CHAPTER IV

DATA ANALYSIS AND RESULT DISCUSSION

4.1 STATISTICAL DESCRIPTION OF USD/IDR REFERENCE RATE DATA

To understand the data used in this study, Table 4.1 below shows several basic statistical descriptions of the data.

Table 4.1 Statistical description of USD/IDR reference rate

	N	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**
JISDOR	544	13,612	16,741	14,423.0882	483.5048

The data is also divided into 505 observations of training dataset and 39 observation of testing dataset which are illustrated in figure 4.1.

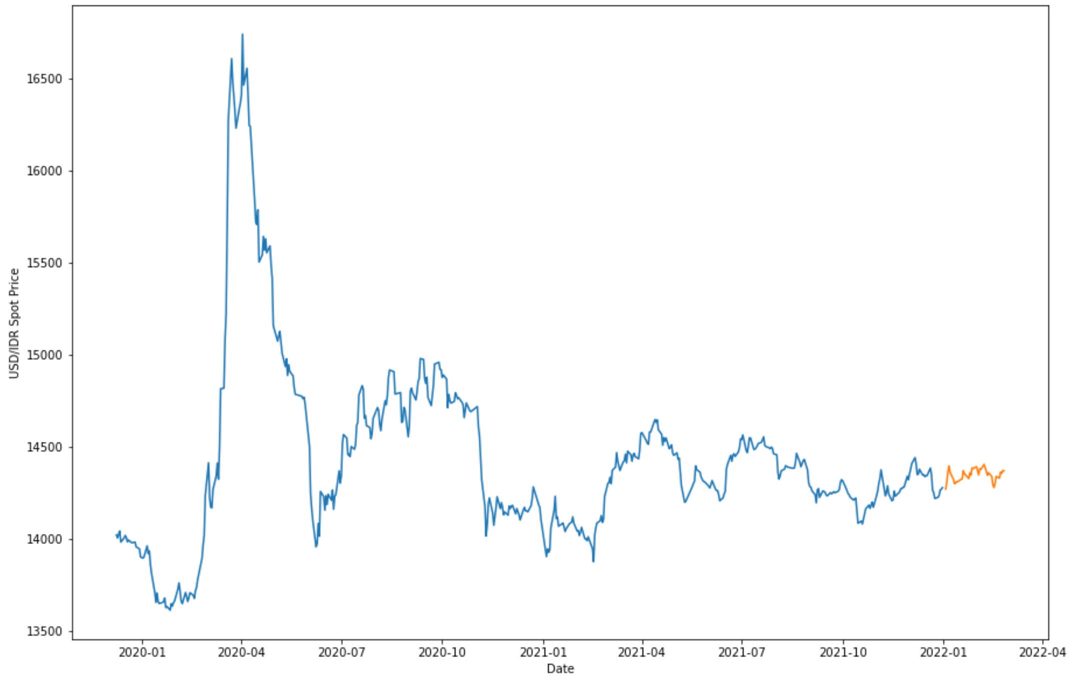

Figure 4.1 Time series plot for training and testing dataset

4.2 ARIMA MODEL

4.2.1 STATIONARY TESTING

The major assumption of the ARIMA model is that the time-series data must be stationary. ARIMA relies on stationarity patterns because the model uses historical information to predict the future. By observing figure 4.1, it can be seen that the data used in this study is not stationary. It is also clearly shown in ACF and PACF plots in figure 4.2. The ACF of data decreases slowly, proving that it is nonstationary data. Meanwhile, the first PACF is equal to one, indicating of nonstationary as well.

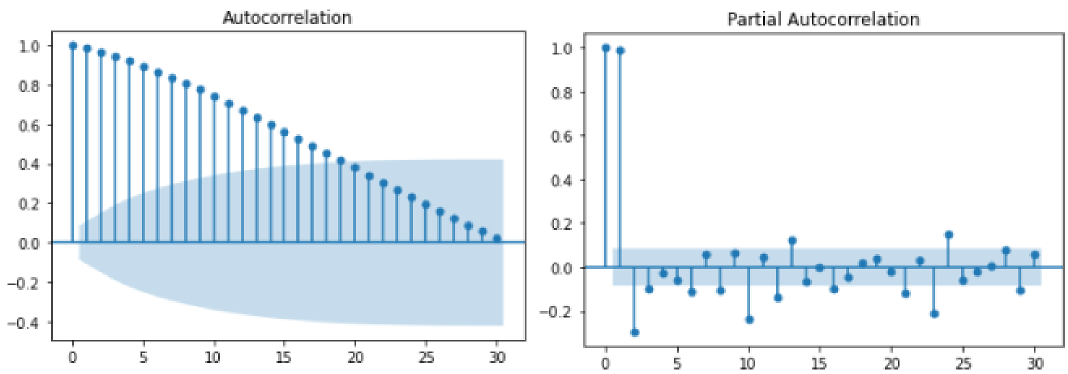

Figure 4.2 ACF and PACF plots

When the original or level data is not stationary, preliminary differencing is necessary to obtain the stationary series. The differencing process can be in the first or the second order. Figures 4.3 and 4.4 provide the comparison graphs of data and ACF plots. From the plot, first-order differencing is enough to make the data stationary. The ACF plot also shows that the sample correlation of first-order differencing data approaches zero or die out gradually. To ensure the stationary of the model, ADF test is also employed. ADF test for first-order differencing data is 1.8528×10^{-6}, which is smaller than the ADF test for second-order differencing, which is 0.00078. Therefore, the first-order differencing data is the best to use for analysis.

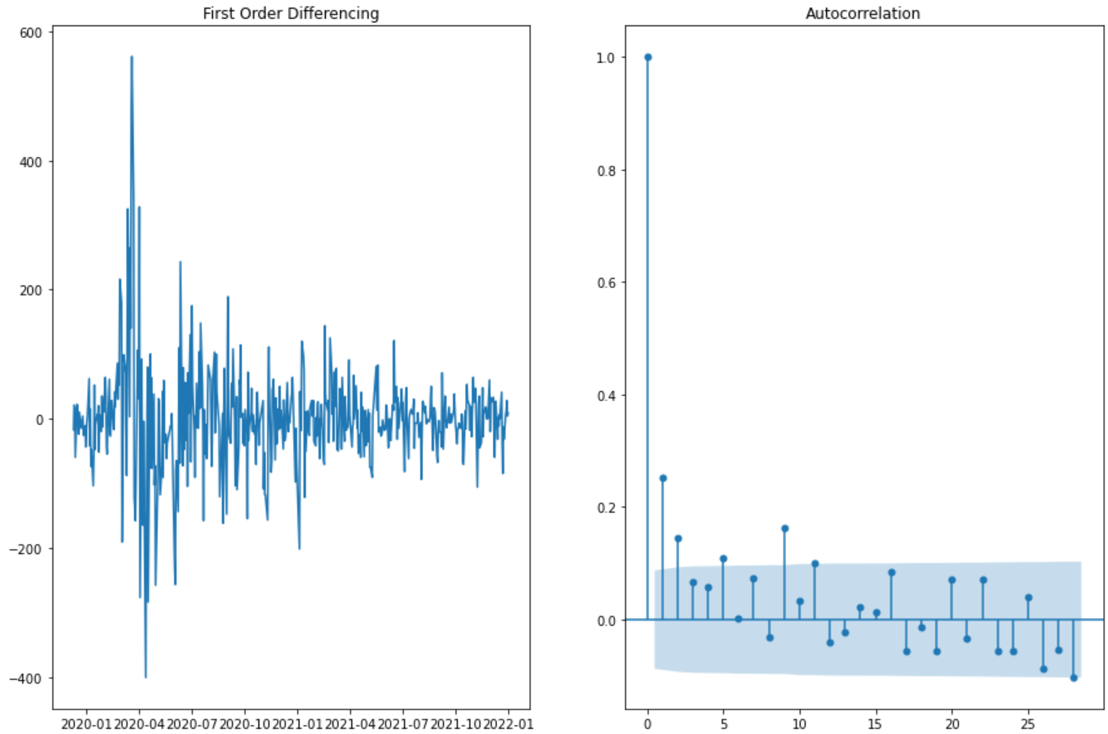

Figure 4.3 First-order differencing data and ACF plot

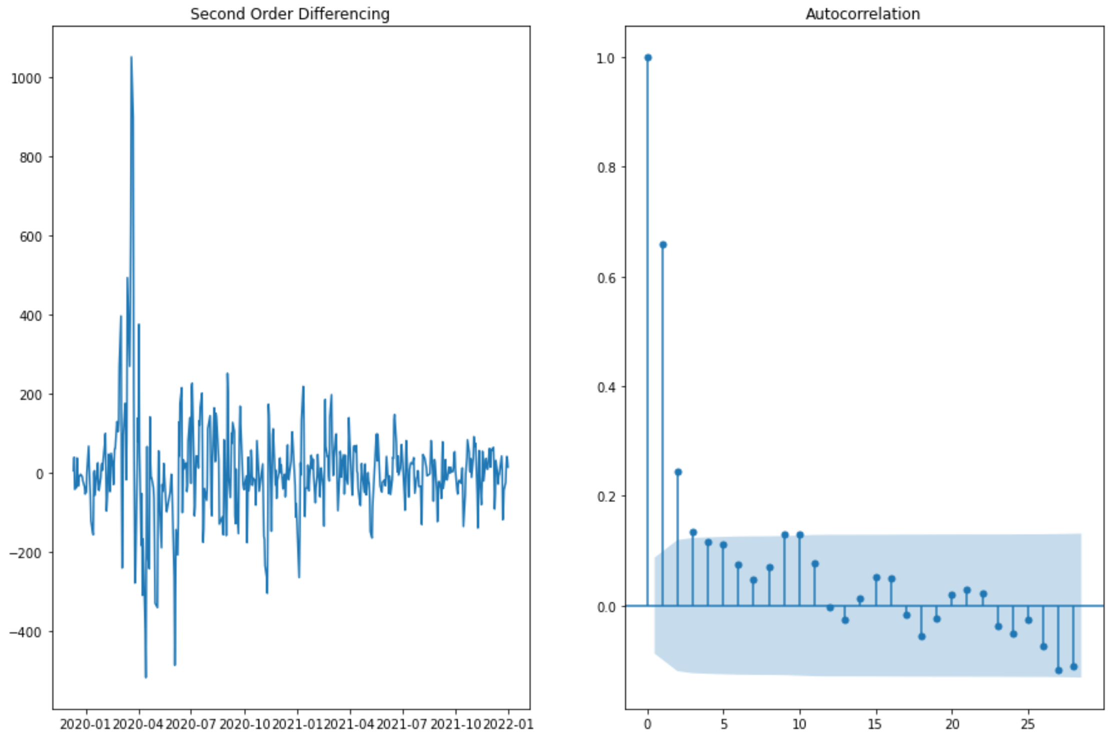

Figure 4.4 Second-order differencing data and ACF plot

4.2.2 MODEL IDENTIFICATION

If the data is stationary in mean and variance, the assumptions of the ARIMA model has been met. The next step is to identify the order of AR (*p*) and MA (*q*) which are fit the data with the order of I (*d*) is equal to one. The result is shown in the table 4.2.

Table 4.2 ARIMA models and AIC values

Model	**AIC value**
ARIMA(0,1,0)	5833.465
ARIMA(0,1,1)	5808.408
ARIMA(1,1,0)	5802.431
ARIMA(1,1,1)	5800.068
ARIMA(1,1,2)	5801.501
ARIMA(2,1,1)	5801.016
ARIMA(2,1,2)	5795.129
ARIMA(2,1,3)	5770.057
ARIMA(3,1,2)	5764.439
ARIMA(3,1,3)	5802.166

The best ARIMA model for the data is ARIMA(3,2,1) with the lowest Akaike's Information Criterion (AIC), which is at 5764.439. AIC helps in determining the strength of linear regression model and penalizes a model for adding parameters. The parameters for ARIMA(3,1,2) is shown in table 4.3.

Table 4.3 Parameters for ARIMA model

Parameters	**Coefficient**	**Std error**
AR1 (ϕ_1)	-1.5429	0.029
AR2 (ϕ_2)	-0.3504	0.043
AR3 (ϕ_3)	0.3076	0.023
MA1 (θ_1)	1.8642	0.027
MA2 (θ_2)	0.9232	0.027

The parameters can be written into equation 4.1.

$$
\begin{aligned}
Y_t = (1 - 1.5429)Y_{t-1} + (-0.3504 + 1.5429)Y_{t-2} \\
+ (0.3076 + 0.3504)Y_{t-3} - 0.3076Y_{t-4} + \varepsilon_t \\
- 1.8642\varepsilon_{t-1} - 0.9232\varepsilon_{t-2}
\end{aligned}
$$

$$
\begin{aligned}
Y_t = -0.5429Y_{t-1} + 1.1925Y_{t-2} + 0.658Y_{t-3} - 0.3076Y_{t-4} + \varepsilon_t \quad \ldots(4.1) \\
- 1.8642\varepsilon_{t-1} - 0.9232\varepsilon_{t-2}
\end{aligned}
$$

4.2.3 DIAGNOSTIC CHECKING

After estimating the right parameters, the next step is to undergo a diagnostic checking. Figure 4.6 provides the statistical tests done to the fitted model. For this study, Ljung-Box test at lag 1 is used and H_0 means white noise residual, with $p = 0.05$. The value of probability is equal to 0.02 and the p-value is 0.90, which is more than the value of p. Hence, the assumption H_0 cannot be rejected.

Other than that, reviewing the residual plots can be an alternative to check the white noise of residual. In figure 4.7, the residual errors seem to fluctuate around zero even though they do not have uniform variance. The correlogram plot shows that the residual errors are not autocorrelated. Hence, it can be concluded that the residual is white noise and the model is fit to forecast.

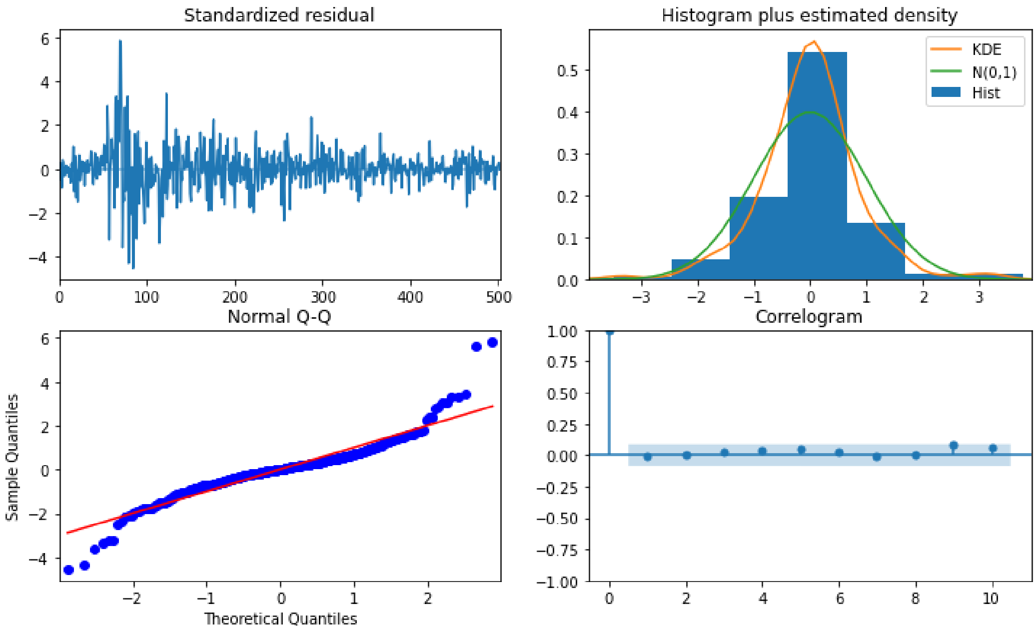

Figure 4.5 Diagnostic checking plots

4.3 LSTM MODEL

LSTM model has different aspects of specification. Each aspect is described as follows.

4.3.1 OPTIMIZING THE WEIGHTS AND BIASES

The weight and biases in the neurons must be optimized before training any Neural Network (NN) models. The optimized values of both will cause the output values generated by NN to be close to the actual values. To do that, the data is partitioned into "windows". The "windows" refer to sets of individual data points adjoining in the time series. This method is also known as the lagging method in statistics.

4.3.2 MULTIPLE PARAMETERS AND METHODS

Unlike any other models, parameters in NN are not estimated by the model. A number of parameters, which are known as hyperparameters, are determined manually. Hyperparameters include the number of neurons and layers inside the network, and the preferred methods that used to train the network (Michelucci 2018). There is no rule in choosing the appropriate hyperparameters.

NN models, especially LSTM, contain several hidden layers. A layer could be enough, but adding more layers could enhance the performance of model, making it a deep NN (Cybenko 1989; Pascanu et al. 2014). LSTM in this study used 3 hidden layers, which consist of 2 LSTM layers and 1 Dense layer.

In training LSTM, several methods are defined. The Mean Square Error (MSE) is a good choice as a measure for loss or loss function for time series based problems. The training data is split again into a validation set to help tune the hyperparameters. Optimizer, an algorithm to minimize the loss function, is performed in a three-step loop. The first step's purpose is to obtain gradient and value via forward and backward propagation. The second step's purpose is to put forward a new step and increment depending on the current step. The last and third step's purpose is to integrate the increment into the original function. This process is repeating in a certain number of times (Lv et al. 2017). The optimizer chosen in this study is Adaptive Moment

Estimation (ADAM). Batch is iteration of one or more samples used in forecasting. Same with hyperparameters, there is no rule in determining the size of batch. The number of epochs, defined as when the whole training dataset is passed through the model once, is decided by the characteristic of data. Regularization function is a function used to improve the generalization in LSTM by mitigating overfitting. Usually, for thinned networks, dropout has been proven to work the best.

4.3.3 PARAMETER TUNING

Literature based search was performed to determine the most appropriate parameters for the LSTM architecture in this study. The data size of 544 observations is considered small to be used in constructing deep learning models. Hence, only several pieces of literature are available. Selected parameters are presented in table 4.3.

Table 4.4 The selected parameters for LSTM model based on literature

Parameters	**Values**
Window size	5
Number of neurons	5, 10, 15
Number of layers	3
Loss function	MSE
Optimizer	ADAM
Batch size	4
Number of epochs	50 and 200
Dropout	0.2

4.3.4 MODEL SELECTION

There are 2 values for the number of epochs and 3 values for the number of neurons. Hence, to choose the fitted model for the data, model training is performed in each value. The first step is selecting the best number of epochs. Table 4.3 provides the result of model training with 2 different numbers of epochs; 50 and 200.

Table 4.5 Result of model training with different number of epochs

Window size	Neurons	Batch size	Epochs	RMSE
5	5	4	50	47.21789
5	5	4	200	57.43257

It is clearly seen that model with 50 epochs has the lowest value of RMSE. Therefore, 50 is chosen parameter for the model. The second step is training model for the right number of neurons. Table 4.4 provides the result of model training with 3 different numbers of neurons; 5, 10 and 15.

Table 4.6 Result of model training with different number of neurons

Window size	Neurons	Batch size	Epochs	RMSE
5	5	4	50	47.21789
5	10	4	50	44.10481
5	15	4	50	52.26863

Model with 10 neurons has the lowest value of RMSE. Hence, 10 is chosen parameter for the model. In conclusion, table 4.5 summarizes the chosen parameters for LSTM model in this study.

Table 4.7 The chosen parameters for LSTM model

Parameters	Values
Window size	5
Number of neurons	10
Number of layers	3
Loss function	MSE
Optimizer	ADAM
Batch size	4
Number of epochs	50
Dropout	0.2

4.4 FORECASTING

4.4.1 EVALUATION METRICS TO MEASURE ERROR

To determine which model makes better forecasts, MSE, RMSE and MAPE are employed to evaluate the error. Mean Squared Error (MSE) is faster to calculate, thus, it is used by many algorithms. The Root Mean Square Error (RMSE) is related to MSE and common metric used in determining the accuracy and error rate for many models. Mean Absolute Percentage Error (MAPE) is also one of the most commonly used metric to measure the accuracy of forecast and is quite well known by many business managers.

MSE, RMSE and MAPE can be computed by following formulas

$$MSE = \sum_{i=1}^{n} \frac{e_t^2}{n} \qquad \ldots(4.2)$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{e_t^2}{n}} \qquad \ldots(4.3)$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \frac{|e_t|}{dt} \qquad \ldots(4.4)$$

where e_t is difference between the forecasted values and the actual values, while n is the number of forecasts. The lowest the value of these evaluation metrics, the better the model.

4.4.2 RESULTS DISCUSSION

Figure 4.7 visualizes the real and predicted reference exchange rate of USD/IDR generated by ARIMA(3,1,2). On the other hand, figure 4.8 illustrates the actual and predicted value of reference rate of USD/IDR by the optimized LSTM model.

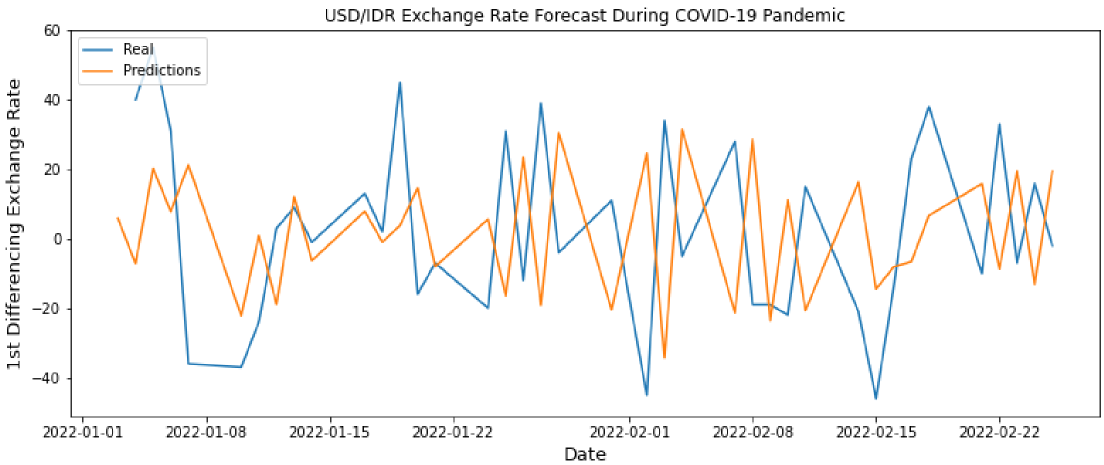

Figure 4.6 Plot of real and predicted JISDOR from ARIMA model

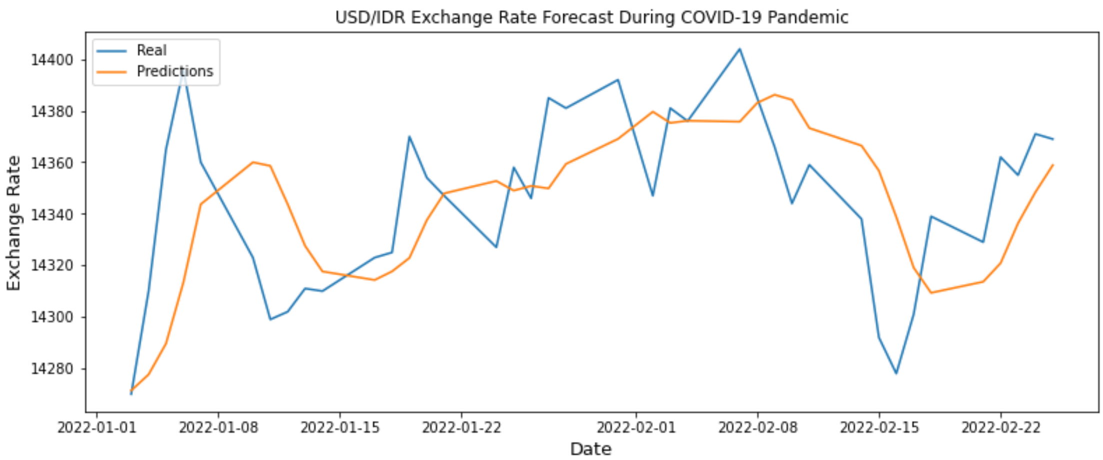

Figure 4.7 Plot of real and predicted JISDOR from LSTM model

Table 4.8 Error measures for ARIMA and LSTM model forecast

Model	**MSE**	**RMSE**	**MAPE**
ARIMA(3,1,2)	4848.9526	69.6344	0.43%
Optimized LSTM	1945.2343	44.1048	0.25%

Table 4.6 provides the values of evaluation metrics for both models. Looking at the plots and table, the result seems to match the initial hypothesis that the LSTM model

outperforms the ARIMA model in forecasting exchange rate of USD/IDR during the pandemic. Looking at figure 4.7, even for short-term forecasting, ARIMA does not have a chance to be able to have predict values close to the actual value. One possible explanation is likely due to the high uncertainty and volatility caused by the pandemic, making the financial market unpredictable and, later, at the time this study is performed, the global crisis caused by war between Ukraine and Russia. This has to be a note for investors. ARIMA is easy to construct and has purpose of obtaining short to mid-term predictions, but during pandemic, the model becomes unreliable in forecasting.

At a glance, incorporating ML into forecasting financial assets seems perfect and powerful. ML can detect signals that other traditional models did not. However, forecasting by time series based models will likely generate inaccurate and inconsistent predicted values. Combining technical analysis like this with fundamental analysis will always be the best strategy to employ when investing or trading in underlying assets. Incorporating macroeconomic indicators and sentiment analysis into model will probably improve the accuracy of the models, specially on the exchange rate forecasting.

LSTM is one of several architectures of DL that has promising future in handling financial time series problems. However, LSTM is one of complex models that the studies on these are difficult to do. One reason is the more clutter added, the less effective the models become. Forecasting exchange rate using this method cannot be the main solution to understanding the movement of the rate but can be used as a temporary reference. Using either model to turn a profit in trading could be a worthless endeavor, considering none of these models make perfectly accurate forecast.