Jurnal TRANSFORMATIKA
Vol.20, No.2, Januari 2023, pp. 30-41
P-ISSN: 1693-3656, E-ISSN: 2460-6731
https://journals.usm.ac.id/index.php/transformatika/

Robusta London Coffee Price Forecasting Analysis Using Recurrent Neural Network – Long Short Term Memory (RNN – LSTM)

Ferzy Tryanda Nosa[1], Dian Kurniasari[2*], Amanto[3], Warsono[4]
[1,2,3,4] Program Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia, email:[1]ferzynosa@gmail.com, [2]dian.kurniasari@fmipa.unila.ac.id,[3]amanto.nuryaman@fmipa.unila.ac.id,[4]warsono.1963@fmipa.unila.ac.id

ARTICLE INFO

History of the article :
Received 30 Oktober 2022
Received in revised form 15 Desember 2022
Accepted 3 Januari 2023
Available online 30 Januari 2023

Keywords:

Coffee price forecasting, LSTM, Units, Dropout, RMSE, MAP.

*** Correspondece:**

Telepon:
-

E-mail:
dian.kurniasari@fmipa.unila.ac.id

ABSTRACT

Coffee price forecasting has a significant role in preventing price fluctuations at a time. Therefore, a method is needed that can be used to forecast the price of coffee. This study discusses the analysis of coffee price forecasting using the Recurrent Neural Network – Long Short-Term Memory (RNN – LSTM) method. This study will be determined the best LSTM model that aims to get the results of forecasting the price of London robusta coffee with the smallest Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) values. Using the LSTM model with units of 128 and dropouts of 0.1, forecasting the price of London robusta coffee has an RMSE value of 1,303 and MAPE of 3.53%. Therefore, the LSTM model can indicate the cost of London robusta coffee with an accuracy rate of 96.47%.

1. INSTRODUCTION

Most of the activities in everyday life are related to mathematics and statistics. Various benefits are caused by this science, one of which is estimating an event that may occur in the future, commonly called prediction or forecasting.

Prediction or forecasting is a science used to estimate future events so that policymakers can use forecasting results in taking strategic policies to solve problems in the future [1]. When forecasting, a proper method is needed to produce the lowest possible error value.

Coffee is a brewed drink of coffee beans that have been roasted and smashed into powder. There are four known coffee groups: arabica coffee, robusta coffee, liberika coffee, and easel coffee [2]. Coffee is also a people's commodity cultivated for a long time and can become a source of income for many coffee farmers worldwide. In addition to being a source of people's income, coffee is the leading export commodity and head of state foreign exchange income. Nevertheless, coffee commodities often experience price fluctuations due to the global market's imbalance between demand and coffee supply. With these fluctuations, a forecasting method is needed to predict the selling price of coffee in the future. This fact is supported based by robusta London Coffee History Data for the period January 2008 - October 2021 obtained from the website investing.com shows that coffee prices are unstable. The decision to determine the selling price is a problem that must be sought to be resolved. Because if we decide the price is too high, it will be challenging to sell, and vice versa. If the price is too low, it can cause losses for coffee farmers.

One method that can be used is Long Short Term Memory (LSTM) which is the development of a recurrent neural network (RNN) based on Gao, Chai. Liu's research on predicting stock movements the following day using the Moving Average (MA) method produces a Root Mean Square Error (RMSE) value of 40.9691, Exponential Moving Average (EMA) delivers an RMSE weight of 24.6726, The Support Vector Machine (SVM) returns an RMSE value of 21.8863. The LSTM returns an RMSE value of 20.4688 [3]. Then, Susanti and Adji's research on the prediction of the Composite Stock Price Index (JCI) using the Autoregressive Integrated Moving Average (ARIMA) method to predict it obtained the conclusion that the model received could not be generalized for too far a period [4]. The purpose of using this LSTM is to avoid these problems and be able to produce the smallest RMSE value to predict the price of coffee.

To solve the problem of predicting difficulties, we can use the Recurrent Neural Network – Long Short-Term Memory (RNN-LSTM) method [5]. This method will focus on making accurate predictions of a variable. The best predictions or forecasting are based on the degree of prediction error, where the smaller the error rate produced, the more precise a method is used in predicting. Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are the error rate calculations.

Based on the above exposure, researchers are interested to find out how accurate the RNN-LSTM method is in predicting the price movement of London Robusta Coffee. Therefore, the researchers raised the study's title, "Analysis of London Robusta Coffee Price Forecasting Using Recurrent Neural Network – Long Short Term Memory (RNN – LSTM)."

2. RESEARCH METHOD

The data used in this study is secondary data obtained from https://id.investing.com/ regarding data on the history of London robusta coffee futures for 13 years, starting from January 2008 to October 2021, on a weekly scale. The data amounts to 722 table-shaped tables with the Date, Last, Opening, Highest, and Lowest columns.

In this study, researchers will display the best LSTM model for predicting the last price or close on London Robusta Coffee using the Recurrent Neural Network – Long Short Term Memory (RNN-LSTM) method with the help of Python software powered by Google Colab. Then, the resulting model will be evaluated based on RMSE and MAPE values.

The steps taken in this research method are as follows:

1. Inputting data on the history of London robusta coffee obtained through https://id.investing.com/ into Google Colab.
2. Visualize data by creating a time series graph to view the data flow.
3. Dividing data preprocessing is by scaling data using MinMaxScaler, then dividing the data into training data and data testing.
4. We are building a model by applying RNN-LSTM.
5. Make predictions to see price predictions on London robusta coffee's "Last" index.
6. Validate models using RMSE and MAPE.
7. View the visualizations resulting from the predicted models obtained.
8. Do forecasting to get prices on the "Last" index of London robusta coffee.
9. View visualizations that result from the predicted data.
10. Make conclusions based on the results obtained.

As for the image form, the steps are taken as follows:

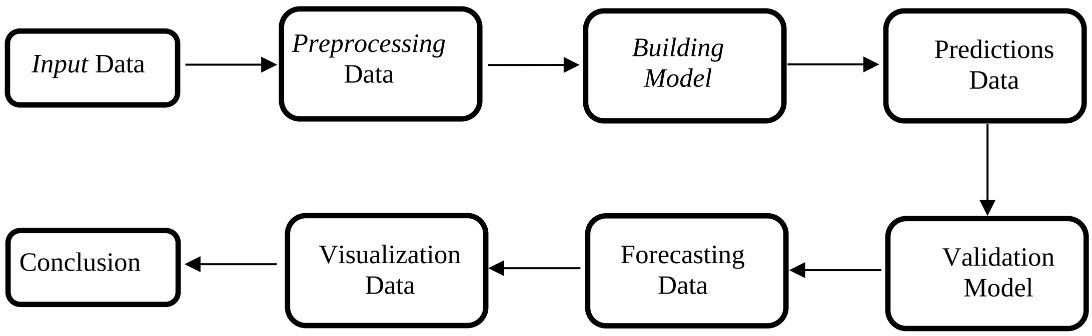

Figure 1. Research Steps

2.1 *Forecasting*

Predictions or forecasting are needed in various fields, such as education, health, development, and economy, to businesses run by a company. Forecasting is usually used to determine future needs and make the right decisions [6].

There are two approaches to forecasting: a quantitative approach and a qualitative one. A quantitative approach is an approach that uses statistical methods such as causal and time series models, while the qualitative approach is an approach based on the opinions or opinions of the parties concerned.

2.2 *Data Mining*

Data mining is finding exciting patterns from vast amounts of data. Knowledge discovery involves data cleaning, integration, selection, transformation, pattern finding, evaluation, and knowledge presentation [7].

Data mining tasks can be grouped into eight groups based on their functionality: classification, regression, clustering association rule learning, anomaly detection, time series forecasting, text mining, and feature selection [8].

2.3 *Machine Learning*

Machine learning is machine learning that aims to solve a problem systematically by studying data. Machine learning can find insights, improve, and learn new things to form a sound system. At the output stage, the machine learning process relies heavily on data for training and testing materials.

According to [9], machine learning has 2 basic learning techniques, namely:

1. Supervised learning is machine learning that can receive information already owned by the data by giving it a specific label.
2. Unsupervised learning is machine learning used on data that previously had no labels. This technique can find hidden structures or patterns in data that do not have labels, usually used to find pattern classifications.

2.4 Artificial Neural Networks (ANN)

ANN is an approach to processing information that works following the working system of the human brain. In the human brain, each neuron is interconnected, and information flows to each neuron [10]. The ANN model has essential elements called layers. This layer can be categorized as follows:

1. The input layer is a layer that has a role as a place for information from data that acts according to the desired output. The input layer has several neurons capable of presenting important

parameters for solving problems. The data will be channelled through this input layer to the hidden and output layers.
2. The hidden layer is a layer that is between the input layer and the output layer, which is in charge of receiving data from the input layer and channelling it to the output layer.
3. The output layer is the layer that receives data and provides calculation results from the input layer using the activation function. The value of the output layer represents the output from X to a Y value.

The architecture of ANN can be seen as follows:

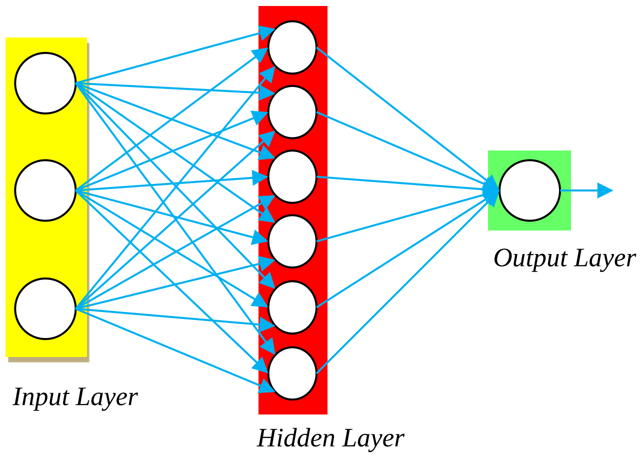

Figure 2. ANN Architecture Model

2.5 Recurrent Neural Networks (RNN)

RNN is a kind of neural network that can see hidden correlations in data in speech recognition applications, natural language processing, and time series predictions [11]. RNNs are especially good for sequence modelling problems that operate on the input information and trace information obtained previously due to repeated connections.

$$S_t = f\big((U * X_t) + (W * S_{t-1})\big) \qquad (2.1)$$

$$O_t = g(V * S_t) \qquad (2.2)$$

Where:

S	= network memory at time t
U, V, dan W	= shared weight matrix in each layer
X_t dan O_t	= represent the input and output at the t time
$f(\ldots)$ dan $g(\ldots)$	= represent nonlinear functions

DOI: http://dx.doi.org/10.26623/transformatika.v20i2.5482

The architectural model of the RNN is as follows:

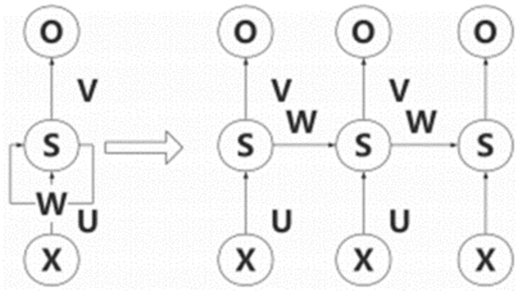

Figure 3. RNN Architecture Model

2.6 Long Short-Term Memory (LSTM)

LSTM uses one of the most common forms of RNN, which is intended to avoid long-term dependency problems and is suitable for processing and predicting time series [12]. The LSTM model consists of a series of individual memory cells that replace neurons in the hidden layer of the RNN, and the key is the state of the memory cells. The LSTM model filters information through the gate structure to maintain and update the state of the memory cells. The door structure includes an input, forget, and output gate. Each memory cell has three sigmoid layers and one tanh layer.

The picture of the LSTM structure is as follows:

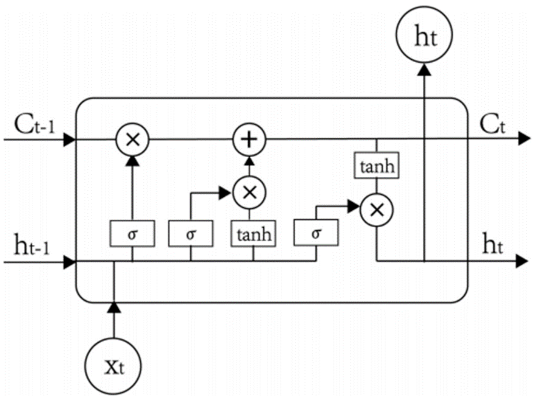

Figure 4. LSTM structure

1. The input gate has two functions, namely the first is to find the state of the cell that must be updated and the second function of the input gate is to update the information to b to be updated to the state of the cell
2. Forget gate in LSTM aims to determine which cell status information will be removed from the model
3. Output gate aims to control how much of the current state of the cell is discarded

The activation functions used in the LSTM model are the tanh function and the sigmoid function. According to [13], the sigmoid function has a value range from 0 to 1.

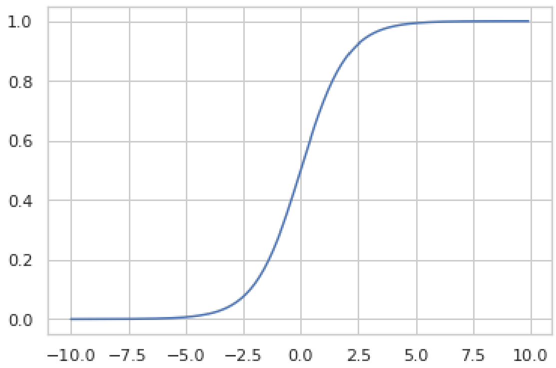

Figure 5. Graph of the Sigmoid Function

The sigmoid function can also be called the logistic function, which is shown as follows:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.3}$$

Then, the tanh function or also known as the hyperbolic tangent, has a range of values from -1 to 1. The tanh function can be said to be almost the same as the sigmoid function. That is, if it is depicted in graphical form, the resulting curve will be shaped like the letter S. However, the function tanh has a broader range of values than the sigmoid function, so it will be more effective for complex nonlinear modelling

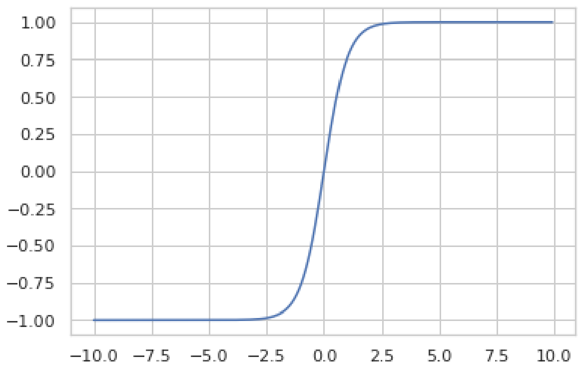

Figure 6. Graph of the Tanh Function

The tanh function is formulated as follows:

$$tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.4}$$

DOI: http://dx.doi.org/10.26623/transformatika.v20i2.5482

$$tanh(x) = \frac{sinh(x)}{cosh(x)} \quad (2.5)$$

2.7. Model Validation

There are two main reasons for looking at the level of accuracy in predicting time series models, namely by looking at the RMSE and Mean Absolute Percentage Error (MAPE) values [14].

RMSE is the square root of the sum of the squared errors, the difference between the actual value and the predicted value, and dividing the sum by the number of forecasting times [15]. For RMSE it is formulated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}{n}} \quad (2.6)$$

Where :

Y_i : Actual data values
$\widehat{Y}_i$: The final value of forecasting data
n : The amount of data

Furthermore, MAPE is the absolute value of the percentage of data error against the mean. MAPE is formulated as follows:

$$MAPE = \frac{\sum_{i=1}^{n}\frac{\left|\widehat{Y}_i - Y_i\right|}{Y_i}}{n} \times 100\% \quad (2.7)$$

Where :
Y_i : Actual data values
$\widehat{Y}_i$: The final value of forecasting data
n : The amount of data

3. Results and Analysis

LSTM is one form of RNN that can be used to predict or predict the current and future coffee prices using coffee history data that has been obtained before. The data obtained cannot be analyzed because it is still raw. The coffee history data has five columns: Date, Last, Opening, Highest, and Lowest. The coffee history data listed in the table below is only part of it. That aims to be used as an overview of the comprehensive data amounting to 722. The value of the currency used in the coffee history data is the US Dollar currency.

The first step in carrying out the forecasting process using LSTM is determining the input to be used. Coffee history data has four indices: Last, Opening, Highest, and Lowest. Then, the last used index of the four indices starts from January 13, 2008, to November 7, 2021, with a weekly period.

After inputting data, it is then scaling the data. Data scaling is a transformation process that aims to change the value of data from the original form to another form so that the data can be used more efficiently and produce the smallest possible error value. The technique used in scaling data this time is the MinMaxScaler technique, which can also be called *Min-Max Scaling*. *Min-Max Scaling* is a data transformation technique that adjusts data values by a range or interval of minimum values to a maximum weight of [0.1]. When written in mathematical form, *Min-Max Scaling* can be written as follows:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{3.1}$$

After scaling the data, do the data sharing into training and testing. Training data is data used to train LSTM algorithms, while data testing is data used to determine the performance of LSTM algorithms introduced before when they want to do forecasting. The data composition to be used is 80% for training data and 20% for data testing. Of the 722 data available, it will be divided as much as 80% into training data (577 training data) and 20% into data testing (145 data testing).

After sharing training data and data testing, the construction or building model of LSTM will be carried out using one input layer and one output layer. Then, the number of neurons can also be called units used in the hidden layer, namely 16, 32, 64, and 128. Epochs will be used at 50 periods and dropouts of 0.1 and 0.2. To find out the exact number of units and dropouts can be determined by applying hyper tuning to units and dropouts that have been predetermined. There are no rules in determining the number of units and dropouts, so the number of units and dropouts is obtained by applying hyper tuning so that the resulting model can be optimal in forecasting the price of coffee. The following are the test results using hyper tuning at the number of units 16, 32.64, and 128, dropouts of 0.1 and 0.2, and epochs of 50:

```
Best: -0.003600 using {'LSTM_unit': 128, 'dropout': 0.1}
-0.006765 (0.004613) with: {'LSTM_unit': 16, 'dropout': 0.1}
-0.006550 (0.004388) with: {'LSTM_unit': 16, 'dropout': 0.2}
-0.006230 (0.004700) with: {'LSTM_unit': 32, 'dropout': 0.1}
-0.005444 (0.004318) with: {'LSTM_unit': 32, 'dropout': 0.2}
-0.004535 (0.002421) with: {'LSTM_unit': 64, 'dropout': 0.1}
-0.004674 (0.002487) with: {'LSTM_unit': 64, 'dropout': 0.2}
-0.003600 (0.001628) with: {'LSTM_unit': 128, 'dropout': 0.1}
-0.004015 (0.001679) with: {'LSTM_unit': 128, 'dropout': 0.2}
```

Figure 7. Building Model LSTM

Figure 7 shows that the best LSTM model produced has units or can also be called LSTM units of 128 and dropouts of 0.1. This model will then be used to forecast the price of coffee.

After building the LSTM model, a price prediction experiment on the "Last" Index of London Robusta Coffee will be conducted. That is done to determine whether the previously formed LSTM model can predict the price of coffee or not. If we can expect the price of coffee, then the LSTM model can also be used to predict the price of coffee in the future. Here are the results of the price predictions on the "Last" Index of London Robusta Coffee:

DOI: http://dx.doi.org/10.26623/transformatika.v20i2.5482

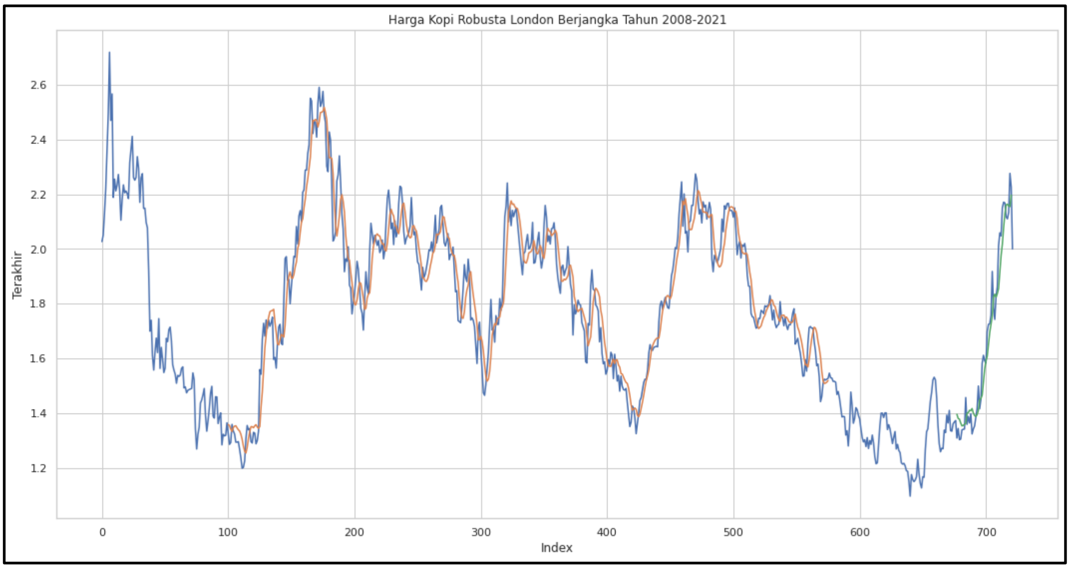

Figure 8. Robusta London Coffee Price Prediction Plot

Based on Figure 8, it can be seen that the lines that appear there are three colours, namely, the lines are blue, orange, and green. The blue line interprets the plot using the original data; the orange line interprets the story of the prediction result using training data, and the green line analyses the field using data testing. It can be concluded that the LSMT model with units of 128 and dropouts of 0.1 can predict prices on the "Last" Index of London Robusta Coffee well because the plot of the orange and green lines moves closer to the blue line plot.

After making a prediction, how much the LSTM model generates an error value will be seen. That is done to further ensure the LSTM model's accuracy in predicting coffee prices. If the accuracy rate is significant and only produces a small error value, then the LSTM model can be used to forecast future coffee prices. Here are the RMSE, MAPE, and accuracy levels:

```
math.sqrt(mean_squared_error(y_test,test_predict))
```

```
1.3028276075025023
```

```
from sklearn.metrics import mean_absolute_percentage_error
MAPE = mean_absolute_percentage_error(y_train, train_predict, sample_weight=None, multioutput='uniform_average')
MAPE
```

```
3.5299598413753484
```

```
# Test Accuracy
Accuracy = 100 - MAPE
print(Accuracy,'%')
```

```
96.47004015862466 %
```

Figure 9. LSTM Model Validation

Based on Figure 9, it can be seen that the resulting RMSE value is 1,303. It can be interpreted that the LSTM model can predict the price of a coffee with a small error rate because it produces a small RMSE value. Then furthermore, a MAPE value of 3.53 can be called 3.53%. As a result, it can be interpreted that the average percentage of absolute errors or MAPE of the LSTM model is

3.53%. Therefore, the LSTM model with units of 128 and dropouts of 0.1 can predict and forecast prices on the "Last" Index of London Robusta Coffee with an accuracy rate of 96.47%.

After knowing the level of accuracy of the LSTM model to be used, then there will be a forecasting of coffee prices for the next 30 weeks as follows:

Table 1. Price Forecasting Results On London Robusta Coffee Last Index

Date	Forecasting Data	Recent Data
11/7/2021	2,146321	2,315
11/14/2021	2,139573	2,270
11/21/2021	2,132030	2,308
11/28/2021	2,124435	2,386
12/5/2021	2,116758	2,376
12/12/2021	2,108985	2,439
12/19/2021	2,101127	2,462
12/26/2021	2,093210	2,488
1/2/2022	2,085267	2,435
1/9/2022	2,077335	2,341
1/16/2022	2,069451	2,335
1/23/2022	2,061648	2,193
1/30/2022	2,053961	2,181
2/6/2022	2,046416	-
2/13/2022	2,039041	-
2/20/2022	2,031857	-
2/27/2022	2,024880	-
3/6/2022	2,018128	-
3/13/2022	2,011610	-
3/20/2022	2,005335	-
3/27/2022	1,999309	-
4/3/2022	1,993536	-
4/10/2022	1,988017	-
4/17/2022	1,982751	-
4/24/2022	1,977736	-
5/8/2022	1,972969	-
5/15/2022	1,968445	-
5/22/2022	1,964158	-
5/29/2022	1,960102	-
6/5/2022	1,956270	-

Based on Table 1, the value generated in the Forecasting Data is not much different from the value in the Latest Data on the London Robusta Coffee "Last" index from November 7, 2021, to January 30, 2022. So this can be interpreted for price forecasting in the "Last" Index of Robusta Coffee London will decrease until June 5, 2022. This price drop chart can be seen as follows:

DOI: http://dx.doi.org/10.26623/transformatika.v20i2.5482

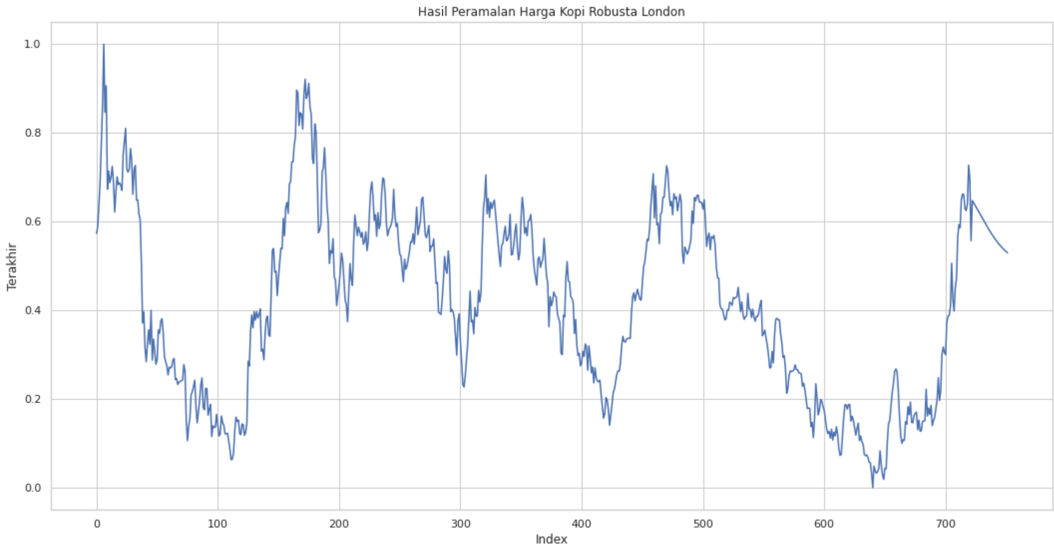

Figure 10. Robusta London Coffee Price Forecasting Plot

.

4. Conclusion

Based on the results and discussions, the following conclusions are obtained:

1. The LSTM method can predict or forecast prices on the "Last" Index of London Robusta Coffee using units of 128 and dropouts of 0.1.
2. Based on RMSE and MAPE values, the LSTM method only produces a small error value, namely RMSE of 1,303 and MAPE of 3.53%. This value can be interpreted that the accuracy rate of this LSTM method being 96.47%. Therefore, this LSTM method will be used as well as possible to predict or forecast the price of other goods because this method only produces a small error value.
3. Based on the results of coffee price forecasting using the LSTM method, it turns out that the price in the "Last" Index of Robusta Coffee London will decrease from November 7, 2021, to June 5, 2022.

References

[1] Utami, T.W. and Darsyah, M.Y. Peramalan Data Saham dengan Model Winter's. *Jurnal Statistika*. 2015; 3(2). https://doi.org/10.26714/jsunimus.3.2.2015.%25p

[2] Rahardjo, P. Kopi. Jakarta : Penebar Swadaya Grup. 2012.

[3] Gao, T., Chai, Y., dan Liu, Y. *Applying Long Short Term Memory Neural Networks for Predicting Stock Closing Price*. Proceeding IEEE International Conference on Software Engineering and Service Science (ICSESS). 2017; 575-578.

[4] Susanti, R., & Adji, A. R. Analisis Peramalan IHSG dengan Time Series Modeling ARIMA. *Jurnal Manajemen Kewirausahaan*. 2020; 17: 97-106.

[5] Wiranda, L., and Sadikin, M. Penerapan Long Short Term Memory Pada Data Time Series untuk Memprediksi Penjualan Produk PT. Metiska Farma. *Jurnal Nasional Pendidikan Teknik Informatika*. 2019; 8(3): 184-196. https://dx.doi.org/10.23887/janapati.v8i3.19139

[6] Auliasari, K., Kertaningtyas, M., dan Kriswantono, M. Penerapan Metode Peramalan untuk Identifikasi Potensi Permintaan Konsumen. *Informatics Journal*. 2019; 4(3): 121-129.

[7] Han, J., Kamber, M., dan Pei, J. Data Mining: Concepts and Techniques. 3rd Edition. San Francisco: Elsevier. 2012.
[8] Kotu, V. dan Deshpande, B. Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner. Waltham: Elsevier. 2015
[9] Pustejovsky, J. dan Stubbs, A. Natural Language Annotation for Machine Learning. California: O'Reilly Media. 2012.
[10] Prathama, A.Y., Aminullah, A., dan Saputra, A. Pendekatan ANN (Artificial Neural Network) Untuk Penentuan Prosentase Bobot Pekerjaan Dan Estimasi Nilai Pekerjaan Struktur Pada Rumah Sakit Pratama. *Teknosains*. 2017; 7(1): 14-25.
[11] Tian, C., Ma, J., Zhang, C., dan Zhan, P. A Deep Neural Network Model for Short Term Load Forecast Based on Long Short-Term Memory Network and Convolutional Neural Network. *Energies*. 2018; 11.
[12] Qiu, J., Wang, B., dan Zhou, C. Forecasting Stock Prices with Long-Short Term Memory Neural Network Based on Attention Mechanism. *Advanced Design and Intelligent Computing*. 2020; 15(1).
[13] Lewis, N. D. Neural Network for Time Series Forecasting with R. US: CrateSpace Independent Publishing Platform. 2017.
[14] Vercellis, C. Business Intelligence: Data Mining and Optimization for Decision Making. United Kingdom: John Wiley & Son. 2009.
[15] Budiman, H. Analisis Dan Perbandingan Akurasi Model Prediksi Rentet Waktu Support Vector Machines Dengan Support Vector Machines Particle Swarm Optimization Untuk Arus Lalu Lintas Jangka Pendek. *Systemic*. 2016; 2(1): 19-24.

DOI: http://dx.doi.org/10.26623/transformatika.v20i2.5482