IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 4, August 2026, pp. 3131~3143
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i4.pp3131-3143

Hybridization of hybrid-ARIMA-EM and XGBoost for enhanced price predictive modeling

Isam Ahmed M. Yaqoob, Khairul Azhar Kasmiran, Teh Noranis Mohd Aris, Nor Azura Husin, Mohd Yunus Sharum
Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Selangor, Malaysia

Article Info

Article history:

Received Feb 24, 2025
Revised May 20, 2026
Accepted Jun 19, 2026

Keywords:

ARIMA
Extreme gradient boosting
Gated recurrent unit
Prediction
Price forecasting

ABSTRACT

Managing finance entails the art and science of distributing available and potential funds among various competing needs. Government expenditures fund programs that provide a wide range of services to different population segments. As a result, the demand for enhanced and additional services often surpasses the government's financial capacity. Firstly, the price forecasting procedures for the extreme gradient boosting (XGBoost), gated recurrent unit (GRU), and hybrid-ARIMA-EM models will be summarized. Secondly, the accuracy of the models will be assessed on two real datasets collected from Kaggle (Crude_Oil_Price and KL_apartment). This study then proposes combining the hybrid-ARIMA-EM model with XGBoost to enhance the price forecasting performance in terms of time series analysis. Experimental results show that the suggested combination outperforms other selected models in price forecasting accuracy.

Corresponding Author:

Khairul Azhar Kasmiran
Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
Selangor, Malaysia
Email: k_azhar@upm.edu.my

1. INTRODUCTION

Price forecasting is essential and critical. However, the key question is what the best prediction method is and whether it can be optimized for greater accuracy. Additionally, which factors or indicators influence price the most and should be prioritized in the calculation steps, given the wide range of complex and dynamic factors. By leveraging artificial intelligence (AI) algorithms, businesses can enhance price forecasting accuracy, mitigate the risk of overestimating or underestimating cash inflows and outflows, and make better decisions regarding investments, financing, and other financial strategies. Furthermore, it is challenging to establish a suitable relationship between these indicators (external or non-financial factors) and prices to evaluate the effectiveness of the mentioned factors on prices [1]–[3], but AI can help. A remarkable model, extreme gradient boosting (XGBoost), has been widely utilized by many authors to achieve high accuracy in forecasting. It is an effective forecasting model with numerous amazing advantages because of its mechanism [4]. Firstly, because of its ensemble nature and implementation optimizations, XGBoost performs well, outperforming conventional machine learning (ML) algorithms. Second, it integrates regularization strategies, including L1 and L2 regularization to reduce overfitting and improve generalization capacity. Additionally, XGBoost offers a method for determining the significance of a feature, aiding in feature selection and model interpretation. It also does a great job of managing missing values, which eliminates the need for imputation or preprocessing. The last feature of XGBoost is its support for parallel processing, which greatly reduces computing time and allows for effective training on big datasets. XGBoost is a recommended option for predicting workloads because of its stability, interpretability, and scalability.

Prakash *et al*. [5] presented a new method that uses an XGBoost-aided ML model to enhance the performance of distributed Raman amplifiers (DRA) in next-generation ultra-long-haul optical networks. The focus of the model is to optimize the flexibility and consistency of DRA performance. It addresses the issue of inconsistent dimensionality between pump frequency and power by utilizing a decision-tree system to accurately predict the signal gain spectrum and bit error rate (BER). Their findings indicate that the method can identify faults within 2.3 milliseconds with 99.6% accuracy, outperforming other ML-based tree models in estimation accuracy and effectiveness. Consequently, the study showcases the successful application of an XGBoost model in estimating the desired gain profile and BER of DRAs in the low-loss optical wavelength region (1260–1650 nm), enabling advancements in ultra-long-haul optical networks.

After the COVID-19 pandemic, many researchers have started focusing on analyzing data and making predictions in various areas. Jin *et al*. [6] aimed to analyze pandemic trends, develop predictive models, and provide insights for controlling future outbreaks effectively. Traditional forecasting methods struggle to accurately predict pandemic trends due to the presence of both linear and nonlinear factors in COVID-19 data. To address this, the study proposed three combined forecasting models: convolutional neural network (CNN)-long short-term memory (LSTM)-autoregressive integrated moving average (ARIMA), temporal convolutional network (TCN)-LSTM-ARIMA, and sparrow search algorithm (SSA)-LSTM-ARIMA, which integrates deep learning and traditional ARIMA techniques to capture both linear and nonlinear factors. These models were initially tested on COVID-19 data from Quebec and Canada, and then validated using data from Italy. Comparative analyses showed that the CNN-LSTM-ARIMA model performed the best in terms of predictive accuracy, stability, and suitability for forecasting COVID-19 pandemics. The study's findings demonstrate the importance of integrating deep learning and traditional techniques for effective pandemic forecasting and highlight the potential of the CNN-LSTM-ARIMA model in guiding future control measures and mitigation strategies. Additionally, the TCN model is designed for processing sequential data in parallel, reducing modeling and forecasting time. TCN models can handle sequences of varying lengths and mapping input sequences directly to output sequences of the same length. The study also emphasized the incorporation of dilated causal convolutions and residual connections within TCN models and highlighted the reliability of an ARIMA-based model in forecasting pandemic trends.

Nowadays, researchers tend to combine prediction algorithms to leverage their unique strengths, enhancing the accuracy and reliability of forecasts [7]–[9]. Integrating two types of models/algorithms helps mitigate the shortcomings and limitations that prediction models often face across different fields [10], [11]. In a previous scientific paper, Yaqoob *et al.* [12] discussed various forecasting methods in the field of data analysis. A new model is introduced for cash forecasting called the hybrid-ARIMA-EM model and compared its accuracy with other well-known cash forecasting methods such as random forest, decision trees, Prophet, linear regression, multivariate adaptive regression splines (MARS), and perceptron. The comparison showed that our proposed model provided superior accuracy in cash forecasting. XGBoost and gated recurrent unit (GRU) are some of the most famous models [13]. Therefore, this paper will consider these models in comparison with hybrid-ARIMA-EM. Ryu *et al.* [14] tackles the challenge of predicting dementia risk, highlighting a deep learning method that demands extensive image data and training time. Recognizing these limitations, the study presents a dementia risk prediction model based on XGBoost, which improves efficiency through variable extraction from numerical dementia data and hyperparameter optimization. Initially, the method extracts variable importance from standard independent variables using gradient boosting to create derived variables. These derived variables are then analyzed for importance in forming Top-N groups. Hyperparameter tuning is subsequently applied to optimize the model's performance based on the characteristics of each Top-N group. The study evaluates the XGBoost model’s performance using goodness-of-fit metrics for ML classification models. The results indicated that the Top-20 model exhibits the best performance, achieving an accuracy of 85.61% and an F1-score of 79.28%. Compared to the decision tree, random forest, support vector machine, and k-nearest neighbors models, the proposed XGBoost model demonstrated superior performance.

2. METHOD

2.1. Dataset

Two Kaggle datasets have been selected and utilized in this paper. The first one contains crude oil prices available at https://www.kaggle.com/datasets/sc231997/crude-oil-price. The second one contains the monthly rent of Kuala Lumpur apartments available at https://www.kaggle.com/datasets/ariewijaya/rent-pricing-kuala-lumpur-malaysi.

2.2. Hybrid-ARIMA-EM forecasting

The hybrid-ARIMA-EM model will be implemented to obtain the forecasting of the years 2017-2021 then a comparison process will be applied between the obtained data and the original data for the same selected period. The model formula is given in (1).

$$\mathrm{Hybrid-ARIMA-EM} = \mathrm{AVG}(\mathrm{AF_i} + \phi_\mathrm{i} + \mathrm{Ơ'_i}) \times \mathrm{W_i} \tag{1}$$

Where AVG is average function; AF is ARIMA forecasting of the forecasted period; Φ is the value of the same period; Ơ is the value of the next period; and W is weight ratio.

The weight will be computed through (2).

$$\alpha = \mathrm{AVG(TS)} \tag{2}$$

Where TS is time series: the whole period of the selected time series. Then, get the sigmoid of α as defined in (3).

$$\mathrm{W} = 1 + \frac{1}{1\mathrm{e}^{-\alpha}} \tag{3}$$

The forecasting methodology leveraging the proposed hybrid-ARIMA-EM model can be concisely summarized as:

i) Prepare the dataset: the relevant data is extracted: two variables start_year and end_year define the range of years to extract from the dataset. Rows will be selected from the dataset where the 'Year' column is greater than or equal to start_year and less than or equal to end_year.
ii) Calculating the average time series length: the number of rows in the dataset is determined and used to calculate the variable AVG(TS).
iii) Perform ARIMA forecasting: an array ARIMA forecasting is created and initialized with a series of numerical values representing the ARIMA forecasting data.
iv) Determine alpha or α: the variable alpha is assigned the value of AVG(TS), representing the average time series length calculated earlier.
v) Calculate weight: applying sigmoid exponential function will make the data range fall between 0 to 1 which is easy to illustrate in the graph and facilitates calculations.
vi) Calculate forecasted values: the 'Crude Oil Price' column values are extracted from the dataset and assigned to the variable Φ.
vii) The process shift (-1) step: where selecting the year to shift the 'Crude Oil Price' column values one step forward, and assigns them to the variable Ơ.
viii) The forecasted values are calculated by applying the suggested equation to the arrays ARIMA forecasting, Φ, and Ơ along with the weight array W.

To sum up the whole progress in short points:

i) The hybrid-ARIMA-EM equation combines the ARIMA forecasting with additional components (ϕi and Ơi) and weights (Wi) to generate forecasts.
ii) The weights are determined by averaging the entire time series and applying a sigmoid transformation to the average.
iii) This approach aims to incorporate both historical data (ARIMA forecasting) and additional factors (ϕi and Ơi) into the forecasting process, with weights adjusted based on the average behavior of time series.

2.2.1. Prediction results comparison and evaluation

When evaluating the effectiveness of different forecasting models, it is essential to quantify their accuracy. This is typically achieved by comparing their forecasts against the actual observed values using statistical metrics. Two common metrics used for this purpose are root mean square error (RMSE) and mean absolute percentage error (MAPE), as shown in (4) and (5).

$$\mathrm{RMSE} = \frac{\sqrt{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{x_i} - \acute{\mathrm{Y}}_\mathrm{i})^2}}{\mathrm{N}} \tag{4}$$

Where i variable values; N is the number of non-missing data points; x is an actual observation time series; and Ý is an estimated time series.

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \quad (5)$$

Where A_t is actual value; F_t is forecast value; and n is number of times the summation iteration happens.

RMSE measures the average magnitude of the errors between predicted and observed values. It is computed by taking the square root of the average of the squared differences between predicted ($\acute{Y}$) and observed (x) values, divided by the total number of observations (N). This provides a single number representing the typical magnitude of errors in the forecast. Whereas MAPE measures the accuracy of a forecasting model as a percentage of the absolute error relative to the actual values. It is calculated by taking the mean of the absolute percentage errors between actual (At) and forecasted (Ft) values, averaging over all observations (n). Using the hybrid-ARIMA-EM prediction process, Figures 1 and 2 illustrate the yearly and monthly forecasting outcomes for the years 2015-2022 applied to the selected dataset (Crude_Oil_Price).

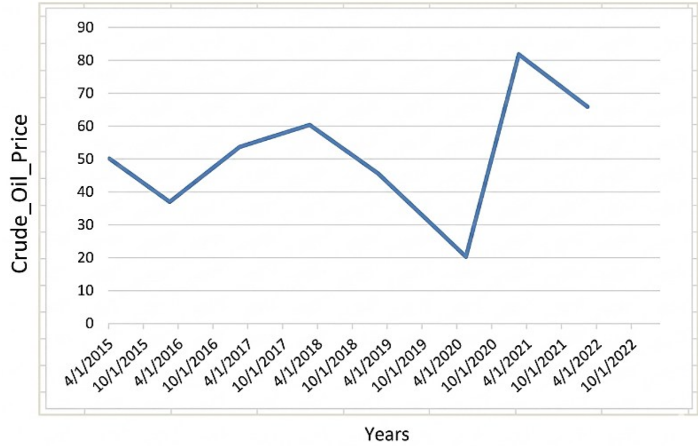

Figure 1. Forecasted years (2015-2022) by using the hybrid-ARIMA-EM model

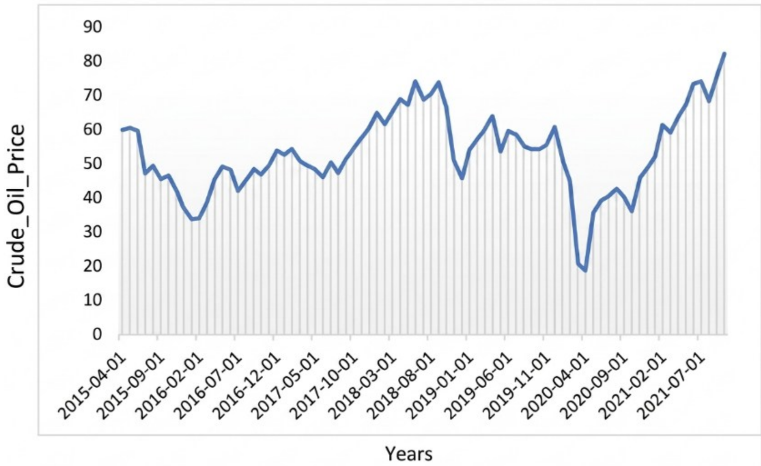

Figure 2. Monthly forecasting for the years 2015-2022 for the Crude_Oil_Price dataset by using hybrid-ARIMA-EM model

2.3. Extreme gradient boosting

XGBoost is a powerful forecasting model with several impressive advantages. Firstly, it delivers superior performance compared to traditional ML algorithms due to its ensemble nature and optimized implementation. Secondly, it uses regularization methods like L1 and L2 to reduce overfitting and enhance its ability to generalize [15]. Additionally, XGBoost includes a feature importance assessment mechanism, which aids in model interpretation and feature selection. It also handles missing values effectively, eliminating the need for data imputation or preprocessing [16]. Lastly, XGBoost supports parallel processing, allowing efficient training on large datasets and significantly cutting down computational time. These characteristics collectively make XGBoost a favored choice for forecasting tasks, offering scalability, robustness, and interpretability [17].

The XGBoost forecasting model has several benefits, but it also has certain limitations. The first is that it is complicated, which could be difficult for inexperienced users to understand and adjust the parameters. Secondly, there is the computationally demanding part of training an XGBoost model, particularly when working with huge datasets or high-dimensional feature spaces, which could call for a substantial amount of processing power [18].

2.3.1. XGBoost in time series prediction

In terms of examining the power of the XGBoost model in time series forecasting, the model will be applied to the selected dataset, which is the Crude_Oil_Price dataset for the years 2015-2023 from the Kaggle website. The steps of the process are as follows that performs a time series forecasting task for the selected dataset by using the XGBoost algorithm to forecast the years 2015-2022, as in Table 1 and Figure 3.

i) Data loading and preparation: loads a dataset containing historical crude oil prices and sets the relevant column (year and crude oil price) as a prime column that will be used to analyze and forecast the new values for the determined result.
ii) Train-test split: the dataset will be split into training and testing sets. The training set contains data up to a certain point (all data except the last 8 years), while the testing set contains the remaining data (the last 8 years).
iii) Model training: an XGBoost regression model will be initialized and fitted using the training data.
iv) Prediction: the model predicts crude oil prices for each month of the years 2015 to 2022. It iterates through each year, predicts prices for each month of that year, and stores the predictions.

Table 1. Forecast of years 2015-2022 for the dataset Crude_Oil_Price by using the XGBoost model

Year	Predicted crude oil price
2015	49.30752563
2016	44.46873856
2017	51.85506439
2018	64.58948517
2019	57.18280029
2020	38.59614563
2021	68.0691452
2022	94.63577271

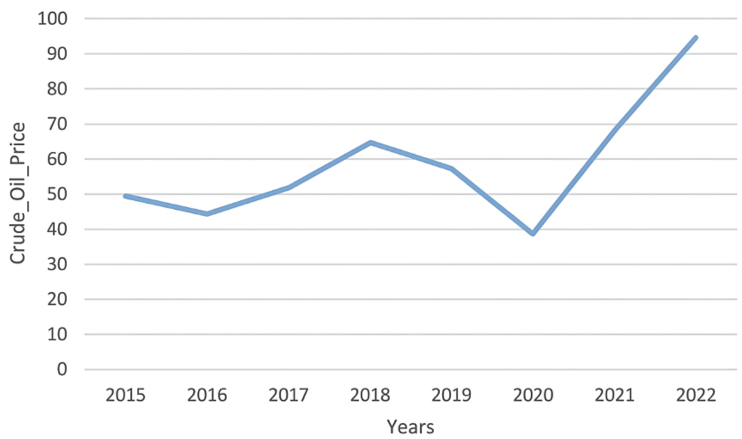

Figure 3. Forecast of years 2015-2022 for the dataset Crude_Oil_Price by using the XGBoost model

2.3.2. Accuracy examination

RMSE and MAPE were implemented to check the level of accuracy of the XGBoost model in terms of time series. The results were then compared with the RMSE and MAPE values of the proposed hybrid-ARIMA-EM presented in section 2.2. Both models were evaluated using the same dataset and period, as shown in Table 2.

Table 2. Results of RMSE and MAPE for the forecasting models XGBoost and hybrid-ARIMA-EM

Forecasting model	RMSE	MAPE
XGBoost	2.67	3.10
Hybrid-ARIMA-EM	1.04	0.009

2.4. Gated recurrent unit

The GRU-based prediction model aims to forecast future points based on previous data points [19]–[21]. It will be applied to predict years 2010-2023 for the dataset Crude_Oil_Price, as seen in Table 3 and Figures 4 and 5. The steps as follows that are needed to demonstrate and shortly explain the whole progress:

i) Data extraction and preparation: loads the targeted columns (years, oil prices) from the dataset (Crude_Oil_Price).
ii) Normalization: the data is normalized using min-max scaling to bring all features to the same scale, typically between 0 and 1.
iii) Train-test split: the dataset is split 80:20 into training and testing sets, without shuffling to preserve temporal order.
iv) Data reshaping: the input data is reshaped to fit the input requirements of the GRU model.
v) Model building: a sequential neural network model is constructed using Keras with the Adam optimizer and mean squared error (MSE) loss function. It consists of a GRU layer with 50 units and rectified linear unit (ReLU) activation function, followed by a dense layer with one unit.
vi) Model training: the model is trained using the training data for 50 epochs with a batch size of 32. Validation data is provided to monitor the model's performance during training.
vii) Prediction: the trained model is used to make predictions on the testing data. Predictions are inverse-transformed to obtain actual crude oil prices.
viii) Evaluation: the model's performance (RMSE and MAPE) is compared with hybrid-ARIMA-EM in Table 4.

To conclude, Figure 6 shows forecasting for all selected models (XGBoost, GRU, and hybrid_ARIMA_EM), on the Crude_Oil_Price dataset for the years 2015-2022.

Table 3. Forecast of years 2015-2022 for the dataset Crude_Oil_Price by using the GRU model

Year	Predicted Crude_Oil_Price
2015	73.38457
2016	84.83762
2017	88.64758
2018	88.04543
2019	88.04543
2020	59.28685
2021	51.25039
2022	61.85501

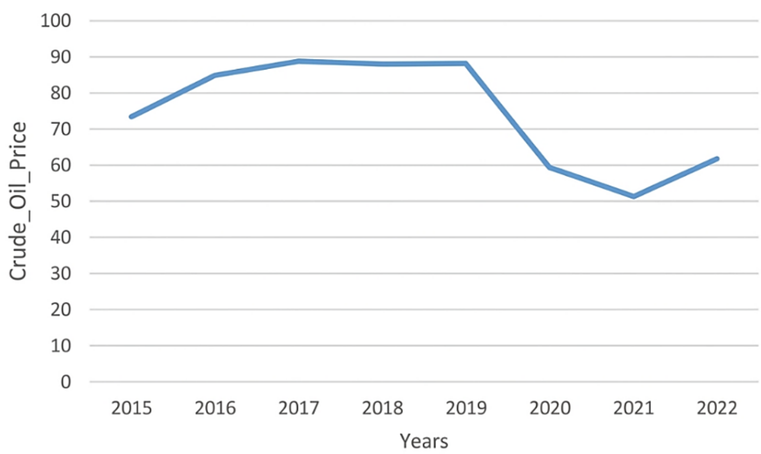

Figure 4. Annual forecast of the years 2015-2022 for the dataset Crude_Oil_Price by using the GRU model

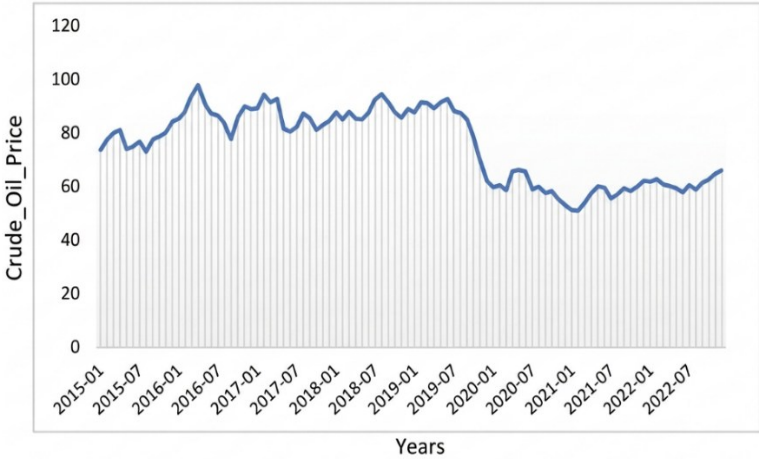

Figure 5. Monthly forecast of the years 2015-2022 for the dataset Crude_Oil_Price by using the GRU model

Table 4. Results of RMSE and MAPE for the forecasting models XGBoost and hybrid-ARIMA-EM

Forecasting model	RMSE	MAPE
GRU	4.10	0.47
Hybrid-ARIMA-EM	1.04	0.009

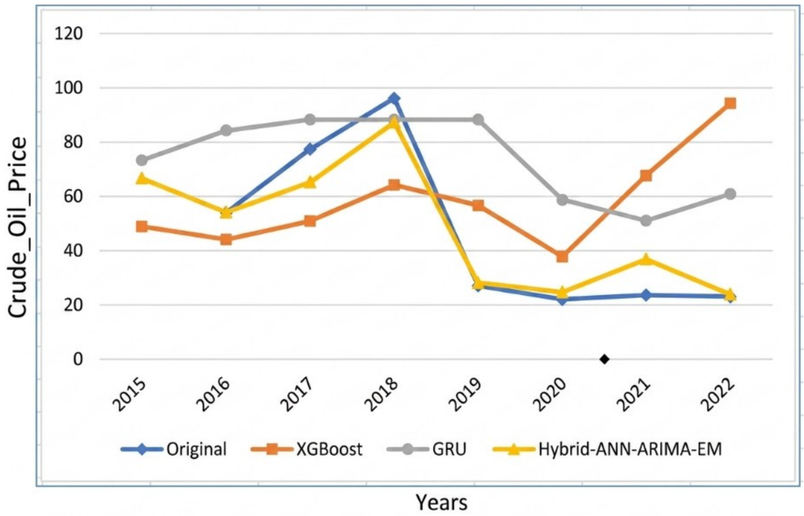

Figure 6. Forecast of the years 2015-2022 for the dataset Crude_Oil_Price by using the selected models

2.5. Hybrid-ARIMA-EM+XGBoost integration

The models presented in this study have been merged sequentially, and the best results were from the merger of the hybrid-ARIMA-EM and XGBoost models. Refer to Tables 5 and 6, and Figure 7. The hybrid-ARIMA-EM model and XGBoost approach forecasting from fundamentally different perspectives, and combining their methodologies can create a synergistic effect that leads to improved forecasting accuracy. The hybrid-ARIMA-EM model is designed to handle temporal data by focusing on the linear patterns and seasonal trends that often characterize time series. It uses ARIMA to establish a baseline forecast for these patterns [22]–[24]. What makes the hybrid-ARIMA-EM model unique is its incorporation of additional components (e.g., ϕ and Ơ), which represent historical and future shifts in the data, along with a weighting mechanism that adjusts predictions dynamically. This weighting is calculated using a sigmoid function applied to the average length of the time series, enabling the model to fine-tune its forecasts based on overall data behavior.

Table 5. Original and forecasted data of crude oil prices for the years 2015-2021 for the combined models

Year	Original	XGBoost-GRU	Hybrid-ARIMA-GRU	Hybrid-ARIMA-XGBoost
2015	48.24	54.12	55.99	51.18
2016	48.33	50.77	55.98	49.68
2017	54.01	56.48	58.62	53.99
2018	74.73	63.52	60.56	61.63
2019	53.79	60.60	58.84	55.42
2020	35.49	53.30	56.40	41.70
2021	52.20	67.53	60.85	61.38

Table 6. Results of RMSE and MAPE for the combined forecasting models

Combining model	RMSE	MAPE
Hybrid-ARIMA-EM + XGBoost	0.81	0.015
Hybrid-ARIMA-EM + GRU	13.40	25.68
XGBoost + GRU	7.086	13.58

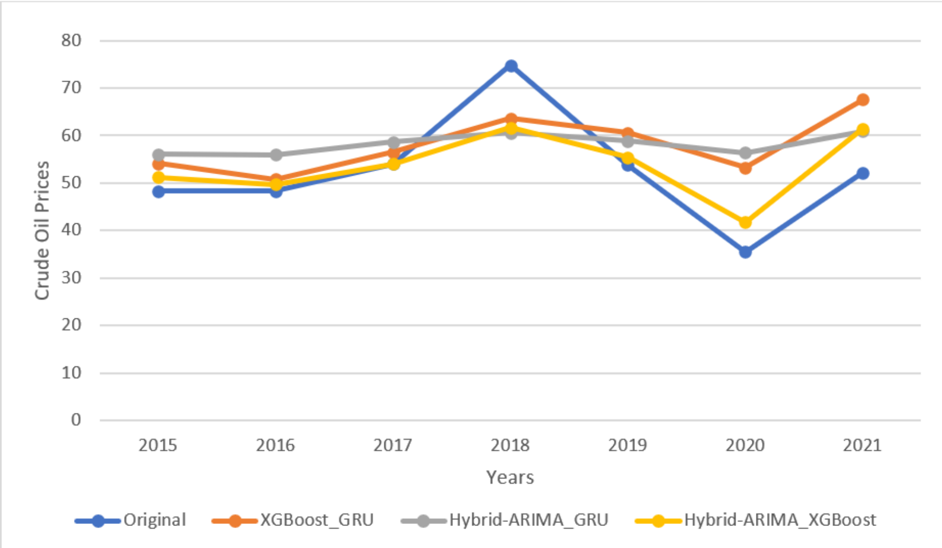

Figure 7. Forecasted crude oil prices for the years 2015-2021 using the combined forecasting models

However, ARIMA-based approaches are limited in their ability to capture nonlinear relationships within the data. This limitation remains even when ARIMA is enhanced with components such as those in hybrid-ARIMA-EM. For datasets like crude oil prices, where nonlinear dynamics (e.g., geopolitical events, market speculation) can significantly influence trends, this limitation becomes a key constraint.

XGBoost, on the other hand, excels in modeling complex, nonlinear relationships. It works by building an ensemble of decision trees that iteratively reduces errors in the predictions, adapting flexibly to a wide range of data behaviors [25]. However, XGBoost lacks intrinsic awareness of the temporal ordering and seasonal dependencies inherent in time-series data. This requires additional preprocessing, such as creating lagged features or capturing rolling averages, to incorporate these temporal dynamics. Refer to Figure 8.

With the integration of the two selected models, the strengths of both approaches will be revealed and will be more beneficial. The ARIMA component of the hybrid-ARIMA-EM can provide a robust foundation by capturing linear trends and seasonal patterns. This output can then be integrated into XGBoost as an engineered feature. Additionally, the ϕ and Ơ variables—representing historical shifts and forward-looking adjustments—could further enrich the feature set for XGBoost, enabling it to account for specific time-series dynamics. To summarize:

i) Enhanced feature engineering: the ϕ and Ơ components from hybrid-ARIMA-EM could be used as engineered features in the XGBoost model to incorporate time-series dynamics.
ii) Nonlinearity and temporal trends: the linear forecasting strength of ARIMA complements the nonlinear pattern recognition of XGBoost. ARIMA's predictions (ARIMA forecasting) can act as an additional feature or benchmark for XGBoost.

iii) Weight optimization: the sigmoid-weighting mechanism can be applied to features fed into XGBoost, emphasizing their influence based on the time-series average.

iv) Robustness to data variability: combining the models would integrate ARIMA's resilience to short-term variability with XGBoost's adaptability to complex relationships.

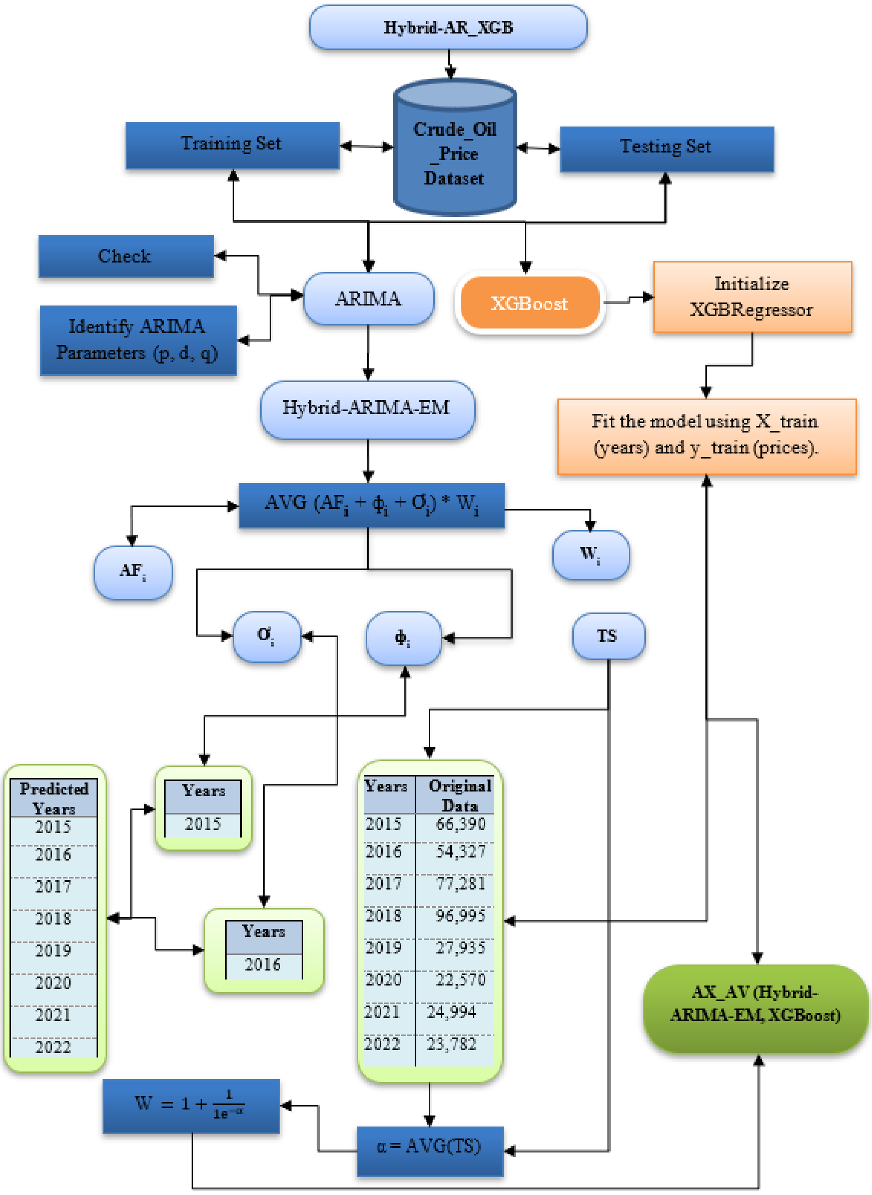

Figure 8. Hybrid-ARIMA-XGBoost prediction process

3. RESULTS AND DISCUSSION

As noted in the previous section, Tables 5 and 6, and Figure 7 show the performance of combined models with the Crude_Oil_Price dataset, with best results coming from hybrid-ARIMA-EM + XGBoost. A second dataset from Kaggle, the KL_apartment dataset, was chosen to show the accuracy of the proposed combined model on a different dataset. The dataset deals with apartment monthly rent prices through the years 1994 to 2024, and the period 2000-2024 was selected for the prediction process. Tables 7 and 8, and Figure 9 show that by combining XGBoost with the hybrid_ARIMA_EM model, the most accurate outcomes are obtained, which are so near to the original values. These results enhance the proposed suggestion of combining the stated models since it is the more precise method to make the predictions.

Table 7. Original and forecasted data of the years 2000-2024 for the combined models

Years	XGBoost + hybrid_ARIMA_EM	GRU + XGBoost	GRU + hybrid_ARIMA_EM	Original data
2000	2,199	2,200	2,099	2,300
2001	1,505	1,722	1,927	1,300
2002	1,506	1,699	1,904	1,200
2003	1,482	1,683	1,866	1,100
2004	1,846	1,932	1,978	1,800
2005	1,844	1,927	1,972	1,800
2006	1,582	1,725	1,908	1,400
2007	1,713	1,825	1,938	1,600
2008	2,634	2,524	2,158	3,000
2009	2,413	2,523	1,937	1,600
2010	1,976	2,023	2,000	2,000
2011	2,134	2,023	2,157	3,000
2012	2,042	2,073	2,015	2,100
2013	2,436	2,373	2,110	2,700
2014	1,845	1,923	1,968	1,800
2015	3,620	3,273	2,393	4,500
2016	1,713	1,823	1,936	1,600
2017	1,792	1,823	2,015	2,100
2018	2,634	2,523	2,157	3,000
2019	1,385	1,573	1,858	1,100
2020	2,108	2,123	2,031	2,200
2021	2,124	2,123	2,047	2,300
2022	2,174	2,173	2,047	2,300
2023	2,305	2,273	2,078	2,500
2024	1,776	2,273	1,549	1,450

Table 8. Results of RMSE and MAPE for the combined forecasting models

Combined model	RMSE	MAPE
Hybrid-ARIMA-EM + XGBoost	542.31	0.37
Hybrid-ARIMA-EM + GRU	868.79	0.59
XGBoost+ GRU	1115.48	0.38

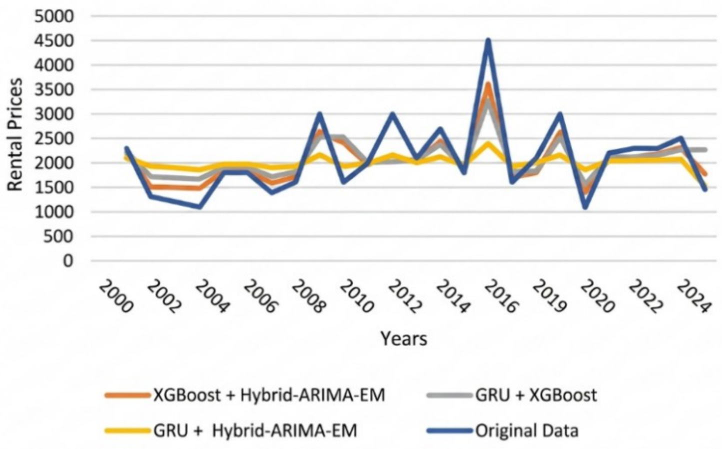

Figure 9. Forecast of the years 2000-2024 using the combined forecasting models

4. CONCLUSION

Price forecasting is a fundamental and critical element of financial analysis and economics. Effective forecasting is essential for informed decision-making. Accurate and reliable decisions hinge on precise predictions, and the use of advanced AI and sophisticated statistical models is the most effective approach to achieving these accurate results. This study meticulously highlights the distinctions among several predictive equations using ML models. The research demonstrates that integrating the hybrid-ARIMA-EM model and XGBoost is the most effective and accurate model for time series prediction that covers the data. While the present study champions the efficacy of the hybrid-ARIMA-EM model in terms of merging it with an effective classifier (XGBoost), it is crucial to acknowledge the multifaceted nature of the forecasting methodology landscape. Each model, encompassing ARIMA and ML models, embodies a distinct spectrum of advantages and limitations. The authors' intention is not to diminish the significance of these established models but rather to show the possible advancements achieved by the integration concept in elevating accuracy and effectiveness. In conclusion, searching for accurate price forecasting is an active journey marked by innovation and adaptation. It is hoped that this research can help and guide decision-makers through the maze of forecasting methodologies.

FUNDING INFORMATION

There is no agency funding or support involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Isam Ahmed M. Yaqoob	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Khairul Azhar Kasmiran		✓				✓		✓	✓	✓	✓	✓		
Teh Noranis Mohd Aris	✓		✓	✓		✓			✓		✓		✓	
Nor Azura Husin				✓		✓				✓			✓	
Mohd Yunus Sharum					✓		✓			✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors declare no conflict of interest whether financial, personal, or professional.

DATA AVAILABILITY

The study used two datasets collected from Kaggle, the first one holds the crude oil prices openly available at https://www.kaggle.com/datasets/sc231997/crude-oil-price and the second one holds the monthly rent of Kuala Lumpur apartments openly available at https://www.kaggle.com/datasets/ariewijaya/rent-pricing-kuala-lumpur-malaysi. Kaggle holds a various and large collection of datasets utilized for data science and machine learning projects.

REFERENCES

[1] N. Maaroufi, M. Najib, and M. Bakhouya, "Predicting the future is like completing a painting: towards a novel method for time-series forecasting," *IEEE Access*, vol. 9, pp. 119918–119938, 2021, doi: 10.1109/ACCESS.2021.3101718.

[2] Y. Yang and Y. Yang, "Hybrid method for short-term time series forecasting based on EEMD," *IEEE Access*, vol. 8, pp. 61915–61928, 2020, doi: 10.1109/ACCESS.2020.2983588.

[3] S. Yang and J. Liu, "Time-series forecasting based on high-order fuzzy cognitive maps and wavelet transform," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 6, pp. 3391–3402, Dec. 2018, doi: 10.1109/TFUZZ.2018.2831640.

[4] X. Yao, X. Fu, and C. Zong, "Short-term load forecasting method based on feature preference strategy and LightGBM-XGBoost," *IEEE Access*, vol. 10, pp. 75257–75268, 2022, doi: 10.1109/ACCESS.2022.3192011.

[5] A. Prakash, J. Thangaraj, S. Roy, S. Srivastav, and J. K. Mishra, "Model-aware XGBoost method towards optimum performance of flexible distributed raman amplifier," *IEEE Photonics Journal*, vol. 15, no. 4, pp. 1–10, Aug. 2023, doi: 10.1109/JPHOT.2023.3286272.

[6] Y.-C. Jin *et al.*, "Models for COVID-19 data prediction based on improved LSTM-ARIMA algorithms," *IEEE Access*, vol. 12, pp. 3981–3991, 2024, doi: 10.1109/ACCESS.2023.3347403.

[7] Y. Wang and Y. Guo, "Forecasting method of stock market volatility in time series data based on mixed model of ARIMA and XGBoost," *China Communications*, vol. 17, no. 3, pp. 205–221, Mar. 2020, doi: 10.23919/JCC.2020.03.017.

[8] A. Inteha and N. Al-Masood, "A GRU-GA hybrid model based technique for short term electrical load forecasting," in *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, Jan. 2021, pp. 515–519, doi: 10.1109/ICREST51555.2021.9331156.

[9] M. Unnisa and V. Ganesan, "An improved XGBoost classifier for micro expression recognition using hybrid optimization algorithm," in *2024 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, Apr. 2024, pp. 1–6, doi: 10.1109/IC3IoT60841.2024.10550313.

[10] Abdullah *et al.*, "Intelligent hybrid ARIMA-NARNET time series model to forecast coconut price," *IEEE Access*, vol. 11, pp. 48568–48577, 2023, doi: 10.1109/ACCESS.2023.3275534.
[11] X. Liu *et al.*, "Multistep-ahead prediction of ocean SSTA based on hybrid empirical mode decomposition and gated recurrent unit model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 7525–7538, 2022, doi: 10.1109/JSTARS.2022.3201228.
[12] I. M. A. Yaqoob, K. A. Kasmiran, T. N. M. Aris, N. A. Husin, and M. Y. Sharum, "Improving cash forecasting performance using a new hybrid-ANN-ARIMA-EM model," *Journal of Theoretical and Applied Information Technology*, vol. 101, no. 18, 2023.
[13] Y. Shi *et al.*, "Prediction of the low-frequency ambient noise level in the Northeast South China sea based on a gated recurrent unit network model," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024, doi: 10.1109/LGRS.2024.3360119.
[14] S.-E. Ryu, D.-H. Shin, and K. Chung, "Prediction model of dementia risk based on XGBoost using derived variable extraction and hyper parameter optimization," *IEEE Access*, vol. 8, pp. 177708–177720, 2020, doi: 10.1109/ACCESS.2020.3025553.
[15] B. Majhi and A. Kashyap, "Development of a hybrid model using deep neural network and XGBoost for mortality prediction of ICU patients," in *2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT)*, Apr. 2022, pp. 324–328, doi: 10.1109/CSNT54456.2022.9787564.
[16] A. H. Syed and T. Khan, "A supervised multi-tree XGBoost model for an earlier COVID-19 diagnosis based on clinical symptoms," in *2022 7th International Conference on Data Science and Machine Learning Applications (CDMA)*, Mar. 2022, pp. 219–223, doi: 10.1109/CDMA54072.2022.00041.
[17] C. Sheng and H. Yu, "An optimized prediction algorithm based on XGBoost," in *2022 International Conference on Networking and Network Applications (NaNA)*, Dec. 2022, pp. 1–6, doi: 10.1109/NaNA56854.2022.00082.
[18] N. Zhai, P. Yao, and X. Zhou, "Multivariate time series forecast in industrial process based on XGBoost and GRU," in *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, Dec. 2020, pp. 1397–1400, doi: 10.1109/ITAIC49862.2020.9338878.
[19] H. Shi, L. Wang, R. Scherer, M. Wozniak, P. Zhang, and W. Wei, "Short-term load forecasting based on adabelief optimized temporal convolutional network and gated recurrent unit hybrid neural network," *IEEE Access*, vol. 9, pp. 66965–66981, 2021, doi: 10.1109/ACCESS.2021.3076313.
[20] C. Wang, Z. Liu, H. Wei, L. Chen, and H. Zhang, "Hybrid deep learning model for short-term wind speed forecasting based on time series decomposition and gated recurrent unit," *Complex System Modeling and Simulation*, vol. 1, no. 4, pp. 308–321, Dec. 2021, doi: 10.23919/CSMS.2021.0026.
[21] M. Subramanian, S. L. Swetha, and V. R. Rajalakshmi, "Deep learning approaches for melody generation: an evaluation using LSTM, BILSTM and GRU models," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Jul. 2023, pp. 1–6, doi: 10.1109/ICCCNT56998.2023.10308344.
[22] J. Nasir, M. Aamir, Z. U. Haq, S. Khan, M. Y. Amin, and M. Naeem, "A new approach for forecasting crude oil prices based on stochastic and deterministic influences of LMD using ARIMA and LSTM models," *IEEE Access*, vol. 11, pp. 14322–14339, 2023, doi: 10.1109/ACCESS.2023.3243232.
[23] H. Yu, L. J. Ming, R. Sumei, and Z. Shuping, "A hybrid model for financial time series forecasting—integration of EWT, ARIMA with the improved ABC optimized ELM," *IEEE Access*, vol. 8, pp. 84501–84518, 2020, doi: 10.1109/ACCESS.2020.2987547.
[24] X. Zheng, J. Cai, and G. Zhang, "Stock trend prediction based on ARIMA-LightGBM hybrid model," in *2022 3rd Information Communication Technologies Conference (ICTC)*, May 2022, pp. 227–231, doi: 10.1109/ICTC55111.2022.9778304.
[25] Y. Mo, Y. Mei, and Y. Chen, "Used sailboat price prediction based on hybrid PSO-XGBoost algorithm," in *2023 International Conference on Algorithms, Computing and Data Processing (ACDP)*, Jun. 2023, pp. 258–262, doi: 10.1109/ACDP59959.2023.00048.

BIOGRAPHIES OF AUTHORS

Isam Ahmed M. Yaqoob received his M.Sc. in Computer Science from the University of Salford, United Kingdom in 2015. He is a data analyst and IT programmer at the Iraqi Ministry of Finance. His research interests include data analytics, statistical analysis, image processing, and programming science. He has many contributions during his work as a specialist in software affairs, data analysis, e-government applications, and the digitization of the Iraqi economy. He has achieved many studies and practical works through participation in high-level committees at the level of the Ministry of Finance and the Iraqi Prime Ministry. Besides, he has given lectures for 10 years in public and private universities in addition to having work experience in the private sector as a software project designer and programmer for a period exceeding 10 years. He can be contacted at email: gs59654@student.upm.edu.my.

Khairul Azhar Kasmiran received his Ph.D. from the University of Sydney, Australia, in 2012. He is a senior lecturer at the Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia. His interests include intelligent systems, optimization, deep learning, reinforcement learning, formal verification, and software development. He can be contacted at email: k_azhar@upm.edu.my.

Teh Noranis Mohd Aris received her Ph.D. in Computer Science from Universiti Kebangsaan Malaysia in 2008. She is an associate professor at the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia. Her research interests include data analytics, fuzzy systems, image processing, software agent, and programming science. Her expertise has led to the contribution of models in tutoring, image, natural disaster, and medical. She is a leader and member of several Malaysia Ministry of Education public grants, and she has been a supervisor for ongoing and graduated masters and Ph.D. students. The outcomes from her research grants are publications in indexed journals and software copyrights. She can be contacted at email: nuranis@upm.edu.my.

Nor Azura Husin received her Ph.D. in Artificial Intelligence from Universiti Putra Malaysia in 2013 and is currently a senior lecturer in the Department of Computer Science, Faculty of Computer Science and Information Technology at Universiti Putra Malaysia. Her research focuses on data mining, deep learning, machine learning, and intelligent computing, with expertise in developing prediction models, and she mainly contributed to the fields of education, medicine, and agriculture. Recently, she was appointed as the Editor-in-Chief of the upcoming Faculty of Computer Science and Information Technology Journal, marking a significant milestone in advancing academic research within the faculty. She can be contacted at email: n_azura@upm.edu.my.

Mohd Yunus Sharum holds a Bachelor of Computer Science (Software Engineering), Master of Science (Real-Time Software Engineering), and Ph.D. in Multimedia Systems. He is currently lecturing core undergraduate courses such as computer programming, computer organization and architecture, and operating systems. His research interests include software security, intelligent computing, natural language processing, and computer systems. He can be contacted at email: m_yunus@upm.edu.my.