

Multistep Ahead Forecasting of WTI Crude Oil Prices Using Time Series and Machine Learning Models

Muhammad Shafiq¹, Mohammad Abiad^{2,*}, Noor Zali Khan¹, Ihtisham ul Haq³

¹*Institute of Numerical Sciences, Kohat University of Science and Technology, Pakistan*

²*College of Business Administration, American University of the Middle East, Kuwait*

³*Department of Econometrics, Tashkent State University of Economics, Tashkent, Uzbekistan*

*Corresponding author: mohammad.abiad@aum.edu.kw

Abstract. Crude oil is a critical energy source and a raw material for various products, including fuels (gasoline, diesel, and jet fuel), lubricants, petrochemicals, and asphalt. It undergoes refining processes to separate and convert it into usable products. Multiple factors, including economic indicators, geopolitical events, supply and demand dynamics, and technological advancements influence crude oil prices. Its global market has prices often benchmarked to grades like Brent Crude and West Texas Intermediate (WTI). In this article, the daily WTI crude oil prices are multi-step ahead forecasted using four different methods which are autoregressive integrated moving average (ARIMA), random forest (RF), k -nearest neighbors (k NN), and autoregressive neural networks (ARNN). The data set consists of daily WTI crude oil prices from February 13, 2014, to February 13, 2024, divided into 80% training set and 20% validation set. The performance of the methods is evaluated by one-step and seven-step ahead forecasting using root mean square error (RMSE) and mean absolute percentage error (MAPE) as the accuracy measures. The results show that ARIMA outperforms the methods for one and seven-step ahead forecasting, followed by ARNN, k NN, and RF. The study provides useful insights for investors and policymakers in the oil market.

1. INTRODUCTION

Crude oil is a naturally occurring, unrefined petroleum substance comprising hydrocarbon deposits and various organic compounds. Drilling is typically used to extract it from underground reservoirs. Crude oil serves as a vital energy source and is the primary raw material for many everyday products such as diesel, gasoline, jet fuel, heating oil, and various plastics and chemicals [1]. Due to its widespread use in transportation, manufacturing, and various industries crude oil has a significant impact on the global economy. Fluctuations in crude oil prices directly influence the prices of gasoline, diesel, and other energy products. Higher oil prices caused an increase in

Received: Aug. 23, 2025.

2020 *Mathematics Subject Classification.* 62M10.

Key words and phrases. ARIMA; k NN; neural networks; time series.

transportation costs, heating, and production costs, impacting consumer spending and business operations. Since oil is a key component in the services and production of goods, changes in oil prices can affect overall price levels, leading to inflationary pressures in the economy. High oil prices can contribute to higher inflation rates, affecting consumer purchasing power and economic stability [2].

Countries that are net exporters of oil enjoy benefits from high prices of oil as they earn more revenue from exports. In contrast, countries that are net importers of oil face increased import costs, which can negatively impact trade balances and currency exchange rates. The oil and gas industry is a significant contributor to employment and investment in many economies. Changes in oil prices can affect investment decisions and employment levels within the industry, along with related sectors such as transportation, manufacturing, and construction [2].

Since most of the oil is produced in specific regions of the world including the Middle East and North America, the geopolitical tensions in that region can significantly affect the prices of crude oil due to interrupting the supply. Political instability, sanctions, or conflicts in those regions that are the major oil producers can have far-reaching consequences on global oil markets and the economy [3]. High demand for crude oil results in a price increase, whereas low demand leads to a decrease in price. Key factors such as production levels, changes in consumption patterns, and economic growth rates can affect both supply and demand dynamics. Economic downturns or recessions typically reduce oil demand, whereas times of strong economic growth can result in heightened demand and elevated prices. Sanctions on major oil-producing nations interrupt the supply chain or concern about future supply leads to an increase in prices. Currency exchange rates, inflation, GDP growth, and interest rates are the major economic factors that affect the demand for crude oil causing a price hike. Horizontal drilling and hydraulic fracturing (fracking) are the advanced technologies that enabled the extraction of oil from once unreachable reserves, impacting global oil supply and prices [4]. By influencing consumer behavior and investment decisions, the energy sector may influence the future outcome for oil demand and pricing through the use of renewable energy sources and the implementation of environmental policies and regulations to reduce greenhouse gas emissions. Overall, a complex interplay between macroeconomic factors, supply and demand dynamics, and geopolitical developments affects crude oil prices, making the oil markets very volatile and unpredictable [5].

Crude oil is forecasted by [6] through a proposed novel hybrid method based on autoregressive integrated moving average (ARIMA), long short-term memory (LSTM), and local mean decomposition (LMD) models. The authors conclude that their method can improve forecasting accuracy by decomposing the original data into stochastic and deterministic components, forecasting each component by ARIMA, and summing the residuals that are forecasted through LSTM. Using various accuracy measures the author forecasts the WTI crude oil prices to test their proposed methods. In short and long-term forecasting of crude oil, the proposed method surpasses the other methods along with reducing computation cost and overfitting problems of neural networks. According to

the authors, the proposed method to forecast crude oil prices provides guidance and insights for policymakers to make decisions.

A blending ensemble learning model is utilized by [7] to forecast WTI and Brent crude oil at different time series frequencies such as daily, weekly, and monthly. The author named their proposed model LKDSR comprised of five machine learning models i.e., linear regression, k-nearest neighbor (KNN) regression, decision tree regression, SVR, and ridge regression. This is known as blending which employs a second-order ensemble method. The author argues that their proposed approach will provide more reliable, accurate, and robust forecasts in contrast to individual models. Various accuracy measures are used to test the model performance along with employing Diebold- Mariano (DM) test to check the significance of performance accuracy. The author used a grid search technique to find the optimum hyperparameter of the model. The results show that their proposed model surpasses all individual models including deep learning-based ensemble models, obtaining an R-square value of 0.99. The authors say that their proposed method is useful for stakeholders and policymakers to make accurate forecasts of oil.

employs a second-order ensemble method known as blending.

LSTM neural networks are used by [8] to forecast WTI and Brent crude oil prices. The data span from February 10, 1986, to May 17, 2021, taken from the USA-based Energy Information Administration website. The proposed model is compared with the classical time series model ARIMA and Artificial Neural Network (ANN). The study highlights how major events, such as the Gulf Wars and the 9/11 terrorist attacks, affected crude oil prices and how well the LSTM model captured these fluctuations. The study concludes that LSTM neural networks are effective for forecasting crude oil prices, particularly for long-term predictions. However, the accuracy decreases with shorter forecast timescales, suggesting the need to consider additional factors for more precise short-term forecasts.

The Brent crude oil in Tanzania is forecasted by [9] through Box Jenkins methodology utilizing monthly data from January 2002 to February 2022. The author tested various ARIMA models whose orders are identified through PACF and ACF plots and made data stationary through first-order lag difference. Finally, ARIMA(0,1,1) outperformed other models in terms of AIC value. The author concludes that COVID-19 and the Ukraine war have a negative impact on crude oil prices. Furthermore, the ARIMA model using Box Jenkins methodology is applied by [10] using yearly data from 1946-2016 and forecasting the yearly price for the period 2017-2021 through the ARIMA(2,1,1) model. The ARIMA and SARIMA model for forecasting daily crude oil is compared by [11]. They use closed data of crude oil from January 27, 2020, to January 25, 2023, downloaded from Yahoo Finance. They conclude that both ARIMA(0,1,0) and SARIMA(0,1,0)(0,0,0)[7] has same performance accuracy.

This work seeks to build various autoregressive integrated moving averages (ARIMA) and machine learning models including ARNN, k NN, and random forest to forecast multi-step forecasting of daily WTI crude oil prices.

2. WEST TEXAS INTERMEDIATE (WTI) CRUDE OIL

WTI crude oil serves as a key benchmark for global oil prices. It is a sweet and light crude oil that is produced mainly in the Permian Basin of Texas and New Mexico. WTI crude oil is traded on the New York Mercantile Exchange (NYMEX) and is used as the reference price for many oil contracts and derivatives [12]. The prices of WTI crude oil are influenced by multiple factors like geopolitics, supply and demand, market sentiment, and weather. The descriptive statistics of WTI crude oil from February 13, 2014, to February 13, 2024, are presented in Table 1 in which the minimum WTI crude oil price is USD -37.63 per barrel which was recorded on April 20, 2020, due to oversupply in COVID-19 which caused an unprecedented drop in prices. The mean is greater than the median which indicates the positive skewness in data varify by skewness greater than zero however the kurtosis of exact 3 shows that the data is mesokurtic. The Augmented Dickey-Fuller (ADF) test statistic value is equal to -2.809 with a corresponding p-value (0.579) surpassing the level of significance 0.05 indicating that the null hypothesis failed to reject that the data is not stationary. Forecasting prices of WTI crude oil is difficult due to the complex and nonlinear nature of the factors involved, high volatility, and uncertainty. Accurate forecasting of WTI crude oil prices can help investors, traders, consumers, and policymakers, to make better decisions and optimize their strategies [13].

The right panel of Figure 1 shows the boxplot of daily WTI crude oil prices for each month. The boxplot shows that the median price is approximately equal for all months but March, May, and June show a significant number of outliers above the third quartile.

TABLE 1. Summary statistics of WTI crude oil

Min	Mean	Median	Max	Skewness	Kurtosis
-37.63	63.49	59.64	123.7	0.481	3

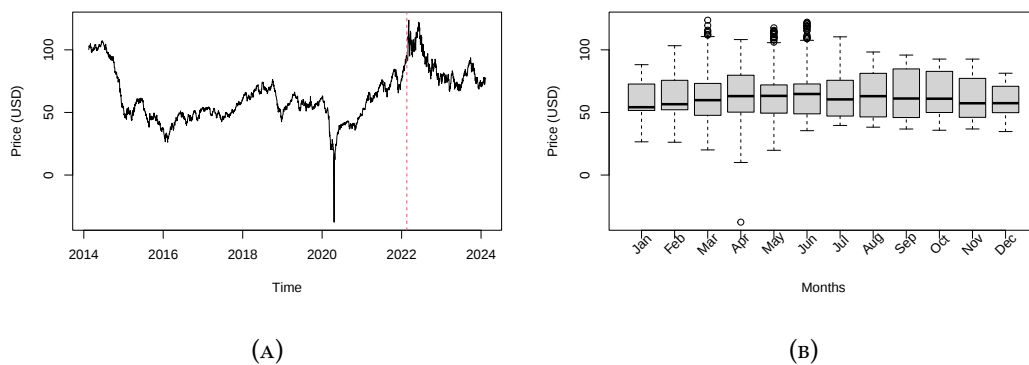


FIGURE 1. (A) Daily historical WTI crude oil prices from February 13, 2014 to February 13, 2024. The red dashed vertical line shows the border of training and validation set on February 18, 2022. (B) The boxplot shows the daily WTI Crude oil prices for each month for all study years.

TABLE 2. Augmented Dickey-Fuller test on level and first-order lag difference

	Dickey-Fuller	Lag order	p-value
Level	-1.998	12	0.579
First difference	-12.452	12	< 0.01

3. MATERIALS AND METHODS

In this section, the methodology to forecast the WTI crude oil price is described. The series is split into 80% training and 20% validation parts. The training set contains 2111 observations from February 13, 2014, to February 18, 2022. The validation set contains 529 observations from February 19, 2022, to February 13, 2024.

3.1. Autoregressive Integrated Moving Average (ARIMA). ARIMA is a composition of autoregressive p and moving average q term along with integrated term d which is a lag difference of series that makes the data stationary therefore the autocorrelation, mean and variance remain constant over time [14]. The coefficients of these parameters are estimated through the least square. It is a specialized model developed for forecasting time series data. The assumption of the model includes stationarity of the series, the linear dependency of future observation on past values, and error term ϵ . The mathematical model of ARMA is obtained by combining the autoregressive and moving average term as shown in Equation 3.1.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 y_{t-1} + \theta_2 y_{t-2} \cdots + \theta_q \epsilon_{t-q} + \epsilon_{t(q)} \quad (3.1)$$

Where y_t is the price of crude oil at time t , c is the intercept of the model, ϕ_i is the parameter of autoregressive, θ_i is the parameter of moving average, $(\epsilon_t, \epsilon_{t(q)})$ is the random error. The orders p and q are determined through the PACF, and ACF plots respectively, selecting the lags where the correlation spikes are greater than 0.05. The optimum order of ARIMA is chosen through the trial and error method. [15]

3.2. Random forest. A machine learning algorithm Random Forest developed by Leo Breiman used for regression and classification tasks [16]. During the training process Random Forest functions by creating numerous decision trees and outputting the average of the predictions of the individual trees as shown in Figure 2. This ensemble approach reduces variance and improves the robustness of the model. The training data is randomly selected with replacement to generate several subsets, each of which trains an individual decision tree. This method is known as bootstrap aggregation or bagging, which helps in reducing overfitting and variance [17].

For time series forecasting, the mathematical foundation of Random Forest involves constructing multiple decision trees $\{T_1, T_2, \dots, T_B\}$, where B is the total number of trees. Each tree T_b is built using a bootstrap sample of the data and a random subset of features at each split.

The forecast for a novel data point is determined by averaging predictions generated by each tree in the ensemble. Mathematically, the Random Forest's prediction for a specific input $\mathbf{x}$ can be formulated as:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(\mathbf{x})$$

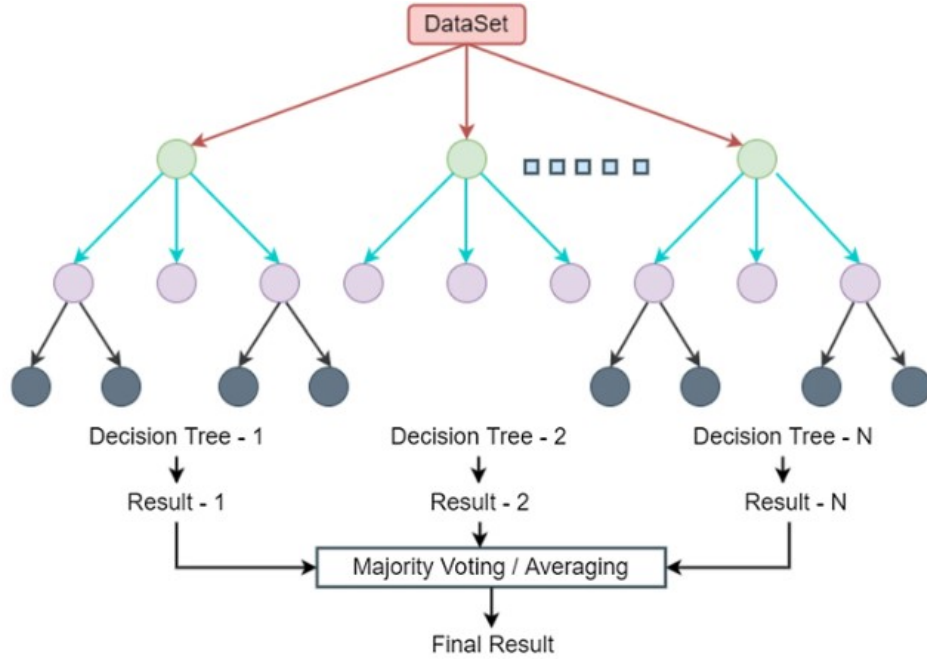


FIGURE 2. Random Forest tree representation. (Source: [7])

In one-step forecasting, the model utilizes the current observation y_t and the prior observation y_{t-1} as input features, while $T_b(\mathbf{x})$ denotes the prediction made by the b -th tree. For seven-step ahead forecasting, the model includes y_t , y_{t-1} , and y_{t-7} as features. The input feature vector $\mathbf{x}_t$ for forecasting can be represented as:

$$\mathbf{x}_t = [y_{t-1}, y_{t-7}]$$

Here, y_{t-1} is the lagged value from one step prior, and y_{t-7} is the lagged value from seven steps prior. Several decision trees are trained using various bootstrap samples derived from the training dataset. Each tree acquires a segment of the pattern within the time series. The model learns to predict the future value y_{t+1} for one-step ahead and y_{t+7} for seven-step ahead forecasting. For one-step ahead forecasting, the forecasted value $\hat{y}_{t+1}$ is:

$$\hat{y}_{t+1} = \frac{1}{B} \sum_{b=1}^B T_b([y_t, y_{t-1}])$$

For seven-step ahead forecasting, the predicted value $\hat{y}_{t+7}$ is:

$$\hat{y}_{t+7} = \frac{1}{B} \sum_{b=1}^B T_b([y_t, y_{t-1}, y_{t-7}])$$

The main hyperparameters including the number of trees ($n_estimators$), the maximum depth of each tree (max_depth), and the number of features considered for splitting ($max_features$) are tuned using cross-validation to optimize the model's performance.

When Random Forest is used for daily WTI crude oil price forecasting, it performs optimized in ensemble learning and decision trees. By including lagged values y_{t-1} and y_{t-7} as features, the model successfully captures temporal dependencies for both one-step and seven-step ahead forecasts. For time series analysis in financial and commodity markets, this methodology is useful by providing accurate and reliable forecasting.

3.3. Autoregressive neural network (ARNN). ARNN effectively captures the temporal dependency and pattern that exist in data when used for time series data such as WTI crude oil prices. It is a powerful machine learning algorithm for modeling non-linear and complex relationships in data [18]. The basic architecture of a neural network consists of an input layer, several hidden layers, and an output layer. The network of connected nodes also known as the neuron builds a layer of neural networks to evaluate the input data and get accurate forecasts. During the training process, the forecasting error is reduced by continuously changing the weights that are assigned to the connections between neurons [19].

The lag value of the series is received as an input value by a neural network. The current value at time t and one order lag value at $t - 1$ is received as input value by the algorithm for one step ahead forecasting while the current value at time t and seven order lag value $t - 7$ is received as an input value for seven-step ahead forecasting by neural network. The input feature vector $\mathbf{x}_t$ may be expressed mathematically as follows:

$$\mathbf{y}_t = [y_{t-1}] \quad (\text{for one-step ahead})$$

$$\mathbf{y}_t = [y_{t-1}, y_{t-7}] \quad (\text{for seven-step ahead})$$

The input data is modified by a hidden layer to apply a consecutive weighted sum throughout the activation function as shown in Figure 3. To understand the complex relationship in the model the non-linear aspect is introduced by activation function. The final output is also processed by an activation function to induce non-linearity, and the linear output is set to FALSE. The activation of the j -th neuron in the hidden layer, denoted as h_j , is determined by the weight w_{ij} connecting the i -th input to the j -th hidden neuron, and the bias b_j of the j -th hidden neuron. This can be formulated as follows:

$$h_j = f\left(\sum_i w_{ij}x_i + b_j\right)$$

Here, f is denoted as an activation function such as the ReLU (Rectified Linear Unit).

The forecasts are provided by the output layer. For regression tasks, this is typically a linear combination of the activations from the hidden layer.

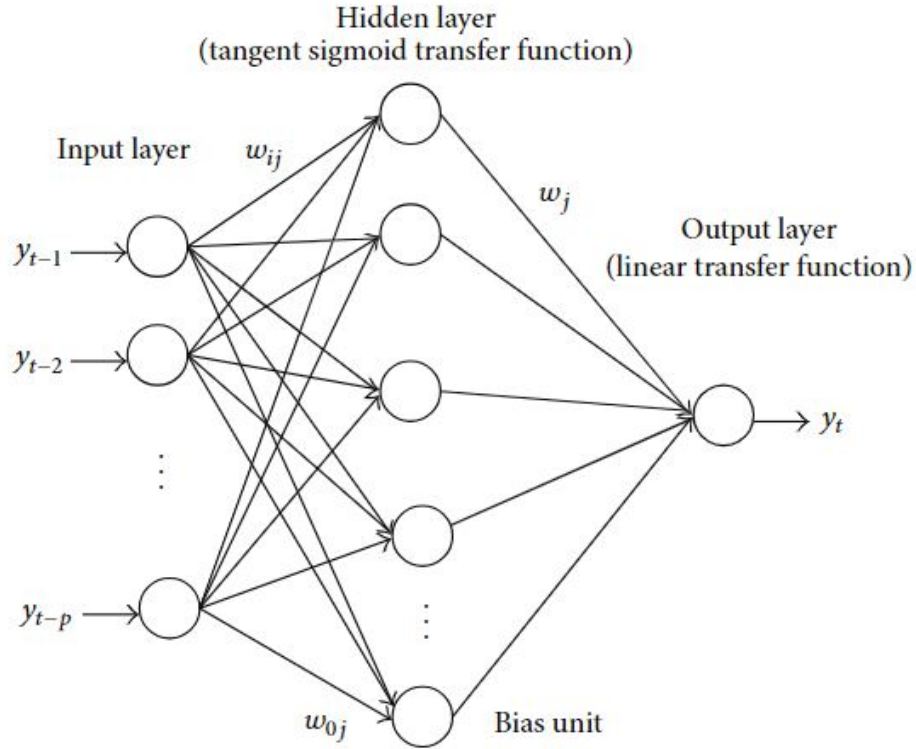


FIGURE 3. The ARNN architecture comprises of three layers i.e., Input, hidden, and output layer along with each weight and bias. (Source: [20])

Let $\hat{y}_{t+k}$ represent the forecasted value for k -step ahead, w_{jk}^{out} the weight between the j -th hidden neuron and the output neuron, and b^{out} the bias of the output neuron [21]. The output can be expressed as:

$$\hat{y}_{t+k} = \sum_j w_{jk}^{\text{out}} h_j + b^{\text{out}}$$

The Min-Max transformation is used to scale the data by normalizing the input feature in the $[0,1]$ range. This normalization enhances the training speed and increases the efficiency of the model. For one step and seven steps ahead forecasting the neural network configuration is made of two hidden layers of five and four neurons. cross-validation technique is employed to choose these parameters and hidden layers to maximize the performance of the algorithm.

To reduce the overfitting in the training output the training process is set with minimum life sign to concentrate the model on the important features only. Therefore the temporal relationship can effectively capture and provide accurate forecasting by scaling the data, developing a neural network model by choosing optimum hidden layers, and lag setup.

3.4. ***k*-Nearest Neighbour (*k*NN)**. *k*NN algorithm is a simple, non-parametric, and robust work efficiently for both regression and classification. Based on how similar the previous observation is, *k*NN forecasts the future unknown observation in a time series. The methodology of *k*NN to predict the outcome involves finding the *k* neighbors (similar instances) of a particular query instance in the training set and combining their outputs shown in Figure 4. A prediction in most regression tasks is made by calculating the weighted average of response value utilizing the *k* closest neighbor. This method uses distance measures such as Euclidean distance to evaluate the similarity between data [22].

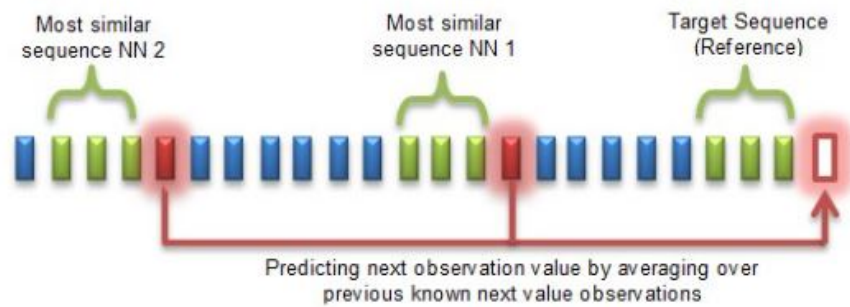


FIGURE 4. Calculating nearest neighbors by *k*NN and forecast future value based on their average. (Source: [23])

*k*NN algorithm, For a given query point $\mathbf{q}$, identifies the *k* training instances that are closest to $\mathbf{q}$ based on the calculated distances. The prediction $\hat{y}$ for the query point $\mathbf{q}$ is computed as the average value of the desired outcomes for the nearest *k* neighbors, considering their respective weights. The weights generally vary in inverse relation to the distances. [24].

$$\hat{y} = \frac{\sum_{i=1}^k \frac{1}{d(\mathbf{q}, \mathbf{x}_i)} y_i}{\sum_{i=1}^k \frac{1}{d(\mathbf{q}, \mathbf{x}_i)}}$$

Here, y_i is the target value of the *i*-th nearest neighbor, and $d(\mathbf{q}, \mathbf{x}_i)$ is the distance between the query point $\mathbf{q}$ and the *i*-th nearest neighbor $\mathbf{x}_i$.

The *k*NN model is built using the training set. The value of *k* is varied from 1 to 12, and the combination function is set to weighted, meaning that the influence of each neighbor is inversely proportional to its distance from the query point. The Euclidean distance is used to calculate the distance matrix, which determines the similarity between different observations. The model is applied recursively to forecast one-step and seven-step ahead values. This means that for each prediction, the model uses previously forecasted values as inputs for subsequent predictions.

To forecast y_{t+1} , the input feature vector is:

$$\mathbf{x}_t = [y_t, y_{t-1}, \dots, y_{t-(n-1)}]$$

The k NN model identifies the k nearest neighbors of $\mathbf{x}_t$ and uses their target values to predict y_{t+1} . For seven-step ahead forecasting a lag value of order seven is needed to incorporate the input vector feature.

$$\mathbf{x}_t = [y_t, y_{t-1}, y_{t-7}, \dots, y_{t-(n-1)}]$$

The k NN model forecasts the y_{t+1} and y_{t+7} using recursive forecasting method. These steps allow the k NN model to reliably forecast daily WTI crude oil prices for both one-step and seven-step ahead. Recursive forecasting updates the input feature set with newly forecasted values regularly, ensuring that the model can handle multiple step-ahead forecasts.

4. OUT-OF-SAMPLE FORECASTING

The 529 training set observations are forecasted through ARIMA and machine learning models. The forecasting accuracy of the model is evaluated through mean absolute percentage error (MAPE), root mean square error (RMSE), mean percentage error (MPE), and mean absolute error (MAE). Figure 5 shows the stationary time series made through first-order lag difference and the second row of the figure shows the ACF and PACF plot of the stationary data. The order of ARMA is identified through these plots.

Table 3 shows the forecasting accuracy of the models on testing data in which ARIMA(4,1,2) outperformed machine learning models on both short and intermediate-term forecasting. The MAPE of ARIMA(4,1,2) is 2.066 for one-step-ahead forecasting and 3.838 for seven-step-ahead forecasting. The MAPE shows the percentage deviation of the forecasted value from the true price. The negative sign in MPE shows that the forecasted value is overestimated on average. The barplot of MAPE of each model is shown in Figure 6.

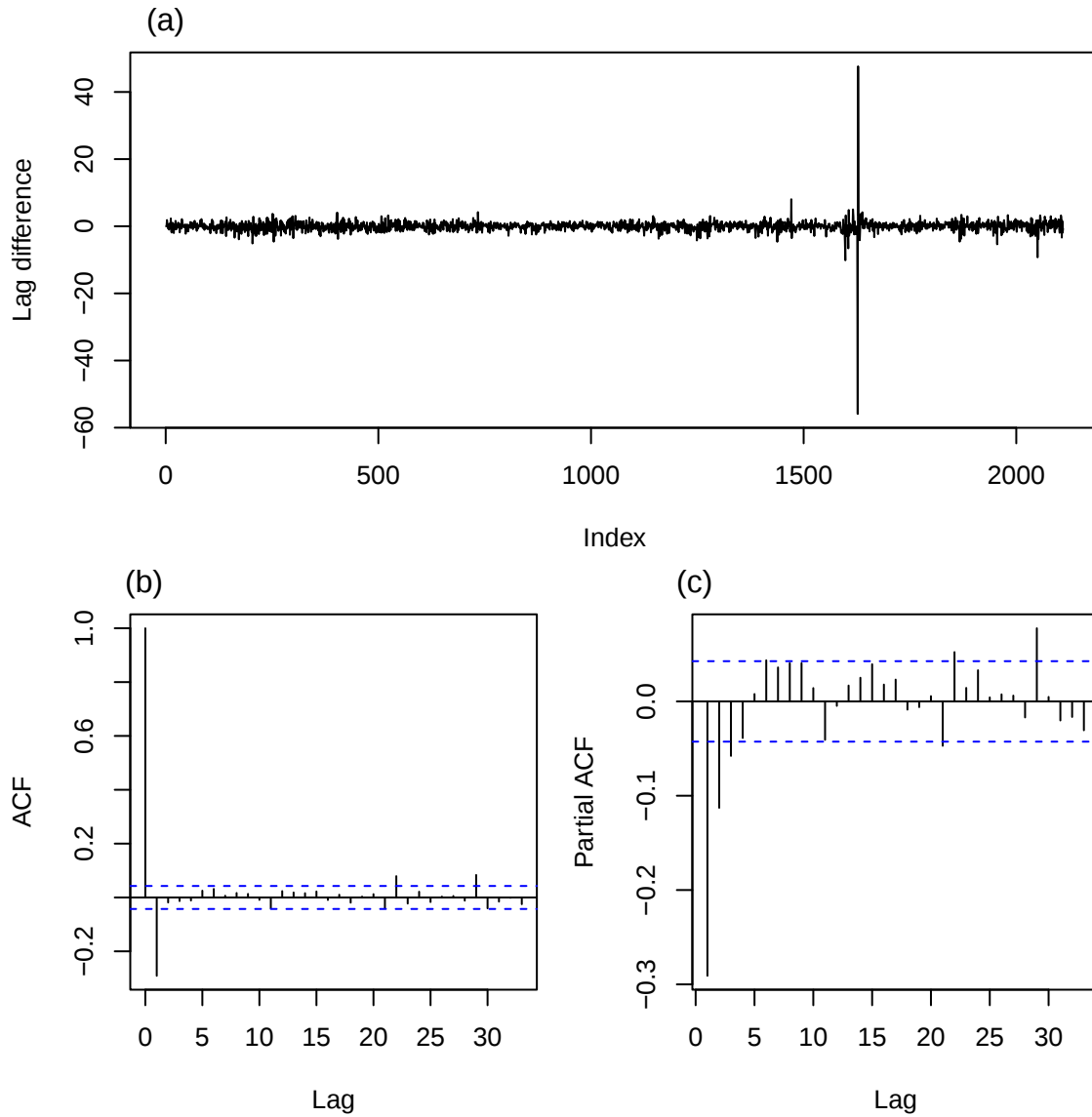


FIGURE 5. (a) The plot shows the stationary series of daily WTI crude oil through first-order lag difference (b) Autocorrelation function (ACF) plot (c) Partial autocorrelation function (PACF) plot.

TABLE 3. Testing error of forecasting models on one step and seven steps ahead forecasting

Models	One-step ahead forecasting				Seven-step ahead forecasting			
	RMSE	MAE	MPE	MAPE	RMSE	MAE	MPE	MAPE
ARIMA(4,1,2)	2.5	1.809	-0.095	2.066	4.662	3.327	-0.2	3.838
kNN	2.874	2.214	0.628	2.317	4.797	3.431	-0.121	3.985
ARNN	2.87	2.012	0.15	2.221	5.512	4.147	0.229	4.762
RF	3.33	2.214	0.628	2.43	6.219	4.731	0.988	5.444

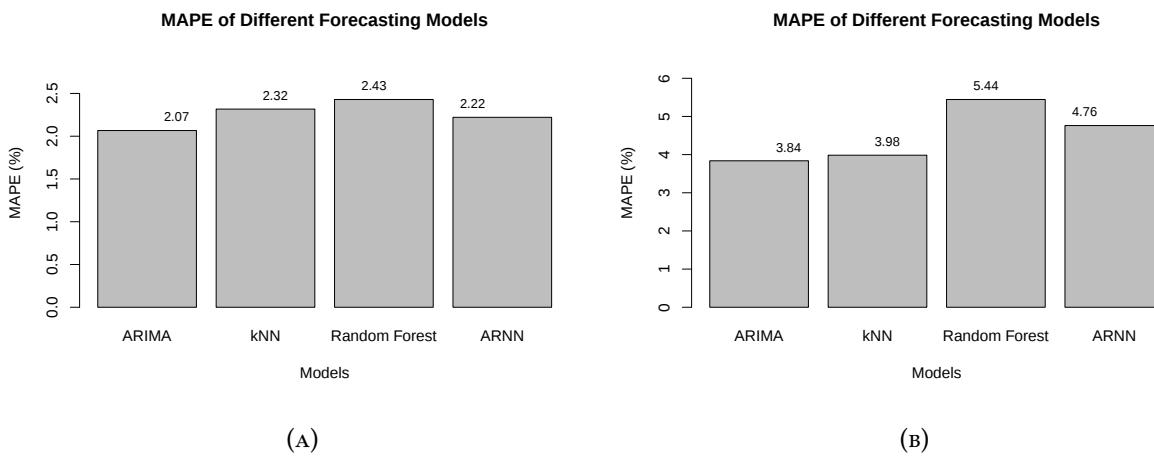


FIGURE 6. (a) MAPE of different models on one-step ahead forecasting (b) MAPE of different models on seven-step ahead forecasting

The adjacent bar plot in Figure 7 and 8 shows the MAPE of each model in each month for both one-step and seven-step ahead forecasting. The forecasting accuracy of models in March and June is poor compared to other months. This is because of the outliers in that month as shown in the right panel of Figure 1.

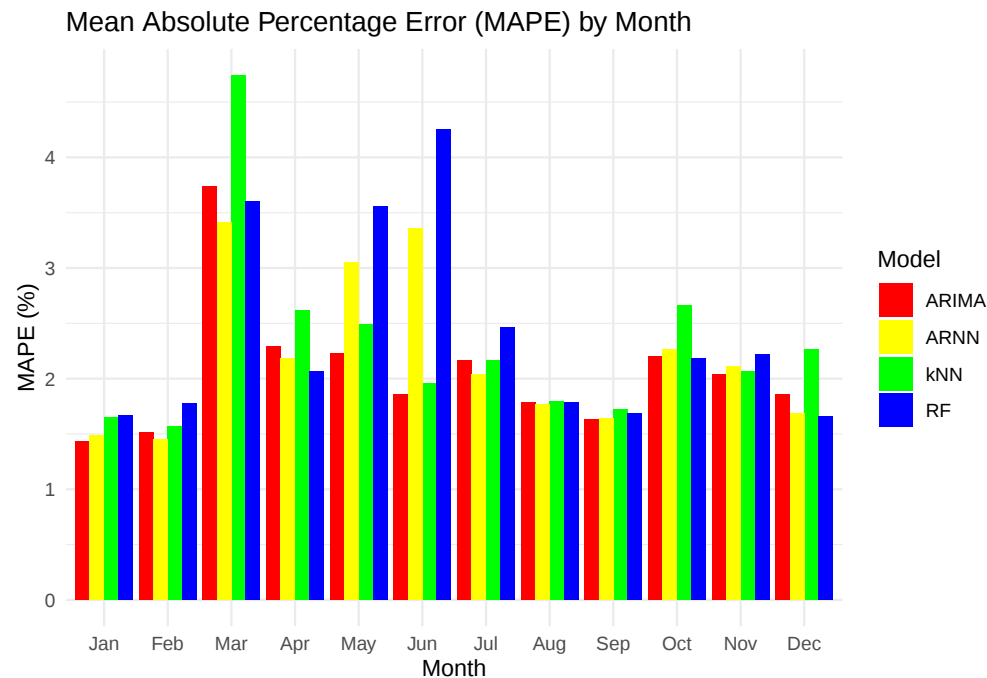


FIGURE 7. Barplot of MAPE for one step ahead forecasting at each month

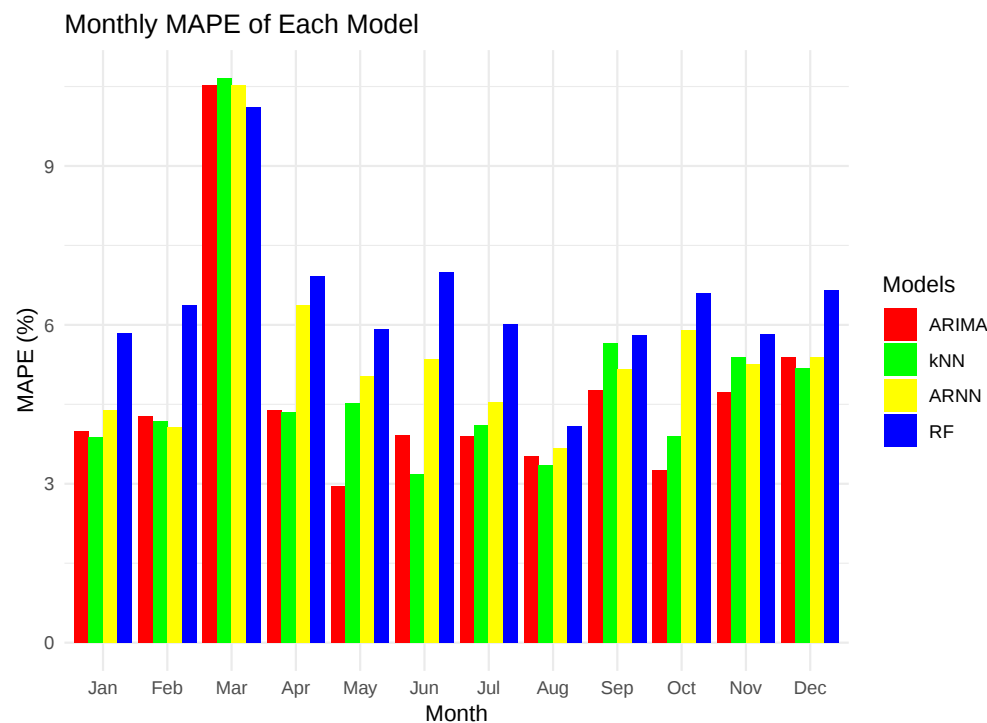


FIGURE 8. Barplot of MAPE for seven steps ahead forecasting at each month

The MAPE of each model on each day of the week is visualized in Figure 9 and 10 which indicate that the model's performance does not vary significantly on each day of the week for both one step and seven step ahead forecasting. For one step ahead forecasting the MAPE of models on Sunday is relatively small but it is due to a small number of observations recorded on Sunday.

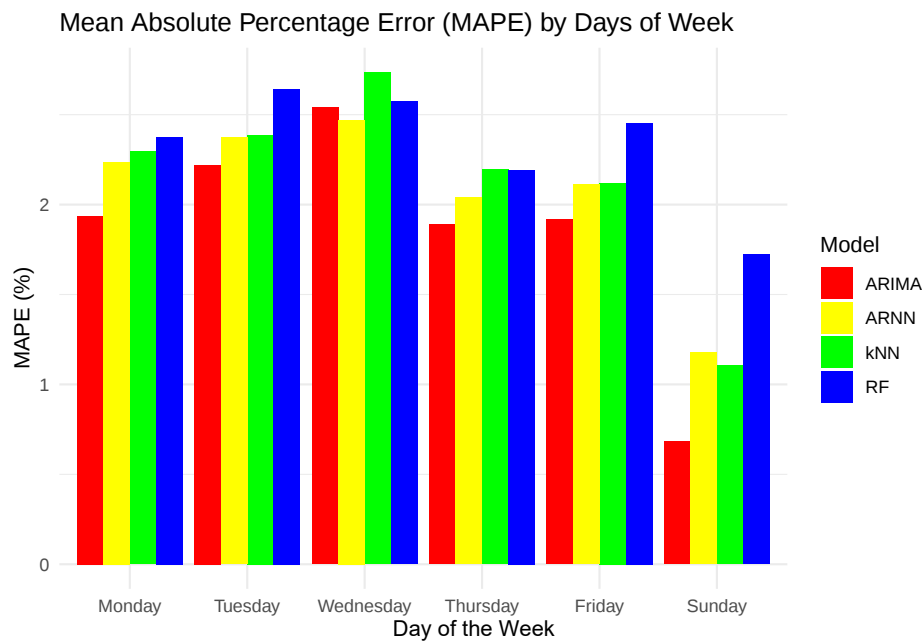


FIGURE 9. Barplot of MAPE for one step ahead forecasting at each day of the week

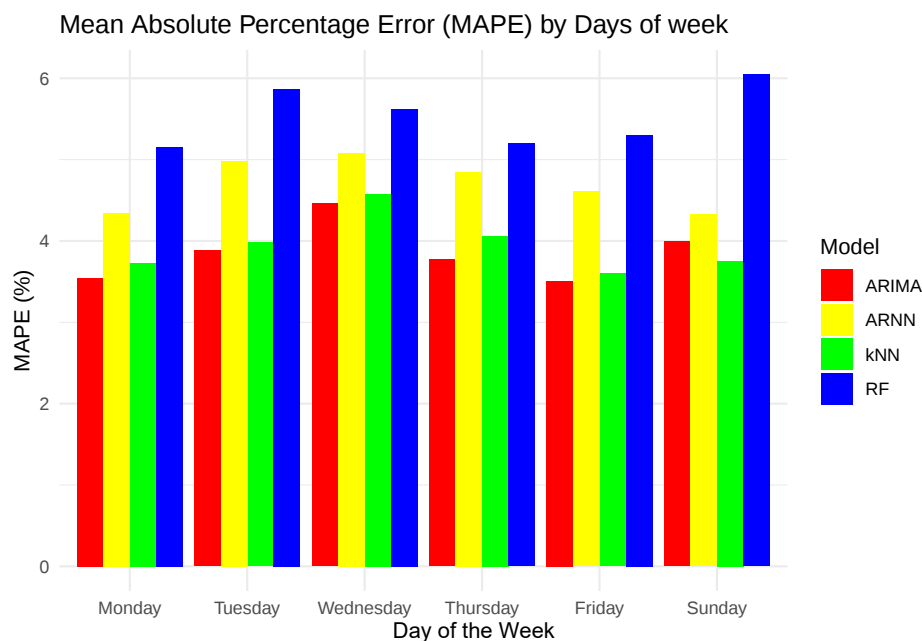


FIGURE 10. Barplot of MAPE for seven steps ahead forecasting at each day of week

The assumption of the ARIMA model is tested through residual analysis. The time series plot of ARIMA, their ACF, and PACF plot for one-step and seven-step ahead forecasting by ARIMA is shown in Figure 11 and 12 respectively. The figures show that the residuals are stationary with mean zero and most of the spikes on different lag in ACF and PACF plots are below the threshold line of 0.05 therefore we conclude that there is no problem of autocorrelation in residuals.

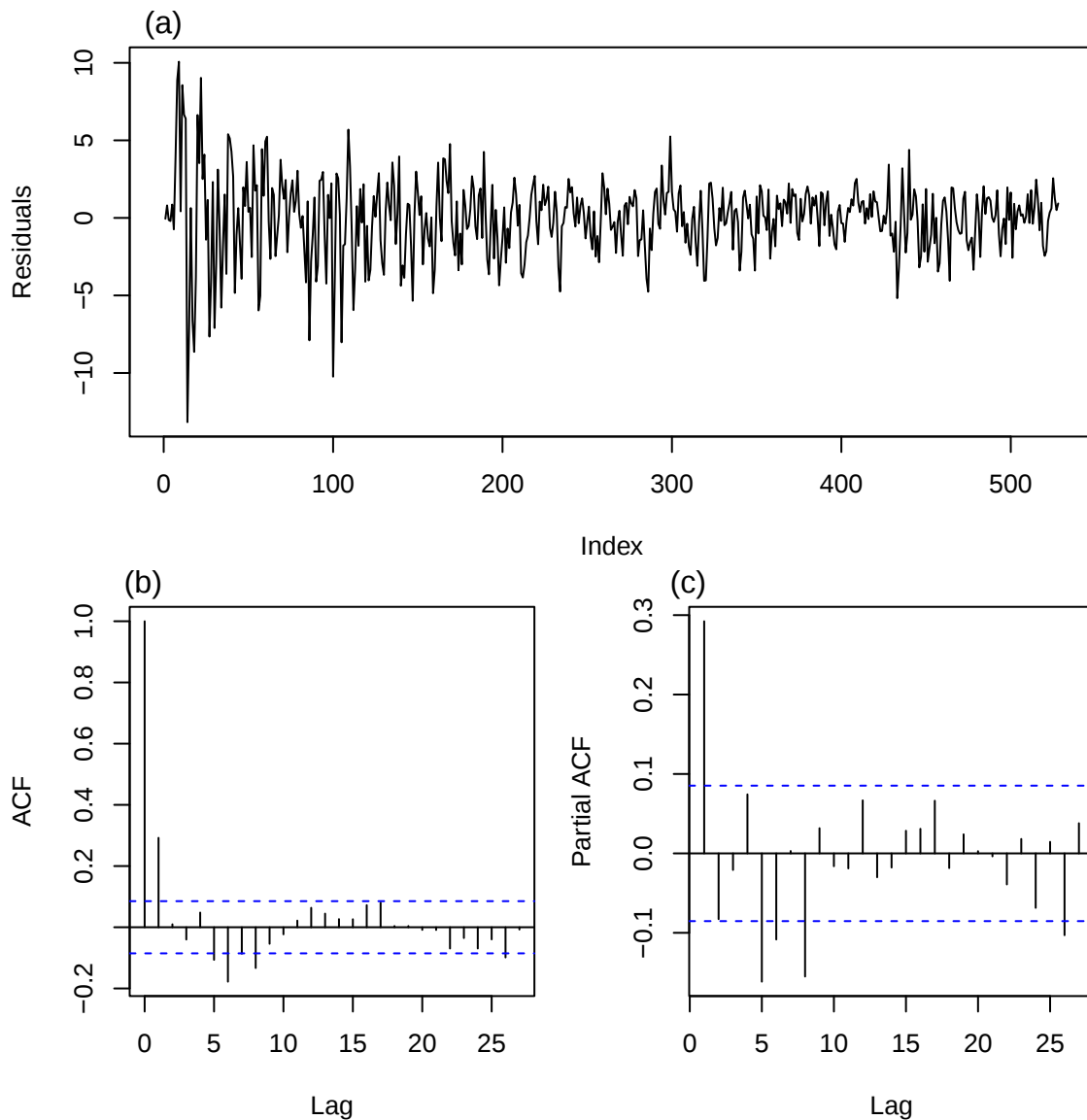


FIGURE 11. (a) Residuals of ARIMA (b) ACF plot of ARIMA residuals (c) PACF plot of ARIMA residuals.

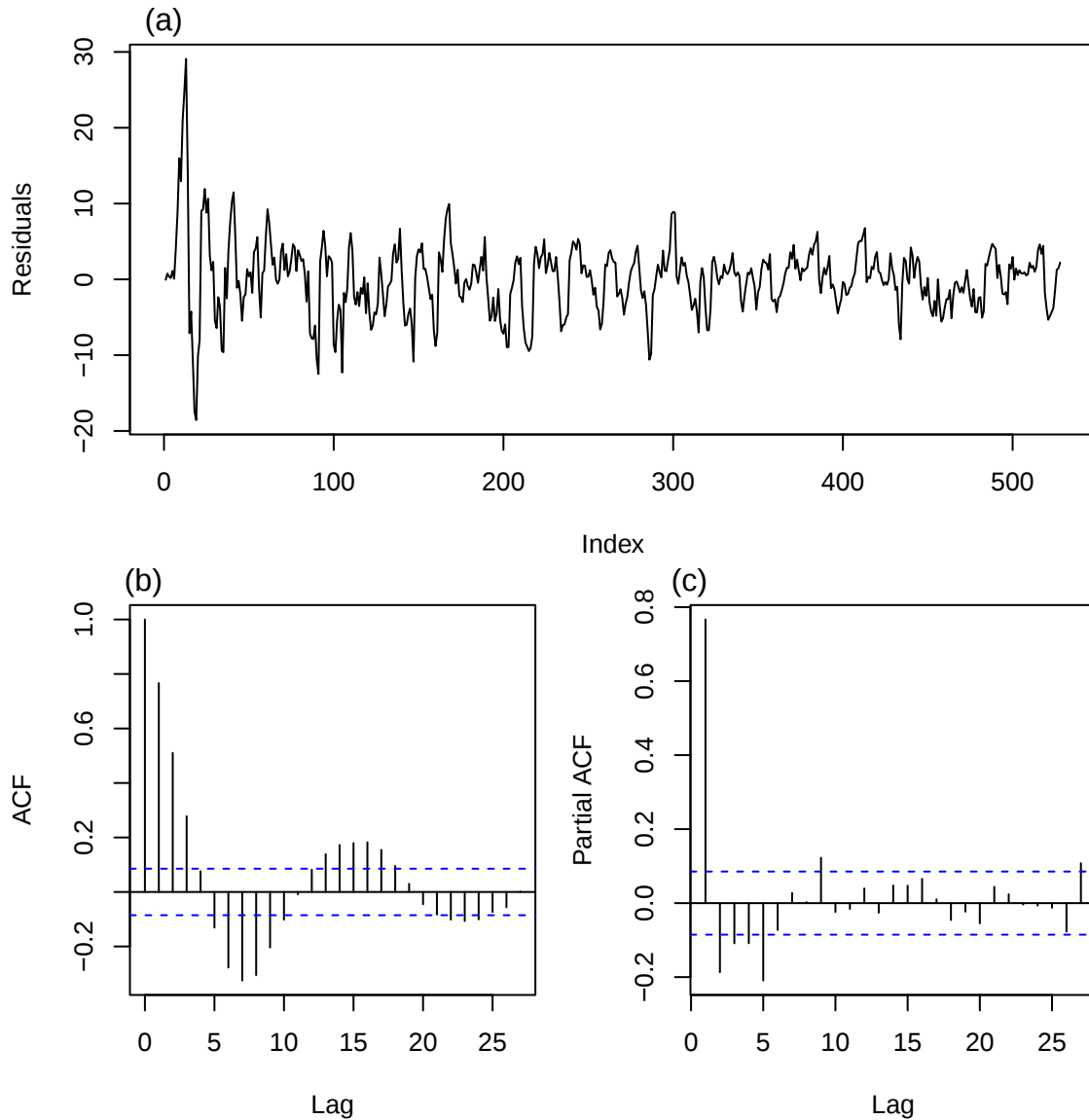


FIGURE 12. (a) Residuals of ARIMA of seven steps ahead forecasting (b) ACF plot of ARIMA residuals of seven steps ahead forecasting (c) PACF plot of ARIMA residuals of seven steps ahead forecasting.

The one-step and seven-step ahead forecasted values for the validation set, as forecasted by each model, are shown in Figures 13 and 14, respectively. The forecasted values are very close to the observed values. However, for the seven-step ahead forecasting depicted in Figure 14, the random forest model (blue dashed line) provides a poor forecast of the WTI crude oil prices, exhibiting high fluctuations in its predictions.

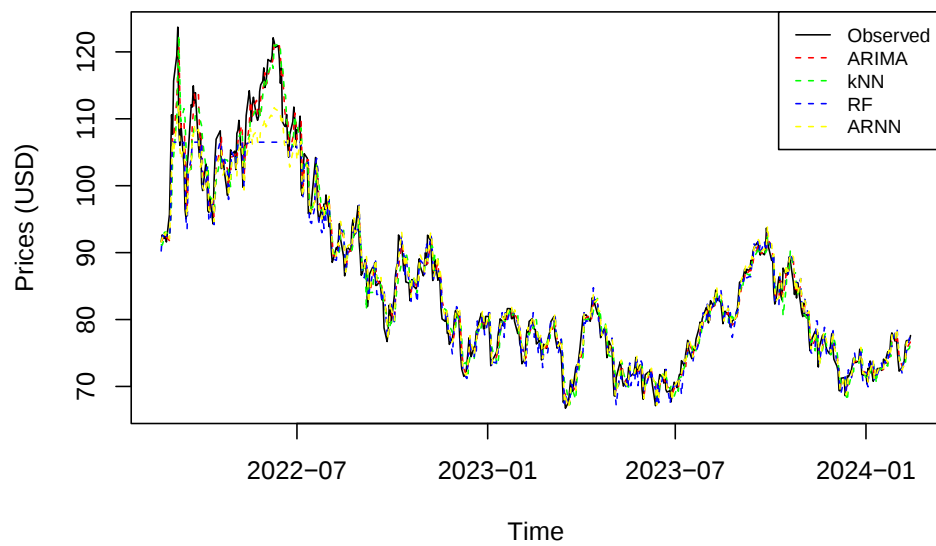


FIGURE 13. Forecasted value of daily WTI crude oil of validation set

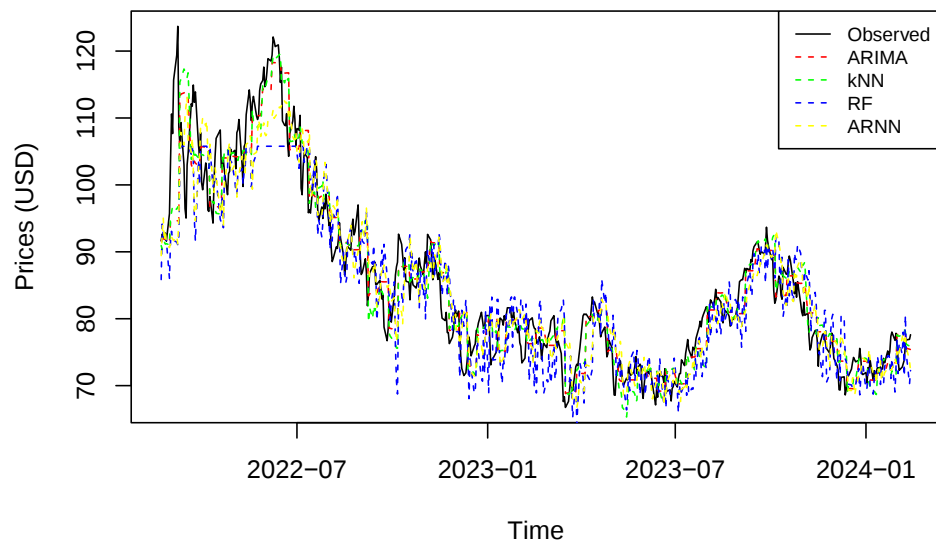


FIGURE 14. Seven steps ahead forecasted value of daily WTI crude oil of validation set

To enhance clarity in comparing the forecasting performance of each model, the observed and true values for the last 30 days, from January 4, 2024, to February 13, 2024, are visualized in Figure 15 for one-step ahead forecasts and in Figure 16 for seven-step ahead forecasts. The plots show

that for one-step ahead forecasting, the ARIMA(4,1,2) and ARNN models produce forecasts closer to the observed values compared to the k NN and RF models. For seven-step ahead forecasting, the ARIMA(4,1,2) and k NN models provide forecasts nearer to the observed values than the ARNN and RF models.

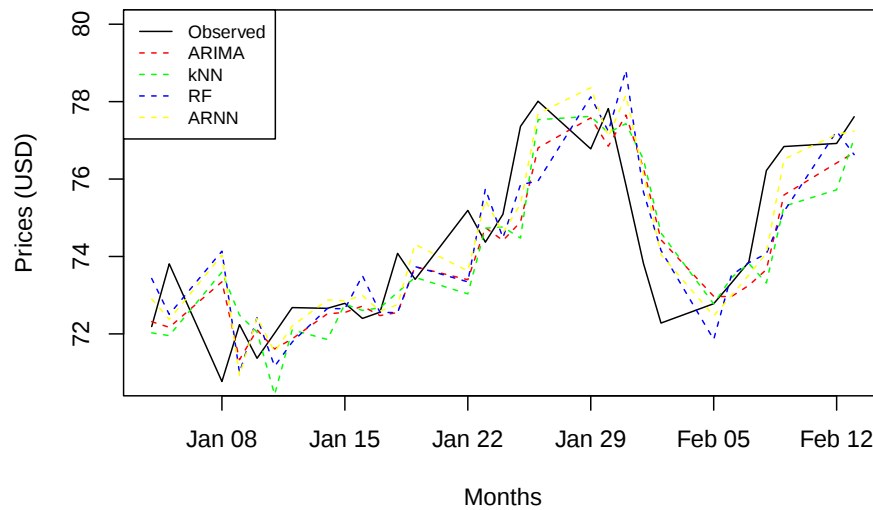


FIGURE 15. Forecasted and observed value of last thirty days of validation set

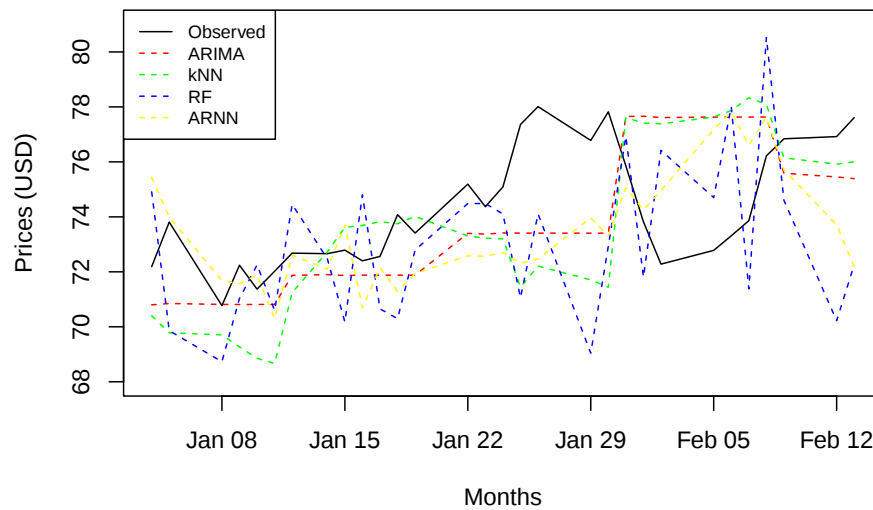


FIGURE 16. Seven steps ahead forecasted and observed value of last thirty days of validation set

The forecasting accuracy of the models is compared through the Diebold-Mariano test. The alternative hypothesis of the test statistics is that the model in columns is superior to the model in rows. The P-values of the test statistic are shown in Table 4 for both one-step and seven-step ahead forecasting. The test concludes that ARIMA(4,1,2) has higher accuracy than the machine learning algorithms for both one-step and seven-step ahead forecasting. In machine learning algorithms for short-term forecasting, ARNN and RF perform better than k NN. For intermediate-term forecasting

TABLE 4. P-values of Diebold-Mariano test

Models	One-step ahead forecasting				Seven-step ahead forecasting			
	ARIMA	RF	k NN	ARNN	ARIMA	RF	k NN	ARNN
ARIMA		0.999	1.000	0.996		1.000	0.955	1.000
RF	0.000		0.022	0.000	0.000		0.000	0.000
k NN	0.000	0.978		0.491	0.045	1.000		1.000
ARNN	0.004	1.000	0.509		0.000	1.000	0.000	

5. DISCUSSION

In this work, we develop various ARIMA and machine learning models to compare the forecasting accuracy of classical and machine learning approaches for multi-step ahead forecasting of daily WTI crude oil prices. The descriptive statistics and visualizations indicate that WTI crude oil prices exhibit a non-linear structure, suggesting that machine learning models should theoretically capture these complex relationships more effectively than the linear ARIMA model. However, our results show that ARIMA significantly outperforms the non-linear machine learning models. This outcome can be attributed to several factors. ARIMA is a specialized linear model designed explicitly for time series analysis, relying on the assumption that future observations depend on past observations, which is a fundamental characteristic of time series data [25]. In contrast, machine learning models, which are generally developed for classification and regression tasks, do not inherently make this assumption. Crude oil prices exhibit significant fluctuation and are impacted by a multitude of external factors. Machine learning algorithms find it challenging to grasp this level of volatility, especially if the noise in the data is high. In these scenarios, the ARIMA model might be more robust as it is designed for time series data with such characteristics.

The time series data is assumed to be stationary in ARIMA models. If the data is not stationary preprocessing techniques like differencing are used to bring the data into stationarity. Since machine learning models do not make the explicit assumption of stationarity, it can occasionally be difficult to understand the underlying patterns. To identify the pertinent patterns in the data, machine learning models frequently need to do substantial feature engineering. If the features are poorly designed, the model cannot function as intended [26]. On the other hand, ARIMA models do not require feature engineering as they directly model the time series data. Machine learning

models require careful hyperparameter tuning, which can be challenging and time-consuming. If the hyperparameters are not optimized properly, the model performance may suffer. ARIMA models have fewer parameters to tune, making the optimization process simpler.

According to the data analysis, there are a lot of outliers between March and June that have a big impact on how accurate the ARIMA model is at forecasting the price of WTI crude oil. In particular, the ARIMA model shows a significant drop in forecasting accuracy over these months. The finding draws attention to the ARIMA model's significant flaw: its vulnerability to outliers. The model performs poorly when these outliers are present because they make it more difficult for the model to recognize and forecast the underlying patterns in the time series. This conclusion emphasizes the need for strong techniques to manage outliers when using ARIMA models for time series forecasting, especially for datasets with high volatility and outliers. [27].

6. CONCLUSIONS

In this article, we compare the forecasting accuracy of machine learning models and ARIMA models by forecasting daily WTI crude oil prices through both one-step and seven-step ahead forecasts. The presence of unit root in data is tested through the Augmented Dickey-Fuller (ADF) test in which it is concluded that the data is non-stationary and the presence of unit root. Using first-order lag difference the series is made stationary to meet the assumption of ARIMA. Multiple ARIMA models are built and tested whose orders are determined through a combination of Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) plots, along with trial-and-error methods. Additionally, we developed machine learning models including k -Nearest Neighbors (k NN), Artificial Neural Networks (ARNN), and Random Forest (RF), with their parameters optimized using k -fold cross-validation.

Our results demonstrate that the ARIMA model outperforms the machine learning models in both one-step and seven-step ahead forecasting. Specifically, ARIMA showed superior accuracy, followed by k NN and ARNN. However, it is important to note that due to the presence of outliers in March and June, the forecasts are less reliable. This result emphasizes how reliable the ARIMA model is, especially when stationary data shows both linear and autoregressive dependencies.

Furthermore, our research emphasizes the challenges that machine learning models have in this scenario. The performance of machine learning algorithms can be negatively impacted by the existence of noise and the difficult, labor- and time-intensive process of parameter adjustment. As a result, even though machine learning models may capture non-linear correlations, careful parameter tuning and management of data anomalies are necessary for them to be successful. This study indicates that ARIMA is still a more dependable and effective option for forecasting when dealing with time series data that include large linear and autoregressive components.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

REFERENCES

- [1] D. Langevin, S. Poteau, I. Hénaut, J.F. Argillier, Crude Oil Emulsion Properties and Their Application to Heavy Oil Transportation, *Oil Gas Sci. Technol.* 59 (2004), 511–521. <https://doi.org/10.2516/ogst:2004036>.
- [2] K. Lang, B.R. Auer, The Economic and Financial Properties of Crude Oil: A Review, *N. Am. J. Econ. Financ.* 52 (2020), 100914. <https://doi.org/10.1016/j.najef.2019.01.011>.
- [3] M. Monge, L.A. Gil-Alana, F. Pérez de Gracia, Crude Oil Price Behaviour Before and After Military Conflicts and Geopolitical Events, *Energy* 120 (2017), 79–91. <https://doi.org/10.1016/j.energy.2016.12.102>.
- [4] J. Zhao, Exploring the Influence of the Main Factors on the Crude Oil Price Volatility: An Analysis Based on Garch-Midas Model with Lasso Approach, *Resour. Policy* 79 (2022), 103031. <https://doi.org/10.1016/j.resourpol.2022.103031>.
- [5] Z. Liu, Z. Ding, T. Lv, J.S. Wu, W. Qiang, Financial Factors Affecting Oil Price Change and Oil-Stock Interactions: A Review and Future Perspectives, *Nat. Hazards* 95 (2018), 207–225. <https://doi.org/10.1007/s11069-018-3473-y>.
- [6] J. Nasir, M. Aamir, Z.U. Haq, S. Khan, M.Y. Amin, et al., A New Approach for Forecasting Crude Oil Prices Based on Stochastic and Deterministic Influences of Lmd Using Arima and Lstm Models, *IEEE Access* 11 (2023), 14322–14339. <https://doi.org/10.1109/access.2023.3243232>.
- [7] M. Hasan, M.Z. Abedin, P. Hajek, K. Coussement, M.N. Sultan, et al., A Blending Ensemble Learning Model for Crude Oil Price Forecasting, *Ann. Oper. Res.* (2024). <https://doi.org/10.1007/s10479-023-05810-8>.
- [8] K. Zhang, M. Hong, Forecasting Crude Oil Price Using Lstm Neural Networks, *Data Sci. Financ. Econ.* 2 (2022), 163–180. <https://doi.org/10.3934/dsfe.2022008>.
- [9] L. Gasper, H. Mbwapbo, Forecasting Crude Oil Prices by Using Arima Model: Evidence from Tanzania, *J. Account. Financ. Audit. Stud.* 9 (2023), 158–175. <https://doi.org/10.32602/jafas.2023.017>.
- [10] J. Selvi, R.K. Shree, J. Krishnan, Forecasting Crude Oil Price Using ARIMA Models, *Int. J. Adv. Res.* 7 (2018), 334–343.
- [11] V.P. Ariyanti, Tristyanti Yusnitasari, Comparison of Arima and Sarima for Forecasting Crude Oil Prices, *J. RESTI* 7 (2023), 405–413. <https://doi.org/10.29207/resti.v7i2.4895>.
- [12] A. Almeida, A.A. Golpe, J.M. Martín-Alvarez, J.C. Vides, Brent Vs. West Texas Intermediate in the Us Petro Derivatives Price Formation, *Pet. Sci.* 21 (2024), 729–739. <https://doi.org/10.1016/j.petsci.2023.09.007>.
- [13] Y. Yang, T. Yang, S. Chen, C. Tong, Exploring the Non-Linearity of West Texas Intermediate Crude Oil Price from Exchange Rate of Us Dollar and West Texas Intermediate Crude Oil Production, *Energy Strat. Rev.* 41 (2022), 100854. <https://doi.org/10.1016/j.esr.2022.100854>.
- [14] R.H. Shumway, D.S. Stoffer, R.H. Shumway, D.S. Stoffer, ARIMA Models, in: *Time Series Analysis and Its Applications: With R Examples*, pp. 75–163, (2017).
- [15] R.A. Ahmed, A.B. Shabri, Daily Crude Oil Price Forecasting Model Using ARIMA, Generalized Autoregressive Conditional Heteroscedastic and Support Vector Machines, *Am. J. Appl. Sci.* 11 (2014), 425–432. <https://doi.org/10.3844/ajassp.2014.425.432>.
- [16] L. Breiman, Random Forests, *Mach. Learn.* 45 (2001), 5–32. <https://doi.org/10.1023/a:1010933404324>.
- [17] S.J. Rigatti, Random Forest, *J. Insur. Med.* 47 (2017), 31–39.
- [18] A. Tealab, H. Hefny, A. Badr, Forecasting of Nonlinear Time Series Using Ann, *Futur. Comput. Inform. J.* 2 (2017), 39–47. <https://doi.org/10.1016/j.fcij.2017.05.001>.
- [19] M. Adil, R. Ullah, S. Noor, N. Gohar, Effect of Number of Neurons and Layers in an Artificial Neural Network for Generalized Concrete Mix Design, *Neural Comput. Appl.* 34 (2020), 8355–8363. <https://doi.org/10.1007/s00521-020-05305-8>.
- [20] A. Shabri, R. Samsudin, Daily Crude Oil Price Forecasting Using Hybridizing Wavelet and Artificial Neural Network Model, *Math. Probl. Eng.* 2014 (2014), 201402. <https://doi.org/10.1155/2014/201402>.

- [21] S. Sharma, S. Sharma, A. Athaiya, Activation Functions in Neural Networks, *Int. J. Eng. Appl. Sci. Technol.* 04 (2020), 310–316. <https://doi.org/10.33564/ijeast.2020.v04i12.054>.
- [22] F. Martínez, M.P. Frías, M.D. Pérez, A.J. Rivera, A Methodology for Applying K-Nearest Neighbor to Time Series Forecasting, *Artif. Intell. Rev.* 52 (2017), 2019–2037. <https://doi.org/10.1007/s10462-017-9593-z>.
- [23] F.H. Al-Qahtani, S.F. Crone, Multivariate K-Nearest Neighbour Regression for Time Series Data – A Novel Algorithm for Forecasting Uk Electricity Demand, in: *The 2013 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2013, pp. 1-8. <https://doi.org/10.1109/ijcnn.2013.6706742>.
- [24] Y. Lee, C. Wei, T. Cheng, C. Yang, Nearest-neighbor-based Approach to Time-Series Classification, *Decis. Support Syst.* 53 (2012), 207–217. <https://doi.org/10.1016/j.dss.2011.12.014>.
- [25] B.K. Nelson, Time Series Analysis Using Autoregressive Integrated Moving Average (ARIMA) Models, *Acad. Emerg. Med.* 5 (1998), 739–744. <https://doi.org/10.1111/j.1553-2712.1998.tb02493.x>.
- [26] R. Hossain, D. Timmer, Machine Learning Model Optimization with Hyper Parameter Tuning Approach, *Glob. J. Comput. Sci. Technol. D Neural Artif. Intell* 21 (2021), 6–13.
- [27] A.C. Petrica, S. Stancu, A. Tindecu, Limitation of Arima Models in Financial and Monetary Economics, *Theor. Appl. Econ.* 23 (2016), 19–42.