

Application of SSA Decomposition in LSTM and GRU Models for Crude Oil Price Forecasting

Annisa Martina ^{1*}, Irwan Girana ², Rini Cahyandari ³

*Mathematics Department, Faculty of Science and Technology, UIN Sunan Gunung Djati Bandung
Jalan A.H. Nasution 105, Bandung, Indonesia*

*annisamartina@uinsgd.ac.id

Abstract

The world price of crude oil is very influential on the rate of inflation and the rate of economic growth; therefore, it is indispensable that the best methods for forecasting future prices with high accuracy are indispensable. To improve forecasting accuracy, various forecasting methods continue to be developed, for example the Hybrid model. This model can improve the forecasting accuracy of a single model. Hybrid Singular Spectrum Analysis (SSA)-Long Short-Term Memory (LSTM) and hybrid Singular Spectrum Analysis (SSA)-Gated Recurrent Unit (GRU) models combine the concept of decomposition with forecasting. Hybrid forecasting works with through two stages. First, SSA breaks the data down into trend, seasonal, noise and residue components. Second, decomposition outcome data were predicted using LSTM and GRU models. This Hybrid model has been shown to improve forecasting accuracy. Using weekly data on world crude oil opening prices, SSA-LSTM can improve forecasting accuracy by 78% over LSTM forecaster models. This is seen from respectively the RMSE values of 4.36 and the MAPE of 5.2% changing to 0.97 and 1.16%. Whereas in SSA-GRU can improve forecasting accuracy by 79% than single forecasting models GRU. This is seen from respectively RMSE values of 4.86 and MAPE 6.33% changing to 1.01 and 1.36%.

Keywords: Forecasting, Singular Spectrum Analysis, Long Short-Term Memory, Gated Recurrent Unit, SSA-LSTM, SSA-GRU

I. INTRODUCTION

Crude oil is a dark-colored liquid formed from the remains of dead organisms that were buried in layers of soil and rock millions of years ago. Crude oil can be processed into energy sources such as vehicle fuel, power plants, and others. Crude oil can also be referred to as petroleum. Crude oil prices often experience spikes on international markets. The surge is due to the fact that this oil has an influence in driving the economy [1]. Crude oil prices are very influential on the rate of inflation and the rate of economic growth [2]. Therefore, it is indispensable that the best methods for forecasting future prices with high accuracy are indispensable.

Forecasting is an activity that can predict future events. There are many classic methods that are often used in forecasting time series, namely Autoregressive (AR), Moving Average (MA), Autoregressive Integrated Moving Average (ARIMA), etc [3]. Some classical forecasting methods rely on the assumption of stationarity, data that remains constant over time. However, there are many situations when data is non-stationary such as having significant trends or random fluctuations. This method assumes a linear relationship. If the forecasted

data has nonlinear properties and complex patterns, this method may not be able to provide accurate and adequate results. This method is also susceptible to the influence of outlier data, namely extreme data points that are significantly different. Some forecasting methods assume that the data depends on a certain distribution such as the normal distribution. If the data does not match these assumptions, the forecast resulting from the method is inaccurate. Therefore, these methods have limitations to some degree.

Over time, other forecasting methods were introduced. One method that is currently popular is ANN forecasting method. ANN can predict linear or nonlinear data and can learn complex patterns. The ANN method has the ability to learn complex relationships between input and output that are difficult to understand by classical forecasting methods. This method can also work well on large datasets with many input variables. In some cases, the resulting forecasts are more accurate than classical forecasting methods. Examples of forecasting methods that use the ANN concept are LSTM and GRU. According to K.E. ArunKumar a, et al (2022) stated that the ARIMA model is suitable and performs well in modeling data that has a linear relationship. Meanwhile, Recurrent Neural Network (RNN)-based models such as LSTM and GRU have better performance on data that tends to be nonlinear [4]. In research by Oxaichiko Arissintaa.I, et al (2022), it was found that forecasting results using the ARIMA method had a greater error value than using LSTM and GRU [5].

Apart from ANN, a new forecasting method was introduced, namely the hybrid model [6]. A Hybrid Model is an approach that combines two or more different forecasting methods. This approach aims to utilize the advantages of each model so that it can overcome the limitations of a single forecasting model. Combining different approaches can improve the results of a single forecast. Some forecasting data is complex, such as containing nonlinear patterns and unstable fluctuations, so hybrid models make it possible to deal with complexity better. Hybrid models can be tailored to specific needs and data characteristics. This flexibility provides the advantage of being able to choose the most suitable modeling approach for each situation. An example of a hybrid model forecasting is combining the concept of decomposition with forecasting. The principle of decomposition is to separate a data series into several simple parts. This separation is useful for increasing accuracy in forecasting and helps understand the behavior of data series better [3].

SSA is a non-parametric data analysis method that functions to separate time series data into several simpler components such as trends, cycles, seasonality and random elements. This method can ignore assumptions in classical time series analysis such as stationarity assumptions and does not need to carry out logarithmic transformations [7]. If the decomposed data contains a strong trend component, then this component is not suitable for modeling with ARIMA because it has non-stationary properties. So, if you forecast using SSA-ARIMA model, it is not very suitable.

In research by Qi Tang, et al (2021), it was stated that forecasting using the hybrid SSA-LSTM model had better accuracy than other models [6]. Based on the research of K. Wang, et al (2023), SSA-LSTM model was built for each intrinsic mode function (IMF) components and one residual (RES) component, which effectively extracted the predicted values [8]. According to research from Y. Yang et al, (2023), SSA-GRU model has a stronger predictive impact to increase the reliability of electric load forecasts[9]. In research, Chao Li, et al (2022) stated that the hybrid SSA-GRU forecasting model had quite good forecasting results compared to the single model [10]. Therefore, based on the problems that have been explained, this research will propose a hybrid SSA-LSTM and SSA-GRU forecasting model which can overcome the limitations and improve the forecasting results of a single forecasting model, especially when faced with nonlinear data. It is hoped that this research can be used as a reference to predict time series data better.

II. LITERATURE REVIEW

A. Singular Spectrum Analysis (SSA)

SSA is a time series data analysis technique used for forecasting by combining elements of classical time series analysis, multivariate geometry, multivariate statistics, dynamic systems and signal processing [11]. The aim of SSA is to decompose data into small, identifiable components such as trend, seasonality and noise.

1) *Embedding*: Embedding is the step of changing time series data into an X path matrix, from initially being one-dimensional data to multidimensional data. For example, a time series $F = (f_1, \dots, f_N)$ with length N . Assume $N > 2$ and F is a nonzero series that has at least one i so that $f_1 \neq 0$. The path matrix is $L \times K$ and L is the window length, which is an integer with the condition $2 < L < N$ and forms $K = N - L + 1$. The path matrix X is called a Hankel matrix because all its anti-diagonal elements have the same elements, written as follows:

$$X = [x_{ij}]_{i,j=1}^{LK} = \begin{bmatrix} f_1 & f_2 & f_3 & \dots & f_K \\ f_2 & f_3 & f_4 & \dots & f_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_L & f_{L+1} & f_{L+2} & \dots & f_{LK} \end{bmatrix} \quad (1)$$

2) *Singular Value Decomposition (SVD)*: SVD aims to obtain component separation from the decomposition of time series data. Given $S = XX^T$, let $\lambda_1, \dots, \lambda_L$ be the eigenvalues of matrix S with $\lambda_1 \geq \dots \geq \lambda_L \geq 0$ and U_1, \dots, U_L be the eigenvectors of matrix S which correspond to the eigenvalues. Let $d = \max \{i, \text{untuk } \lambda_i > 0\}$, if $O_i = X^T U_i / \sqrt{\lambda_i}$ for $i = 1, 2, \dots, d$, then the SVD of the path matrix X is obtained:

$$X = X_1 + X_2 + \dots + X_d \quad (2)$$

Where $X_i = \sqrt{\lambda_i} U_i O_i^T$. The set (λ_i, U_i, O_i) is called the i th eigentriple of SVD.

3) *Grouping*: This step is groups the X_i matrix to separate the eigentriple components into several subgroups, trend, seasonality and noise. This process groups the index set $\{i, 2, \dots, d\}$ into m mutually exclusive subsets, I_1, I_2, \dots, I_m with $m = d$. For example, $I = \{i_1, i_2, \dots, i_m\}$, then the matrix X_I corresponding to group I is defined as $I = I_1, I_2, \dots, I_m$. Then $X_i = X_1 + X_2 + \dots + X_d$ can be expanded to become:

$$X_I = X_{I_1} + X_{I_2} + \dots + X_{I_m} \quad (3)$$

The selecting the set $I = \{I_1, I_2, \dots, I_m\}$ is called eigentriple grouping. Determining group members can be seen with the scatter plot of X_I , where those who have almost the same shape are classified into similar groups.

4) *Diagonal Averaging*: The final step is to change each matrix from equation (3) into a new series with length N . For example, matrix Y is an $L \times K$ matrix with elements $y_{i,j}$, $1 \leq i \leq L$, $1 \leq j \leq K$ for $L \leq K$. Given $L^* = \min(L, K)$, $K^* = \max(L, K)$, $n = L + K - 1$, $y_{ij}^* = y_{ij}$ if $L < K$ and $y_{ij}^* = y_{ji}$ if $L > K$.

$$Y = \begin{bmatrix} y_{11} & y_{12} & \dots & y_K \\ y_{21} & y_{22} & \dots & y_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ Y_L & Y_{L+1} & Y_{L+2} & Y_{LK} \end{bmatrix} \quad (4)$$

Matrix Y will be converted into the series g_0, g_1, \dots, g_{n-1} in the following way:

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m,k-m+2}^*, & 0 \leq k < L^* - 1 \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m,k-m+2}^*, & L^* - 1 \leq k < K^* \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m,k-m+2}^*, & K^* \leq k < N \end{cases} \quad (5)$$

For $k = 0$, $g_0 = y_{1,1}$, for $k = 1$, $g_2 = \frac{(y_{1,2} + y_{2,1})}{2}$ and so on.

B. Long Short-Term Memory (LSTM)

LSTM is a variation of RNN [13]. LSTM was developed specifically to overcome the problem of gaining a good understanding of the long-term context in time series data. Fig. 1 is an illustration of the LSTM architecture. LSTM can capture complex nonlinear relationship patterns in time series data, making it possible to make more accurate forecasts. Long-term memory cells in LSTM can “remember” important information from the past that can influence future forecasts, so they can learn and remember long-term patterns in the data. LSTM uses gates to control the flow of information in the network so that it can help with vanishing gradient and exploding gradient problems. LSTM has four main structures, i.e:

1) *Forget Gate*: Gate to determine what information will be removed from the cell state.

$$f_t = \sigma(x_t \cdot W_f + h_{t-1} \cdot U_f + b_f) \quad (6)$$

2) *Input Gate*: Gate to determine which part will be updated.

$$I_t = \sigma(x_t \cdot W_i + h_{t-1} \cdot U_i + b_i) \quad (7)$$

$$\hat{C}_t = \tanh(x_t \cdot W_c + h_{t-1} \cdot U_c + b_c) \quad (8)$$

3) *Cell State*: Important information is stored and accessed when the network sees new data.

$$c_t = f_t * c_{t-1} + I_t * \hat{C}_t \quad (9)$$

4) *Output Gate*: Functions to control how information in the cell state will be used to produce output

$$o_t = \sigma(x_t \cdot W_o + h_{t-1} \cdot U_o + b_o) \quad (10)$$

$$h_t = o_t * \tanh(c_t) \quad (11)$$

C. Gated Recurrent Unit (GRU)

GRU is a simpler variation of LSTM. GRU can combine many of the same concepts, but has a much simpler structure so it can train more quickly on hidden layers [12]. The purpose of creating a GRU is to sort all the information correctly that will be used in making decisions because not all information from the past can be used in making decisions. Fig. 2 is an illustration of the GRU architecture.

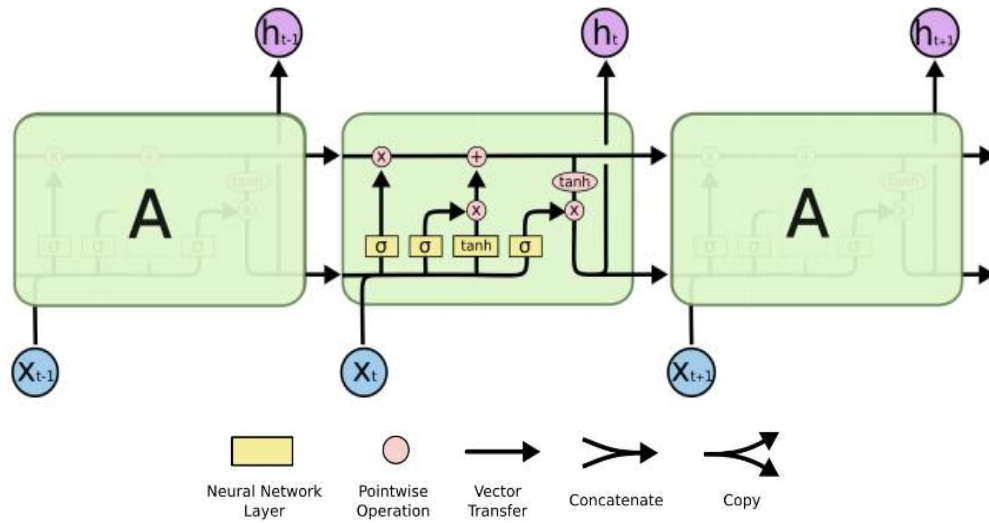


Fig. 1. LSTM Architecture [14]

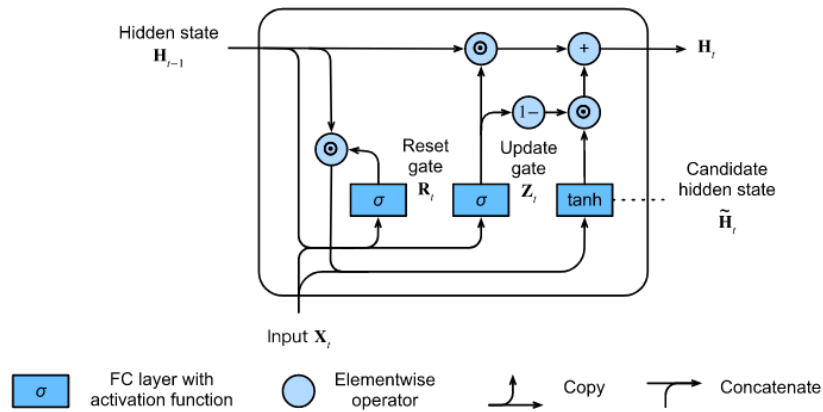


Fig. 2. GRU Architecture [15]

GRU can capture long-term patterns of time series data so that it can deal with forecasting involving complex relationships. The GRU gate mechanism allows organizing information in the network well so that it can learn long-term dependencies and avoid vanishing gradients and exploding gradients. GRU has two main structures, i.e:

1) *Update Gate*. A gate that functions to decide how much past information will be stored.

$$Z_t = \sigma(x_t \cdot W_z + h_{t-1} \cdot U_z + b_z) \tag{12}$$

2) *Reset Gate*. Useful in determining the combination of new input information with past information.

$$R_t = \sigma(x_t \cdot W_r + h_{t-1} \cdot U_r + b_r) \tag{13}$$

$$\hat{C}_t = \tanh(x_t \cdot W_h + (R_t * h_{t-1}) \cdot U_h + b_h) \tag{14}$$

$$h_t = (1 - Z_t) * \hat{C}_t + h_{t-1} * Z_t \tag{15}$$

Where:

f_t	: forget gate	h_t	: hidden state	\tanh	: tanh function
I_t	: input gate	Z_t	: update gate	W	: weight
\hat{C}_t	: calon cell state	R_t	: reset gate	U	: recurrent weight
c_t	: cell state	x_t	: data input	b	: biased
o_t	: output gate	σ	: sigmoid function		

D. Hybrid SSA–LSTM and Hybrid SSA–GRU

The Hybrid SSA–LSTM and Hybrid SSA–GRU model is a combination of the SSA decomposition concept with LSTM and GRU forecasting. The SSA decomposition method can break down a data series into several components such as trend, seasonality, and noise. The aim of the decomposition stage is so that the forecasting model can learn patterns better because the data has been broken down based on its respective characteristics. The SSA method is used as a preprocessing stage before entering the forecasting stage with LSTM and GRU [16]. Next, each SSA component will be forecasted using LSTM and GRU. The forecasting results for each component will be reconstructed by adding them according to the decomposition principle. Mathematically, the final result of forecasting can be represented as follows.

$$\hat{Y}_t = \hat{T}_t + \hat{I}_t + \hat{E}_t + \hat{\varepsilon}_t \quad (16)$$

Where:

\hat{Y}_t	: Hybrid forecasting in period t
\hat{T}_t	: forecast of the SSA trend component in period t
\hat{I}_t	: forecast of SSA seasonal components in period t
\hat{E}_t	: forecast of the SSA noise component in period t
$\hat{\varepsilon}_t$: forecast of SSA residual components in period t

III. RESEARCH METHOD

The method used is hybrid SSA-LSTM and hybrid SSA-GRU forecasting model. In Fig. 3 is a research flowchart, which consists of hybrid SSA-LSTM, hybrid SSA-GRU, and a single method without hybrid which is LSTM and GRU. Then, the study will be carried out in three stages, namely the first stage of data description analysis, the second stage of decomposition, and the third stage of forecasting.

A. Descriptive Analysis

This stage is carried out to see in detail the data to be processed. Flowchart descriptive analysis can be seen in Fig. 4. As for this stage, several steps will be carried out as follows:

1) *Data Description*. The data used will be described in the form of graphs, in addition, simple measurements such as mean, standard deviation, maximum and minimum will be carried out.

2) *Stationarity Test*. This stage will check the data used to be stationary or nonlinear with graph test and ADF test. The graph test can be viewed visually by performing a plot between the observation value and the time value. Stationary data occurs when the graph tends to be constant and has no interesting or decreasing trend. The unit root test is a formal concept used to determine stationarity.

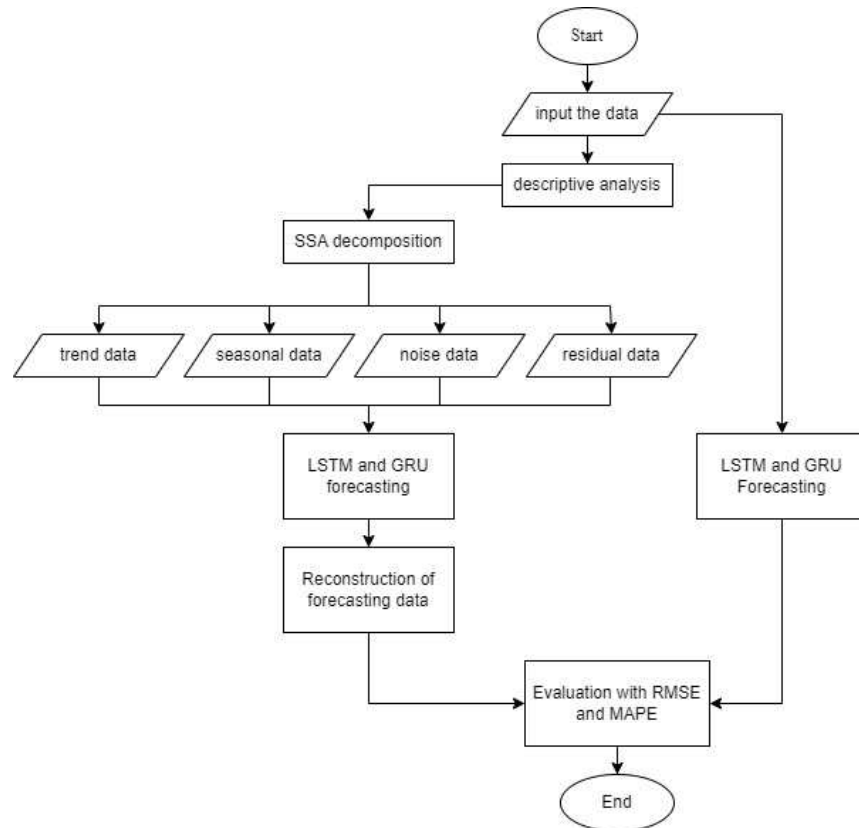


Fig. 3. Research Flowchart

The test was developed by D.A. Dickey and W.A. Fuller and became known as the Augmented Dickey-Fuller Test (ADF Test). ADF test is the most widely used type of test today because it can take into account the presence of autocorrelation problems that often occur in data time series [17]. The ADF test is determined by comparing the statistical value of the ADF with its critical value. It can also be compared with alpha (α) confidence degrees with a value of 1%, 5%, or 10%. If the result obtained is that the statistical value of ADF is lower than the degree of confidence, the null hypothesis is rejected. It can be concluded that the data is stationary.

Hypothesis:

H_0 : non-stationary data

H_1 : stationary data

If $p\text{-value} < \alpha$ then H_0 is rejected, the opposite is true.

3) *Terasvirta Test*. This stage will check the data used to be linear or nonlinear with the Terasvirta test. Nonlinear refers to the context of a relationship pattern that does not follow an aligned relationship pattern. Nonlinear data show irregular complex patterns, do not follow specific patterns or no disproportionate change between input and output variables [18]. Nonlinearity testing aims to ascertain whether a data follows a linear or nonlinear pattern. The Terasvirta test is one of the tests often used in nonlinearity testing. This test is better than the White test and Ramsey test in detecting nonlinearities [19]. The Terasvirta test belongs to the Lagrange

Multiplier (LM) test group, in addition, this test is developed from neural networks. The steps in the Terasvirta test are as follows:

- i. Regressing Y_t at $1, Y_{t-1}, \dots, Y_{t-p}$ then computes $\hat{u} = Y_t - \hat{Y}_t$.
- ii. Regressing \hat{u} at $1, 1, Y_{t-1}, \dots, Y_{t-p}$ and m additional predictors of the quadratic tribe from the Taylor expansion approximation results.
- iii. Determine the coefficient of determination (R^2) of the regression in the previous stage.
- iv. Determine the test statistic $X^2 = nR^2$ where n is the number of observations.

Hypothesis:

H_0 : linear model

H_1 : nonlinear model

If $p\text{-value} < \text{degree of confidence } (\alpha)$, then H_0 is rejected, true the opposite.

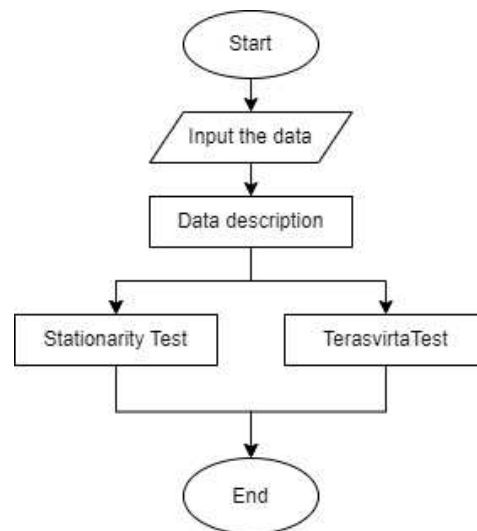


Fig. 4. Flowchart Descriptive Analysis

B. Decomposition

This stage is the process of decomposition to break data into several components based on their nature. The decomposition stage uses SSA method. Flowchart decomposition can be seen in fig. 5. In general, the stages of SSA are as follows:

1) *Embedding*. Stage converts the time series data into a trajectory matrix X . The window length parameter (L) is also specified.

2) *Singular Value Decomposition (SVD)*. Stage to obtain the sorting of components from the decomposition of time series data. The result is eigentriple.

3) *Grouping*. Stage for grouping X_i matrices with the aim of separating eigentriple components into several subgroups namely trend, seasonality, and noise.

4) *Diagonal Averaging*. Stage transforms the matrix from the previous grouping result into a new series with Length N . The final result of the decomposition process is data broken down by parts based on the properties of its components.

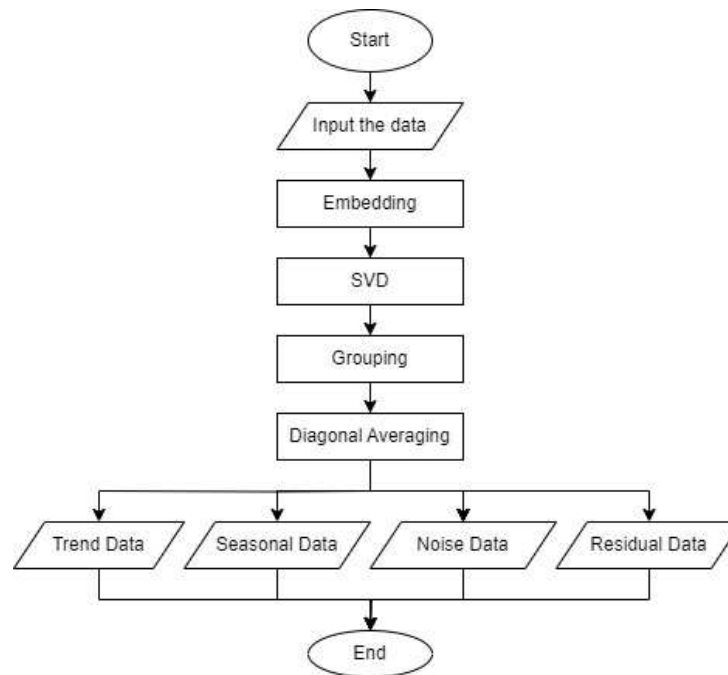


Fig. 5. Flowchart Descomposition

C. Forecasting

After going through the decomposition stage, the data obtained will enter the forecasting stage which is using LSTM and GRU models. In this forecasting it will go through several stages namely:

- 1) *Normalization*. This process is done to equalize all the data to be predicted into a scale of zero to one.
- 2) *Segmentation*. This data will be separated and grouped into series. Suppose a sequence length of 50 data is selected for each group. Then a series of 50 data from the past will be used to predict the 51st data.
- 3) *Data Division*. The data division will be divided into two i.e. training data used in the formation of training models and test data for useful Implement this model. In general, the data is divided into an 80:20 ratio, with 80% for the training data and 20% for the test data.
- 4) *Model Construction and Parameter Selection*. In this forecasting, two models will be used, namely LSTM and GRU models. As for the required parameters, they are as follows:
 - i. Units, the number of memory cells in the model layer. The number of units will affect the complexity of the model.
 - ii. Batch size, the number of data samples processed simultaneously in one iteration when training or testing a model. Batch size determines how many time steps will be processed simultaneously before weight adjustment.
 - iii. Dropout, to reduce overfitting that randomly disables a number of units in the network during the training process. Dropout level selection is the presentation of the disabled unit.
 - iv. Sequence length is the length of the time step used in a batch of data that is fed into the model as input to each iteration.

- v. Epoch, refers to one full round of the entire training data used to train the model. Each epoch, the model will receive all batches in sequence and perform calculations to update the weights.
- vi. Learning rate, is used to control how large the change, fast slow weight and bias at each renewal step during model training in order to achieve the desired gradient calculation.
- vii. Feature, refers to variables that are used as inputs to train the model.

5) *Calculation.* This is a data training process performing a training process based on forecasting models and predefined parameters. After the training process is completed, the model data will be stored to be applied into the data testing so that the final results will be obtained from the forecasting results.

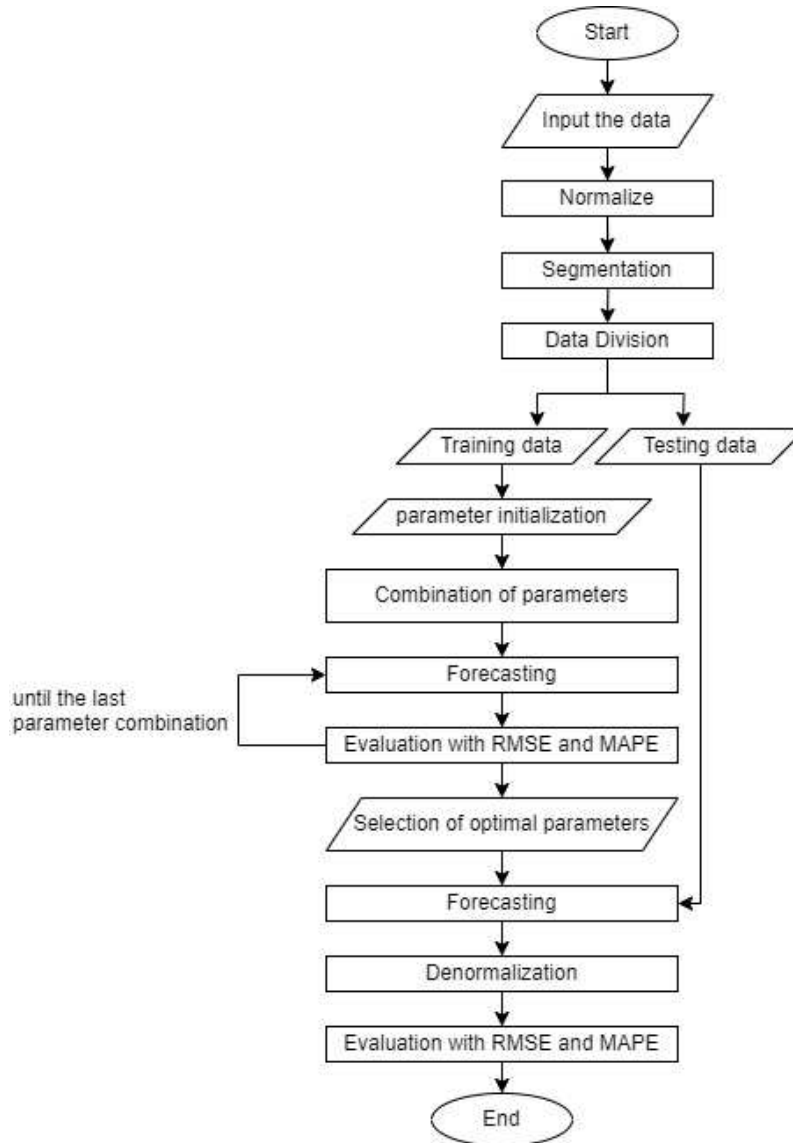


Fig. 6. LSTM and GRU Forecasting Flowchart

6) *Denormalisation.* This process is the inverse of normalisation. After the completion of the forecast, the data is still in the state of the value range between zero to one. Therefore, this process will return the data to its original state.

7) *Evaluation.* of forecasting results with data testing calculated their accuracy with RMSE and MAPE.

After going through all stages, the process will return to stages 4 to 6 by updating the parameters. The final result will be taken by looking at the smallest RMSE and MAPE values. LSTM and GRU forecasting flowchart can be seen in Fig. 6.

The data used is world crude oil price data. World oil prices are taken from the West Texas Intermediate (WTI) price which is the standard. The data taken is 1113 weekly opening prices from January 6 2003 to April 30 2023 [20]. This data is nonstationary and nonlinear, meaning it has a trend component and has an unstable movement pattern. When decomposing SSA, the parameter window length (L) = 10 will be used with the grouping process divided into 3 so that the final result is data that is broken down into groups of trend, seasonality, noise and residues created from the remains of the decomposition. At the forecasting stage, four scenarios will be used, namely using the LSTM, GRU, Hybrid SSA-LSTM, and Hybrid SSA-GRU models. The final forecasting results will measure the accuracy of the prediction error using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The forecasting process will use the hyperparameter tuning grid search process to determine optimal initialization parameters. The parameters used are: units (5, 25, 50), batch sizes (1, 4, 16, 64), dropout (0.0), sequence length (15, 30), epochs (25, 50, 100, 150), learning rate (0.1, 0.01), and features (1).

IV. RESULTS AND DISCUSSION

In the SSA decomposition process, 10 components are obtained based on the previously determined L parameters. In the SVD stage, triple eigen is obtained. The eigenvector will be plotted to see the proportion of data sharing from the 10 components (Fig. 7). These components will enter the grouping process to be united based on their similar characteristics. To see components that have similarities more clearly, a W-correlation graph will be shown. W-correlation is useful for seeing the correlation between different components. A higher correlation value indicates a strong relationship between two or more components. The higher the relationship, the darker the color obtained, the opposite applies (Fig. 8). The results of the W-correlation value will be divided into three components, namely trend, seasonality and noise. The results are obtained in Table I.

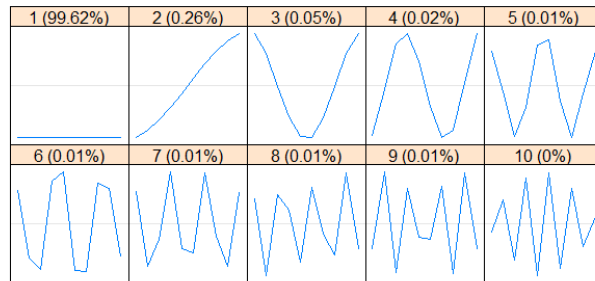


Fig. 7. Eigenvector Graph of World Crude Oil Data

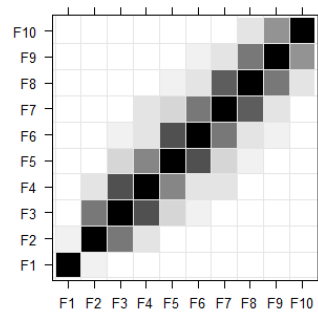


Fig. 8. W-correlation Graph for World Crude Oil Data

TABLE I
GROUPING BASED ON CHARACTERISTICS

Characteristics	Group members	X_I	Influence on data
Trends	1	$X_{I1} = X_1$	99.62 %
Seasonal	2, 3, dan 4	$X_{I2} = X_2 + X_3 + X_4$	0.33 %
Noise	5, 6, 7, 8, 9, dan 10	$X_{I3} = X_5 + \dots + X_{10}$	0.05 %

The final results of the SSA decomposition process, obtained four data breakdowns with the characteristics of trend, seasonality, noise, and the residual data remaining from the decomposition was also taken as seen in Fig. 9. Red is for actual data, green is trend data, black is seasonal data, blue is noise data, and orange residual data. In the forecasting process, the optimal parameters obtained from hyperparameter tuning are as shown in Table II. These parameters are used to predict data for the next five weeks based on the respective models and data characteristics. The prediction results are obtained in Table III.

TABLE II
OPTIMAL PARAMETERS FOR PREDICTING WORLD CRUDE OIL OPENING PRICES

Parameter Type	LSTM	GRU	Hybrid SSA-LSTM				Hybrid SSA-GRU			
	Model	Model	T*	S*	N*	R*	T*	S*	N*	R*
Unit	5	25	50	5	50	5	25	5	50	25
Batch size	64	64	4	4	4	64	64	64	16	4
Dropout	0	0	0	0	0	0	0	0	0	0
Sequence length	15	15	30	30	30	15	15	30	30	15
Epoch	100	150	100	150	50	150	150	150	50	50
Learning rate	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Feature	1	1	1	1	1	1	1	1	1	1

*T: trends, S: seasonal, N: noise, R: residue

TABLE III
PREDICTION OF THE OPENING PRICE OF WORLD CRUDE OIL IN THE NEXT 5 WEEKS

No	Date	Actual Data	LSTM	GRU	Hybrid SSA-LSTM	Hybrid SSA-GRU
1	07/05/2023	71.35	78.4769210	77.8991928	73.3830236	73.5997327
2	14/05/2023	70.04	79.1388855	79.0573043	72.3042032	71.6999303
3	21/05/2023	71.88	80.2953796	80.1281433	72.1377718	74.2174441
4	28/05/2023	73.23	81.5222930	81.1835861	74.0797035	74.6545886
5	04/06/2023	73.86	82.7767868	82.2436599	73.8611880	75.0029702

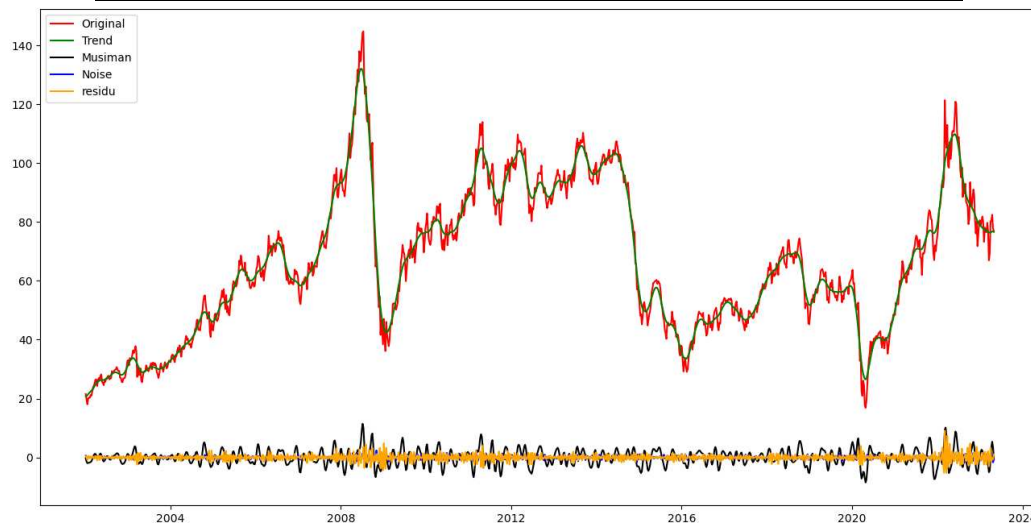


Fig. 9. Graph of SSA Decomposition Results of World Crude Oil Data

Forecasting accuracy measurements are calculated to see the level of accuracy of the prediction results obtained as shown in Table IV. In Table IV, it can be seen that overall, the type of forecasting using the LSTM model has better results than the GRU model. This can be seen from the respective RMSE and MAPE values in forecasting the opening price of world crude oil.

TABLE IV
PREDICTION OF THE OPENING PRICE OF WORLD CRUDE OIL IN THE NEXT 5 WEEKS

Measure of Accuracy	LSTM	GRU	Hybrid SSA – LSTM	Hybrid SSA-GRU
RMSE	4.36	4.86	0.97	1.01
MAPE (%)	5.2	6.33	1.16	1.36

The Hybrid SSA-LSTM and Hybrid SSA-GRU model improves forecasting accuracy. This can be compared between the Hybrid SSA-LSTM, and Hybrid SSA-GRU models with the LSTM and GRU models. The Hybrid SSA-LSTM model can improve forecasting accuracy by as much as 78% compared to a single LSTM forecasting model. This can be seen from the respective values of RMSE 4.36 and MAPE 5.2% changing to 0.97 and 1.16%. Meanwhile, the Hybrid SSA-GRU model can increase forecasting accuracy by 79% compared to the single GRU forecasting model. This can be seen from the respective values of RMSE 4.86 and MAPE 6.33% changing to 1.01 and 1.36%. In a case study using weekly data on world crude oil opening prices, SSA decomposition can increase forecasting accuracy by 78-79% in LSTM and GRU forecasting.

In Fig. 10, you can see a graphic summary of the four forecasting models. Blue indicates actual data, green for LSTM forecasting results, black for GRU forecasting results, red for Hybrid SSA-LSTM forecasting results, and orange for Hybrid SSA-GRU forecasting results. It is estimated that the weekly opening price of world crude oil in the next five weeks will experience a slight decline, after that it will experience a slight increase as shown in table 3. However, the results of this prediction could be slightly off due to the fluctuating and difficult to predict movement of the opening price of world crude oil. because there may be other factors that could cause this price change.

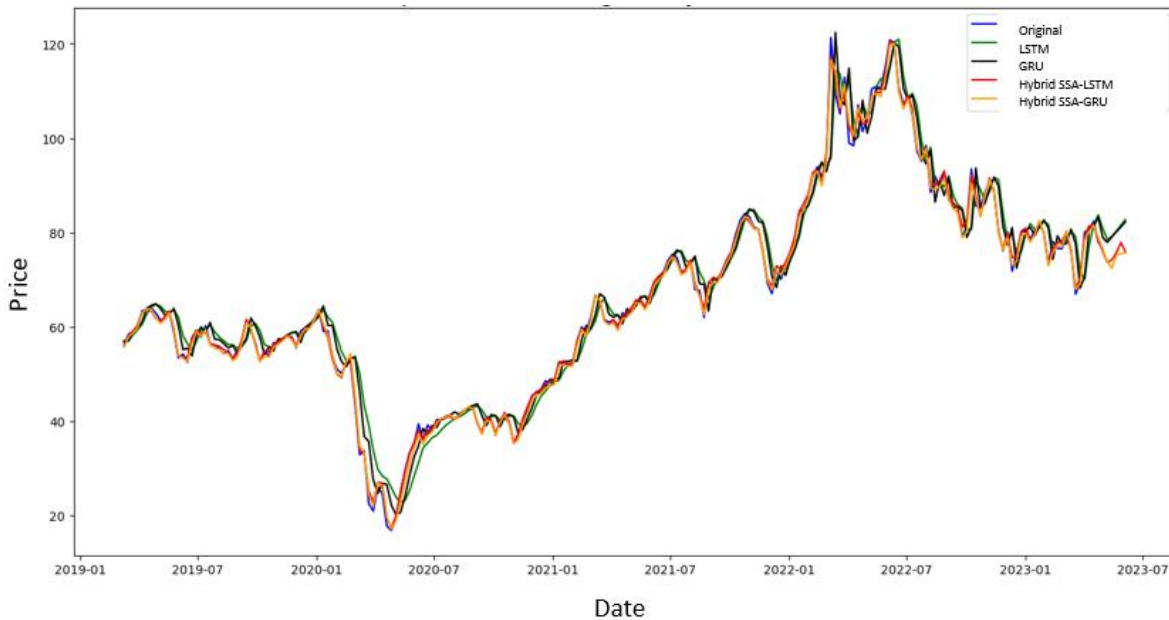


Fig. 10. Graph of Summary of World Crude Oil Price Predictions

V. CONCLUSION

Hybrid SSA-LSTM and Hybrid SSA-GRU forecasting works in two stages. First, SSA decomposition is useful for breaking down data into trend, seasonal, noise and residual components. Second, the decomposed data is predicted using LSTM and GRU models. The Hybrid SSA-LSTM and Hybrid SSA-GRU model is proven to be able to increase forecasting accuracy. This can be compared between the Hybrid SSA-LSTM, and Hybrid SSA-GRU models with the LSTM and GRU models. In this study, the data used are data on world crude oil prices. World oil prices are taken from the price of West Texas Intermediate (WTI) which is the standard. The data taken is the weekly opening price from January 06, 2003 to April 30, 2023 with 1113 data. Then generated the predicted value of the next 5 weeks of the world crude oil opening price. The Hybrid SSA-LSTM model can improve forecasting accuracy by as much as 78% compared to a single LSTM forecasting model. This can be seen from the respective values of RMSE 4.36 and MAPE 5.2% changing to 0.97 and 1.16%. Meanwhile, the Hybrid SSA-GRU model can increase forecasting accuracy by 79% compared to the single GRU forecasting model. This can be seen from the respective values of RMSE 4.86 and MAPE 6.33% changing to 1.01 and 1.36%. Thus, in a case study using weekly data on world crude oil opening prices, the application of SSA decomposition can increase forecasting accuracy by 78-79% in LSTM and GRU forecasting.

There are several things that cause this component to have poor forecasting. First, seasonal components, noise, and residues are stationary. Stationary data are less likely to have trends and have a pattern of long-term limitations. LSTM and GRU models are designed to study and remember long-term patterns on data. Secondly, excess complexity, LSTM and GRU models have complex structures with many parameters that need to be estimated. Incongruous initialization of parameters can also be the cause. Thus, when used to forecast stationary data, this complexity can likely lead to overfitting, where the model performs well on training data, but results in poor performance on test data or new data. It is recommended that further research consider the type of method that is more effective in predicting each component of decomposition results, especially data with stationary properties. In addition, it is worth paying attention in the hyperparameter tuning stage to think of the most optimal method in maximizing the parameters of initialization results that have a shorter computational time.

REFERENCES

- [1] M. A. Nizar, "Dampak Fluktuasi Harga Minyak Dunia Terhadap Perekonomian Indonesia," *Buletin Ilmiah Litbang Perdagangan*, Vol 6 No 2, 2012, doi.org/10.30908/bilp.v6i2
- [2] S. D. Purnomo, I. Istiqomah, and L. S. Badriah, "Pengaruh Harga Minyak Dunia Terhadap Pertumbuhan Ekonomi, Inflasi, Dan Pengangguran Di Indonesia," *Jurnal PROFIT Kajian Pendidikan Ekonomi dan Ilmu Ekonomi*, vol. 7, no. 1, pp. 13–24, Apr. 2020, doi: 10.36706/jp.v7i1.11025.
- [3] S. Makridakis, S. C. Wheelwright, and V. E. McGee, *Metode dan Aplikasi Peramalan*. Jilid 1. Jakarta: Erlangga, 1999.
- [4] K. E. ArunKumar, D. V. Kalaga, C. Mohan Sai Kumar, M. Kawaji, and T. M. Brenza, "Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends," *Alexandria Engineering Journal*, vol. 61, no. 10, pp. 7585–7603, Oct. 2022, doi: 10.1016/j.aej.2022.01.011.
- [5] I. Oxaichiko Arissinta, I. Dwi Sulistiyawati, and D. Kurnianto Iqbal Kharisudin, "Pemodelan Time Series untuk Peramalan Web Traffic Menggunakan Algoritma Arima," *Prosiding Seminar Nasional Matematika*, vol. 5, pp. 693–700, 2022.
- [6] Q. Tang, R. Shi, T. Fan, Y. Ma, and J. Huang, "Prediction of Financial Time Series Based on LSTM Using Wavelet Transform and Singular Spectrum Analysis," *Math Probl Eng*, vol. 2021, 2021, doi:

- 10.1155/2021/9942410.
- [7] D. Alif Shafira and T. Wahyu Utami, "Peramalan Wisatawan Mancanegara Berkunjung Ke Bali Menggunakan Metode Singular Spectrum Analysis (SSA)," *Prosiding Seminar Nasional Indonesian R Summit*, vol.1, no 1. 2020.
- [8] K. Wang, X. Fan, X. Yang, Z. Zhou, "An AQI Decomposition Ensemble Model Based on SSA-LSTM Using Improved AMSSA-VMD Decomposition Reconstruction Technique," *Environmental Research Elsevier*, 2023, doi.org/10.1016/j.envres.2023.116365.
- [9] Y. Yang, C. Li, X. Xiang, " Electric Load Forecasting Based on SSA-GRU," *IEEE Conference Publication*, 2023, ISBN:979-8-3503-1468-7.
- [10] C. Li, Q. Guo, L. Shao, J. Li, and H. Wu, "Research on Short-Term Load Forecasting Based on Optimized GRU Neural Network," *Electronics (Switzerland)*, vol. 11, no. 22, Nov. 2022, doi: 10.3390/electronics11223834.
- [11] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Series Structure: SSA and Related Techniques*. 2001.
- [12] LazyProgrammer, *Deep Learning: Recurrent Neural Networks in Python LSTM, GRU, and more RNN machine learning architectures in Python and Theano*, 2016.
- [13] J. Brownlee Disclaimer, *Long Short-Term Memory Networks With Python Develop Sequence Prediction Models With Deep Learning*. 2017.
- [14] C. Olah, "Understanding LSTM Networks",2015. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/> A.Zhang, M.Li, Z.C. Lipton, and A.J. Smola. (2019). Hans-On Learning Deep Learning. [Online]. Available: <https://zh.d2l.ai>
- [15] A. Zhang, M.Li, Z.C. Lipton, and A.J. Smola. (2019). Hans-On Learning Deep Learning. [Online]. Available: <https://zh.d2l.ai>
- [16] S. L. Zubaidi, J. Dooley, R. M. Alkhaddar, M. Abdellatif, H. Al-Bugharbee, and S. Ortega-Martorell, "A Novel approach for predicting monthly water demand by combining singular spectrum analysis with neural networks," *J Hydrol (Amst)*, vol. 561, pp. 136–145, Jun. 2018, doi: 10.1016/j.jhydrol.2018.03.047.
- [17] S. Aktivani, S. Ahli, P. Badan, P. Statistik, and K. Padang, "Uji Stasioneritas Data Inflasi Kota Padang," 2020.
- [18] George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel, *Time Series Analysis Forecasting and Control*. Hoboken, New Jersey: John Wiley & Sons, Inc, 2016.
- [19] Hendri Prabowo, Suhartono, and Dedy Dwi Prastyo, "The Performance of Ramsey Test, White Test and Terasvirta Test in Detecting Nonlinearity," *INFERENCE*, vol. 3(1), pp. 2721–3862, 2020.
- [20] Data Historis Minyak Mentah WTI Berjangka. <https://id.investing.com/commodities/crude-oil-historical-data>