>28/12/2020</td>
<td>6075</td>
</tr>
<tr>
<td>29/12/2020</td>
<td>6050</td>
</tr>
<tr>
<td>30/12/2020</td>
<td>6025</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Close Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/1/2015</td>
<td>7400</td>
</tr>
<tr>
<td>5/1/2015</td>
<td>7225</td>
</tr>
<tr>
<td>6/1/2015</td>
<td>7050</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>28/12/2020</td>
<td>6075</td>
</tr>
<tr>
<td>29/12/2020</td>
<td>6050</td>
</tr>
<tr>
<td>30/12/2020</td>
<td>6025</td>
</tr>
</tbody>
</table>

The first step of preprocessing data is cleaning the dataset. The dataset has 1506 data and there is one null data on June 19th, 2019. So, it has to be removed and will continue the process with 1505 data. Figure 5 shows a graph of the closing stock price after cleaning the dataset.

After that, split the dataset. The data is divided into two, 80:20, for training and testing data, respectively. The training data is carried out from January 2nd, 2015 to October 4th, 2019 while the testing data is from October 7th, 2019 to December 30th, 2020. Table 2 shows the split dataset.

![Graph of Closing Stock Price after Cleaning](image)

**Table 1. Date of Closing Stock Price PT Astra International Tbk**
The last step of preprocessing is normalization. Normalization is needed to minimize the errors that are in the range (0,1). The process will use MinMaxScaling in Equation (1), and Table 3 shows the normalization result. After this step, all values of stock prices become in the interval 0 and 1.

Table 3. Normalization of Closing Stock Price

<table>
<thead>
<tr>
<th>Date</th>
<th>Normalized Close Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/1/2015</td>
<td>0.7018</td>
</tr>
<tr>
<td>5/1/2015</td>
<td>0.6720</td>
</tr>
<tr>
<td>6/1/2015</td>
<td>0.6422</td>
</tr>
</tbody>
</table>

The data will be processed with Python software 3.9 using the keras library, ADAM optimizer, and mean squared error as a loss function. The optimization is utilized to minimize errors and determine optimum weight so that it can maximize the accuracy of the models. This research will use changing of parameters in training data, there are Batch Sizes of 4, 16, and 64, and Epoch of 50, 100, 200, and 400.

LSTM architecture uses two layers that are hidden where each of them is 50 nodes, while the output layer uses one node to generate one output, with a default dropout of 0.2. Dropout is to reduce overfitting. Testing the model using parameter batch sizes of 4, 16, and 64 and epochs of 50, 100, 200, and 400 as well as the time for each calculation parameter is calculated to know which is the best model that has the highest accuracy and fastest process. Table 4, Table 5, Table 6, and Table 7 presents calculation results for various epochs, namely 50, 100, 200, and 400, respectively.

Table 4 shows Model 1 has the smallest value other than models 2 and 3, with RMSE, MSE, MAE, and MAPE are 151.910, 23076.561, 118.128, and 2.3%, respectively. From this table, the smaller the batch size, the longer it takes to process it.

Table 5 shows that Model 5 is the best model because their performance measurements are the smallest other than models 4 and 6, with RMSE, MSE, MAE, and MAPE are 166.198, 27621.713, 120.226, and 2.4%, respectively. Although Model 6 has less time required, Model 5 has the smallest MAPE and only differs by 0.2% from Model 4.

Table 6 shows that Model 9 is the best model because of MAE, MAPE values, and time to obtain are the smallest other than Models 7 and 8. Model 9 has MAE and MAPE are 183.817 and 3.6%, thus it takes time 9 minutes 29 seconds. Although Models 8 and 9 only have different RMSE and MSE by 1.428 and 697.829, Model 9 has less time required.

Table 7 shows that Model 12 is the best model because their performance measurements are the smallest other than models 10 and 11, with RMSE, MSE, MAE, and MAPE are 156.615, 24528.571, 113.257, and 2.2%, respectively.
The four tables show that the larger the batch sizes, the faster the time obtained. Model 1 is the best compared to the other 11 models, as seen from the 4 performance measurements and its time. Model 12 also has lower MAPE values than Model 1, but its values for RMSE, MSE, and the time required are less than those of Model 12, which only has MAPE and MAE values.

Testing data with 301 data are used to forecast stock prices. 4 Batch sizes and 50 epochs were selected for the best model such as Model 1. This model was chosen because it generates the smallest value from four performance measurements. Even though, the MAE and MAPE values in Model 12 are lower the time to obtain was longer and only had a difference of 4.871 for MAE and 0.01 for MAPE with Model 1. Figures 6 and 7 exhibits LSTM data with Model 1 and the graph of comparison between the close price and predicted price, where the orange line is the close price and the green line is the predicted price.

<table>
<thead>
<tr>
<th>Model</th>
<th>Batch Size</th>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4</td>
<td>749.718</td>
<td>562077.670</td>
<td>483.011</td>
<td>9%</td>
</tr>
<tr>
<td>11</td>
<td>16</td>
<td>1273.110</td>
<td>1620808.052</td>
<td>530.121</td>
<td>14.8%</td>
</tr>
<tr>
<td>12</td>
<td>64</td>
<td>156.615</td>
<td>24528.571</td>
<td>113257</td>
<td>2.2%</td>
</tr>
</tbody>
</table>

Table 7. Calculation Result using Epoch = 400

4. Conclusion

This study forecasts the stock price of PT Astra International Tbk (ASII.JK) using historical data from January 2nd, 2015, to December 30th, 2020, totalling 1506 observations. The data is split into an 80% training dataset and a 20% testing dataset. Twelve models are evaluated using combinations of hyperparameters such as epochs (50, 100, 200, and 400) and batch sizes (4, 16, and 64). The best model, Model 1, has a batch size of 4 and an epoch count of 50. This model's performance metrics are RMSE: 151910, MSE: 23076.561, MAE: 118.128, and MAPE: 2.3%. The processing time for this model is 14 minutes and 6 seconds. When compared with the result of the previous work in [21], where the MAE value was 2.98%, the MAPE value of this work is 2.3%. Thus, the accuracy of this prediction is not significantly different. The suggestions for future research are using different data and a longer time span, applying other methods like Gate Recurrent Unit (GRU), and using other parameter combinations with different amounts, such as hyperparameters number, epoch and batch size.

References

B. Aprilia, S. Martha, and N. Imro’ah, “PERAMALAN
HARGA SAHAM DENGAN METODE MODWT-ARIMA,”
D. A. Rezaldi and Sugiman, “Peramalan Metode ARIMA Data
Available: http://journal.unnes.ac.id/sju/prisma/article/view/45036
W. Russida and V. Pratama, “Prediksi Harga Saham Garuda
Indonesia di Tengah Pandemi Covid-19 Menggunakan Metode
W. W. K. Wardani, “Prediksi Harga Saham Syariah
menggunakan Metode Recurrent Neural Network-Long Short
Term Memory,” UIN Sunan Ampel Surabaya, 2021. [Online].
Available: http://digilib.unsa.ac.id/49542/
F. A. Pohan, “Forecasting harga saham pada PT Astra
International Tbk menggunakan metode ARIMA,” IAIN
Padangsidimpuan, 2022.
N. Selle, N. Yudistira, and C. Dewi, “Perbandingan Prediksi
Penggunaan Listrik dengan Menggunakan Metode Long Short
Term Memory (LSTM) dan Recurrent Neural Network
(RNN),” J. Teknol. Inf. dan Ilmu Komput., vol. 9, no. 1 SE-
A. R. T. Rully, “Penerapan Metode Long Short Term Memory
Dalam Memprediksi Jumlah Kasus Covid-19,” Universitas
Yayasan Prima Agus Tek, vol. 8, no. 1 SE-Judul Buku, Feb.
w/271