



Comparison of ARIMA and SARIMA for Forecasting Crude Oil Prices

Vika Putri Ariyanti¹, Tristyanti Yusnitasari²

^{1,2}Department of Computer Science, Gunadarma University

¹vikaputriariyanti@gmail.com, ²tyusnita@staff.gunadarma.ac.id

Abstract

Crude oil price fluctuations affect the business cycle due to affecting the ups and downs of the growth of the economy, which one of the indicators of the economic business cycle phenomenon. The importance of oil price prediction requires a model that can predict future oil prices quickly, easily, and accurately so that it can be used as a reference in determining future policies. Machine learning is an accurate method that can be used in predicting and makes it easier to predict because there is no need to program computers manually. ARIMA is a machine learning algorithm while ARIMA that uses a seasonal component is called SARIMA. Based on background, research purpose is modeling crude oil price forecasting by ARIMA and SARIMA. Forecasting is done on daily crude oil price data taken from Yahoo Finance from January 27, 2020 to January 25, 2023. The evaluation results show the RMSE value of ARIMA and SARIMA is 1.905. The forecast result of 7 days ahead with ARIMA is 86.230003 while SARIMA is 86.260002. The research results are expected to be helpful for policy makers to adopt policies and make the right decisions in the use of crude oil.

Keywords: ARIMA, SARIMA, forecasting, crude oil price

1. Introduction

One of the commodities that is quite important for the Indonesian economy is oil [1]. Oil price fluctuations impact business cycles. This is because fluctuations in crude oil prices affect the rise and fall of economic growth, that is an indicator of economic business cycle phenomena. Oil price fluctuations also have an impact on state revenue in the form of non-tax oil revenue and oil tax revenue [2].

The short and long-term impacts of world oil price fluctuations on macroeconomic variables and fuel subsidies require appropriate government policies to avoid economic and social instability in society [3].

As one of the most important strategic resources in the world, the fluctuation cycle of oil prices is critical to the survival and development of oil companies. Accurate oil price forecasts will help oil companies adjust business strategies, optimize asset structures, avoid oil price risks, and help investors understand the laws of market volatility and make profits in trading [4].

Given that oil price forecasts are important, a model is required which can be used to forecast future oil prices quickly, easily, and accurately that can be used as a reference in determining future policies.

Machine Learning is an accurate method can be used in predicting prices [5]. The use of machine learning, also makes it easier to make predictions because it does not have to manually program the computer continuously. Machine learning must be trained by studying pre-existing data to be able to do something automatically. This aims to make a prediction based on these data [6].

One of the machine learning algorithms is ARIMA. This algorithm is flexible (following data patterns), has a relatively high level of precision in forecasting and is suitable for forecasting several variables quickly, easily, accurately, and cheaply as it requires only historical data to perform forecasting [7].

An extension of the ARIMA model that has seasonal effects is SARIMA [8]. According to the background above, this research will do crude oil price forecasting with ARIMA and SARIMA model. A Forecast is a prediction of what will occur in the future [9].

There are many studies that have used ARIMA for forecasting. Such as Bitcoin price forecasting [10], Litecoin and Dogecoin price forecasting [11], and forecasting third wave of COVID-19 effect in India [12].

This ARIMA model was chosen because the performance of ARIMA is better when used for the short term [10].

Meanwhile, forecasting with SARIMA was done for forecasting the flow of German tourists to Croatia. The research results in sample MAPE of 9.329873 and out of sample MAPE of 3.800651. It indicates the forecasting model performance is very accurate because MAPE is less than 10% [13].

Forecasting is performed on daily crude oil price data which is extracted from Yahoo Finance starting from January 27, 2020 to January 25, 2023. The forecasting model performance by ARIMA can be known by computing the RMSE value. Model results with the best RMSE value are then used to forecast oil prices for one week. This research is expected to help policy makers to adopt the right policies and make the right.

2. Research Methods

2.1 Data Extration

Extracting crude oil price data from Yahoo Finance is conducted with yfinance. It's an open source python package used for downloading of historical market data from Yahoo! Finance [14]. The data retrieved is crude oil price data from January 27, 2020 to January 25, 2023.

2.2 Preprocessing

Preprocessing is used to get data as required. The data required for this research is Close. Close is selected to present the price of the index to be forecasted. Close was selected as it represents the entire index activity on the trading day [15]. So that unnecessary data such as Open, High, Low, Adj Close, and Volume need to be deleted.need to be deleted.

2.3 Parameter Selection

ARIMA parameters are p , d , and q while SARIMA hyperparameters are p , d , and q as well as P , D , Q , and S [16]. The selection of optimal parameter values can be done by plotting ACF and PACF graphs.

In addition, it can also be done by auto arima function. This `auto_arima` function works similar to a grid lookup that tests different sets of parameters p and q for non-seasonal model also P and Q for seasonal models. Model selection is performed using minimized AIC or BIC or selected information criteria [17][12].

Akaike's Information Criterion (AIC), which can help define ARIMA model order, can be formulated in equation 1 [18]:

$$AIC = -2 \log(L) + 2(p + q + k + 1) \quad (1)$$

With L is data likelihood, $k=1$ when $c \neq 0$ and $k=0$ when $c=0$ and final term in brackets represents the sum of the models parameters (σ^2 of residual variance included).

The AIC and Bayesian Information Criterion (BIC) of the ARIMA model in equation 2 and 3 [18]:

$$AICc = AIC + \frac{2(p+q+k+1)(p+q+k+2)}{T-p-q-k-2} \quad (2)$$

$$BIC = AIC + [\log(T) - 2](p + q + k + 1) \quad (3)$$

The best model is found if the AIC, AICc or BIC is minimized. Criteria of information are not likely the best guide to select the correct d value of a model, but just to select p and q values. It is due to different transformations from the data that the likelihood is calculated, so that AIC values of the models with different orders of difference cannot be compared. Therefore, it is necessary to use another method to select d , and then select p and q using AICc [18].

In auto arima to choose differencing using stationarity tests as the Dickey-Fuller and seasonal differencing as the Canova-Hansen for Seasonal models [17].

Stationarity holds an essential position in time series analysis, as to apply the ARIMA model for forecasting past cases, it first needs to be confirmed that time series data cannot have an increasing or decreasing or seasonal trend through the test of stationarity. Time series data that does not have an upward or downward trend or seasonality is what is called stationary [12]. Data stationarity is required to minimize model error [19].

A time series that is stationary is a series without a particular pattern [12]. A series is stationary when the following conditions are [20]: Constant Mean by Time; Constant Variance by Time, and Constant Autocorrelation by Time.

A comparison of stationary and non-stationary plots of time series is shown in Figure 1 [20]. Based on Figure 1, the plot on the left is stationary while the plot on the right shows non-stationary.

ADF test which stands for Augmented Dickey-Fuller is one such unit root test. This test was developed by Dickey and Fuller [21], [22] This test is performed to identify whether a time series is stationary by using the p -value, when [12], [23]: Value of $p > 0.05$: Failing to reject H_0 means that the time series has unit roots and isn't stationary; Value of $p \leq 0.05$: Reject the H_0 , meaning there isn't unit root and the time series is stationary.

The hypotheses of ADF test are as follows [12], [21], [23] : Null hypothesis (H_0): the time series has unit root and isn't stationary (has a time dependent structure); Alternative hypothesis (H_1): the time series hasn't unit root and is stationary.

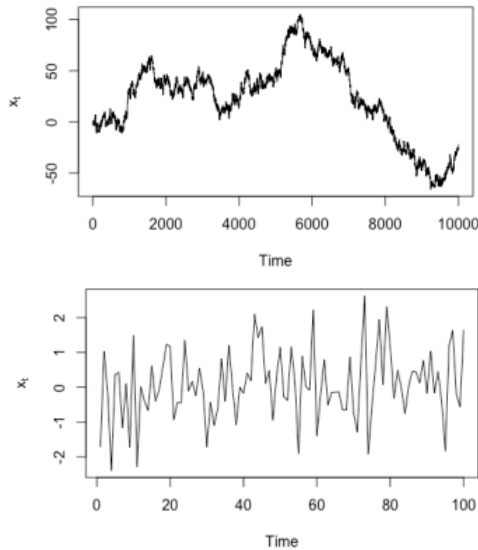


Figure 1. Stationary and Nonstationary Plot Examples

Table 1. Series Periodicity According to Data Collection Rhythm

Data	Period (Count of Observations by Cycle)						
	Minute	Hour	Day	Week	Month	Quarter	Year
Annual							1
Quarterly							4
Monthly						3	12
Weekly					4.348125	13.044375	52.1775
Daily				7	30.436875	91.310625	365.2425
Hourly			24	168	730.485	2191.455	8765.82
Half-hourly			48	336	1460.97	4382.91	17531.64
Minute		60	1440	10080	43829.1	131487.3	525949.2
Seconds	60	3600	86400	604800	2629746	7889238	31556952

2.4 Modeling

Modeling is conducted by ARIMA which was is a method that George Box and Gwilyn Jenkins developed at 1970 and commonly called Box-Jenkins method. This method can be split into AR, I and MA [11].

The AR model can be called autoregressive as this model re-models the variable's past value. This method is utilized to decide the value of the p order coefficient that indicates a value's dependence on the nearest previous value [10]

I is the acronym for Integrated that is utilized to get the time series stationary via the difference in observed at various times (d) [26].

MA (Moving Average) is a method to decide the value of the q coefficient that describes the previous change in the residual value [10].

In summary, ARIMA is a method of statistical perspective represented by three parameters, the first is AR data from past processes retrieved and stored and integrated processing (I) making the process of prediction easier [10].

The ARMA model is a union between the AR and MA models. The common formula of AR and MA models or ARIMA (p, 0, q) in equation 4 [10], [26].

A KPSS test is the inverse of the ADF test, which is a stationarity test that was first considered as a complement to the unit root test [21]. A KPSS test hypothesis is as follows [21], [22], [24]: Null hypothesis (H0): the time series hasn't unit root and is stationary; Alternative hypothesis (H1): the time series has unit root and is not stationary (has a time dependent structure).

For the non-stationary time series, then differencing is necessary to convert it to stationary [12], [23]. It is done by calculation of observation value changes or differences. Besides differencing for the ARIMA model, seasonal differencing is available for the SARIMA model. Meanwhile, the choice of S value is done by referring to table 1 [25]:

$$x_t = c + \sum_{i=1}^p \varphi_i x_{t-i} + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (4)$$

In this case, $\varphi_i \neq 0$, $\theta_i \neq 0$, and $\sigma_\epsilon^2 > 0$. Order p is referred to as AR while order q is referred to as MA. ARMA models can be applied when stationary time series, i.e. the mean and variance are more homogeneous. However, most real time series cannot meet this condition. The ARIMA model does not have such restrictions. For time series that are not stationary, the step of initial differencing is applied for making the time series obtained stationary thus ARMA model is applied [27] This model is presented in equations 5 and 6 [18].

$$y'_t = c + \sum_{i=1}^p \varphi_i y'_{t-i} + \epsilon_t + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (5)$$

$$y'_t = c + \varphi_1 y'_{t-1} + \dots + \varphi_p y'_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \quad (6)$$

For simplicity, complex models can be written in backshift notation as in equation 7 [18]:

$$(1 - \varphi_1 B - \dots - \varphi_p B^p) + (1 - B)^d y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) \epsilon_t \quad (7)$$

$$\begin{array}{ccc} \uparrow & & \uparrow \\ \text{AR}(p) & & \text{d differences} \end{array}$$

For handling seasonality, the process of ARIMA is generalized and SARIMA models is formulated. The SARIMA process was developed for modeling time series by short-term trends, seasonal patterns, and correlations [28].

This process extends standard model of Box and Jenkins also adds seasonal autoregressive factors and moving averages in the modeling process. SARIMA model combines the factors of non-seasonality and seasonality in a multiplicative model as in equation 8 [28].

$$\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^pY_t = \theta_0 + \theta(B)\Theta(B^S)\epsilon_t \quad (8)$$

where:

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \dots - \phi_p B^p \\ &= p \text{ order from AR non-seasonal model} \\ \theta(B) &= 1 + \theta_1 B + \dots + \theta_q B^q \\ &= q \text{ order from MA non-seasonal model} \\ \Phi(B^S) &= 1 - \Phi_1 B^S \dots - \Phi_P B^{PS} \\ &= P \text{ order from AR seasonal model} \\ \Theta(B^S) &= 1 - \Theta_1 B^S \dots - \Theta_Q B^{QS} \\ &= Q \text{ order from MA seasonal model} \\ (1-B)^d &= d \text{ order from non-seasonal differencing} \\ (1-B^S)^D &= D \text{ order from seasonal differencing} \\ \epsilon_t &= \text{error term} \\ B &= \text{backshift operator} \\ S &= \text{seasonal order} \end{aligned}$$

2.5 Evaluation

A forecast is never completely accurate. The difference between forecast and actual is called forecast error [9].

A commonly used method for measuring error in time series models is RMSE [26]. RMSE is utilized for evaluating prediction accuracy of a model. It calculates the difference or residual between the real and forecasted values. RMSE compares the forecast errors

of the various models for a specific dataset and not between the datasets. The formula for calculating RMSE is shown in equation 9 [26]

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2} \quad (9)$$

Here N represents total number of observations, x_i represents the actual value while $\hat{x}_i$ represents the forecast value.

RMSE values close to zero indicate good model predictability and reliability. The range of RMSE values for model evaluation can be seen in Table 2 [29]

Table 2. Ranges of RMSE for Models Performance

Ranges of RMSE	Performance
< 0.009	Excellent prediction accuracy
0.009 < RMSE < 0.09	Good prediction accuracy
0.09 RMSE < 0.5	Reasonable prediction
> 0.5	Inaccurate prediction

2.6 Prediction

Predictions are made to forecast crude oil prices for the next 7 days. This prediction is based on the daily dataset and using the model with the best RMSE value.

3. Results and Discussions

The data used for this research is daily historical crude oil price data starting from January 27, 2020 to January 25, 2023 with a total of 755 data. The dataset extracted from yahoo finance consists of Date, Open, High, Low, Close, Adj Close, and Volume as shown in Table 3.

Table 3. Crude Oil Price Dataset

Date	Open	High	Low	Close	Adj Close	Volume
2020-01-27 00:00:00-05:00	60.279999	60.490002	58.500000	59.320000	59.320000	39367
2020-01-28 00:00:00-05:00	58.950001	60.110001	58.500000	59.509998	59.509998	41537
2020-01-29 00:00:00-05:00	59.669998	60.610001	59.250000	59.810001	59.810001	21453
2020-01-30 00:00:00-05:00	59.610001	59.610001	57.730000	58.290001	58.290001	17435
2020-01-31 00:00:00-05:00	59.070000	59.529999	57.980000	58.160000	58.160000	578
...	...	...	...	...	...	...
2023-01-19 00:00:00-05:00	84.589996	86.849998	83.739998	86.160004	86.160004	28822
2023-01-20 00:00:00-05:00	86.360001	87.769997	85.580002	87.629997	87.629997	24374
2023-01-23 00:00:00-05:00	87.809998	89.099998	86.930000	88.190002	88.190002	21194
2023-01-24 00:00:00-05:00	88.029999	88.709999	85.809998	86.129997	86.129997	21194

This research only uses Close, because it represents all the action of the index during the day of trading. This means preprocessing is performed by removing unneeded data. The results of preprocessing are in Table 4. The results of the preprocessing data are then stationary test. The stationary test is done to get the value of d and as a prerequisite for implementing ARIMA model, the data being stationary. Stationary is needed to minimize model error.

The most basic method to detect stationarity is the graph test, then determine visually whether the graph results qualify as stationary or non-stationary data.

The results of graph test are in Figure 2.

Table 4. Data after Preprocessing

Date	Close
2020-01-27 00:00:00-05:00	59.320000
2020-01-28 00:00:00-05:00	59.509998
2020-01-29 00:00:00-05:00	59.810001
2020-01-30 00:00:00-05:00	58.290001
2020-01-31 00:00:00-05:00	58.160000
...	...
2023-01-19 00:00:00-05:00	86.160004
2023-01-20 00:00:00-05:00	87.629997
2023-01-23 00:00:00-05:00	88.190002
2023-01-24 00:00:00-05:00	86.129997

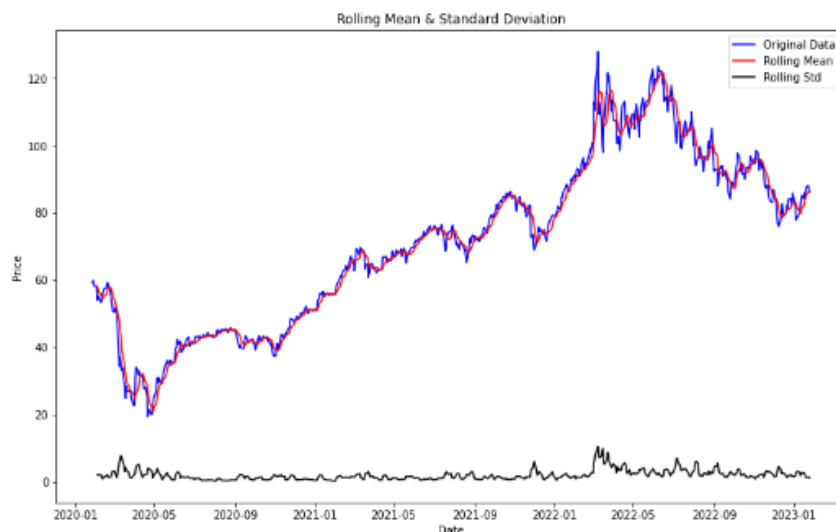


Figure 2. Results of Stationary Test Plot

Figure 2 shows a trend in the crude oil price dataset. Rolling Mean of the data varies with time and the Rolling Standard Deviation has several peaks - troughs that vary with time. The stationarity requirement is that μ (mean) is equal for all t and σ (variance) is equal for all t . Based on Figure 2, the data does not meet the stationarity requirements because the Rolling Mean and Standard Deviation vary and there is a trend. This research also uses ADF and KPSS tests to test for stationarity.

When a time series isn't stationary, it needs to be transformed to stationary or stationarity transformation. It is conducted by counting the difference or shift in the value of the observations. Counting changes or differences in observation values can also be done using the ADF and KPSS methods. The results of ADF and KPSS in Table 5.

Table 5. ADF and KPSS Test Results

	ADF	KPSS
p-value	0.758298	0.01
conclusion	non-stationary	not stationary
Recommended d value	1	1

Table 4 shows that the data is a non-stationary time series. Because the p value in ADF > 0.05 which means it fails to reject H_0 , meaning the data is non-stationary. Similarly, with the KPSS method, the p value > 0.05 . So that the p value fails to reject H_0 , meaning data are not stationary. According to recommendations of both ADF and KPSS tests, the d value for the ARIMA model is 1.

Transformed data to stationarity is checked again until the data has a stationary time series. The results are in Figure 3 and Table 6.

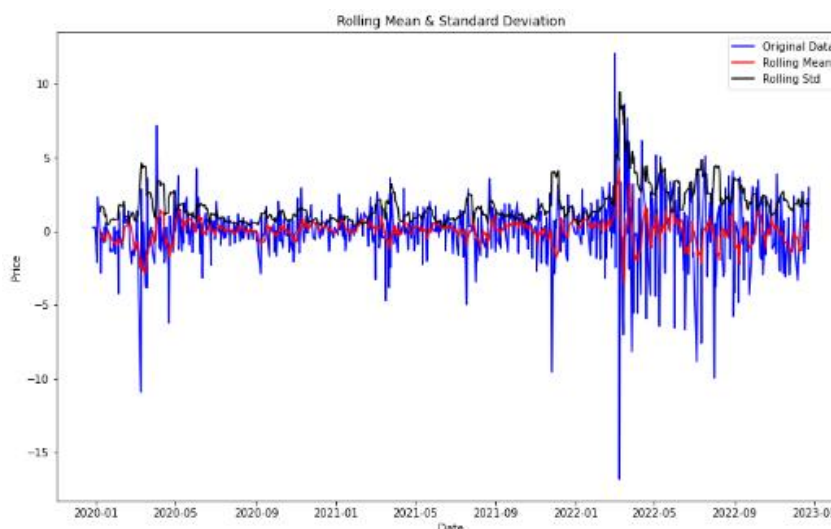


Figure 3. Results of Stationary Test Plot After Differencing

Figure 3 shows the Rolling Mean and Standard Deviation display a slight upward trend over time.

Table 6. ADF and KPSS Test Results After Differencing

	ADF	KPSS
p-value	0.01	0.1
conclusion	stationary	stationary
Recommended d value	0	0

Table 5 shows that the data is a stationary time series. This can be seen from the results of the p value on ADF ≤ 0.05 , thus rejecting H_0 and the data is stationary. Similarly, with the KPSS method, the value of $p < 0.05$, thus the value of p is rejecting H_0 and data is stationary.

Because data is stationary, thus d value that is to be implemented in the ARIMA model is 1.

However, for the SARIMA model, stationarity transformation is also done by finding the seasonal differencing value. The seasonal differencing value is done with the Canova Hansen and Osborn Chui Smith Birchenhall (OCSB) test methods, results of which can be seen in table 7.

Table 7. Recommended D values by Canova Hansen and OCSB

OCSB	nCH
0	0

Based on Table 6, the recommendations of the Canova Hansen test and OCSB test, the D value for the SARIMA model is 0.

The selection of p, q, P, and Q in this research is done using plots of ACF and PACF also auto arima. The plot results of ACF and PACF are shown in Figure 4.

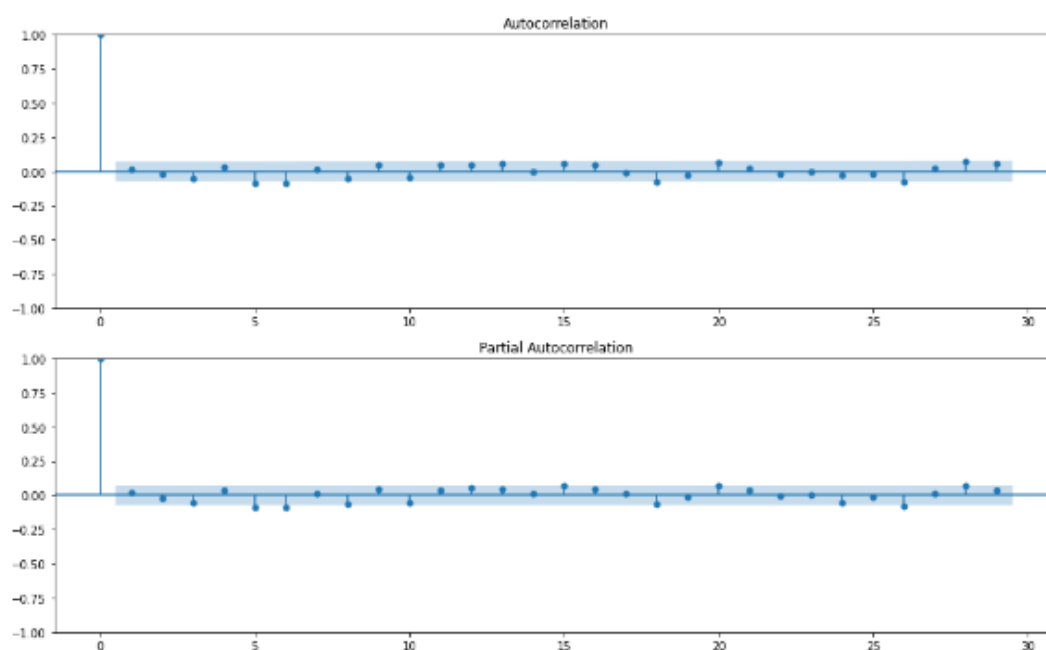


Figure 4. ACF and PACF Plot Results

The lollipop plot in Figure 4 is the result of ACF and PACF. The estimated AR value can be seen by plotting the PACF.

First, ignore the lag 0 value. Lag 0 would indicate a perfect correlation, as the plot will estimate the correlation between the today's value and self. The blue area on the plot represents the confidence interval. The estimated AR value that should be used is to count the number of "lollipops" that are at the top or bottom of a trust interval before the next lollipop enters the blue area.

Thus, by looking the PACF plot in Figure 4, it can be estimated that the AR value is 0. Similarly to the PACF function, estimating the MA value can be done by looking at the ACF plot.

The same logic is applied here, i.e. the number of lollipops that are either at the top or the bottom of the trust interval until the next lollipop enters the blue area.

Thus, by looking the PACF plot in Figure 4, it can be estimated that the MA value is 0.

Based on the results of the d value recommendation with ADF and KPSS as well as the results of the ACF and PCAF plots, an optimal ARIMA model is the ARIMA (0,1,0) model.

This study also uses auto_arima to validate that the optimal ARIMA model for this study is ARIMA (0,1,0). The results of finding the best parameters in the ARIMA model with auto arima in table 8.

Table 8. AIC values of ARIMA Models

Model	AIC
ARIMA (1,1,0)	3324.098
ARIMA (0,1,1)	3322.087
ARIMA (0,1,0)	3320.496
ARIMA (1,1,1)	3324.242

Based on Table 7, ARIMA (0,1,0) indicates that AIC value is least. So, the model is regarded as the best forecasting model.

The SARIMA model with the best parameters in this study can also be found by ACF and PACF plots and auto_arima.

The plot of ACF and PACF to search for the optimal P and Q for the SARIMA model is seen in Figure 5.

The plots of ACF and PACF are not only to find AR and MA, but also to find the values of P and Q of the SARIMA model. The predicted P value is plotted by the PACF plot while the Q value is plotted by the ACF plot. So, by looking at the PACF plot in Figure 5, it can be estimated that the value of P is 0. Meanwhile, by looking at the ACF plot, it can be estimated that the Q value is 0.

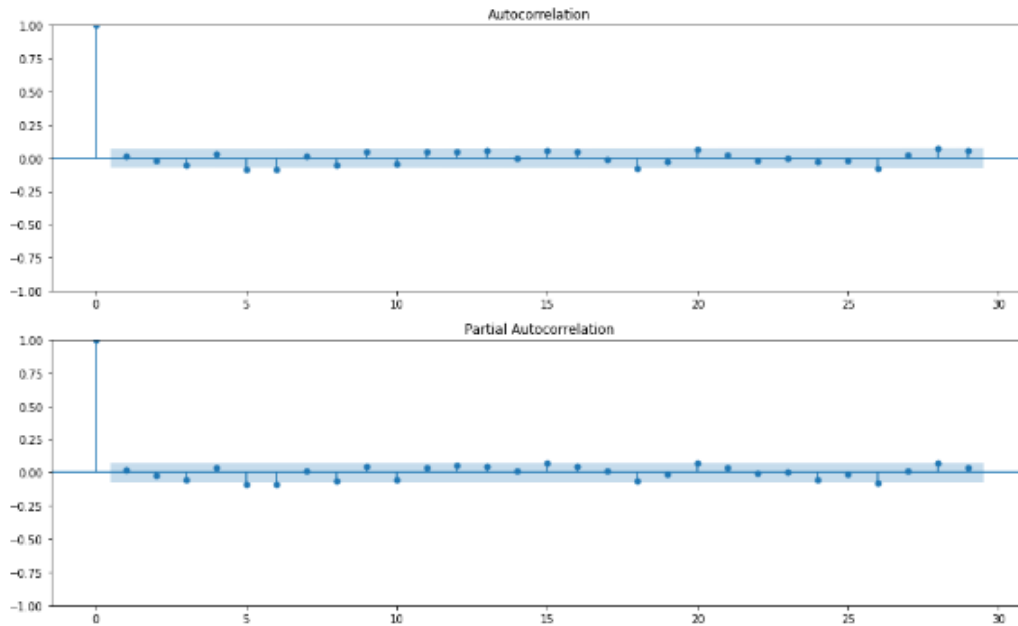


Figure 5. ACF and PACF for Seasonal Plot Results

According to the results of recommendations for the value of d with ADF and KPSS, the value of D with Canova Hansen and OCSB and the results of ACF and PCAF plots. The best SARIMA model is SARIMA (0,1,0) (0,0,0) [7]. As for the S parameter used, it depends on the data and the chosen period. Since this research data is daily data and the chosen time period is weekly, the S value used is 7.

Auto arima is also used to validate that the best SARIMA model for this study is SARIMA (0,1,0) (0,0,0) [7] model.

The results of finding the optimal parameters for the SARIMA model with auto_arima are in Table 9.

From Table 8, SARIMA (0,1,0) (0,0,0) [7] indicates that AIC value is least. So, the model is regarded as the best forecasting model.

After selecting the parameters, the next step is modeling. The modeling results with ARIMA (0,1,0) and SARIMA (0,1,0) (0,0,0) [7] are shown in Figure 6.

Table 9. AIC values of SARIMA Models

Model	AIC
SARIMA (0,1,0) (0,0,0) [7]	3320.500
SARIMA (1,1,0) (1,0,0) [7]	3325.986
SARIMA (0,1,1) (0,0,1) [7]	3325.973
SARIMA (0,1,0) (1,0,0) [7]	3324.221
SARIMA (0,1,0) (0,0,1) [7]	3324.221
SARIMA (0,1,0) (1,0,0) [7]	3326.224
SARIMA (1,1,0) (0,0,0) [7]	3324.101
SARIMA (0,1,1) (0,0,0) [7]	3324.091
SARIMA (1,1,1) (0,0,0) [7]	3325.187

and 7. The RMSE value obtained from this model is 1.905.



Figure 6. Model ARIMA Model Results

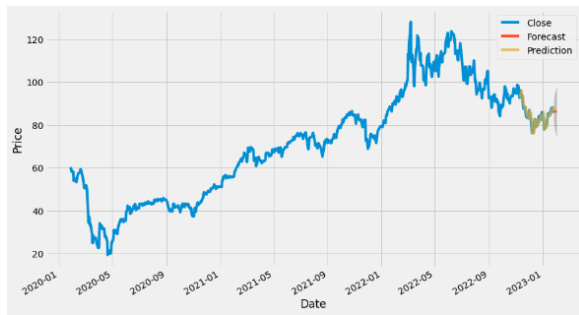


Figure 7. Model SARIMA Model Results

The best model results are then implemented to predict the next 7 days. Prediction can be seen in Table 10.

Table 10. Prediction Results with ARIMA and SARIMA

Date	ARIMA	SARIMA
2023-01-26	86.230003	86.260002
2023-01-27	86.230003	86.260002
2023-01-28	86.230003	86.260002
2023-01-29	86.230003	86.260002
2023-01-30	86.230003	86.260002
2023-01-31	86.230003	86.260002
2023-02-01	86.230003	86.260002

4. Conclusion

The research compared 2 models for crude oil price forecasting, i.e. ARIMA and SARIMA. The research results in best models based on optimal parameter value selection for ARIMA, i.e. ARIMA (0,1,0) while SARIMA is SARIMA (0,1,0) (0,0,0) [7]. Both models have the same evaluation value results of 0.01905% RMSE, it means the model has excellent prediction accuracy. The prediction results for the next 7 days with ARIMA and SARIMA are 86.230003 and 86.260002.

References

- [1] U. Sartika, "Pengaruh inflasi, tingkat suku bunga, kurs, harga minyak dunia dan harga emas dunia terhadap IHSG dan JII di bursa efek Indonesia," *Balance: Jurnal Akuntansi Dan Bisnis*, vol. 2, no. 2, pp. 285–294, 2017.
- [2] S. Sultan, A. M. Soesilo, S. A. T. Rahayu, L. Hakim, and T. Mulyaningsih, "Siklus Bisnis Perekonomian Indonesia (Studi Empiris Harga Minyak Dan Pertumbuhan Ekonomi Periode 1970-2019)," *Buletin Ekonomi: Manajemen, Ekonomi Pembangunan, Akuntansi*, vol. 19, no. 2, pp. 87–120, 2022.
- [3] B. Agus Setiono, "Fluktuasi Harga Minyak dan Pengaruhnya bagi Ekonomi Indonesia," 2014.
- [4] Y. Nan, R. Sun, and N. Yu, "A study on the correlation between international crude oil price fluctuation cycle measure and economic cycle fluctuation," *Available at SSRN 4003148*.
- [5] R. Maulid, "Penerapan Contoh Machine Learning dalam Memprediksi Harga Saham," *Dqlab*. 2021. [Online]. Available: <https://dqlab.id/penerapan-contoh-machine-learning-dalam-memprediksi-harga-saham>
- [6] G. D. Nursyafitri, "4 Contoh Penerapan Machine Learning di Kehidupan Sehari-Hari," *Dqlab*. 2022. [Online]. Available: dqlab.id/4-contoh-penerapan-machine-learning-di-kehidupan-sehari-hari#:~:text=Machine%20Learning%20ini%20memang%20merupakan,ini%20dinamakan%20dengan%20proses%20trainin g
- [7] A. H. Hutasuhut, W. Anggraeni, and R. Tyasnurita, "Pembuatan aplikasi pendukung keputusan untuk peramalan persediaan bahan baku produksi plastik blowing dan inject menggunakan metode ARIMA (Autoregressive Integrated Moving Average) di CV. Asia," *Jurnal Teknik ITS*, vol. 3, no. 2, pp. A169–A174, 2014.
- [8] D. K. Silalahi, "Forecasting of Poverty Data Using Seasonal ARIMA Modeling in West Java Province," *JTAM (Jurnal Teori dan Aplikasi Matematika)*, vol. 4, no. 1, pp. 76–86, 2020.
- [9] B. W. Taylor and R. S. Russell, "Operations management: creating value along the supply chain," 2011.
- [10] I. M. Wirawan, T. Widiyaningtyas, and M. M. Hasan, "Short term prediction on bitcoin price using ARIMA method," in *2019 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2019, pp. 260–265.
- [11] E. Lansiaux, N. Tchagaspian, and J. Forget, "Community impact on a cryptocurrency: Twitter comparison example between Dogecoin and Litecoin," *Frontiers in Blockchain*, vol. 5, p. 829865, 2022.
- [12] S. Mohan, A. K. Solanki, H. K. Taluja, A. Singh, and others, "Predicting the impact of the third wave of COVID-19 in India using hybrid statistical machine learning models: A time series forecasting and sentiment analysis approach," *Comput Biol Med*, vol. 144, p. 105354, 2022.
- [13] B. Agarwal, P. Harjule, L. Chouhan, U. Saraswat, H. Airan, and P. Agarwal, "Prediction of dogecoin price using deep learning and social media trends," *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 8, no. 29, pp. e2–e2, 2021.
- [14] R. Aroussi, "yfinance: Yahoo! Finance market data downloader," *PyPI*. [Online]. Available: <https://pypi.org/project/yfinance/>
- [15] A. A. Adebiyi, O. A. Adewumi, and C. K. Ayo, "Stock price prediction using the ARIMA model," in *International Conference on Computer Modelling and Simulation*, 2014.
- [16] J. Singh and S. Shyamsundar, "AIOPS Framework For Alerting Performance Issues In Microservices Using Time Series Forecasting".
- [17] T. G. Smith, "Pyramid: ARIMA estimators for Python - pyramid 0.9.0 documentation," *Rob J Hyndman*. 2017. [Online]. Available: <https://alkaline-ml.com/pmdarima/0.9.0/index.html>
- [18] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.
- [19] W. Y. Rusyida and V. Y. Pratama, "Prediksi harga saham Garuda Indonesia di tengah pandemi COVID-19 menggunakan metode ARIMA," *Square: Journal of Mathematics and Mathematics Education*, vol. 2, no. 1, pp. 73–81, 2020.
- [20] A. L. Moreno, "Stationary Time Series," *CRAN*. [Online]. Available: <https://cran.r-project.org/web/packages/TSTutorial/vignettes/Stationary.pdf>
- [21] L. Sjösten, "A Comparative Study of the KPSS and ADF Tests in terms of Size and Power." 2022.
- [22] D. Fedorová and others, "Selection of unit root test on the basis of length of the time series and value of ar (1) parameter," *Statistika*, vol. 96, no. 3, p. 3, 2016.
- [23] A. Sudumbrekar, R. Kale, T. Kaurwar, V. Mule, and A. Devkar, "Feasibility Study of ARIMA Model for PM2. 5 Prediction using Real-world Data Gathered from Pune Region".
- [24] S. Oikonomopoulos, "Cryptocurrency Price Prediction using Social Media Sentiment Analysis," 2022.
- [25] D. Ladiray and G. Mazzi, "X11-like seasonal adjustment of daily data," in *2nd International Conference on Advanced Research Methods and Analytics*, 2018.
- [26] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A comparison of ARIMA and LSTM in forecasting time series," in *2018 17th IEEE international conference on machine learning and applications (ICMLA)*, 2018, pp. 1394–1401.
- [27] F. Wang, Y. Zou, H. Zhang, and H. Shi, "House price prediction approach based on deep learning and ARIMA

- model,” in *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, 2019, pp. 303–307.
- [28] T. Baldigara and M. Mamula, “Modelling international tourism demand using seasonal ARIMA models,” *Tourism and hospitality management*, vol. 21, no. 1, pp. 19–31, 2015.
- [29] E. O. Oke *et al.*, “Phenolic compound extraction from Nigerian *Azadirachta indica* leaves: response surface and neuro-fuzzy modelling performance evaluation with Cuckoo search multi-objective optimization,” *Results in Engineering*, vol. 8, p. 100160, 2020.