Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/190601

Machine Learning Techniques in a Hybrid Forecasting Model for Oil Prices Combining ARIMA

Imad Y. Hamid[1], *Sulima M. Awad Yousif*[2], *Saud Aljaloud*[3], *Abdelgalal O. I. Abaker*[4], *Halla Z. S. Elemam*[5], *Khalil A. Alruwaitee*[6,7], *and Manahill I. A. Anja*[8,*]

[1] Department of Mathematics, Faculty of Science and Humanities, Shaqra University, Dawadmi, Saudi Arabia
[2] Department of Basic Science, College of Preparatory year, Najran University, Najran 61441, Saudi Arabia
[3] Department of Computer science, College of Computer Science and Engineering, University of Ha'il, Saudi Arabia
[4] Department of Administrative Sciences, Applied College in Khamis Mushait, King Khalid University, Abha, Saudi Arabia
[5] Department of Administrative Sciences, Applied College in Abha, King Khalid University, Abha, Saudi Arabia
[6] Department of Economics and Finance, College of Business Administration, Taif University, P. O Box 11099, Taif 21944, Saudi Arabia
[7] Applied College, Taif University, P. O Box 11099, Taif 21944, Saudi Arabia
[8] Computer Sciences Program, Department of Mathematics, Turabah University College, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

Received: 15 Feb. 2024, Revised: 10 Apr. 2025, Accepted: 21 Aug. 2025
Published online: 1 Nov. 2025

Abstract: Accurately forecasting world oil prices is essential for formulating effective economic strategies and informed energy policies. While traditional time-series models such as the Autoregressive Integrated Moving Average (ARIMA) are widely used for capturing linear patterns in historical data, they often struggle with the nonlinear and complex dynamics characteristic of oil price movements. Conversely, Decision Tree models effectively detect nonlinear relationships but may lack the ability to model long-term dependencies. This paper proposes a hybrid forecasting model that combines ARIMA and Decision Tree techniques to leverage the strengths of both approaches. The ARIMA model is first employed to capture linear trends and produce baseline forecasts, which are then refined by Decision Tree algorithms to address residual nonlinearities. Experimental evaluations conducted on historical oil price data demonstrate that the hybrid ARIMA2013 Decision Tree model consistently outperforms its individual components in predictive accuracy. Model performance is assessed using statistical measures including Mean Absolute Percentage Error (MAPE) and Symmetric Mean Absolute Percentage Error (sMAPE). The results highlight the robustness and effectiveness of the proposed hybrid method in modeling the intricate behavior of global oil prices. This approach provides a reliable and comprehensive decision-support tool for policymakers and industry stakeholders. Future work may extend this framework to other domains where accurate time-series forecasting is critical.

Keywords: World Oil Price, Hybrid Decision Tree–ARIMA, Time-Series Forecasting, Statistical Criteria, sMAPE, MAPE

1 Introduction

The economic and social development of industrialized nations has long been accompanied by a growing dependence on raw commodities, particularly crude oil. As the backbone of modern energy systems and a key input in industrial production and transportation, oil remains a strategic resource in both developed and developing economies. Notably, the developing world now accounts for over 70% of global oil production, making the analysis of oil price dynamics a matter of global economic significance.

Crude oil prices are inherently volatile due to the complex interplay of global supply and demand, geopolitical tensions, speculation in financial markets, currency fluctuations—particularly the U.S. dollar—and exceptional events such as natural disasters or pandemics. Among these, the law of supply and demand remains the core economic principle driving oil price formation, yet it alone cannot fully explain the irregular fluctuations

* Corresponding author e-mail: meanja@tu.edu.sa

observed in real-world markets. This complexity highlights the importance of robust forecasting models capable of capturing both linear and nonlinear patterns in oil price behavior.

Accurate forecasting of crude oil prices is essential for strategic planning, risk management, and informed policy-making within the energy sector. The present study addresses this challenge by evaluating and comparing the performance of three forecasting approaches: the traditional Autoregressive Integrated Moving Average (ARIMA) model, the Decision Tree (DT) algorithm, and a hybrid ARIMA–Decision Tree model. The hybrid method is designed to leverage the strengths of both linear and nonlinear modeling, offering a more comprehensive predictive framework.

To structure the analysis, Section 2 provides a review of relevant literature on time-series forecasting and hybrid models. Section 3 introduces the forecasting methodologies used in this study, including both individual and hybrid approaches. Section 4 presents the experimental setup, results, and a detailed discussion of model performance. Finally, Section 5 concludes the paper with a summary of the main findings and suggestions for future research directions.

2 Literature Review

Given the critical role of oil prices in shaping inflation and overall economic activity, accurate forecasting of oil prices is indispensable for policymakers. However, developing models that consistently outperform the random walk benchmark over policy-relevant time horizons remains a challenging task [1]. Establishing a reliable forecasting model for crude oil prices is essential. In this study, we initially aimed to develop such a dependable predictive model for crude oil prices [2]. Fluctuations in crude oil prices have profound effects on global economies, making accurate price forecasting crucial for mitigating the risks associated with oil price volatility. Reliable forecasts are vital for a wide range of stakeholders, including governments, public and private institutions, policymakers, and investors [3]. Forecasting the price trajectory and volatility of crude oil has long presented a significant challenge for traders in the oil markets. As the world's primary energy source, crude oil serves as a crucial pillar of the global economy [4]. The volatility of oil prices has attracted considerable attention from scholars, traders, and regulators alike. Previous studies have established a strong correlation between crude oil prices and real GDP, with many researchers arguing that incorporating energy prices can enhance the accuracy of real GDP measurements [5]. Given the critical role of crude oil in a nation's economic development, crude oil futures are closely linked to various other markets. Accurate forecasting provides investors with reliable guidance for decision-making. In response to the rapid evolution of these markets, numerous studies have explored the development of novel social network metrics to improve forecasting models. The authors propose an integrated approach that combines investor sentiment and interest to enhance the prediction of crude oil futures [6]. The global energy crisis has introduced an unprecedented level of volatility and uncertainty to financial markets, significantly impacting the stock prices of companies in the energy sector. Accurate prediction of stock price trends during such turbulent periods is crucial for informed decision-making by investors, policymakers, and industry stakeholders [7]. Forecasting oil prices is a complex task due to the multitude of influencing factors and the inherent volatility of the market. Achieving a comprehensive understanding of oil price fluctuations often requires combining machine learning techniques with traditional statistical models such as GARCH and ARIMA, alongside methods like neural networks, random forests, or gradient boosting. Hybrid or ensemble approaches that integrate the strengths of different models generally provide more accurate and robust forecasts. Moreover, it is essential to acknowledge the limitations and uncertainties inherent in any forecasting methodology to support informed decision-making [8]. The increasing reliance on petroleum is primarily due to the lack of alternative energy sources capable of replacing it. To meet domestic oil demand under these circumstances, effective strategies and policies are essential. This study employs historical data from 1990 to 2022 to forecast Indonesia's oil exports, production, and consumption for the period 2023–2027, using the ARIMA modeling approach [9]. Price forecasting plays a crucial role in the fields of economics and finance. This article presents a comparative analysis of gasoline and diesel price trends in Ghana using ARIMA and SARIMA methodologies [10]. This study suggests that while ARIMA models are effective for short-term forecasting, their limitations—such as sensitivity to non-stationary data—highlight the need for future research to incorporate macroeconomic variables or hybrid modeling approaches to improve predictive accuracy [11]. The researcher in [12] compared two analytical methods—Fourier series and the ARIMA estimator—for forecasting crude oil prices in Indonesia. The results indicate that the Fourier series estimator, utilizing the Cos-Sin function, outperforms ARIMA, achieving RMSE and MAPE values of 7.93 and 8.4%, respectively. Decision trees are favored for interpreting machine learning model outputs thanks to their straightforward structure and ease of understanding [13]. The study assessed the influence of decision trees on data mining across various professional domains [14]. This report delineates numerous characteristics associated with cardiac disorders and explores models utilizing supervised learning methods, including Naive Bayes, Decision Tree, Nearest Neighbor, and Random Forest algorithms [15]. Recent advancements in computational power and storage capabilities have enabled the collection of ordinal, nominal, binary, and univariate customer-centric demographic and psychographic data, resulting in extensive and diverse datasets with varied metrics. Consequently, there has been significant progress in applying data-driven techniques such as decision trees (DTs) [16]. The ARIMA-Decision Tree algorithm, applied to reservoir pressure depletion data, has demonstrated enhanced performance in simulation and forecasting tasks, improving both speed and accuracy. This research emphasizes the importance of integrating Big Data Analytics algorithms in reservoir management and reporting, and suggests that future work could incorporate deliverability calculations to detect and correct anomalous reservoir behavior [17]. Additionally, this paper proposes a combination of regression and machine learning techniques—including Auto-Regressive Integrated Moving Average (ARIMA), Random Forest (RF), Bagging Classification and Regression Trees (BCART)—along with two hybrid models, ARIMA-RF and ARIMA-BCART, to predict wind power generation one, two, and seven days ahead [18].

3 Prediction Mode

3.1 Autoregressive Integrated Moving Average (ARIMA)

The ARIMA model is a widely used approach for time series forecasting that integrates three key components:

– Autoregressive (AR): Captures the relationship between an observation and a specified number of its past values.
– Integrated (I): Ensures stationarity by differencing the data to remove trends and seasonality.
– Moving Average (MA): Models the dependency between an observation and past forecast errors.

The ARIMA model is denoted as ARIMA(p, d, q), where:

– p (AR order) represents the number of lagged observations included in the model.
– d (degree of differencing) indicates how many times the data is differenced to achieve stationarity.
– q (MA order) corresponds to the number of lagged forecast errors incorporated into the model.

The expression for the general ARIMA model is:

$$Y_t = C + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t, \tag{1}$$

Y_t is the predicted value at time t, C is constant, ϕ_i are the autoregressive coefficients, θ_j are the coefficients of the moving average and ε_t is the error term at time t.

3.2 Decision Tree model

A decision tree is a versatile method applicable to both regression and classification tasks. It recursively partitions the dataset into subsets based on optimal attribute selection, aiming to maximize information gain or minimize impurity [19]. Mathematically, a decision tree can be characterized by the following components:

– Root Node: The root node represents the starting point of the decision process, containing the entire dataset. It acts as the topmost node in the hierarchical structure [20].
– Internal Nodes (Decision Nodes): Each internal node corresponds to a feature within the dataset, selected based on criteria that best split the data, such as Information Gain, Gini Index, or Variance Reduction [21]. Branches emerging from these nodes represent outcomes of tests or decisions: if the test condition is met, the path follows the left branch; otherwise, it moves to the right branch [22].
– Leaf Nodes (Terminal Nodes): Leaf nodes signify the final output of the decision tree. In classification problems, they represent class labels; in regression problems, they denote the predicted continuous values [23].

3.3 Hybrid

The Hybrid Decision Tree-ARIMA model combines the strengths of both Decision Trees and ARIMA to enhance forecasting performance [23]. In this approach, ARIMA primarily handles time-series prediction by modeling linear temporal dependencies, while the Decision Tree complements it by capturing nonlinear patterns through classification or regression tasks [24], as illustrated in Figure 1.

Figure 1 depicts the Hybrid Decision Tree–ARIMA model used for time series forecasting.

The hybrid model can be succinctly summarized as follows:

$$\hat{y}_t^{ARIMA} = C + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}\ , \tag{2}$$

$$e_t = y_t - \hat{y}_t^{ARIMA}, \tag{3}$$

$$e_t = f(X_t), \tag{4}$$

$$\hat{y}_t^{Hybrid} = C + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + f(X_t), \tag{5}$$

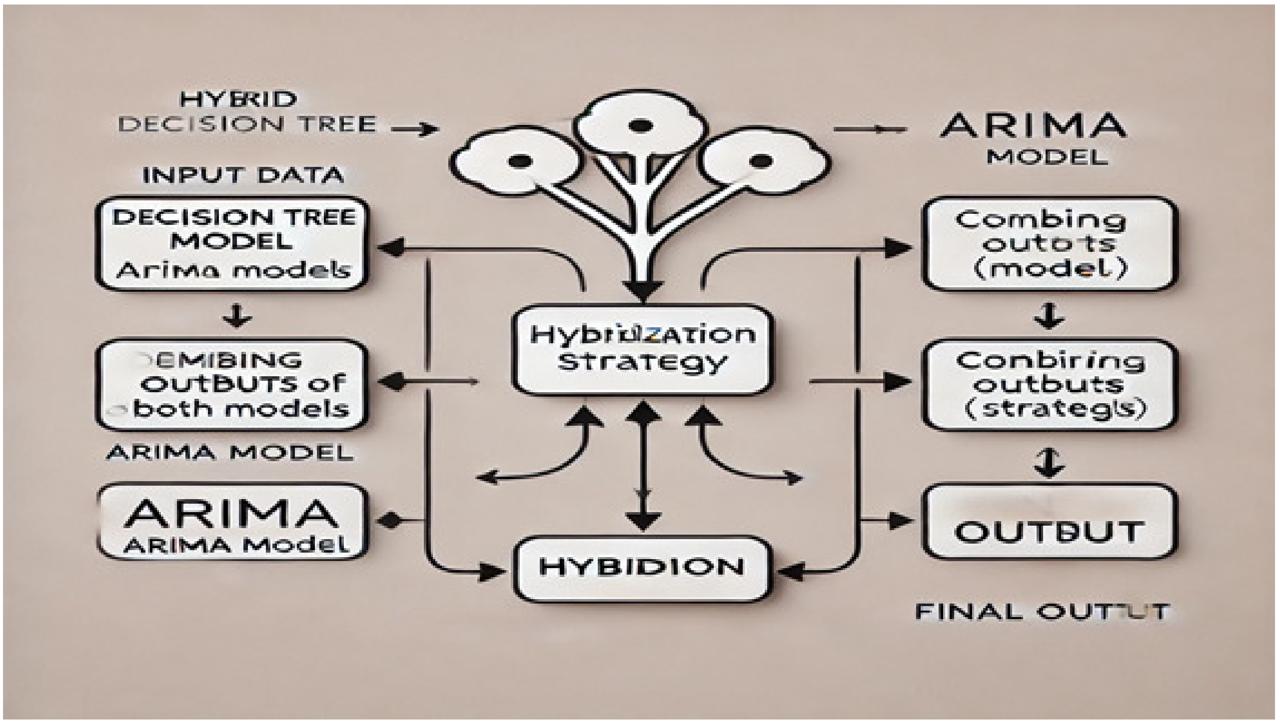


Fig. 1: Hybrid Decision Tree and ARIMA Model for time series forecasting

$\hat{y}_t^{Hybrid}$ is the final prediction, $\hat{y}_t^{ARIMA}$ is ARIMA based prediction, X_t comprises the components incorporated into the Decision Tree, including historical forecast errors, lagged values, and external variables. $f(X_t)$ is the result of the Decision Tree model (error correction or feature-driven modification).

The hybrid modeling process begins with fitting the ARIMA model to the original time series data to capture its linear structure. Once the ARIMA model is trained, the residuals—representing the nonlinear patterns not explained by ARIMA—are extracted. These residuals are then modeled using a Decision Tree, which is well-suited for capturing complex, nonlinear relationships. Finally, the predictions from both models are combined: the ARIMA forecast captures the linear trend, while the Decision Tree accounts for the remaining nonlinearities, resulting in a more accurate and robust final prediction.

4 Numerical Results and Discussion

For our analysis, we utilized historical oil price data obtained from the DataStream database. The dataset clearly illustrates the relationship between crude oil prices and major foreign exchange rates over the period from January 2018 to August 30, 2024 (see Table 1).

The dataset spans the period from January 2, 2018, to January 22, 2018, capturing daily fluctuations in oil prices and key foreign exchange rates. During this period, crude oil prices ranged from 10.01 USD to 123.7 USD per barrel, with an average of 67.92 USD, indicating an overall upward trend. The EUR/USD exchange rate varied between 0.79 and 1.05, reflecting shifts in the Euro's strength relative to the US Dollar. The USD/JPY rate, representing the value of one US Dollar in Japanese Yen, fluctuated between 102.23 and 161.91, highlighting significant volatility in the yen. Similarly, the USD/CAD exchange rate ranged from 1.20 to 1.79, underscoring the well-documented correlation between oil prices and the Canadian Dollar, given Canada's role as a major oil exporter.

Figure 2 presents a dot plot illustrating the distribution of oil prices, where each dot represents a recorded price along a numerical axis. The values range from 10 to 123 USD per barrel. The distribution appears continuous and approximately normal, reflecting the overall variability and central tendency in the observed oil prices.

Figure 3 displays an interval plot of oil prices with a 95% confidence interval (CI) for the mean. The central blue dot represents the sample mean of oil prices, while the vertical bars denote the confidence interval, illustrating the range within which the true mean is expected to lie with 95% probability. A note indicates that individual standard deviations were used to compute these

Table 1: Oil price and principal foreign exchange

Crude Oil Pricing	The cost of crude oil in USD perbarrel on the specifieddate
Euro/USD ollar (USD)	The exchan gerate of the Euro (EUR) relative to the US Dollar
British Pound to USD	The exchange rate of the British Pound (GBP) relative to the (USD)
USD /Swiss Franc	The exchange rate of the USD relative to the Swiss Franc (CHF)
USD /Japanese Yen	The exchange rate of the USD relative to the JapaneseYen (JPY)
AUD to USD	The exchange rate of the Australian Dollar (AUD)relative to the USD
New Zealand Dollar	The exchange rate of the NZD relative to the USD
USD/Canadian Dollar CAD	The exchange rate of the United States Dollar (USD)relative to the Canadian Dollar(CAD)

Table 2: Descriptive statistics for study variables

Date	Mean	StDev	Minimum	Maximum	Range
Crude Oil Pricing	67.926	18.457	10.010	123.700	113.69
Euro/USD ollar (USD).	0.89318	0.04629	0.79981	1.04548	0.24567
British Pound to USD	0.77718	0.03902	0.69784	0.94373	0.24567
USD/SwissFranc	0.94713	0.03902	0.83696	1.08220	0.24524
USD /JapaneseYen	121.39	16.95	102.23	161.91	59.67
AUD to USD	1.428	0.0853	1.2349	1.7253	0.4904
New Zealand Dollar	1.5334	0.0982	1.3422	1.7912	0.4489
USD/Canadian Dollar CAD	1.3164	0.436	1.2022	1.4494	0.2472

intervals. This visualization helps assess the uncertainty and variability in oil price estimates — a narrower interval implies higher precision in the mean estimation.

Grubbs' test was applied to identify potential outliers in crude oil prices. The computed G-statistic was 3.14 with a corresponding p-value of 1.000, indicating insufficient evidence to reject the null hypothesis. This suggests that all observations originate from a single normal distribution. Therefore, no significant outliers were detected at the 5% significance level ($\alpha = 0.05$). (See Fig. 4).

The decision tree model is employed to forecast oil prices using key foreign exchange rates, including USD/JPY, AUD/USD, USD/CHF, GBP/USD, and USD/CAD. The tree initiates its first split with the USD/JPY exchange rate, underscoring its dominant influence on oil price fluctuations. The initial division occurs at the threshold value of 108.042. If USD/JPY is below this value, the next split is determined by the AUD/USD rate; conversely, if it exceeds 108.042, the model considers USD/CHF as the next most influential variable. Further splits are made based on additional exchange rates, such as GBP/USD and USD/CAD, to progressively enhance the accuracy of the prediction. Each bifurcation is guided by statistically significant thresholds (P-value ¡ 0.05), confirming the relevance of each variable in the model's structure (see Figure 5).

Figure 6 presents four time-series plots comparing the performance of several forecasting models for global oil prices over the period from 2016 to 2023. Each subplot illustrates the forecasted values produced by a specific predictive model, while the final plot displays the actual observed oil prices during the same timeframe. The models under evaluation are:

– Predictive Decision Tree: This plot shows the forecasts generated solely by the Decision Tree model, which segments the data based on rule-based splits derived from explanatory variables.
– Predictive DT-ARIMA (Hybrid Model): This plot represents the results from a hybrid model that combines ARIMA's time series modeling capabilities with Decision Trees, thereby capturing both temporal dependencies and complex variable interactions.
– Predictive ARIMA: This plot depicts the predictions from the classic ARIMA model, which models autoregressive, integrated, and moving average components to capture trends and seasonality in oil prices.
– Oil Price (Actual): The final plot displays the actual historical oil prices, serving as the benchmark against which the forecasting models are evaluated.

A visual comparison of these curves enables a clear assessment of the relative strengths and limitations of each model in tracking real fluctuations, providing valuable insight into their predictive accuracy over an extended period.

Analysis of Results

The Predictive Decision Tree (DT) model effectively captures the overall direction and major movements of oil prices, demonstrating its ability to identify key patterns based on explanatory variables. However, it exhibits notable volatility in its forecasts, reflecting a sensitivity to short-term fluctuations and noise within the data. This behavior suggests that while the DT model can adapt quickly to changes, it may also overfit transient irregularities, leading to less stable predictions.

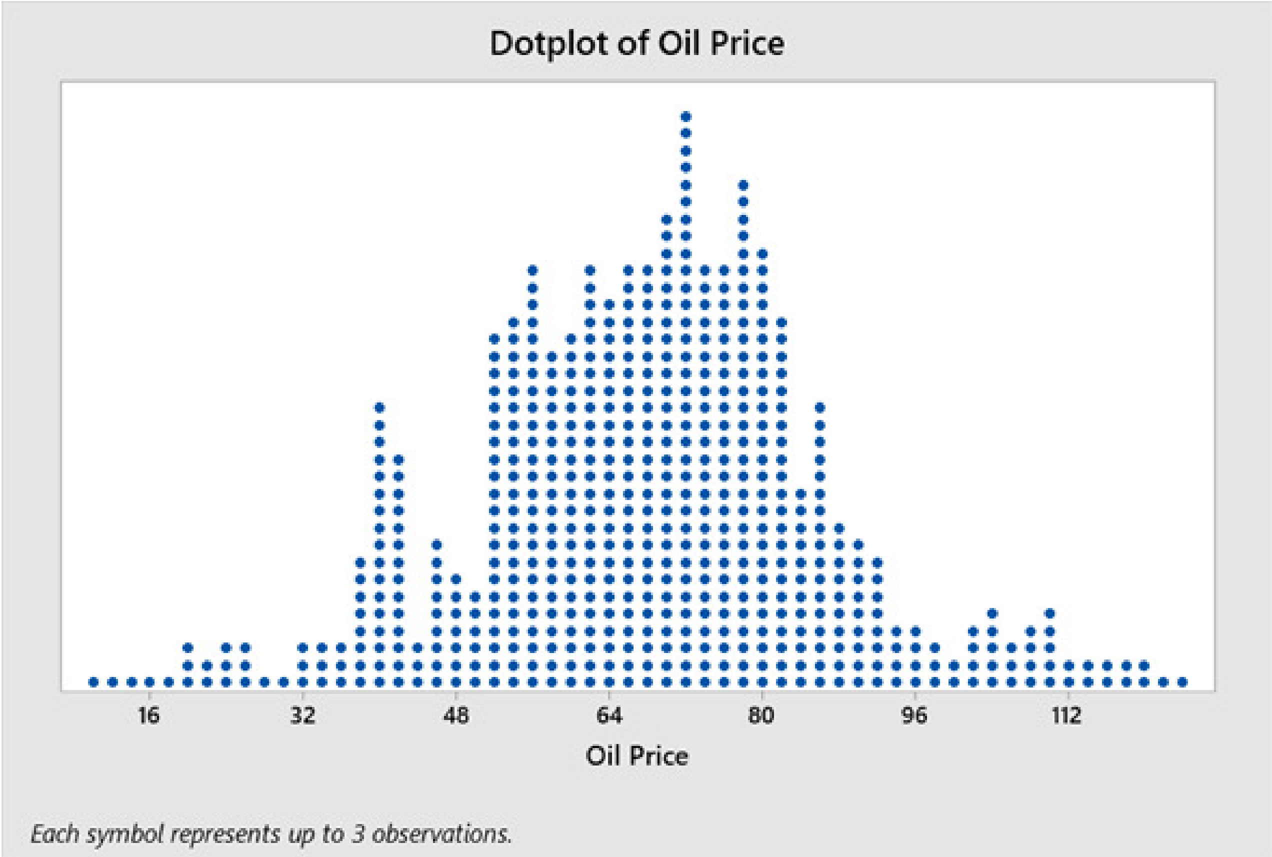


Fig. 2: Dotplot of Oil Price

In contrast, the Predictive DT-ARIMA hybrid model provides a more balanced and refined forecasting output. By integrating the strengths of the ARIMA model—known for modeling temporal dependencies and smoothing trends—with the Decision Tree's capacity to handle nonlinear relationships and complex interactions, this hybrid approach yields forecasts that closely align with the actual oil price trajectory. This synergy substantially enhances predictive accuracy, especially in periods marked by moderate volatility, confirming the benefit of combining statistical and machine learning methodologies.

The Predictive ARIMA model captures the general trend and seasonality of oil prices effectively, offering smooth and stable predictions. However, it tends to underperform when faced with sudden or sharp price spikes, such as those observed in the 2019 to 2022 period. This limitation stems from ARIMA's reliance on linear assumptions and lag structures, which may not fully adapt to abrupt market shocks or nonlinear behavior inherent in oil price dynamics.

Finally, the Oil Price subplot depicts the real historical fluctuations, highlighting notable peaks around 2022 followed by a significant decline. This actual price behavior provides a crucial benchmark, illustrating the challenges forecasting models face in accurately predicting sharp market movements amid periods of heightened uncertainty and volatility.

Key Observations

The hybrid DT-ARIMA model outperforms the standalone Decision Tree and ARIMA models, demonstrating a stronger correlation with actual oil price movements. While the Decision Tree model exhibits greater sensitivity to short-term fluctuations, which may result in overfitting to transient noise, the ARIMA model provides relatively stable forecasts but struggles to adapt to sudden nonlinear price shocks. By combining the strengths of both approaches, the hybrid model effectively balances stability and responsiveness, accurately capturing both long-term trends and short-term volatility in oil prices.

The Root Mean Squared Percentage Error is given by:

$$RMSPE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left(\frac{y_i-\hat{y}_i}{y_i}\right)^2} \times 100 \qquad (6)$$

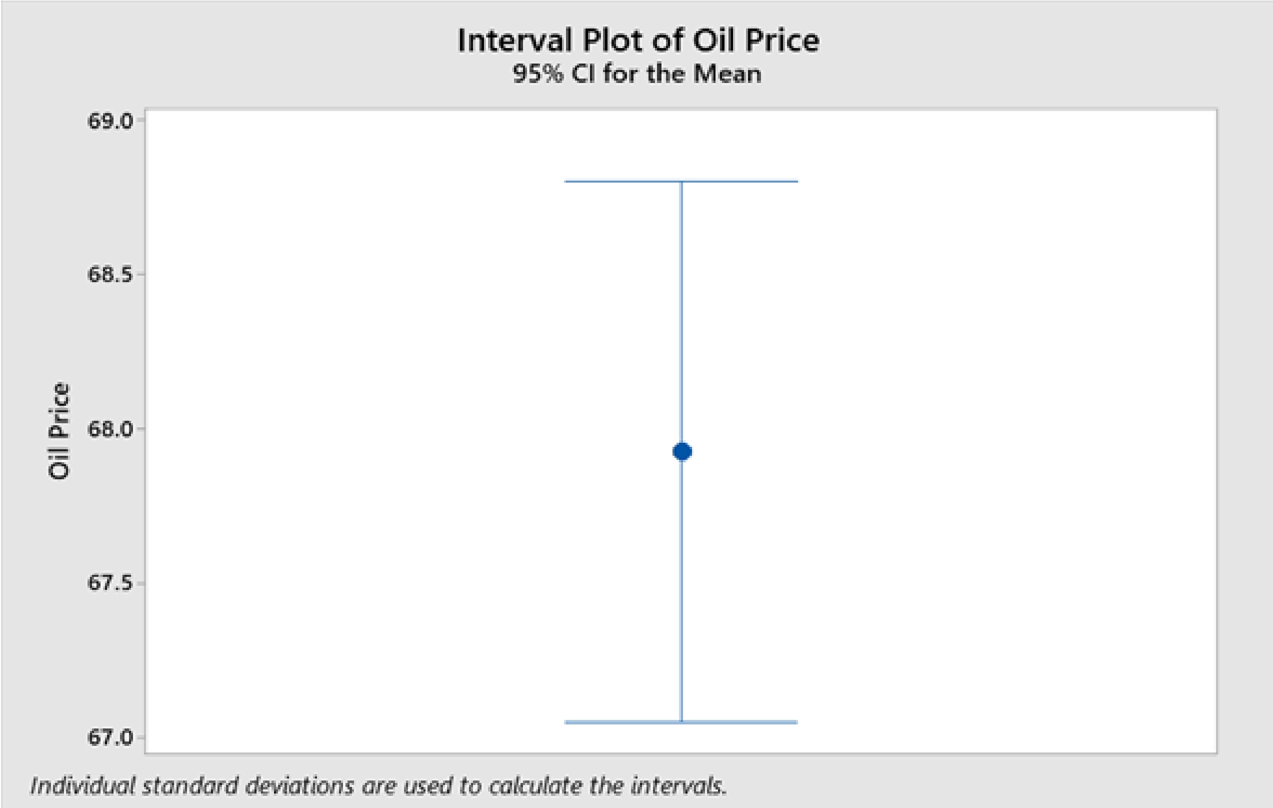


Fig. 3: Interval Plot of Oil Price

It measures the accuracy of a forecasting model, with lower values indicating better performance.

The RMSPE values for the different models are as follows:

Hybrid	0.0013
DT	0.053204
ARIMA	0.053416

The hybrid DT-ARIMA model achieves the lowest Root Mean Squared Percentage Error (RMSPE) of 0.0013, demonstrating a markedly superior forecasting accuracy compared to the individual models. This exceptional performance highlights the effectiveness of combining the Decision Tree's ability to capture complex nonlinear patterns with ARIMA's strength in modeling linear temporal dependencies. By leveraging the complementary advantages of both methods, the hybrid model delivers a more robust and nuanced prediction framework capable of adapting to the multifaceted nature of oil price movements.

–Evaluation of the Decision Tree Model's Efficacy:

The Decision Tree (DT) model records a substantially higher RMSPE of 0.053204, signaling limitations in its predictive capabilities for this dataset. Although DTs are powerful in detecting nonlinear relationships and interactions among variables, they tend to be vulnerable to overfitting, especially when applied to noisy or highly volatile time series data such as oil prices. This propensity results in model instability and poorer generalization, which undermines the accuracy of longer-term forecasts. Consequently, while Decision Trees can be valuable for short-term pattern recognition, their standalone use may lead to inflated prediction errors when extended over broader time horizons.

–Evaluation of the ARIMA Model's Efficacy:

The ARIMA model exhibits a slightly higher RMSPE of 0.053416 than the Decision Tree model. ARIMA is well-established for its proficiency in capturing linear trends and autoregressive structures within time series data. However, its underlying assumption of linearity restricts its effectiveness in contexts characterized by sudden nonlinear shifts and complex market dynamics inherent in crude oil prices. The inability of ARIMA to adequately model abrupt fluctuations or regime changes contributes to elevated forecasting errors, particularly during volatile periods.

–Evaluation of the Hybrid Decision Tree - ARIMA Model's Efficacy:

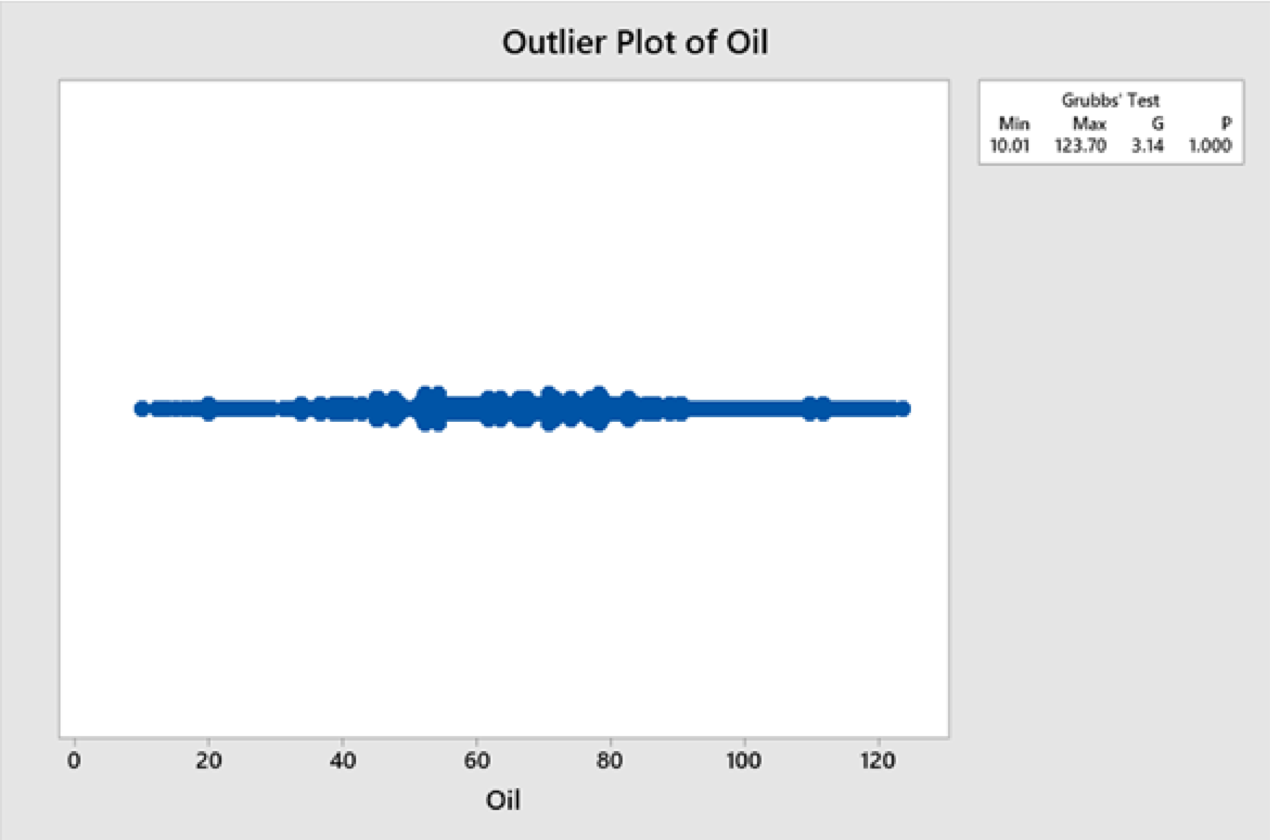


Fig. 4: Outlier Plot of Oil price

The hybrid DT-ARIMA model considerably outperforms the individual models, underscoring the value of integrating statistical and machine learning techniques. By combining ARIMA's capacity for modeling stable linear components with the Decision Tree's flexibility in identifying nonlinear patterns and interactions, the hybrid approach attains enhanced adaptability and precision. This synergy enables the model to better capture both gradual trends and abrupt market movements, resulting in significantly reduced RMSPE values and improved overall forecast reliability. Such findings emphasize the promise of hybrid models as superior tools for complex time series forecasting in financial and economic domains.

5 Conclusion

Accurate oil price forecasting is paramount for the development of effective economic strategies and informed decision-making within the global energy sector. This study leveraged historical data encompassing crude oil prices alongside key global currency exchange rates to investigate the drivers of oil price volatility and enhance prediction accuracy.

Descriptive statistics, given in Table 1, reveal a period characterized by significant price volatility, with an average barrel price of $67.926, ranging from a low of $10.010 to a high of $123.700, reflecting marked market fluctuations over the study horizon. Furthermore, substantial exchange rate variations—particularly in the EUR/USD pair, fluctuating between 0.79981 and 1.04548 (Table 2)—underscore the strong interconnection between currency markets and oil prices.

Comparative model analysis demonstrates that the hybrid DT-ARIMA model outperforms its standalone counterparts by effectively capturing both long-term trends and short-term volatility. The Decision Tree component highlights the critical influence of exchange rates such as USD/JPY and USD/CHF on oil price dynamics, suggesting that the interplay of these currency cycles exerts considerable impact on price fluctuations. This interpretable and predictive approach provides policymakers and analysts with valuable insights into the mechanisms linking foreign exchange markets and oil price behavior.

Root Mean Squared Percentage Error (RMSPE) results validate the superiority of the hybrid model, which achieves a remarkably low error of 0.0013—substantially

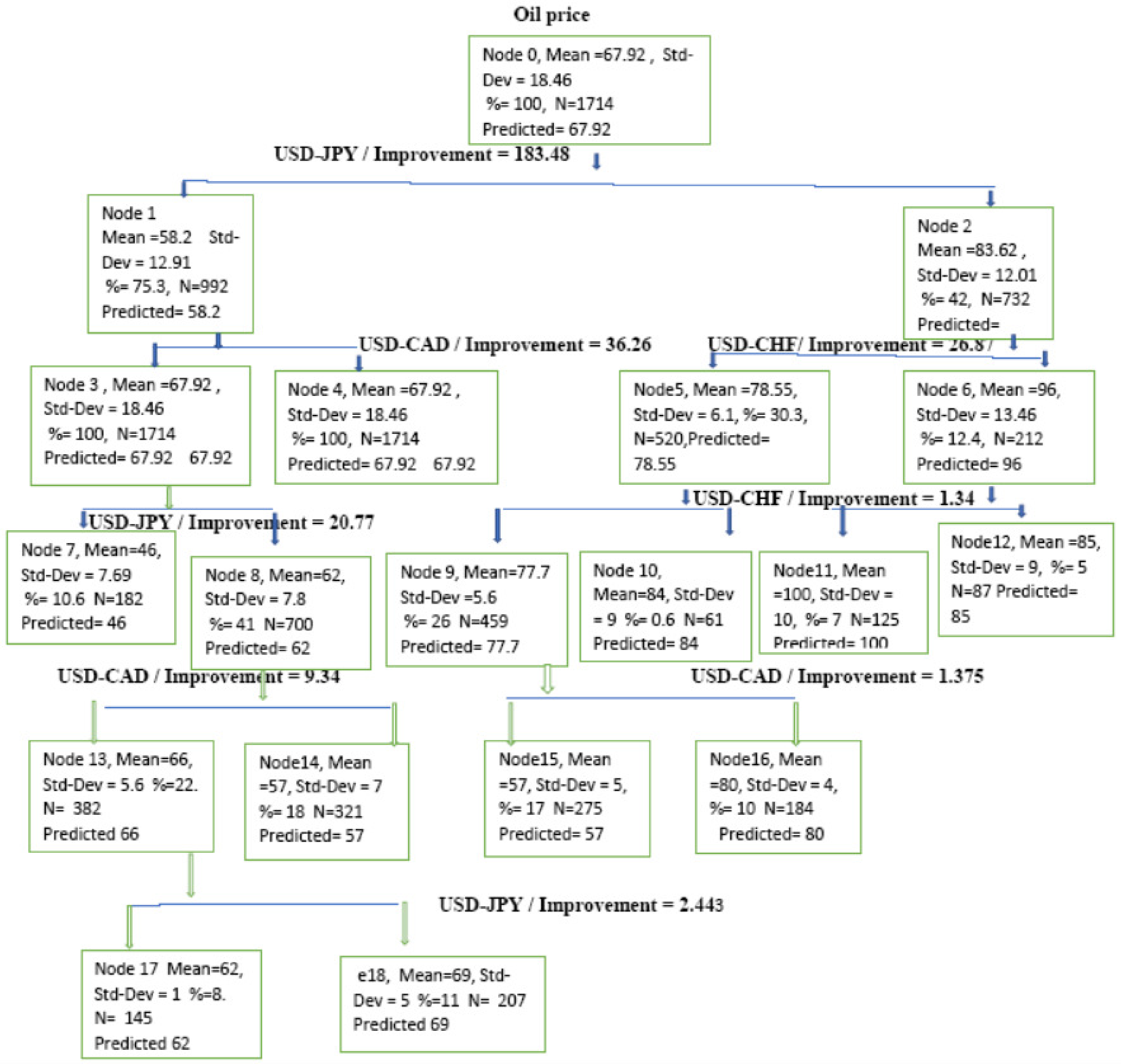


Fig. 5: Decision Tree Model for Predicting Oil Prices Based on Foreign Exchange Rates

better than the Decision Tree (0.053204) and ARIMA (0.053416) models. This enhanced performance arises from the complementary strengths of the non-linear pattern recognition capacity of the Decision Tree and the linear trend modeling power of ARIMA, overcoming the limitations of each individual method.

The core contribution of this research lies in the innovative fusion of a machine learning technique (Decision Tree) with a classical time series model (ARIMA), striking an optimal balance between leveraging historical data patterns and adapting to contemporary market fluctuations. This hybrid framework paves the way for more robust and reliable predictive models in financial markets and energy economics.

Future Directions and Research Opportunities:

– Incorporating geopolitical factors, such as international crises, trade tensions, and political events, to better capture their effects on oil price volatility.
– Integrating additional macroeconomic variables—including inflation rates, interest rates, and renewable energy production data—to further refine forecast precision.

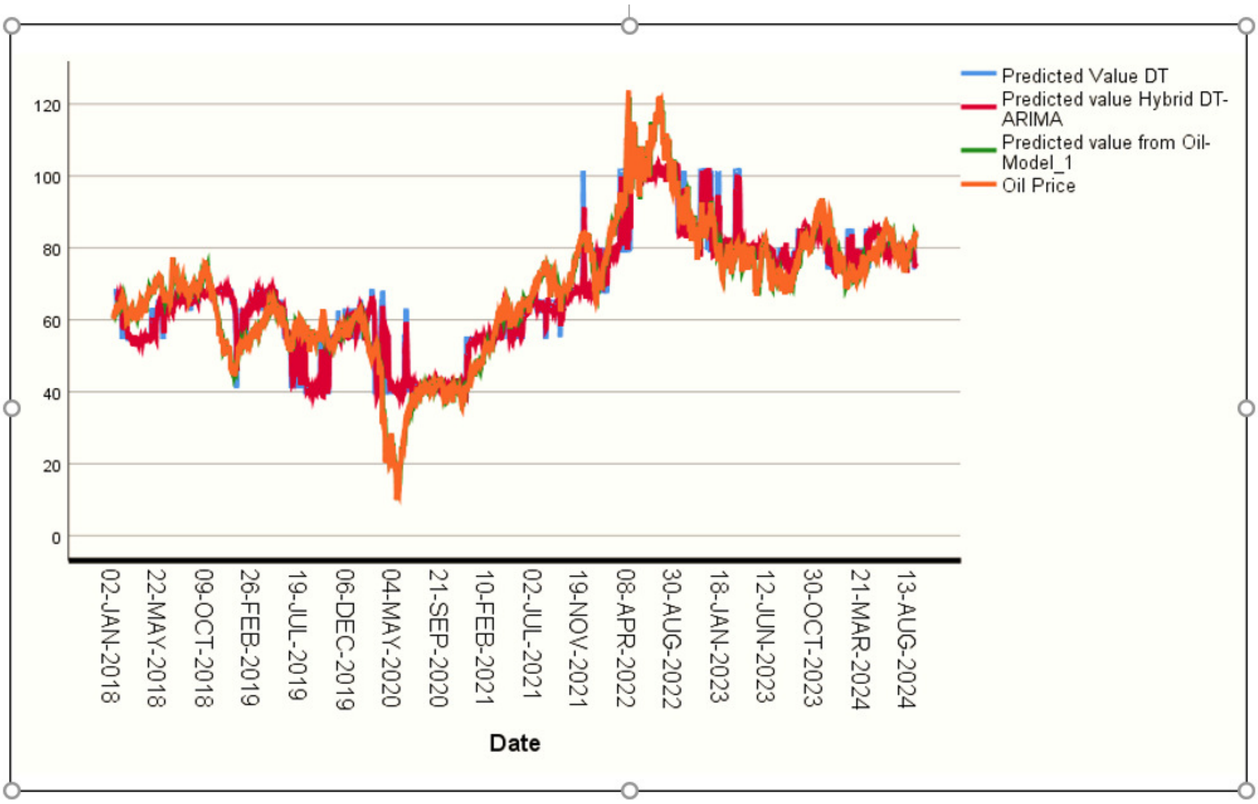


Fig. 6: Evaluation of DT, ARIMA, and Hybrid DT-ARIMA Models for Predicting Oil Prices

–Exploring advanced deep learning architectures, such as recurrent neural networks (RNNs) and transformer models, to enhance hybrid models' capability to process large-scale, complex temporal data.
–Developing adaptive real-time hybrid models capable of dynamically updating parameters in response to evolving market conditions, thereby delivering timely and accurate forecasts in an ever-changing financial landscape.

In conclusion, the proposed DT-ARIMA hybrid model represents a significant advancement in financial forecasting, offering a powerful and flexible tool for analyzing oil markets and informing strategic decision-making in an increasingly volatile and uncertain global context.

Acknowledgement: The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Research Project under grant number RGP2/492/46.

References

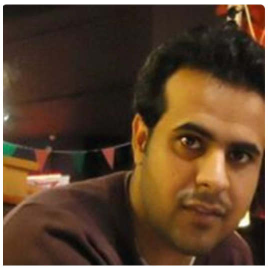
[1] Alquist, R., Kilian, L., & Vigfusson, R. J. (2013). Forecasting the price of oil. In Handbook of economic forecasting (Vol. 2, pp. 427-507). Elsevier.
[2] Guliyev, H., & Mustafayev, E. (2022). Predicting the changes in the WTI crude oil price dynamics using machine learning models. Resources Policy, 77, 102664.
[3] Albahooth, B. (2024). Oil price movements predictions in Kingdom of Saudi Arabia using financial and macro-economic variables. Energy Exploration & Exploitation, 42(2), 747-771.
[4] Heidari, A., Daghighiasli, A., Abbassi, E., & Damankeshideh, M. (2024). Presenting a model for predicting global crude oil prices based on artificial intelligence algorithm. Journal of Petroleum Geomechanics; Vol, 7(3).
[5] Peipei, W., & James, W. (2024). Predicting oil price fluctuations: Integrating external indicators and advanced regression techniques. Resources Policy, 97, 105263.
[6] Yao, X., & Yang, X. (2024). Forecasting crude oil futures using an ensemble model including investor sentiment and attention. Kybernetes, 53(12), 6114-6138.
[7] Rahman, M. K., Dalim, H. M., Reza, S. A., Ahmed, A., Zeeshan, M. A. F., Jui, A. H., & Nayeem, M. B. (2025). Assessing the Effectiveness of Machine Learning Models in Predicting Stock Price Movements During Energy Crisis:

Insights from Shell's Market Dynamics. Journal of Business and Management Studies, 7(1), 44-61.
[8] ZAMRI, F., & Nor, M. E. (2024). Forecasting Oil Price in Malaysia using ARIMA and Artificial Neural Network. Enhanced Knowledge in Sciences and Technology, 4(2), 273-282.
[9] Sarfiah, S. N., & Aripulis, A. S. (2023). THE IMPLEMENTATION OF ARIMA METHOD IN FORCASTING INDONESIAN CRUDE OIL EXPORT, PRODUCTION AND CONSUMPTION. Jurnal REP (Riset Ekonomi Pembangunan), 8(1), 122-141.
[10] 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.
[11] Chamalwa, H. A., Abdulsalam, B., & Abbas, M. (2024). Modeling and Forecasting Crude Oil Prices in Nigeria Using ARIMA: A Time Series Analysis from 2013-2022. UMYU Scientifica, 3(3), 313-321.
[12] Rahma, A. K., Abidin, Q. Z., Prasetyo, J. K., Larasati, B., Amelia, D., & Chamidah, N. (2024). Prediction Of Crude Oil Prices In Indonesia Using Fourier Series Estimator And Arima Method. BAREKENG: Jurnal Ilmu Matematika dan Terapan, 18(3), 1673-1682.
[13] Abd Rahman, M. S., Jamaludin, N. A. A., Zainol, Z., & Sembok, T. M. T. (2023). The application of decision tree classification algorithm on decision-making for upstream business. International Journal of Advanced Computer Science and Applications, 14(8).
[14] The study assessed the influence of decision trees on data mining across various professional domains.
[15] Komarasamy, G., & Ravishankar, T. N. (2022). The Application of Decision Tree Method for Data Mining. Technoarete Transactions on Intelligent Data Mining and Knowledge Discovery, 2(3).
[16] Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. SN Computer Science, 1(6), 345.
[17] Siamba, S., Otieno, A., & Koech, J. (2023). Application of ARIMA, and hybrid ARIMA Models in predicting and forecasting tuberculosis incidences among children in Homa Bay and Turkana Counties, Kenya. PLOS digital health, 2(2), e0000084.
[18] Siamba, S., Otieno, A., & Koech, J. (2023). Application of ARIMA, and hybrid ARIMA Models in predicting and forecasting tuberculosis incidences among children in Homa Bay and Turkana Counties, Kenya. PLOS digital health, 2(2), e0000084.
[19] Krishna, R., Sathish, H., & Zhou, N. (2023). Forecasting uncertainty parameters of virtual power plants using decision tree algorithm. Electric Power Components and Systems, 51(16), 1756-1769.
[20] Duan, M., Zheng, T., Gao, Y., Wang, G., Feng, Z., & Wang, X. (2024, March). Dga-gnn: Dynamic grouping aggregation gnn for fraud detection. In Proceedings of the AAAI conference on artificial intelligence (Vol. 38, No. 10, pp. 11820-11828).
[21] Di Teodoro, G. (2024). Toward personalized medicine and transparent decision-making with Machine Learning Models for complex diseases: a focus on HIV-1 and type 2 Diabetes.
[22] Cankaya, B., Topuz, K., Delen, D., & Glassman, A. (2023). Evidence-based managerial decision-making with machine learning: The case of Bayesian inference in aviation incidents. Omega, 120, 102906.
[23] Liang, Z., & Ismail, M. T. (2025). Advanced ceemd hybrid model for vix forecasting: optimized decision trees and arima integration. Evolutionary Intelligence, 18(1), 12.
[24] Shahriar, S. A., Kayes, I., Hasan, K., Hasan, M., Islam, R., Awang, N. R., ... & Salam, M. A. (2021). Potential of ARIMA-ANN, ARIMA-SVM, DT and CatBoost for atmospheric PM2. 5 forecasting in Bangladesh. Atmosphere, 12(1), 100.

Imad Yagoub Hamid is an Assistant professor of Statistics, Department of Mathematics, Faculty of Sciences and Humanities, Shaqra University, Saudi Arabia. He received a PhD in Applied Statistic from Sudan University of Science and Technology in May 2009. M.Sc. in Science of Statistics from Juba University, Sudan, in March 2003. B.Sc. in Statistics from Omdurman Islamic University in July 1999. Major Specialization: Statistics. Subspecialty: Applied Statistics. Research areas of interest: Applied statistical models, Artificial Neural Networks and Machine Learning Techniques.

Sulima M. Awad Yousif Assistance Professor in Statistics and Mathematics at Department of Basic Science, Deanship of the Preparatory Year of Female - University of Najran, Kingdom of Saudi Arabia. She received a PhD in Statistics and Demography, M.Sc. in Population and Development and B.Sc. (Honors) in Applied Social Statistics and Demography from University of Gezira- Sudan. Major Specialization: Statistics. Subspecialty: Applied Statistics. Research areas of interest: Applied statistical models, Applications of Machine techniques in predication and classification. Assistance Professor in Statistics and Mathematics Department of Mathematics - Deanship of the Preparatory Year of Female - University of Najran, Kingdom of Saudi Arabia. She received a PhD in Statistics and Demography, M.Sc. in Population and Development and B.Sc. (Honors) in Applied Social Statistics and Demography from University of Gezira- Sudan. Major Specialization: Statistics. Subspecialty: Applied Statistics. Research

areas of interest: Applied statistical models, Applications of Machine techniques in predication and classification.

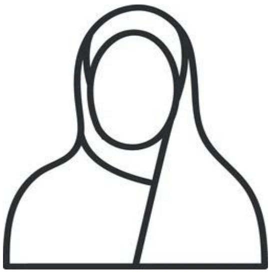


Saud Zamil Aljaloud is an Assistant Professor at the College of Computer Science and Engineering, University of Hail, Saudi Arabia. He holds a Ph.D. in Computer Science and an M.Sc. in Software Engineering from the University of Southampton, UK, along with a Bachelor's degree in Computer Science from the University of Hail. His research focuses on artificial intelligence, machine learning, and web technologies. He has contributed to international projects in MOOCs and social computing, with publications and presentations in renowned conferences such as ICWE and ESWC. Dr. Aljaloud is proficient in a range of programming languages and technologies, and has served as a reviewer for leading academic conferences in his field.

Abdelgalal O I Abaker is an associate Professor of applied Statistics at Applied College, King Khalid University, Saudi Arabia. He received a PhD from Aljazera University Sudan. His research interests are in the area of Applied Statistics. His research articles have been published in international Applied Statistics journals of good repute.

Halla Z. S. Elemam is an associate Professor at department of business administration, applied College, King Khalid University, Saudi Arabia. Her research interests are in Providing training courses in the field of statistical analysis, measurement and evaluation, quality learning outcomes.

Khalil A. Alruwaitee is an Assistant Professor at Taif University. He received a PhD in Finance from Victoria University, participated in academic events such as the HDR Student Conference 2020. Involvement includes the 4th ICIEBSH and the 6th Annual General Business Conference. His research focuses on financial markets, volatility, and portfolio risk.



Manahill Idriss Adrob Anja is an Assistant Professor of computer science, College of Turabah, Taif University. Research interests: spatial analysis of health services and GIS-based healthcare systems Focused on enhancing healthcare planning through spatial data technologies,web programming