ENHANCED KNOWLEDGE IN SCIENCES AND TECHNOLOGY
e-ISSN: 2773-6385

EKST

Vol. 4 No. 2 (2024) 273-282
https://publisher.uthm.edu.my/periodicals/index.php/ekst

Forecasting Oil Price in Malaysia using ARIMA and Artificial Neural Network

Nur Farhana Zamri[1], Maria Elena Nor[1]*

[1] *Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hab Pendidikan Tinggi Pagoh, KM 1, Jalan Panchor, 84600 Pagoh, Muar, Johor, MALAYSIA*

*Corresponding Author: maria@uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2024.04.02.031

Article Info

Received: 27 December 2023
Accepted: 11 January 2024
Available online:12 December 2024

Keywords

Forecasting Oil Price, ARIMA, Artificial Neural Network

Abstract

The importance and difficulty of accurately predicting future oil prices stem from the fact that these prices impact so many different economic and non-economic variables. Numerous factors, such as economic growth, political events, and individual ambitions, contribute to the high degree of unpredictability inherent in oil price forecasts. This aim for this study was to forecast the oil price in Malaysia using the ARIMA and ANN models and identify the model that would be most suitable to forecast the future oil price In Malaysia. This study forecasts oil prices from September 2003 to June 2023. The ARIMA (1,1,0) model was determined to be the best model based on its lowest p-value and MSE (Mean Square Error). Moreover, the forecasting performance of ARIMA and ANN models have been compared by three prediction accuracy metrics which are Mean Absolute Error (MAE), the Mean Absolute Percentage Error (MAPE), and the Root Mean Square Error (RMSE). The results show that the ANN models have a better forecast performance compare with ARIMA (1,1,0) due to its lower MAE, MAPE and RMSE which are 36.30, 14.23% and 48.81 respectively. As a result of their higher precision in forecasting, Artificial Neural Networks (ANN) are more suitable for their ability to anticipate the price of energy in Malaysia. Consequently, ANN were utilised in order to make predictions for the subsequent six months, which span from July 2023 to December 2023.

1. Introduction

The viscous liquid known as crude oil is extracted from oil reservoirs thousands of meters below the earth's surface. Nothing has been done to improve it. For nations that use oil as a primary economic driver, crude oil is an essential commodity. Consequently, accurate forecasts of real oil prices are an essential part of the field of economic and financial research [1]. In response to the climate disaster, there is a concerted effort underway around the world to progressively phase out fossil fuels. Global politics and international relations revolve around petroleum since it is crucial to industry, civilization, and a significant amount of the world's energy usage [2]. Due to the fact that it is responsible for approximately one-third of the total energy consumption in the world, crude oil has a significant influence on the economy of the entire world. The economies of nations that both produce and consume oil are quite sensitive to swings in oil prices [3].

Policymakers could make better energy judgements and policies if they had accurate oil price projections. It would benefit the nation. Predicting crude oil prices is difficult in predicting research since several factors affect them. Oil prices are affected by many factors. Oil prices are affected by economic growth, disputes, wars, and news. These elements are in addition to supply, demand, and inventory [4]. Price changes in a vital energy supplier can affect economic development, financial markets, and national security. A significant amount of attention has been drawn to the process of projecting the price of oil as a consequence of this. Oil prices are primarily determined by supply and demand, however, which are the two fundamental factors. Not only that, but non-fundamental factors such as international politics, huge power games, and speculative trading also play an essential role in the market [5].

Oil prices, and energy costs more generally, affect the global income distribution, inflation rates, economic growth rates, and balance of payments for all nations. This is due to the fact that crude oil and its byproducts are indispensable to a wide range of businesses. Among these, the oil price is one of the most important determinants of Egypt's capacity for sustainable development [6]. In both the short and long run, changes in the value of the currency and the cost of crude oil can impact the rate of economic growth, according to research [7].

Investment, agriculture, and industrialization can help the government diversify its revenue streams and reduce its dependence on crude oil and its income volatility caused by crude oil price swings. According to studies [8], global demand shocks produce 35% of oil price volatility, whereas oil supply shocks contribute 50%. Oil price drops caused by supply shocks boost economic activity in industrialised nations but slow it in emerging nations.

The objective of this study is to make a prediction about the price of oil in Malaysia by employing the ARIMA and Artificial Neural Network approaches. In the subsequent step, the accuracy of a number of different forecasting models have been evaluated and compared through the three different parameters which are RMSE, MAPE, and MAE that are commonly used in statistical analysis. The final stage is to apply the method that has proven to be the most accurate in predicting the future price of oil in Malaysia to the time period that ranges from July 2023 to December 2023.

2. Methodology

The information regarding the price of oil was retrieved from the IndexMundi website (https://www.indexmundi.com/commodities/?commodity=crudeoil&months=120&currency=myr). The data for each month were collected starting in September 2003 and continuing through June 2023. ARIMA and ANN are being utilised to provide a forecast for the price of oil. The data have been partitioned into two distinct sets which are the training set and the test set. There are 45 data in the testing set from October 2019 to June 2023, and 193 data in the training set from September 2003 to September 2019. Following data collection, Minitab and Python statistical software were used to analyse the data.

2.1 ARIMA Model

The Box-Jenkins model, a version of the autoregressive integrated moving average (ARIMA) model, shows the relative importance of one dependant variable to other dynamic variables. The Box-Jenkins approach involves model identification, parameter estimations, and diagnosis checking. According to [9], the Box-Jenkins technique begins with a data visualisation and stationary variance testing. The Box-Cox Transformation should be used to convert data with inconsistent variance. If data mean is not stationary, differencing is needed.

After that, the ACF and PACF plots will decide the model order, AR(p) and MA(q) values, as shown in Table 1. The fifth stage determines preliminary models using estimated parameter values. The model is most accurate if the Ljung-Box test p-value is bigger than the parameter p-value. The Ljung-Box test act as a tool to test the lack of fit of time series model. The hypothesis testing for Ljung-Box test is:

H_o : The model does not exhibit lack of fit (the model is adequate)

H_1 : The model exhibit lack of fit (the model is inadequate)

Table 1 Model Identification

ACF	PACF	Model
Decay to zero with exponential pattern	Cuts off after lag q	AR(p)
Cuts off after lag q	Decay to zero with exponential pattern	MA(q)
Decay to zero with exponential pattern Cuts off after lag q	Decay to zero with exponential pattern Cuts off after lag q	ARMA(p,q) AR(p) or MA (q)

After selecting a model, it matters to determine whether it is suitable. The selected model of residuals can be examined using this method by plotting the residuals of autocorrelation function (ACF). It is necessary to make adjustments to the models if they do not behave like white noise. It is possible to use the model for prediction if the residuals show properties comparable to white noise.

2.2 Artificial Neural Network

Neural network has three layers which are input, hidden, and output. The amount of features in the data collection affects the input layer's dimensionality or nodes. Through synaptic connections with the nodes that were formed in the subterranean layer, the nodes are linked. At the input layer, individual synaptic linkages are assigned specific weights. The weights function as a discriminating element that establishes the permissibility or prohibition of particular signals or inputs. Additionally, the weights serve to denote the extent or magnitude of impact on the concealed stratum. In order to acquire knowledge, a neural network modifies the weight assigned to each component iteratively.

The LSTM feature was employed in this study. Cells, which are the building blocks of an LSTM, are responsible for receiving and storing data streams. The top line of every cell forms a structure that links modules together, much like a conveyor belt. It collects data from one module and transfers it to another. The use of gates in each cell enables the removal, filtering, or addition of data to subsequent cells. Therefore, the cells are able to determine if data is acceptable or not through the gates, which employ a sigmoidal neural network layer [10]. The Long Short-Term Memory (LSTM) theory is illustrated in Figure 1.

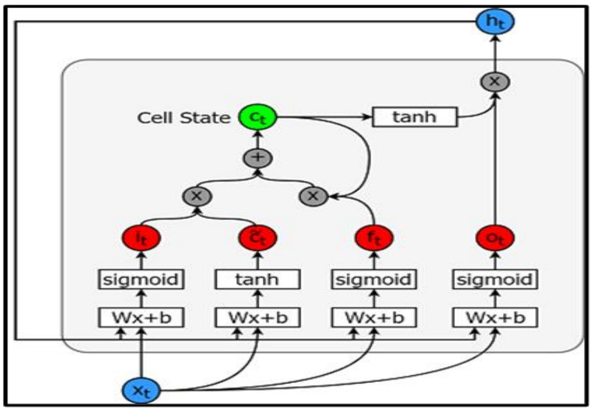

Fig. 1 Long Short Term Memory (LSTM) model

2.3 Forecast Accuracy Measures

To determine the most accurate forecasting model from five techniques, MAE, RMSE, and MAPE were employed. As these three accuracy criteria decrease, forecasting methods are regarded more accurate [11]. Eq. 1, Eq. 2, and Eq. 3 each show the results of the calculations for the MAE, MAPE, and RMSE, respectively.

$$\text{MAE} = \frac{\sum_{t=1}^{n}|y_t - \hat{y}_t|}{n} \tag{1}$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100\% \tag{2}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}} \tag{3}$$

3. Result and Discussion

Fig. 2 shows a time series plot of the oil price in Malaysia from September 2003 to June 2023. The plot includes a total of 238 data points. The graph clearly shows the data has a discernible trend but lacks any seasonal patterns.

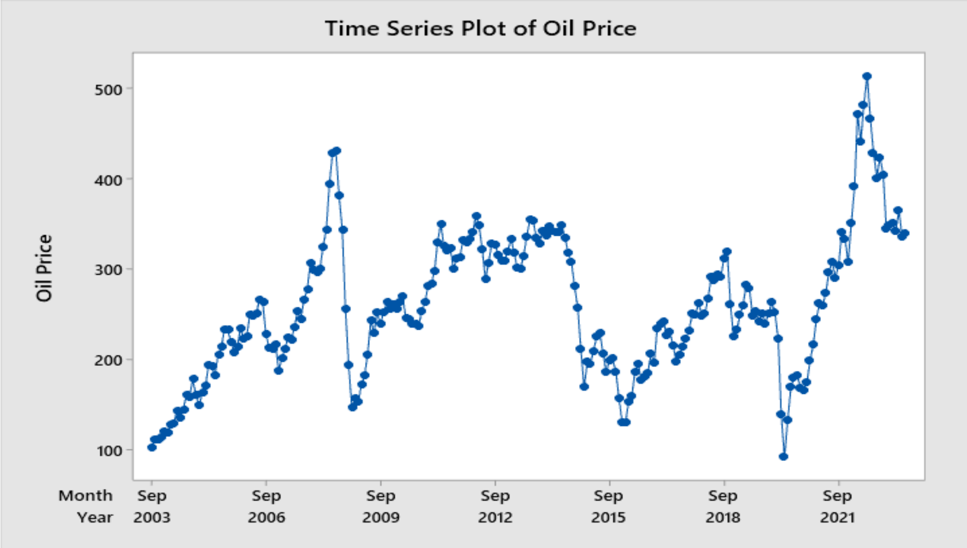

Fig. 2 Time Series of Oil Price in Malaysia from September 2003 to June 2023

3.1 ARIMA Model

Fig. 3 shows time series plot of oil price from September 2003 to September 2019. The figure shows a decrease in trend that is not influenced by seasonality, spanning the period from July 2008 to September 2009.

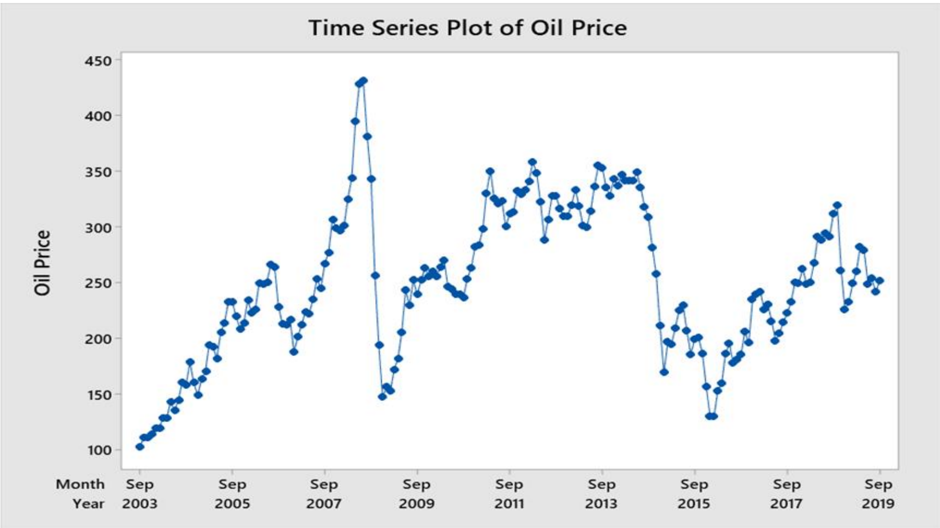

Fig. 3 Time Series of Oil Price in Malaysia from September 2003 to September 2019

Fig. 4 shows that the Box-Cox plot for Malaysian oil prices has a lower control limit of 0.14 and an upper control limit of 1.07. No additional transformation is required since the interval between the lower and upper control limits includes the value of 1, indicating that the plot is stable in variance.

Enhanced Knowledge in Sciences and Technology Vol. 4 No. 2 (2024) p. 273-282

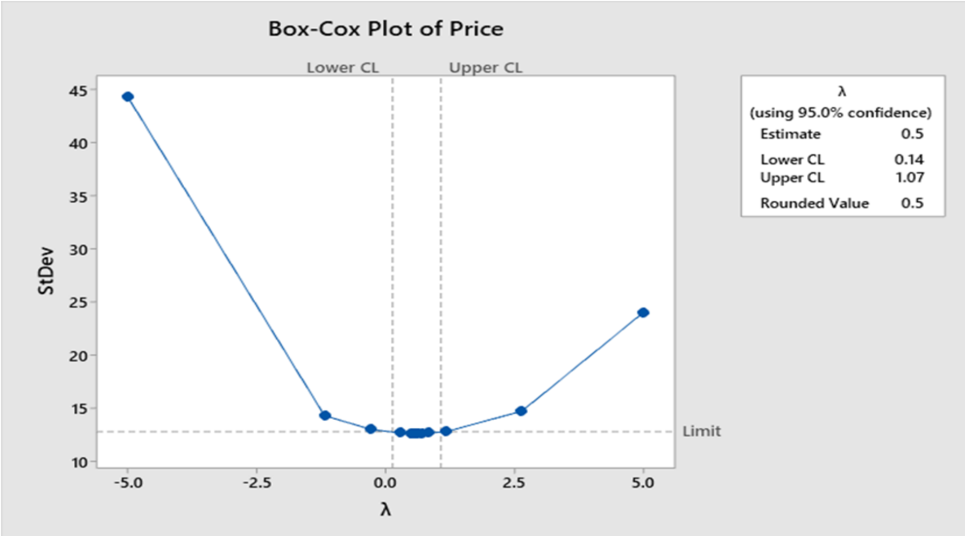

Fig. 4 Box-Cox Plot of Oil Price in Malaysia

Fig. 5 and Fig. 6 show the differenced data ACF and PACF plots. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) graphs can reveal the ARIMA model. The ACF and PACF plots reveal a rise that peaks at lag 1. Thus, ARIMA (1,1,0), (1,1,1), and (0,1,1) models were chosen. Table 2 compares ARIMA model parameters with Ljung-Box tests.

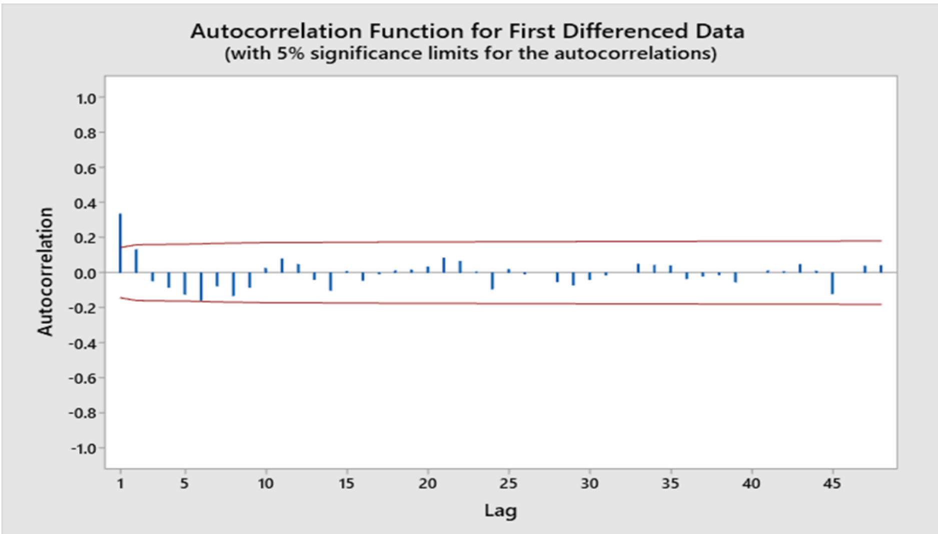

Fig. 5 The Autocorrelation Function for the first differenced data

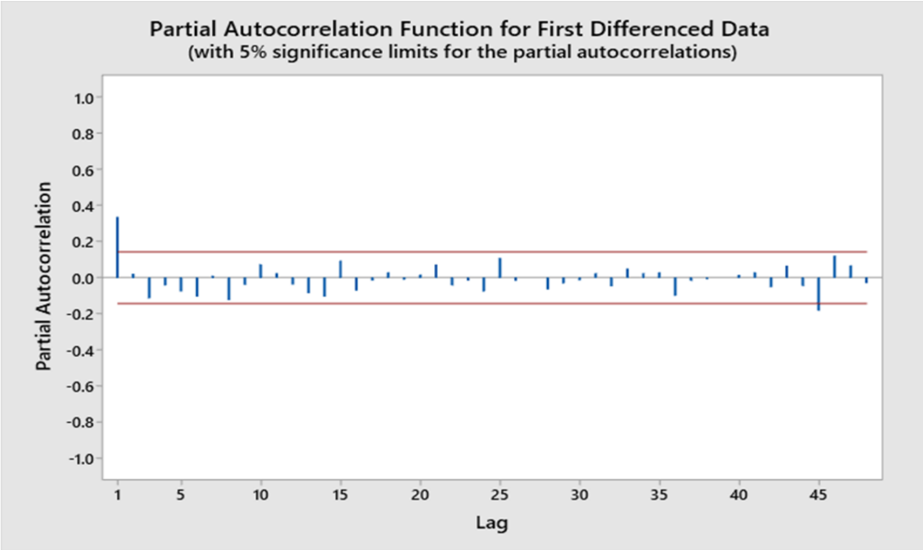

Fig. 6 The Partial Autocorrelation Function for the first differenced data

Table 2 ARIMA Model Parameters and Ljung-Box Chi-Square Statistics

Model	p-value	Ljung-Box				MSE
		Lags 12	Lags 24	Lags 36	Lags 48	
ARIMA (1,1,0)	AR (1) : 0.00	**0.359**	**0.520**	**0.858**	**0.865**	0.336116
ARIMA (1,1,1)	AR (1) : 0.069 MA (1) : 0.868	0.613	0.228	0.394	0.539	0.337830
ARIMA (0,1,1)	MA (1) : 0.00	0.101	0.287	0.682	0.789	0.342901

Since AR (1) and MA (1) have p-values over 0.05 for the ARIMA (1, 1, 1) model, there is no statistical significance. Thus, the model must be rejected. AR (1) and MA (1) for ARIMA (1,1,0) have p-values less than 0.05, and the Ljung-Box test lags are all bigger than 0.05. Results show both models are valid and significant. Table 2 shows that model (1,1,0) has a smaller Mean Squared Error (MSE) than model (0,1,1). Thus, the optimal model is ARIMA (1,1,0). Fig. 7 shows the time series plot of residuals. The residuals time series plan makes it clear that the series does not show any trend. So, this means that the variance of residual plot remains constant.

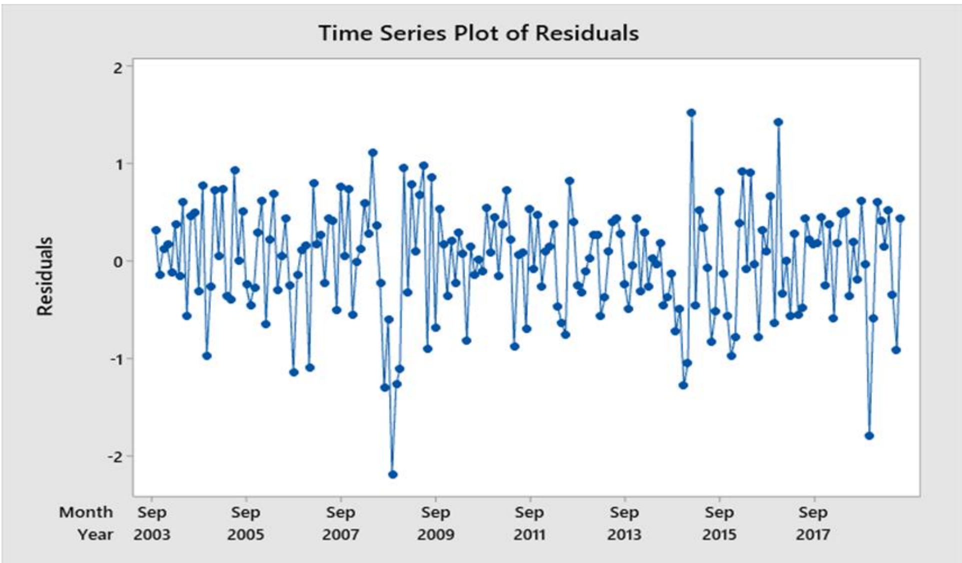

Fig. 7 Time series plot of residuals

Fig. 8 and Fig. 9 show that the Autocorrelation Function and the Partial Autocorrelation Function of the oil price residuals are both within the 95% confidence range. Because of this, we can conclude that the residuals are independent and that there is no substantial association. This means that the model can be used to make predictions.

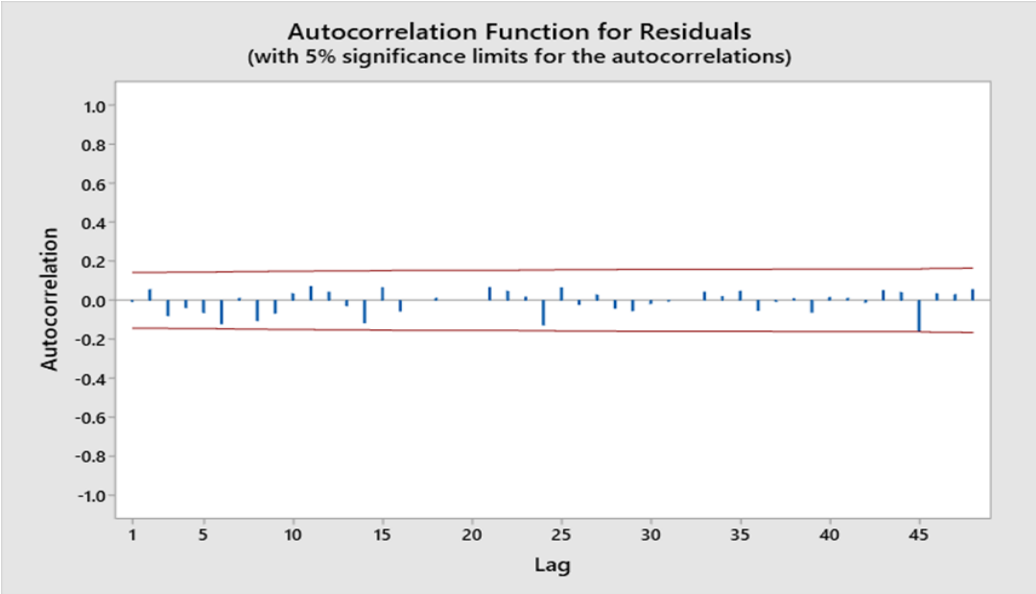

Fig. 8 The Autocorrelation Function (ACF) of Residuals for oil price

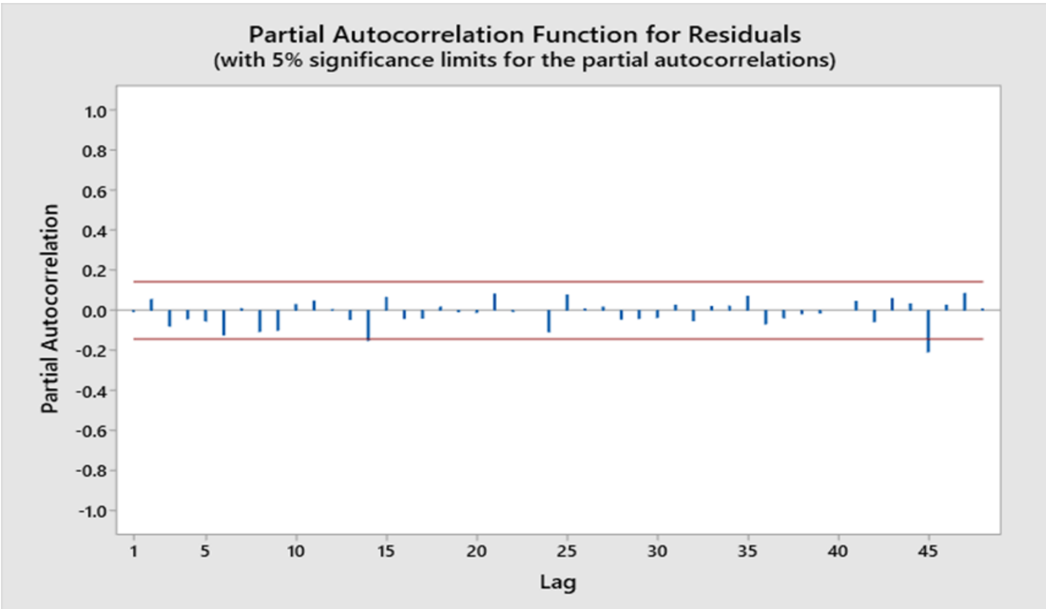

Fig. 9 The Partial Autocorrelation Function (ACF) of Residuals for oil price

3.2 Artificial Neural Network

This project utilised TensorFlow's Keras API to create a simple LSTM neural network. Python, TensorFlow, and Keras were needed to run the algorithm. The Keras package created the LSTM model. Then NumPy, matplotlib, pylot, and panda were called. The algorithm divides the data into 80% for training and 20% for testing to match the ARIMA approach and make comparisons fairer. Scaling the dataset with Min-Max ensures that all numbers are between 0 and 1. How crucial this component is affects neural network learning.

Models were created using the "Adam" algorithm and mean squared error loss function. After making scaled predictions on the test data with the trained model, apply an inverse transformation to get their original price values. The algorithm's "fit lstm" function trains and builds the LSTM model. The function needs the training dataset, epochs, and neurons. The distributed LSTM training system will run 100 training epochs.

Table 3 : Model Summary

Layer (type)	Output Shape	Parameters
Lstm (LSTM)	(None, 50)	10400
Dense (Dense)	(None, 1)	51

A sequential model that is comprised of two different kinds of layers is shown in Table 3. In the first layer, there is an LSTM layer that has an output shape of (None, 50). This indicates that it generates an output of shape (None, 50) after processing using the LSTM layer with 50 units. The following layer is a Dense Layer, and its output shape is (None, 1), which indicates that the output form that comes after the Dense layer will have one unit. There are a total of 10,451 parameters in the model, which includes an LSTM layer that has 10,400 parameters and a Dense layer that has 51 parameters.

3.3 Comparison of Forecasting Performance

The comparison between the predicted values and the actual values of the ARIMA (1,1,0) and ANN is depicted in Fig. 10 and Table 4. The neural network closely approximates the original data, as shown in the graph, and the MAE, MAPE, and RMSE values are also lower than ARIMA (1,1,0). Overall, the ANN is the most accurate model for forecasting energy prices in Malaysia.

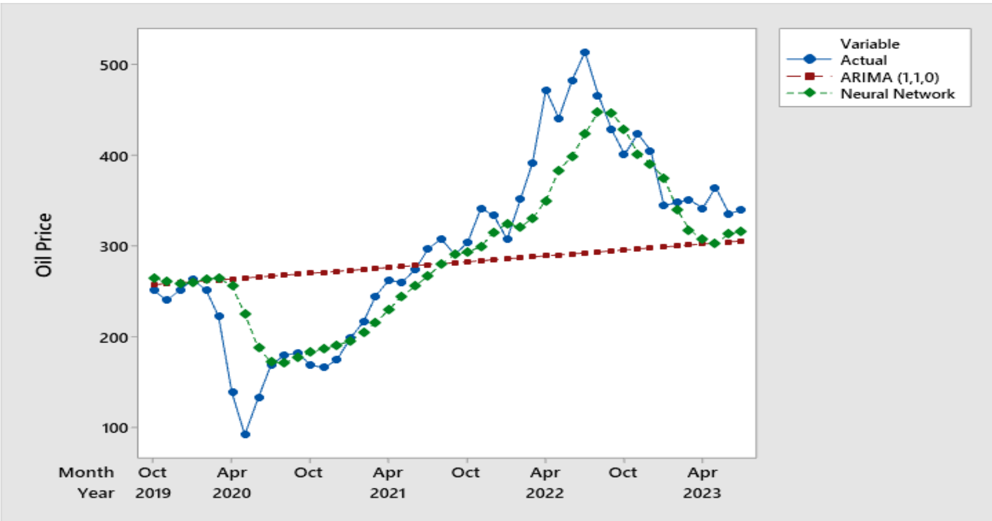

Fig. 10 Time Series Plot of actual and forecast value of Oil Price In Malaysia

Table 4 Accuracy measurement between ARIMA and Neural Network

Method	MAE	MAPE	RMSE
ARIMA (1,1,0)	72.03	28.79 %	48.83
Artificial Neural Network	36.30	14.23 %	48.81

3.4 Forecast the future oil price in Malaysia using the best method

Fig. 11 and Table 5 show the forecast value of oil price in Malaysia from July 2023 to December 2023 by using ANN. It is estimated that the oil price in Malaysia from July 2023 to December 2023 will have a decrease from 319.03 to 278.44

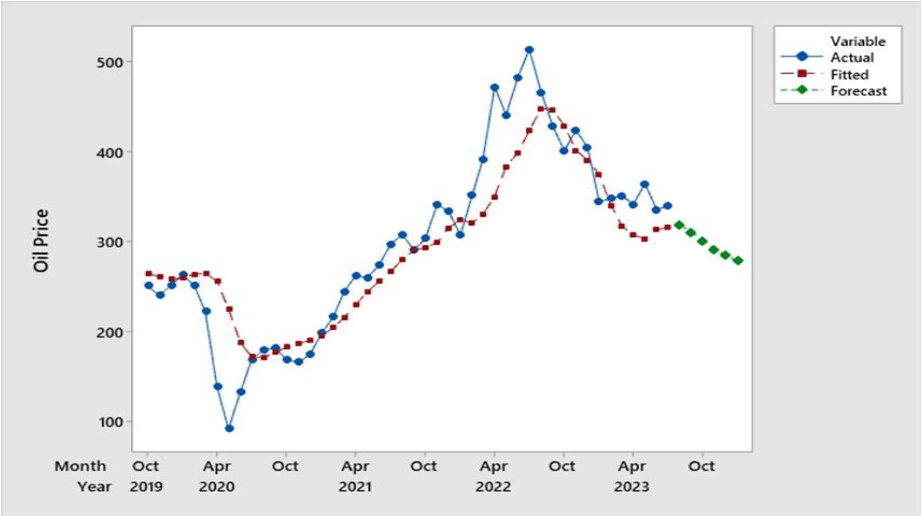

Fig. 11 Time series plot of oil price in Malaysia with forecast values from July 2023 to December 2023

Table 5 : Forecast value of oil price in Malaysia from July 2023 to December 2023 by using neural network

Month	Forecast Value
July	319.03
August	310.38
September	300.81
October	291.37
November	284.53
December	278.44

4. Conclusion

The Malaysian oil price was predicted using an ARIMA model and ANN in this study and found that ARIMA (1,1,0) was the best model due to p-value that is smaller than alpha which is 0.05 and Ljung-Box test value that are greater than 0.05. It was also the model with the smallest mean squared error (MSE) when compared to the ARIMA (1,1,1) and ARIMA (0,1,1) models. Subsequently, the LSTM algorithm was utilised for the ANN to forecast with a total of 100 epochs and 32 batch size.

The mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) are the three measurements that will be used in this investigation. To achieve the second objective of this study, the accuracy of the forecasts obtained by different forecasting models will be analysed and compared. When compared to the ARIMA (1,1,0) approach, ANN obtains lower values for MAE, MAPE, and RMSE which are 36.30, 14.23%, and 48.81, respectively.

The final objective of this study, which was to anticipate the future price of oil in Malaysia using the best method, was effectively accomplished. ANN were used in this study to forecast the next 6 months ahead because of their higher performance in compared with ARIMA (1,1,0), which will span from July 2023 to December 2023.

Finally, it must be said that predicting oil prices is no easy feat, what with all the moving parts and inherent unpredictability in the market. In order to achieve a comprehensive understanding of oil price fluctuations, it is recommended to utilize a combination of machine learning and statistical models such as GARCH and ARIMA in conjunction with neural networks, random forests, or gradient boosting. It is common for more accurate predictions to be generated via ensemble approaches or hybrid models that combine the best features of many procedures. In addition, to make educated decisions, one must be aware of the limitations and unpredictability of any forecasting method.

Acknowledgement

The authors would thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia for its support.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Nur Farhana Zamri, Maria Elena Binti Nor;* ***data collection:*** *Nur Farhana Zamri, Maria Elena Binti Nor;* ***analysis and interpretation of results:*** *Nur Farhana Zamri, Maria Elena Binti Nor;* ***draft manuscript preparation:*** *Nur Farhana Zamri, Maria Elena Binti Nor. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Yi, Y., Ma, F., Zhang, Y., & Huang, D. (2018). Forecasting the prices of crude oil using the predictor, economic and combined constraints. Economic Modelling, 75, 237–245. https://doi.org/10.1016/j.econmod.2018.06.020

[2] Gupta, N., & Nigam, S. (2020b). Crude Oil Price Prediction using Artificial Neural Network. Procedia Computer Science, 170, 642–647. https://doi.org/10.1016/j.procs.2020.03.136

[3] Abdollahi, H. (2020). A novel hybrid model for forecasting crude oil price based on time series decomposition. Applied Energy, 267, 115035. https://doi.org/10.1016/j.apenergy.2020.115035

[4] Wu, B., Wang, L., Lv, S., & Zeng, Y. (2021). Effective crude oil price forecasting using new text-based and big-data-driven model. Measurement, 168, 108468. https://doi.org/10.1016/j.measurement.2020.108468

[5] Vo, D. H., Vu, T. N., Vo, A., & McAleer, M. (2019). Modeling the Relationship between Crude Oil and Agricultural Commodity Prices. Energies, 12(7), 1344. https://doi.org/10.3390/en12071344

[6] Mahmoud, I. M. (2022). Energy prices and their importance in the economy as applied to Egypt. Maǧallaẗ Kulliyyaẗ Al-maʻārif Al-ǧāmiʻaẗ, 33(3), 198–219. https://doi.org/10.51345/.v33i3.507.g290

[7] Musa, K. S., Maijama'a, R., Shaibu, H. U., & Muhammad, A. (2019). Crude Oil Price and Exchange Rate on Economic Growth: ARDL Approach. OAlib, 06(12), 1–5. https://doi.org/10.4236/oalib.1105930

[8] Caldara, D., Cavallo, M., & Iacoviello, M. (2016). Oil price elasticities and oil price fluctuations. International Finance Discussion Papers, 2016(1173), 1–59. https://doi.org/10.17016/ifdp.2016.1173

[9] Hyndman, R.J. & Athanasopoulus, G. (2014). Forecasting principles and practice. http://otexts .com/fpp

[10] Gibson, A., & Patterson, J. (2015). Deep learning: DL4J and Beyond. O'Reilly Media.

[11] Sahu, P.K., & Kumar, R (2014). The Evaluation of Forecasting Methods for Sales of Sterilized Flavoured Milk in Chhattisgarh. International Journal of Engineering Trends and Technology (IJETT) ,8(2), 98-104