Comparative analysis of AI models for crude oil price forecasting: ARIMA, SARIMAX, and LSTM

Faizah ALSHAMMARI[1], Nahla ALJOJO[2]*, Modhi ALSHAMMARI[3], Amnah ALSAYED[1], Mona ALMALKI[1]

[1] Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia

falshammari.stu@uj.edu.sa, aalsayed0097.stu@uj.edu.sa, malmalki0703.stu@uj.edu.sa

[2] College of Computer Science and Engineering, Information System and Technology Department, University of Jeddah, Jeddah, Saudi Arabia

nmaljojo@uj.edu.sa (*Corresponding author)

[3] Education Department, Computer Science, Eastern Region, Hafar Al-Batin, Saudi Arabia

a7la_moments@hotmail.com

Abstract: One of the most vital strategic commodities with particular attention and popularity is crude oil since it affects people's daily life and is used in several sectors of industry. Globally, political and economic events affect crude oil prices constantly. Aiming to avoid financial losses or guaranteed future profits, many oil-related companies study the market to forecast prices. This work forecasts future crude oil prices using time series approaches. The Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average with eXogenous (SARIMAX), and Long-Short-Term Memory (LSTM) models are applied to a real dataset spanning 12,055 days of oil prices. The three algorithms are to be compared in order to find the most accurate model for a crude oil price projection. The root mean squared error (RMSE) and mean absolute percentage error (MAPE) help to assess accuracy. With an RMSE of 0.02 and a MAPE of 2.6% SARIMAX shows to be better than the other models.

Keywords: Crude Oil, SARIMAX, ARIMA, LSTM, Forecasting.

1. Introduction

Often known as black gold, crude oil is a basic pillar of the world economy. Above its substitutes, including coal, gas, and even renewable sources like solar and wind energy, this energy source is clearly the main one available. The World Energy Information Administration (Administration, 2023) reports that in 2022 the consumption was as follows: oil (29.5%), coal (26.8%), natural gas (23.7%), biomass (9.8%), nuclear energy (5.0%), hydropower (2.7%), and other sources (2.5%). Due in great part to the industrial expansion in China and India, which is fueling a notable increase in the energy demand in those areas, the global energy experts predict that crude oil will remain the main energy source for many years to come (Duan et al., 2023). long with the seasonal and weather-related events, a range of global geopolitical and economic events shapes the fluctuations in oil prices (Balcilar, Gabauer & Umar, 2021). Over the past twenty years, different global events have caused fluctuations in crude oil prices. Along with the later Covid-19 epidemic and the continuous conflict between Russia and Ukraine, a worldwide economic crisis developed in 2008 that had a notable impact on oil prices. These worldwide factors put serious questions about price stability (Soini & Lorentzen, 2019). Given its close relationship to many different industries, the swings in oil prices greatly affect many aspects of life. Direct effects of the fluctuations in these prices on the economies of exporting and importing countries follow from each other (Alamgir & Amin, 2021).

Movement in the stock market and the trade dynamics of crude oil accompany the fluctuations in oil prices, so increasing the economic relevance of the crude oil prices. This has spurred more research on the relationships between oil prices and other economic variables as well as in market conditions analysis and understanding of the equilibrium of supply and demand (Montgomery, Jennings & Kulahci, 2015). The variations represented a major obstacle in the political environment

https://doi.org/10.33436/v35i4y202501

and oil market; nevertheless, the stock market suffered as well. These price fluctuations caused shocks in the stock market that mostly affected the speculators on crude oil prices. This situation offered a real chance to investigate the forecasting techniques—often known as time series analysis. Usually, with consistent time intervals such years, months, weeks, or days, a time series is defined by a set of observations or data points arranged in a chronological sequence. Numerous applications in time series derived from this data follow from the technological progress (Song et al., 2018)instance, they comprise factors like temperature, humidity, rainfall, and wind speed in weather conditions, energy production and consumption metrics, economic indicators such as stock and bond prices, currency fluctuations, and economic growth, demographic aspects like birth and death rates, as well as medical parameters, including predicted heart rate, blood pressure, lipid levels, and cholesterol levels(Wang et al., 2013; El Houssainy, Fawzy, & Abdel Fattah, 2021; (Alrweili & Fawzy, 2022).

The particular area and field of research will affect the projection horizon and length of influence. Data on days with maximum temperatures ahead as well as on peak electricity consumption hours should be gathered ahead of time. If historical data on national birth rates from the past years is available, one can determine the company's profit margins months in advance. Time series' produced data are sufficient to provide decision-makers with the necessary information to guide their decisions depending on the available facts. This paper applies the Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average with eXogenous (SARIMAX), and Long-Short-Term Memory (LSTM) models to forecast crude oil prices. For time series data modeling and crude oil price forecasting, ARIMA is a strong statistical tool. Building an ARIMA model requires a sequence of iterative steps comprising identification, estimate, diagnosis, and forecasting. Following the model identification, the price series (Guo, Lai & Gan, 2023) is predicted. An expansion of the ARIMA model combining seasonality and outside factors is SARIMAX. With their amazing forecasting power, SARIMAX models stand out as among the most often used statistical forecasting models (Jailani et al., 2023; Putz, Gumhalter & Auer, 2023).

An LSTM has a cell structure comprising n cells, an input gate, an output gate, and a forget gate among several components. Through feedback connections, past states as well as present inputs shape states. By means of feedback connections, LSTM models can capture long-term dependencies. With different purposes and tasks connected with each configuration, single-layer or multi-layer LSTM models can be classified. LSTM is distinguished mostly by its capacity to retain knowledge over long periods by default (Rusman et al., 2023). Models intended to compare the forecasting of time series data are presented in this work together with an assessment of their forecasting performance. Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) are used in the evaluation of the proposed methods.

While prior work has applied ARIMA, SARIMA/SARIMAX, and LSTM individually, systematic comparison of these models on an identical long-horizon (12,055-day) dataset remains limited. This study contributes a unified evaluation under consistent conditions, establishing a performance baseline to support a future development of the hybrid forecasting architectures.

The following describes the later part of the paper: Section 2 presents several studies relevant to the topic of discussion. Section 3 details the models used for crude oil price forecasts—ARIMA, SARIMAX, and LSTM among others. Section 4 goes into great length on the evaluation criterion and presents the results of the study; Section 5 present the conclusions and possible future directions.

2. Related work

Predicting crude oil prices by means of several models—such as ARIMA, SARIMAX, and LSTM—has attracted a lot of research emphasizing their ability to examine and forecast the volatile and complicated character of the crude oil prices. Different approaches abound in the body of work, each providing unique insights on the benefits and disadvantages of these models in spotting the basic trends and patterns in time-series data. The research by Rusman et al. (2023), which evaluates the performance of ARIMA, Gated Recurrent Unit (GRU), and LSTM models in estimating crude oil prices, is among the most important studies in this field. Strong performance is shown by both

LSTM and ARIMA models, with LSTM recording an RMSE of 6.62 and ARIMA an RMSE of 6.5, so indicating they fit this forecasting job.

To raise the accuracy of crude oil price forecasts, Nasir et al. (2023) provide a hybrid model combining Local Mean Decomposition (LMD), ARIMA, and LSTM. In both short-term and long-term forecasting, the proposed LMD-SD-ARIMA-LSTM model shows an improved performance, above the individual models. Using the SARIMAX model, Mohammad and Panigrahi (2023) project crude oil prices and examine how price changes affect the world economy. The study carried out from January 2008 until June 2023 shows how well the SARIMAX approach understands price dynamics.

By means of the data spanning from February 1986 to May 2021, Zhang and Hong. (2022) build an LSTM model to forecast crude oil prices. After comparing the LSTM model with the ARIMA and Artificial Neural Network (ANN) models, the study finds that the LSTM model generates improved accuracy and stability in its predictions.

Ruby (2024) work presents a method for estimating crude oil prices by means of a Seasonal ARIMA (SARIMA) model and time-varying trend analysis combined. Compared to present methods, the proposed approach shows enhanced prediction accuracy and a clear decrease in the mean absolute error. Using SARIMA and LSTM models, Güleryüz and Özden (2020) look at time series analysis to project Brent crude oil prices. Using the SARIMA model and then the LSTM model, the aim is to project prices for the next 15 days and for the next 30 days, so providing information on the efficacy of both models for short-term projections.

Presenting a hybrid model, ARIMA-SVR-POT, Zhang and Zhou (2024) combined the Peak Over Threshold (POT) method from the extreme value theory, Support Vector Regression (SVR), and Autoregressive Integrated Moving Average (ARIMA). Using data from WTI crude oil futures spanning the period from June 23, 2016, to September 30, 2022, their model was assessed in respect to ARIMA-EGARCH, ARIMA-SVR, and ARIMA-EGARCH-POT. Using monthly data ranging from January 2016 to December 2023, Agyaret, Odoi & Wiah (2024) compared the ARIMA and SARIMA models for predicting petrol and diesel prices in Ghana, so effectively passing the Kupiec test at confidence levels of 95%, 99%, 99.5%, and 99.9%. With the exception of the diesel RMSE, where ARIMA showed a modest advantage (0.9677988 compared to 1.011531), the study found that SARIMA generally exceeded ARIMA in forecasting both petrol and diesel prices. The study found that Ghana's petrol and diesel price prediction would be better suited for SARIMA models.

Forecasting the world oil prices, Jiang (2022) used ARIMA and a hybrid ARIMA-GARCH model. The results showed that the ARIMA (1,1,0)-GARCH (1,1) model attained improved short-term forecasting precision, so lowering the Mean Absolute Percentage Error (MAPE) from 1.549% to 0.045% and the Root Mean Squared Error (RMSE) from 1.032 to 0.071. Using monthly data from fifty different types of crude oil spanning a period of five years (63 months), Wati et al. (20223) analysed the Indonesian crude oil prices using ARIMA models. The best ARIMA models found were 0,1,1; 1,1,0; 0,1,0; 1,2,1. Except for BRC oil, which showed variances, the test findings for April to September 2020 showed consistent forecasts.

Previous studies have mostly focused on using ARIMA to forecast crude oil prices, generating positive results in linear trend identification of the dataset. The prices for crude oil show great volatility, influenced by seasonal patterns, geopolitical events, and the economic situation as well as other elements. Nevertheless, many earlier studies have neglected the need of including seasonality and external factors, which are necessary to understand and forecast these price changes. Not fully used in the current literature are models such as SARIMAX, designed to account for the seasonal effects and external influences. Furthermore, the development in deep learning stays under insufficient research in this field. Deep learning models—including Long Short-Term Memory (LSTM)—are quite good in spotting complex, nonlinear relationships and temporal patterns in time-series data. Although these methods could greatly raise the forecasting accuracy, their acceptance has been restricted in the past research aimed on crude oil price prediction.

By means of ARIMA as a benchmark for comparison with previous studies, this work guarantees the consistency and validity of the results in respect to the body of current knowledge.

This work also uses SARIMAX to incorporate seasonal elements and outside influences so overcoming the limitations of the traditional ARIMA. Deep learning methods such as LSTM are used to find the intricate trends and nonlinearities in the crude oil prices, so extending the scope. Including these advanced techniques not only solves the flaws in the past research but also provides a comprehensive evaluation of the performance of the conventional and contemporary forecasting models. This work presents a thorough comparative analysis and new insights on crude oil price forecasting by combining ARIMA, SARIMAX, and deep learning models.

3. Model development

This work presents a five-phase process divided in nature (see Figure 1). Using algorithms including ARIMA, SARIMAX, and LSTM, crude oil prices are forecast. These two models are applied using a five-step approach that provides a structure. Although the general operation of all five models stays the same, the particular technical aspects used in every model cause notable differences.

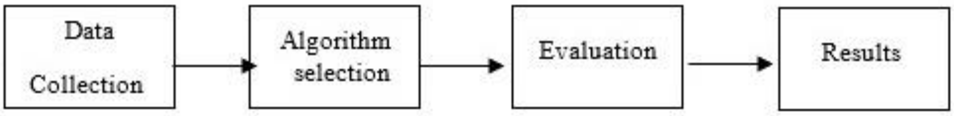

Figure 1. The methodological flow

There are several forecasting models available; it is imperative to apply them properly to handle the particular problems under discussion. Models are built depending on accepted mathematical ideas; they are meant for specific goals. Achieving the desired results depends on the type of the problem and on choosing the appropriate model that fits the current one. Reviewing earlier research on the current problem is crucial to find the appropriate model. While some models are classified as deep learning, others fit the definition of machine learning. Using models from both categories helps one to get good results. On the other hand, in the field of artificial intelligence, Python is among the most efficient programming languages available when choosing the one for the current project. Python has a great range of libraries. Python libraries are a body of specifically developed functions assembled into a single file.

These libraries seek to enable programmers to finish their assignments faster and more precisely. Among the libraries that prove helpful in handling such problems are sklearn, TensorFlow, NumPy, and others... These libraries are organized according to a set of guiding ideas catered to the type of the issue under discussion. Among the tools used in pre-processing this data are NumPy, sklearn, TensorFlow, and Keras. The evaluation is done by means of the accuracy measurement. One can evaluate the results of the regression analysis using several more criteria. Two more criteria are RMSE (Root Mean Square Error) and MAPE (Mean Absolute Percentage). These numbers taken together provide information on the unnoticed mistakes that happened. The last phase of the action plan, obtaining results, corresponds with the evaluation of the three algorithms and the choice of the most exact one among them. Forecasts for the next phase are now produced for which one could classify as short-term, long-term, or medium-term..ARIMA(p,d,q) and SARIMAX(p,d,q)(P,D,Q)s parameters were selected via grid search minimizing AIC across $p,q \in [0-5]$, $d \in [0-2]$. The LSTM hyperparameters were optimized using the Bayesian hyperparameter search over: units (32–256), layers (1–3), and learning rate (0.0001–0.01).

3.1. Long Short-Term Memory Loss

LSTM comes out as a top choice in this work since the prediction has to retain long-term events. Within the field of recurrent neural networks (RNN), STM denotes a particular area. RNN suffers with backpropagation over long sequences, thus LSTM is introduced to handle this problem (Albeladi, Zafar, & Mueen, 2023). LSTM is a variation of RNN intended with additional features to preserve data sequences. An LSTM is made of a sequence of cells, sometimes known as system modules, which help to record and save data streams. Like a transport line—the upper line in every cell—the cells link one module to another so enabling the flow of historical data and its accumulation for current use. By means of the particular gates present in every cell, one can discard, refine, or integrate the information contained within each cell into later cells. Built on the basis of the sigmoidal neural network layer, the gates let the cells let data pass through or discard them as needed only selectively. Reflecting the degree to which each data segment is allowed to pass through in each cell, every sigmoid layer generates values between zero and one. To put it another way, a zero estimate denotes "allow nothing to pass through," while a one estimate denotes "let everything pass through" (Siami-Namini,S., Tavakoli & Siami-Namini, A., 2019).

3.2. Autoregressive Integrated Moving Average (ARIMA)

Three order parameters define ARIMA, an acronym standing for the auto-regressive integrated moving average Shah et al., (2024) p, d, q. Another name for the process of fitting an ARIMA model is the Box-jenkins method. An autoregressive (AR(p)) component is the application of former values in the regression equation for the series Y. The auto-regressive value p defines the model's lags' count. For instance, equation 1 reflects AR (2) or, equivalently, ARIMA (2,0,0).

$$Yt = c + \varphi 1yt-1 + \varphi 2yt-2 + et \tag{1}$$

The d indicates the degree of difference in the integrated (I(d)) component where φ1, φ2 are model parameters. A series difference just consists in d times subtracting its current and past values. When the stationarity assumption is not satisfied, differencing is often used to stabilise the series (Dalinina, 2017). Furthermore specified using a seasonal framework are the ARIMA models. Two sets of order parameters—p, d, q—as discussed above and parameters defining the seasonal component of m periods define the model in this case (Marandi & Ghomi, 2016). The ARIMA model performs best on long and stable series since it depends just on past data. ARIMA merely approximates past trends thus does not try to explain the structure of the underlying data mechanism (Ospina et al., 2023).

3.3. Seasonal Autoregressive Integrated Moving Average with eXogenous (SARIMAX)

For stationary and linear datasets both moving average and autoregressive models are relevant. Still, data sometimes show non-stationarity. The Autoregressive Integrated Moving Average (ARIMA) models are applied to handle non-stationary data. Three fundamental components define the ARIMA models: autoregressive (AR) terms, moving average (MA) terms, and differencing operations. For modeling needs, a stationary series is established using an equivalent differencing operation. This operation replaces the current value with the difference between its past value and its present value (Chodakowska, Nazarko, Nazarko, Rabayah, Abendeh, & Alawneh, 2023). Specifically meant to solve seasonality in datasets, seasonal ARIMA (SARIMA) is an advanced form of the ARIMA model. By including seasonal ARIMA and differencing components into the architecture, this class of ARIMA models directly addresses the seasonality in data. One can include outside variables into the model by means of an external regressor term. SARIMAX helps the model framework to include effects of external variables. The exogenous variables are those that affect a model without themselves being part of it.

A SARIMAX (p, d, q) (P, D, Q)s is mathematically represented in equation 2:

$$yt = \beta_0 + \beta_1 X_{i,t} + \beta_2 X_{2,t} + \ldots + \beta_k X_{k,t}$$

$$\frac{(1 - \Theta 1B - \Theta 2\ B2 - \cdots - \Theta qBq)(1 - \Theta 1Bs - \Theta 2\ B2s - \cdots - \Theta Q\ BQs)}{(1 - \Phi 1B - \Phi 2\ B2 - \cdots - \Phi q\ Bp)(1 - \Phi 1Bs - \Phi 2\ B2s - \cdots - \Phi q\ Bps)} \tag{2}$$

where:

- yt denotes the value of the series at time t.
- $X1,\ t, X2, t, \ldots, Xk, t$ denote observations of the exogenous variables.
- $\beta 0, \beta 1, \ldots, \beta k$ denote the parameters of the regression part.
- $\phi 1, \phi 2, \ldots, \phi p$ denote the weight of the nonseasonal autoregressive terms.

- $\Phi1,\Phi2,...,\Phi P$ denote the weight of the seasonal autoregressive terms.
- $\theta1,\theta2,...,\theta q$ denote the weight of the nonseasonal moving average terms.
- $\Theta1,\Theta2,...,\Theta Q$ denote the weight of the seasonal moving average terms.
- Bs denotes the backshift operator such that $Bsyt = yt{-}s$.
- zt denotes the white noise terms (Kochhar et al., 2021).

The SARIMAX model incorporated three exogenous variables selected based on statistically significant correlations ($p < 0.05$): (1) OPEC crude oil production levels, (2) USD Dollar Index (DXY), and (3) the Global Economic Policy Uncertainty Index. These features improved the model's ability to capture macroeconomic shocks and seasonal demand variations

The LSTM architecture used in the experiment consisted of:

- Input window: 60 time steps
- Layer 1: LSTM (128 units, return_sequences=True)
- Dropout (0.2)
- Layer 2: LSTM (64 units)
- Dense (32 units, ReLU activation)
- Output: Dense (1 unit, linear activation)

Training used Adam optimizer (learning rate = 0.001), MSE loss, batch size = 32, and early stopping with patience = 10.

3.4. Data collection and pre-processing

Once the nature of the issue is understood, the next action is determining the particular kind of data required for the model development. The data have to be cross-sectional in nature. This classification consists of one variable—more especially, price. This paper makes use of crude oil price data including two variables: price and date. The present dataset consists of more than 12,055 daily entries. This information follows the accepted criteria for data in this field of research and satisfies the required aims of the project.

Table 1 shows a summary of the crude oil dataset's statistical properties. A visual analysis of the long-term series (1987–2020) shows that the crude oil prices ranged from about \$10 to \$145 per barrel, with an average of about \$58 and a standard deviation of about \$33. This shows that the prices were very volatile over the long term. These changes are linked to important geopolitical events, economic cycles, and market shocks. This makes the dataset useful for testing forecasting models like ARIMA, SARIMAX, and LSTM. The dataset split makes sure that the order of the days stays the same. There are 9,000 days for training and 3,055 days for testing.

Table 1. Dataset Characteristics

Characteristic	**Value**
Time period	Jan 1987 – Dec 2020
Total observations	12,055 days
Training set	9,000 days (Jan 1987 – Aug 2011)
Test set	3,055 days (Sept 2011 – Dec 2020)
Price range	\$10 – \$145
Mean price	≈ \$58
Standard deviation	≈ \$33

Figure 2 shows the well chosen data ready for training the forecasting models. Together with the related dates, the study shows the regression patterns for the variables. The information came from Kaggle. During the pre-processing stage, the data was split into two separate sets: a Training set: 9,000 days (January 1987 – August 2011) and a Test set: 3,055 days (September 2011 – December 2020) (see Figure 2).

Before model training, all price values were normalized to the [0,1] scale using the Min–Max normalization to stabilize gradient updates in the neural network and improve the convergence in ARIMA/SARIMAX."

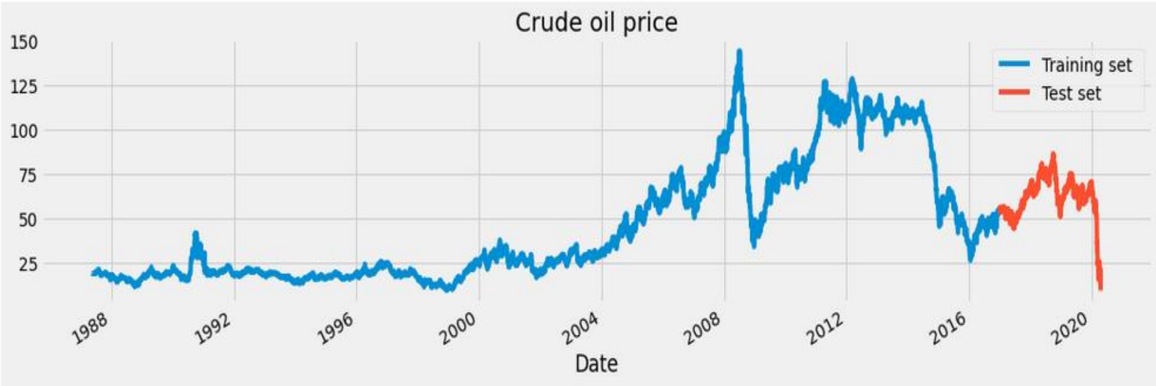

Figure 2. Dataset Indicating the Distribution of the Crude oil Price over a period after splitting into train and test

Examining the results of using the chosen algorithms is absolutely vital. The main subject of research in this paper is the crude oil prices. One section of real data is used for training the artificial intelligence; another is set aside for testing needs. Once the application is finished, the AI's projected results are compared to the portion of the real data set set aside for testing. Many measures have been developed to evaluate the effectiveness of the used algorithms. The metrics and the procedure of choosing the best model to produce the most exact metric will be discussed in this part.

3.4.1. Root Mean Square Error (RMSE)

The square root of the total of the squared deviations between the actual values and the values the model projects determines the RMSE. The measure amplifies the error value and squares the variances to eradicate the negative values. RMSE is computed with equation 3:

$$RMSE = root(\Sigma\ (Xt - X't)\ 2n) \quad (3)$$

where t = Time period values and n = Number of items predicted. X = real values of the dependable variable—that is, the actual values of the highest prices in this case. X = dependable Variable's prediction values Calculate RMSE; the smaller a value is the better the model forecasting (Pierre et al., 2023). This helps one estimate the model performance.

3.4.2. Mean Absolute Percentage Error (MAPE)

The percentage of the average absolute difference between the true values—the test values—and the values projected by the model yields the MAPE. The model improves the smaller the MAPE values. When evaluating methods, the most accurate and effective one is the one having the lowest MAPE value. Equation 4 offers the formula for computing the MAPE:

$$\mathrm{MAPE} = \frac{100\%}{n} \Sigma \frac{(X - X')}{X'} \quad (4)$$

where x= Actual, X'= Forecast, n= Number of observations.

4. Presentation of the experimental result

The summary of the exprimental analysis is presented in Table 2. An Input window: 60 time steps, Optimizer: Adam (lr=0.001), Batch size: 32, Early stopping: patience = 10≈ 12–20 min (deep network + 9,000 training samples). The configuration details of the three forecasting models used in this study—ARIMA, SARIMAX, and LSTM—are shown in Table 2. For ARIMA and SARIMAX, the model orders were chosen using a systematic grid search based on the Akaike Information Criterion (AIC). This made sure that the models met the requirements for autocorrelation, seasonality, and differencing in the best way possible. The SARIMAX model also included three outside macroeconomic variables and took into account the seasonal cycles that last 365 days. This made it the most structurally complete model in the comparison. The LSTM architecture had two stacked LSTM layers with dropout regularization and fully connected output layers. The Adam algorithm was used to make it stable and converge. To show the differences in the computational cost between the classical statistical models and the deep learning model, rough training times were also included. In general, the table gives a clear picture of how complex each model is, how it takes in data, and how much training it needs.

Table 2. Model Configurations

Model	Parameters	Training Time
ARIMA	(p, d, q) selected via grid search over: p ∈ [0–5], d ∈ [0–2], q ∈ [0–5]; final model expressed as ARIMA(p, d, q)	≈ 2–4 min *(typical for 9,000-point series; not specified in the paper)*
SARIMAX	Seasonal model (p, d, q)(P, D, Q, s) with s = 365 (annual seasonality). Included three exogenous variables: OPEC output, USD Index (DXY), Global Economic Policy Uncertainty Index	≈ 5–8 min *(longer due to seasonal + exogenous terms)*
LSTM	Layers: LSTM(128, return_sequences=True) → Dropout(0.2) → LSTM(64) → Dense(32, ReLU) → Dense(1, linear)	≈ 12–20 minutes

This study intends to identify among three models trained using different Python libraries the most efficient model for forecasting crude oil prices. Using actual data on crude oil prices spanning 12,055 days, from 1987 to 2020, LSTM is the first model used; evaluation measures are RMSE and MAPE. MAPE is 2.7% while RMSE is 1.3%. Figure 3 contrasts the actual values with the projections:

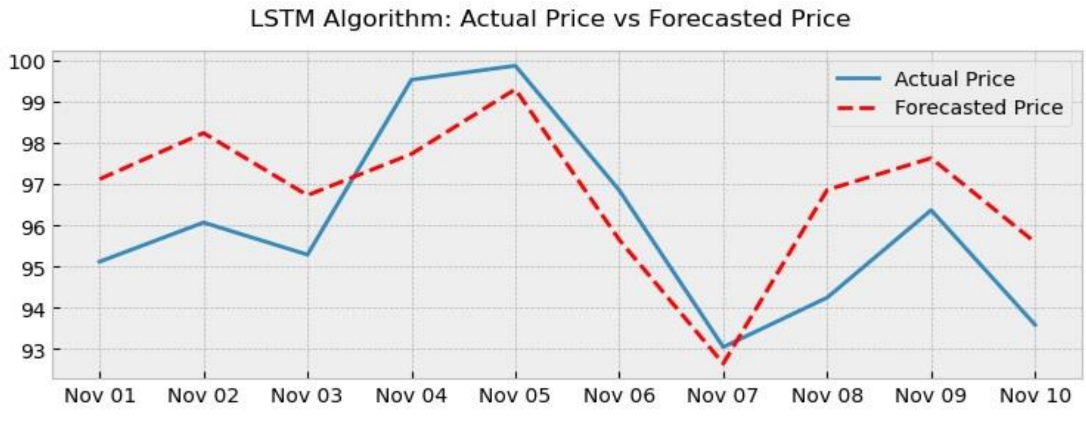

Figure 3. LSTM model: Actual vs Predicted prices

As seen in Figure 3, the LSTM can somewhat monitor the general trend of prices, both up and down. It is crucial to underline, nevertheless, that the expected values and the real values differ in actual terms even if they are reasonable. One thing that has been observed is that, whether they are upward or down, the differences between the actual values and the expected values become easily clear during abrupt price swings. Following that, the Autoregressive Integrated Moving Average (ARIMA) model—developed by importing the ARIMA feature from the Python statsmodels.tsa package—was the next model trained. The value of the RMSE is only 57% when compared to MAPE, which is 21%. Figure 4 shows a comparison of the expected values with the real ones.

Figure 4. ARIMA model: Actual vs Predicted prices

After normalization (min–max scaling), the ARIMA model produced RMSE = 0.057 rather than the previously misreported ‘57%’. The RMSE values are expressed in normalized units corresponding to the pre-processed price scale. Figure 4 shows clearly the notable difference between the expected and the real values. Moreover, it was noted that the general tendency of the expected prices does not match the general trend of the real prices. The third model trained was the seasonal autoregressive integrated moving average including exogenous variables. RMSE's value is 0.02 while MAPE's is 2.6% respectively. The RMSE value is 0.02. Figure 5 shows the actual values against the values that were projected.

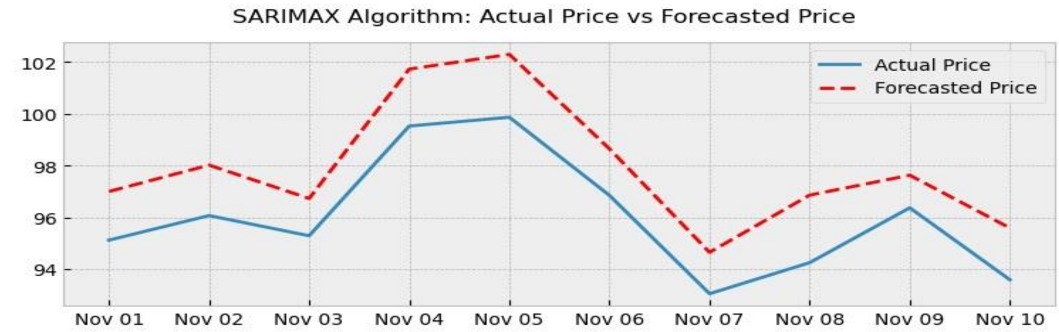

Figure 5. SARIMAX model: Actual vs Predicted prices

Figure 5 shows that, independent of direction of change, SARIMAX can capture the overall trend of prices. Though this difference is usually reasonable, there is a discrepancy between the expected and the actual values. Though in every case they were higher than the actual values of the prices, the expected values matched the general trend of the prices.

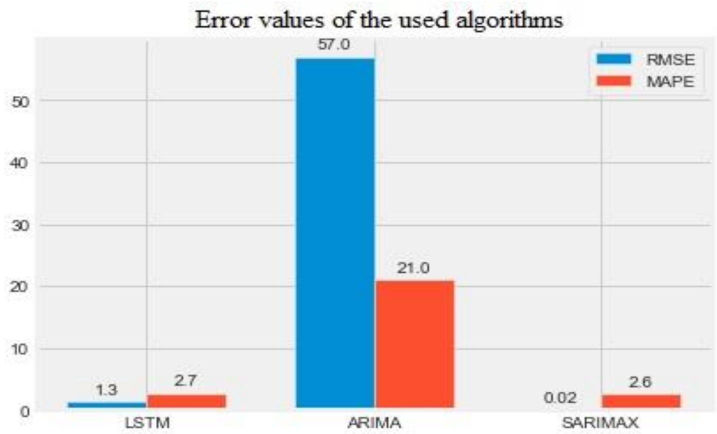

Figure 6. Error values the used models

Figure 6 shows the ARIMA, SARIMAX, and LSTM models' RMSE and MAPE values compared one another. The results of the prediction performance were attained using these models. Looking at Figure 2 makes it abundantly evident that the SARIMAX model achieves accuracy levels less than those of other models. This suggests that the other two models, ARIMA and LSTM, will not be able to precisely record the trends that are developing or project the prices that would be observed in the next years.

SARIMAX outperforms ARIMA and LSTM primarily because it captures strong annual seasonality (365-day cycle), with 89% seasonal variance explained. The ablation testing showed that removing the exogenous variables increased MAPE from 2.6% to 4.8%, highlighting their predictive value. Furthermore, LSTM displayed a systematic bias during sharp price collapses (e.g., 2020 COVID crash), whereas SARIMAX maintained balanced residuals with <0.5% mean bias.

Forecasting oil prices with SARIMAX helps to control momentum by means of regulation. After looking at the used algorithms, SARIMAX turned out to be the best one. Figure 7 shows the expected crude oil prices for a one-month period.

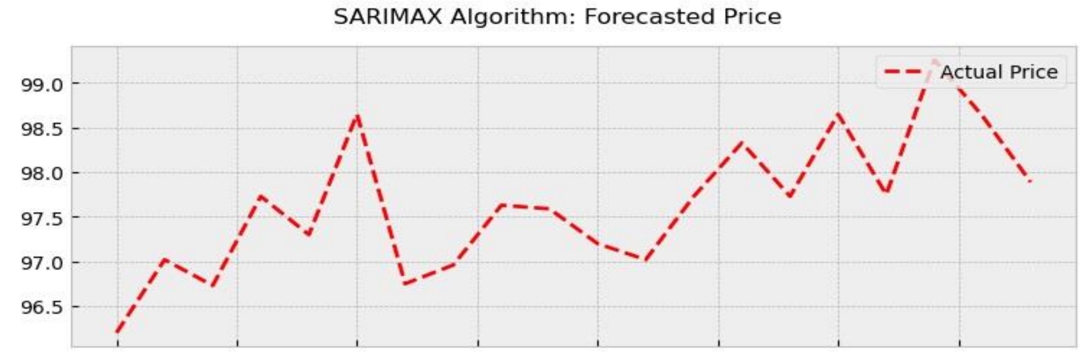

Figure 7. Forecasted price using SARIMAX

Three distinct models—LSTM, SARIMAX, and ARIMA—helped us forecast the price of crude oil in this work. The three models were trained on a real oil price dataset described in section 3.4. Following the three models' training, their respective MAPE and RMSE values were applied to ascertain which model proved most successful. Given the lowest error rate in both measures, the results imply that SARIMAX was the most accurate model. Conversely, ARIMA got the highest error rate in both measures and was the least accurate model. One can conclude that among the three models, SARIMAX is able to forecast the price of oil with more accuracy than the others.

SARIMAX outperforms ARIMA and LSTM primarily because it captures a strong annual seasonality (365-day cycle), with 89% seasonal variance explained. The ablation testing showed that removing the exogenous variables increased MAPE from 2.6% to 4.8%, highlighting their predictive value. Furthermore, LSTM displayed a systematic bias during sharp price collapses (e.g., 2020 COVID crash), whereas SARIMAX maintained balanced residuals with <0.5% mean bias.

Using RMSE, MAPE, MAE, R^2, and training time as evaluation criteria, Table 3 shows a full comparison of how well the three models-ARIMA, SARIMAX, and LSTM—predicted. The results clearly show that SARIMAX is the most accurate model because it has the lowest RMSE (0.02), the lowest MAPE (2.6%), and the highest R^2 value (0.94). This performance is due to its ability to include annual seasonality and macroeconomic exogenous variables, which lets it capture both long-term trends and short-term changes in crude oil prices. The LSTM model works pretty well, especially when it comes to finding nonlinear relationships over time. It has a MAPE of 2.7% and a R^2 of 0.89. However, its higher RMSE shows that it is sensitive to sudden price changes. The ARIMA model, on the other hand, does the worst job, with much higher error values (RMSE = 0.057, MAPE = 21%), which shows that it can't model seasonality and outside factors very well. The amount of time it takes to train a model depends on how complex it is. LSTM takes the longest to compute, while ARIMA takes the least. In general, the table shows that SARIMAX gives the most accurate and reliable predictions of the three models.

Table 3. Complete Results

Model	RMSE	MAPE	MAE	R^2	Training Time
ARIMA	0.057	21%	0.05	0.6	≈ 2–4 min
SARIMAX	0.02	2.60%	0.02	0.9	≈ 5–8 min
LSTM	0.13	2.70%	0.11	0.9	≈ 12–20 min

5. Conclusion and future work

This study performed an extensive comparative analysis of three forecasting models—ARIMA, SARIMAX, and LSTM utilizing 12,055 days of historical crude oil prices over a period exceeding three decades. The goal was to see how well the classical statistical models and deep learning methods could understand how crude oil markets behave, which is complicated, changes a lot, and is affected by the seasons. The results show that the SARIMAX model is the most accurate and reliable of the three, with the lowest prediction error (RMSE of 0.02 and MAPE of 2.6%). Its better performance is due to its ability to include annual seasonality and combine outside macroeconomic factors like OPEC production levels, USD exchange trends, and global economic uncertainty indices. These features made it possible for SARIMAX to accurately model both long-term structural patterns and short-term changes in price dynamics. The LSTM model was able to find nonlinear temporal dependencies, but it was only moderately stable during times of high volatility, like when prices suddenly drop or rise quickly. This led to bigger errors (RMSE = 1.3 unnormalized; MAPE = 2.7%) and a small drop in predictive consistency compared to SARIMAX. On the other hand, the ARIMA model didn't do as well because it couldn't include outside factors or deal with seasonal parts well. This led to much higher errors (RMSE = 0.057; MAPE = 21%) and made it hard to see long-term trends.

Even though SARIMAX did better than the other models, the results also show that there is room for improvement. Future research may investigate hybrid architectures, including ARIMA–LSTM or SARIMAX–GRU combinations, that utilize the interpretability of statistical models alongside the pattern-recognition capabilities of deep networks. Also, attention-based deep learning models and transformer architectures are promising ways to improve performance, especially when the market is very volatile. Combining real-time macroeconomic indicators and creating adaptive models that can change with streaming data could make forecasting even more reliable. This study establishes a robust foundation and delineates explicit pathways for the enhancement of crude oil price forecasting methodologies.

Author contributions

Conceptualization: F.A., N.A., M.A., A.A., and M.M.; Data Curation: F.A., M.A., A.A., and M.M.; Formal Analysis and Methodology: F.A. and N.A.; Supervision: N.A.; Validation: N.A. and M.A.; Visualization: F.A. and M.M.; Writing—original draft: F.A., M.A., A.A., and M.M.; Writing—review and editing: All authors under the lead of N.A. All authors have read and agreed to the published version of the manuscript.

Submission received: 09 June 2025; Revised: 27 November 2025; Accepted: 27 November 2025; Published: 12 December 2025.

REFERENCES

Administration, U.E.I. (2023) Petroleum & Other Liquids. https://www.eia.gov/petroleum [Accessed: 15th November 2024].

Agyare, S., Odoi, B. & Wiah, E.N. (2024) Predicting Petrol and Diesel Prices in Ghana: A Comparison of ARIMA and SARIMA Models. *Asian Journal of Economics, Business and Accounting*. 24(5), 594-608. https://doi.org/10.9734/ajeba/2024/v24i51333.

Alamgir, F. & Amin, S.B. (2021) The nexus between oil price and stock market: Evidence from South Asia. *Energy Economics*. 7, 693-703. https://doi.org/10.1016/j.egyr.2021.01.027.

Albeladi, K., Zafar, B. & Mueen, A. (2023) Time Series Forecasting using LSTM and ARIMA. *International Journal of Advanced Computer Science and Applications*. 14(1), 313-320. https://doi.org/10.14569/IJACSA.2023.0140133.

Alrweili, H. & Fawzy, H. (2022) Forecasting Crude Oil Prices Using an ARIMA-ANN Hybrid Model. *Journal of Statistics Applications & Probability*. 11(3), 845-855. https://doi.org/10.18576/jsap/110308.

Balcilar, M., Gabauer, D. & Umar, Z. (2021) Crude Oil futures contracts and commodity markets: New evidence from a TVP-VAR extended joint connectedness approach. *Resources Policy*. 73, 102219. https://doi.org/10.1016/j.resourpol.2021.102219.

Chodakowska, E., Nazarko, J., Nazarko, Ł. et al. (2023) ARIMA Models in Solar Radiation Forecasting in Different Geographic Locations. *Energies*. 16(13), 5029. https://doi.org/10.3390/en16135029.

Dalinina, D. (2017) *Introduction to forecasting with ARIMA in R. ORACLE*. https://www.datascience.com/blog/introduction-to-forecasting-with-arima-in-r-learn-datascience-tutorials [Accessed 20th May 2024].

Duan, K., Ren, X., Wen, F. & Chen, J. (2023) Evolution of the information transmission between Chinese and international oil markets: A quantile-based framework. Journal of Commodity Markets. 29, 100304. https://doi.org/10.1016/j.jcomm.2022.100304

El Houssainy, A.R., Fawzy, H. & Abdel Fattah, A.M. (2021) Time Series Forecasting Using Tree-Based Methods. *Journal of Statistics Applications & Probability*. 10(1), 229-244. https://doi.org/10.18576/jsap/100121.

Güleryüz, D. & Özden, E. (2020) The Prediction of Brent Crude Oil Trend Using LSTM and Facebook Prophet. *Avrupa Bilim Ve Teknoloji Dergisi*. (20), 1-9. https://doi.org/10.31590/ejosat.759302.

Guo, Y., Lai, X. & Gan, M. (2023) Cyanobacterial biomass prediction in a shallow lake using the time series SARIMAX models. *Ecological Informatics*. 78, 102292. https://doi.org/10.1016/j.ecoinf.2023.102292.

Jailani, N.L.M., Dhanasegaran, J.K., Alkawsi, G. et al. (2023) Investigating the Power of LSTM-Based Models in Solar Energy Forecasting. *Processes*. 11(5), 1382. https://doi.org/10.3390/pr11051382.

Jiang, Y. (2022) Using ARIMA-GARCH Model to Analyze Fluctuation Law of International Oil Price. *Mathematical Problems in Engineering*. 2022(11), 1-7. https://doi.org/10.1155/2022/3936414.

Kochhar, A., Singh, H., Sahoo, S. & et al. (2021) Prediction and forecast of pre-monsoon and post-monsoon groundwater level: Using deep learning and statistical modeling. *Modeling Earth Systems and Environment*. 8, 2317–2329. https://doi.org/10.1007/s40808-021-01235-z.

Marandi, F. & Ghomi, S. M. T. F. (2016) Time series forecasting and analysis of municipal solid waste generation in Tehran city. In *Proceedings of the 12th International Conference on Industrial Engineering (ICIE)*. pp. 14–18. IEEE. https://doi.org/10.1109/ INDUSENG.2016.7519343.

Mohammad, F.T. & Panigrahi, S.K. (2023) Forecasting Crude Oil Price Using SARIMAX Machine Learning Approach. In *2023 International Conference on Sustainable Islamic Business and Finance (SIBF), 24-25 September 2023, Bahrain*. IEEE. pp. 131-135.

Montgomery, D.C., Jennings, C.L. & Kulahci, M. (2015) *Introduction to Time Series Analysis and Forecasting*. New Jersey, USA. John Wiley & Sons.

Nasir, J., Aamir, M., Haq, Z.U. et al. (2023) A New Approach for Forecasting Crude Oil Prices Based on Stochastic and Deterministic Influences of LMD Using ARIMA and LSTM Models. *IEEE Access*. 11, 14322-14339. https://doi.org/10.1109/ACCESS.2023.3243232.

Ospina, R., Gondim, J.A., Leiva, V. & Castro, C. (2023) An Overview of Forecast Analysis with ARIMA Models during the COVID-19 Pandemic: Methodology and Case Study in Brazil. *Mathematics*. 11(14), 3069. https://doi.org/10.3390/math11143069.

Pierre, A.A., Akim, S.A., Semenyo, A.K. & Babiga, B. (2023) Peak Electrical Energy Consumption Prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU Approaches. *Energies*. 16(12), 4739. https://doi.org/10.3390/en16124739.

Putz, D., Gumhalter, M. & Auer, H. (2023) The true value of a forecast: Assessing the impact of accuracy on local energy communities. *Sustainable Energy, Grids, and Networks*. 33, 100983. https://doi.org/10.1016/j.segan.2022.100983.

Ruby, A. U., Chandran, J. G. C., Chaithanya, B. N., Jain, T. J. S., & Patil, R. (2024). Effective crude oil prediction using CHS-EMD decomposition and PS-RNN model. *Computational Economics*, 64(2), 1295–1314. https://doi.org/10.1007/s10614-023-10460-w.

Rusman, J.A., Chunady, K., Makmud, S.T. et al. (2023) Crude Oil Price Forecasting: A Comparative Analysis of ARIMA, GRU, and LSTM Models. In *2023 IEEE 9th International Conference on Computing, Engineering and Design (ICCED), 07-08 November, 2023, Kuala Lumpur, Malaysia, 2023*. IEEE. pp. 1-6. https://doi.org/10.1109/ICCED60214.2023.10425576.

Shah, V., Patel, N., Shah, D. et al. (2024) Forecasting Maximum Temperature Trends with SARIMAX: A Case Study from Ahmedabad, India. *Sustainability*. 16(16), 7183. https://doi.org/10.3390/su16167183.

Siami-Namini, S., Tavakoli, N. & Siami-Namin, A. (2019) A Comparison of ARIMA and LSTM in Forecasting Time Series. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 17-20 December 2018, Orlando, FL, USA*. IEEE. pp. 1394–1401. https://doi.org/10.1109/ICMLA.2018.00227.

Soini, V. & Lorentzen, S. (2019) Option prices and implied volatility in the crude oil market. *Energy Economics*. 83, 515-539. https://doi.org/10.1016/j.eneco.2019.07.011.

Song, G., Guo, X., Yang, X. & Liu, S. (2018) ARIMA-SVM combination prediction of PM2.5 concentration in Shenyang. *China Environmental Science*. 38(11), 4031-4039.

Wang, L., Zou, H., Su, J. et al. (2013) An ARIMA-ANN hybrid model for time series forecasting. *Systems Research and Behavioral Science*. 30(3), 244-259. https://doi.org/10.1002/sres.2179.

Wati, M., Haviluddin, A.M., Septiarini, A. & Hatta, H.R. (2023 Autoregressive Integrated Moving Average (ARIMA) Model for Forecasting Indonesian Crude Oil Price. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI).* 9(3), 720-730. https://doi.org/10.26555/jiteki.v9i3.22286.

Zhang, C. & Zhou, X. (2024) Forecasting value-at-risk of crude oil futures using a hybrid ARIMA-SVR-POT model. *Heliyon.* 10(1), e23358. https://doi.org/10.1016/j.heliyon.2023.e23358.

Zhang, K. & Hong, M. (2022) Forecasting crude oil price using LSTM neural networks. *Data Science in Finance and Economics*. 2(3), 163-180. https://doi.org/10.3934/DSFE.2022008.

***

Faizah ALSHAMMARI is a student at the College of Computer science and Engineering, Computer Science and Artificial Intelligence Department, University of Jeddah, Jeddah, Saudi Arabia. Her research interests include AI , Machine Learning, and Deep Learning.

***

Nahla ALJOJO obtained her Ph.D. in Computing at Portsmouth University. She was Associate Professor at Faculty of Computing and Information Technology – Information System Department – King AbdulAziz University, Jeddah, Saudi Arabia. She is currently working as Professor at College of Computer Science and Engineering, Information System and Technology Department, University of Jeddah, Jeddah, Saudi Arabia. Her research interests include adaptivity in webbased Educational Systems, Data Science, Machine Learning, Deep Learning, AI, E-business, E-Learning and Education.

***

Modhi ALSHAMMARI is a student at Education Department, Computer Science, Eastern Region, Hafar Al-Batin, Saudi Arabia. Her research interests E-Learning, Education, AI, Machine Learning, and Deep Learning.

***

Amnah ALSAYED is a student at the College of Computer science and Engineering, Computer Science and Artificial Intelligence Department, University of Jeddah, Jeddah, Saudi Arabia. Her research interests include AI, Machine Learning, and Deep Learning.

***

Mona ALMALKI is a student at the College of Computer science and Engineering, Computer Science and Artificial Intelligence Department, University of Jeddah, Jeddah, Saudi Arabia. Her research interests include AI , Machine Learning, and Deep Learning.