International Journal of Social Statistics - USJ
Volume 02 | Issue 02 | November, 2025

Machine Learning Approaches for Predicting United States Crude Oil Prices: Model Selection and Key Determinants

D.M.D.S.D. Dissanayake[1], P.P.G.D. Asanka[2]
[1]School of Computing, Engineering and Technology, Robert Gordon University, Aberdeen, UK
[2]Department of Industrial Management, Faculty of Science, University of Kelaniya, Sri Lanka
[1]*d.s.diwyanjalee@gmail.com,* [2]*dineshasanka@gmail.com*

Abstract

This study applies machine learning algorithms to predict U.S. crude oil prices using monthly data from January 2010 to December 2023, incorporating 25 independent variables spanning supply, demand, inventory, speculation, monetary markets, stock markets, commodity markets, and technology indicators. A total of three regularization techniques, six ensemble methods, a Neural Network, and two hybrid models were implemented to evaluate predictive performance. Models were initially run with all variables; significant factors were identified, and parameter tuning was applied to enhance predictive accuracy. Ridge Regression, XGBoost, and a hybrid Neural Network-XGBoost model emerged as the top performers, with XGBoost identifying 21 significant contributors. The hybrid Neural Network-XGBoost model achieved the highest predictive accuracy, with an R^2 of 99.46% and a Mean Squared Error of 2.31, demonstrating its robustness and effectiveness in forecasting crude oil prices.

Keywords: Crude oil price prediction, Machine learning algorithms, Feature importance, United States

Introduction

Crude oil is a liquid fossil fuel found in porous rock layers beneath the Earth's surface, extracted for use as a combustible fuel or refined into various chemical products. It plays a critical role in the modern economy, serving as the primary energy source for transportation, generating electricity, and acting as a key input in the production of plastics, chemicals, and numerous consumer goods. The global economy is heavily dependent on crude oil, with fluctuations in its price influencing job markets, government revenues, international trade, and overall economic stability.

As one of the most traded commodities worldwide, crude oil prices are subject to frequent and sometimes extreme volatility. In recent decades, the international oil market has experienced significant price swings (Chen et al., 2020; Zhou et al., 2021). Even small fluctuations in national crude oil prices can significantly affect both importing and exporting countries (Gupta & Pandey, 2018). Sharp price shocks affect economic activity, monetary and fiscal policies, and energy strategies globally (Chai et al., 2011). For instance, the COVID-19 pandemic led to a steep decline in oil prices in early 2020, while the Russia-Ukraine conflict caused a substantial price increase in the first half of 2022.

The United States, the world's largest oil consumer (20.31% of global consumption) (Bajpai, 2022) and top petroleum liquids producer (18.9 million barrels per day in 2022) (U.S. Energy Information Administration, 2022), is particularly sensitive to oil price fluctuations. Such changes can influence U.S. real GDP through various channels, including aggregate demand shocks, oil-market specific demand shocks, and supply shocks, each with distinct effects on GDP, inflation, and consumer prices (Kilian & Murphy, 2014; Baumeister & Hamilton, 2019). Historical episodes, such as the 2008-2009 global financial crisis and the 2020 oil price collapse, illustrate how price volatility can disrupt production, trade, and broader economic performance (Mhalla, 2020). Accurate prediction of crude oil prices is therefore of significant importance for policymakers, investors, and industry stakeholders.

Numerous studies have explored crude oil price forecasting using statistical, econometric, and machine-learning methods. For example, Keerthan et al. (2019) applied multiple supervised machine learning algorithms to predict petroleum and diesel prices in India, where oil prices increased due to global inflation and rising crude import costs from Dubai and Saudi Arabia. Their model comparisons included the linear model, radial basis function model, polynomial model, random forest, and linear regression. Among these, linear regression achieved the highest accuracy. However, their study was limited to three financial variables and focused solely on the Indian context.

While such studies have yielded promising results, many have relied primarily on financial indicators, thereby limiting the breadth of factors considered. To enhance predictive accuracy, it is necessary to incorporate both financial and non-financial variables that capture broader economic, geopolitical, and market dynamics.

Time-series models remain a major strand of research on crude oil price forecasting. Approaches such as the Random Walk model (Safari & Davallou, 2018), Autoregressive Integrated Moving Average (ARIMA) model (Ding, 2018), and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) family models (Chen et al., 2018) have been shown to perform well in predicting price trends based on historical data. However, these models often require strong assumptions about data stationarity and linearity, which may limit their ability to capture complex nonlinear relationships (Huang & Wang, 2018).

Econometric models have also been widely applied, including Multivariate Regression (Wang et al., 2020), both with subset selection (Busari & Lim, 2021) and without (Zhou et al., 2021), as well as regularization-based methods such as Ridge Regression (Gallo et al., 2010) and Lasso Regression (Mensi et al., 2015). Although these models provide interpretable relationships between variables, they can suffer from overfitting, sensitivity to parameter selection (Wang et al., 2018), and limited capacity to model dynamic market interactions. Gallo et al. highlighted that supply is the primary driver of crude oil price changes, with demand fluctuations often following in response. Additionally, crude oil prices are influenced by prices for other commodities, reflecting interconnected market dynamics. Mensi et al. demonstrated a significant correlation between silver and crude oil prices, indicating that movements in precious metals markets can affect energy commodity prices.

Recent advancements have led to the application of Support Vector Regression (SVR) and other non-linear machine learning models, which can outperform traditional econometric methods in capturing intricate relationships among variables. Hybrid models combining econometric and machine learning approaches have also been explored to improve forecast accuracy.

Evaluating model performance is another critical aspect of forecasting research. Metrics such as the coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) have been widely used (Jha et al., 2024; Bildirici & Turkmen, 2015) to assess predictive capability from multiple perspectives. Combining these metrics provides a more comprehensive understanding of each model's strengths and limitations. Further, Bildirici and Turkmen identified a bidirectional causality between crude oil prices and precious metals such as silver and gold. However, their study revealed that the impact of precious metal price shocks on crude oil prices exhibits nonlinear characteristics. These findings underscore the importance of incorporating diverse financial determinants, including commodity price interactions, when developing predictive models for crude oil prices.

Despite these advances, many existing studies rely heavily on time-series approaches or a narrow set of variables. There remains a gap in research that systematically evaluates multiple supervised machine learning algorithms using a diverse set of financial and non-financial predictors. This study addresses that gap by applying and comparing several machine learning techniques to predict crude oil prices in the United States, aiming to identify the most accurate model and determine the key influencing factors.

Methodology

This study utilized a range of reputable secondary data sources to collect the data required for analysis. Out of the 30 independent variables considered, data for 5 variables, China Oil Import, EU 28 Countries PPI, Petroleum Inventory Total OECD, WTI Crack Spread, and Brent Crack Spread, were unavailable from the accessible secondary sources. The dataset covers the period from 2010 to 2023 and includes monthly data. All monthly data were de-seasonalized prior to analysis. For the dependent variable, WTI crude oil price, a box plot was created to examine its distribution, and a Kolmogorov-Smirnov test was conducted to assess normality. Distance correlation was used as a bivariate analysis to assess the strength of the relationship between the independent and dependent variables.

The study employed a series of advanced machine learning techniques, including regularization techniques, ensemble methods, neural networks, and hybrid modeling. Firstly, all available variables were included in the models, and their predictive performance was evaluated. Secondly, only statistically significant variables were retained, and the models were returned to test performance improvements. Finally, the significant variables were used in hyperparameter tuning to optimize model performance and improve predictive power.

The Python programming language was used to implement machine learning models, preprocess data, and conduct the main analysis. All coding, model development, training, and testing were performed in Google Colab, while Microsoft Excel was used for data management. The dataset was split into 80% for training and 20% for testing. A random seed was applied to ensure reproducibility of results across all experiments. The conceptual framework is illustrated by Figure 1.

Constructs	**Measurements**	**Response Variable**
Supply	Crude oil production, Total OPEC	
	Crude oil production, Total Non-OPEC	
	Crude oil production, World	
	Henry Hub natural gas spot price	
	U.S. tight oil production	
Demand	Petroleum consumption, Total OECD	
	China oil import	
	Fed fund effective	
	Kilian global economic index	
	US CPI index	
	US CPI energy	
	US PPI index	
	US PPI energy	
	EU 28 countries PPI	WTI crude oil price
	US PMI index	
Inventory	Petroleum inventory, Total OECD	
	Crude oil inventory, Total	
	Crude oil inventory, SPR	
	Crude oil inventory, non-SPR	
Speculation	Crude oil non-commercial net long ratio	
Monetary market	Real dollar index	
	The exchange rate of euro against US dollar	
Stock market	S&P 500 index	
	Dow Jones industrial index	
	NASDAQ index	
Commodity market	COMEX: Gold future closing price	
	LME: Copper future closing price	
Technology indicators	WTI-Brent spot price spread	
	WTI crack spread actual value	
	Brent crack spread actual value	

Figure 1: Conceptual framework of the study

Results and Discussion

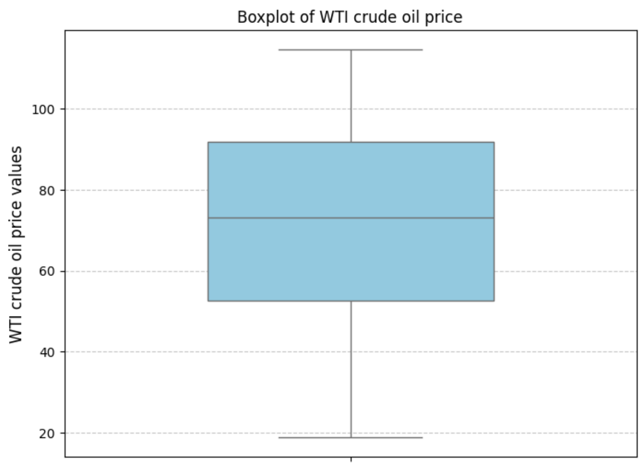

Figure 2: Boxplot of WTI crude oil price

According to Figure 2, the boxplot of WTI crude oil prices shows a median value slightly above 70. The minimum price is approximately 19, while the maximum reaches around 115, indicating the overall range of the data. The interquartile range spans from just above 50 (25th percentile) to just above 90 (75th percentile), suggesting that the middle 50% of the data falls within this interval. The whiskers extend to these values, with no extreme outliers, indicating that most data points lie within the range of approximately 19 to 115.

Kolmogorov-Smirnov Test statistic: 0.09132607159402639, p-value: 0.11395931111855928

Figure 3: Normality test for the WTI crude oil price

According to Figure 3, the Kolmogorov-Smirnov test yielded a p-value of 0.11396, which is greater than the significance level of 0.05. Therefore, fail to reject the null hypothesis and do not have sufficient evidence to conclude that the data deviates from a normal distribution. Based on these results, the dependent variable is likely normally distributed.

Independent Variable	Distance Correlation
COPTO	0.339275
COPTNO	0.241131
COPW	0.483113
HHNGSP	0.588562
USTOP	0.496846
PCTOECD	0.487678
FFER	0.308742
KGEI	0.349369
USCPII	0.491630
USCPIE	0.425185
USPPII	0.458792
USPPIE	0.796509
USPMII	0.370097
COIT	0.653063
COISPR	0.326095
COINSPR	0.763929
CONCNLR	0.465605
RDI	0.674489
EDER	0.663635
SP500IP	0.449789
DJIIP	0.445939
NASDAQIP	0.413822
GFP	0.293621
CFP	0.692242
BSP	0.966000

Figure 4: Distance correlation

According to Figure 4, a bivariate analysis using distance correlation revealed that the Brent spot price is exceptionally strongly associated with the target variable (0.966), suggesting it is a dominant factor influencing crude oil prices. US PPI energy (0.797) and crude oil inventory non-SPR (0.764) also showed strong correlations, suggesting substantial influence on the target. Moderate correlations were observed for the copper futures price (0.692), real dollar index (0.674), euro-USD exchange rate (0.664), and total crude oil inventory (0.653). Variables with weak relationships included the federal funds effective rate (0.309), the gold futures price (0.294), and total non-OPEC crude oil production (0.241).

Table 1: Performance comparison of regularization techniques

Technique	All Variables		Significant factors		Significant Factors After Final Tuning (Test)		Significant Factors After Final Tuning (Train)	
	R^2 (%)	*MSE*	*R^2 (%)*	*MSE*	*R^2 (%)*	*MSE*	*R^2 (%)*	MSE
Lasso	95.21	20.51	95.21	20.51	94.82	21.15	95.41	19.64
Ridge	95.83	17.84	95.83	17.84	95.55	18.02	96.01	17.06
Elastic Net	94.95	21.62	94.95	21.62	94.88	20.34	95.43	19.56

Table 1 summarizes the performance of Lasso, Ridge, and Elastic Net regression models across different stages of model development. After final parameter tuning and feature selection, all models show slight but meaningful improvements in training performance. For instance, Lasso Regression's training R^2 increased by approximately 0.21% (from 95.21% to 95.41%) with a corresponding MSE reduction of about 4.24% (from 20.51 to 19.64). Elastic Net showed a training R2 increase of approximately 0.51% (from 94.95% to 95.43%) and a substantial 9.53% decrease in MSE (from 21.62 to 19.56). Ridge Regression training R^2 increases by approximately 0.19% (from 95.83% to 96.01%), and its MSE decreases by about 4.37% (from 17.84 to 17.06), outperforming both, achieving the highest training R2 and the lowest MSE, demonstrating superior predictive accuracy and fit.

Out of the regularization techniques, Ridge Regression demonstrated the best performance, which may explain why its significant features remain unchanged after feature selection. In Ridge Regression, the model does not eliminate features by setting their coefficients to zero. Instead, it applies to an L2 penalty that shrinks the coefficients of less influential features toward zero while retaining them in the model. Consequently, all 25 independent variables may be considered significant in Ridge Regression, though with varying degrees of influence. The algorithm adjusts the magnitude of each coefficient, allowing some features to have a stronger effect on the prediction. In contrast, others contribute minimally, yet all remain part of the final model.

	Actual	Predicted
0	73.47	73.260693
1	88.06	88.386459
2	61.06	62.819789
3	84.96	81.233258
4	66.18	78.487288
5	70.64	73.498156
6	90.79	86.159947
7	99.74	98.599674
8	65.31	66.960268
9	48.24	45.226076
10	113.93	107.837446
11	80.11	80.491731
12	61.50	64.697978
13	96.38	97.955915
14	45.09	45.977786
15	54.01	52.177323

Figure 5: Actual vs predicted crude oil prices of Ridge Regression after tuning

According to Figure 5, the actual and predicted values for the initial data points are closely aligned, with only a slight difference observed between them.

Table 2: Performance comparison of ensemble techniques

Technique	All Variables		Significant factors		Significant Factors After Final Tuning (Test)		Significant Factors After Final Tuning (Train)	
	R^2 (%)	*MSE*	*R^2 (%)*	*MSE*	*R^2 (%)*	*MSE*	*R^2 (%)*	*MSE*
XGBoost	93.61	27.36	93.42	28.18	93.85	23.71	94.66	22.86
CatBoost	94.1	25.23	93.07	29.64	93.54	25.43	94.23	24.68
Gradient Boosting	93	29.96	92.58	31.75	92.56	29.72	93.26	28.84
Light Gradient Boosting	92.29	32.99	92.09	33.84	91.71	33.51	92.39	32.58
Random Forest	92.87	30.52	92.64	31.48	92.08	31.32	92.89	30.41
AdaBoost	93.17	29.21	93.12	29.43	93.15	26.74	93.96	25.83

Table 2 presents the performance metrics for six ensemble techniques evaluated in this study. Results are shown for models trained on all variables, models using significant factors, and models after final parameter tuning. Both training and testing R^2 (%) and MSE values are reported for the tuned models. Overall, all models exhibit strong predictive power, with R^2 values exceeding 90% across scenarios. After final tuning, all models demonstrate measurable improvements in training performance compared to using significant factors alone. For example, XGBoost's training R^2 increased from 93.42% to 94.66% (an improvement of approximately 1.33%), while its MSE decreased from 28.18 to 22.86 (a reduction of about 18.87%). CatBoost and AdaBoost also demonstrate competitive performance. The close alignment between training and testing metrics suggests that the models are well-regularized, avoiding overfitting or underfitting. Feature selection, combined with hyperparameter tuning, consistently improves model accuracy, underscoring the importance of these steps in crude oil price prediction.

```
Significant Features:
COPTO
COPW
HHNGSP
USTOP
FFER
KGEI
USCPII
USPPII
USPPIE
USPMII
COIT
COISPR
COINSPR
CONCNLR
RDI
EDER
SP500IP
DJIIP
GFP
CFP
BSP
```

Figure 6: Significant features in XGBoost

According to Figure 6, only the significant features identified by XGBoost were included, as it demonstrated the best predictive performance among the evaluated ensemble techniques. XGBoost uses feature importance metrics during training to automatically weight and select relevant variables, making it a robust feature selection method. The model identified 21 features as significant predictors, while 4 features: crude oil production, total non-OPEC, petroleum consumption, total OECD, US CPI energy, and NASDAQ index were determined to be insignificant. All other variables were retained as important factors influencing crude oil price predictions.

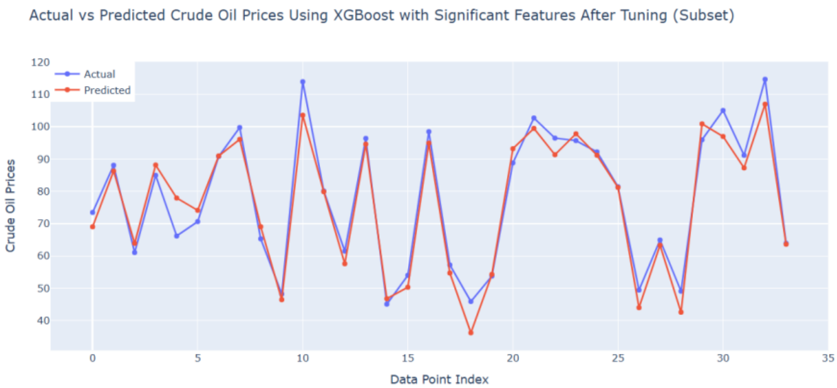

Figure 7: Interactive plot of actual vs predicted values for XGBoost with the significant features after tuning

According to Figure 7, a subset of data with index values ranging from 0 to 35 was selected, and the interactive plot shows that after tuning, the actual and predicted values align closely, indicating high model accuracy.

Table 3: Performance comparison of Neural Network

All Variables		All Variables After Final Tuning (Test)		All Variables After Final Tuning (Train)	
R^2 (%)	*MSE*	*R^2 (%)*	*MSE*	*R^2 (%)*	*MSE*
87.79	52.27	92	32	93.12	29.43

According to Table 3, the Neural Network model demonstrates a notable improvement after final tuning. The training R^2 increases from 87.79% to 93.12%, representing an improvement of approximately 6.07 percentage points. Correspondingly, the training MSE decreases from 52.27 to 29.43, which is a reduction of about 43.69%, indicating a significantly better fit. Unlike other models, Neural Networks are not ideal for feature selection; therefore, all variables were retained for the Neural Network model to ensure it utilizes the full dataset for optimal prediction accuracy.

Table 4: Performance comparison of Hybrid Model I

Technique	All Variables		All Variables After Final Tuning (Test)		All Variables After Final Tuning (Train)	
	R^2 (%)	*MSE*	*R^2 (%)*	*MSE*	*R^2 (%)*	*MSE*
Ridge and Neural Network	93.83	26.41	95	18.50	96.09	16.73

Firstly, Hybrid Model I was developed by combining the best-performing regularization technique, Ridge Regression, with a neural network. Since Ridge Regression does not eliminate features, but shrinks the coefficients of less important variables, all variables were retained and included in Hybrid Model I. This approach utilizes Ridge Regression's regularization capabilities alongside the Neural Network's strength in modeling complex nonlinear relationships.

According to Table 4, the Hybrid Model I demonstrated significant improvement after final tuning, with the training R^2 increasing from 93.83% to 96.09%, representing an improvement of approximately 2.41%, and the training MSE decreasing from 26.41 to 16.73, a reduction of about 36.66%. These results show that combining Ridge Regression and a Neural Network effectively captures both linear and nonlinear patterns in the data, improving crude oil price prediction accuracy.

Table 5: Performance comparison of Hybrid Model II

Technique	All Variables		Significant factors		Significant Factors After Final Tuning (Test)		Significant Factors After Final Tuning (Train)	
	R^2 (%)	*MSE*	*R^2 (%)*	*MSE*	*R^2 (%)*	*MSE*	*R^2 (%)*	*MSE*
XGBoost and Neural Network	92.34	32.79	97.15	12.19	98	3.50	99.46	2.31

Secondly, Hybrid Model II was developed by combining the best-performing ensemble technique, XGBoost, with a neural network. The significant factors identified by the XGBoost model were used as inputs to the Neural Network, leveraging XGBoost's effective feature selection alongside the Neural Network's ability to model complex nonlinear relationships.

According to Table 5, Hybrid Model II, which combines XGBoost and a Neural Network, shows remarkable performance improvements after feature selection and final tuning. Starting with the significant factors, the model achieves an R2 of 97.15% and an MSE of 12.19, indicating strong baseline performance. After final tuning, the training R^2 further improves to 99.46%, an increase of approximately 2.38% over the significant factors without tuning. Correspondingly, the training MSE decreases from 12.19 to 2.31, a substantial 81.04% reduction, indicating a much better fit to the training data. On the testing set, the model maintains strong generalization, with an R^2 of 98% and an MSE of 3.50, indicating only a slight decrease compared to the training results, demonstrating the model's robustness in predicting unseen data.

These results highlight the effectiveness of combining XGBoost's feature selection with the Neural Network's ability to model complex nonlinear relationships, achieving superior accuracy in crude oil price prediction.

	Actual	Hybrid Predicted
0	73.47	72.446747
1	88.06	86.701149
2	61.06	60.104122
3	84.96	83.895103
4	66.18	66.515511
5	70.64	69.983040
6	90.79	89.055946
7	99.74	99.346451
8	65.31	64.712212
9	48.24	47.449299
10	113.93	109.117699
11	80.11	79.380638
12	61.50	59.493340
13	96.38	95.842979
14	45.09	44.228355
15	54.01	52.803288

Figure 8: Actual vs predicted crude oil prices of Hybrid Model II after tuning

According to Figure 8, a detailed comparison of predicted and actual values is presented, highlighting the enhanced predictive accuracy achieved by Hybrid Model II with parameter tuning.

Conclusion and Suggestions

This study demonstrated that the Hybrid Model integrating XGBoost and Neural Network (Hybrid Model II) achieved the highest predictive accuracy for crude oil prices, with an R^2 of 99.46% and a minimal Mean Squared Error of 2.31. By utilizing XGBoost's ability to handle high-dimensional, complex data and Neural Network's capacity to capture nonlinear patterns, this hybrid approach outperformed all other tested models. Using the XGBoost model for feature selection, 21 out of 25 variables were identified as significant, spanning supply indicators (crude oil production total OPEC and world, Henry Hub natural gas spot price, U.S. tight oil production), demand indicators (fed fund effective rate, Kilian global economic index, US CPI and PPI indices including energy specific PPI, US PMI index), inventory indicators (crude oil inventory total, strategic petroleum reserve, non-SPR inventories), speculation indicator (crude oil non-commercial net long ratio), money market indicators (real dollar index, euro to USD exchange rate), stock market indicators (S&P 500 index price, Dow Jones industrial average price), commodity market indicators (gold and copper futures price), and technology indicator (brent spot price). Four variables were found to be insignificant supply indicators (total crude oil production, non-OPEC); demand indicators (total OECD petroleum consumption, US CPI energy); and stock market indicators (NASDAQ index price).

The Ridge Regression-Neural Network hybrid also delivered strong results (R^2 = 96.09%, MSE = 16.73), benefiting from Ridge Regression's regularization capability and the Neural Network's nonlinear modeling power. In contrast, the Light Gradient Boosting Machine recorded the lowest performance (R^2 = 92.39%, MSE = 32.58). Based on these results, Ridge Regression is recommended for regularization, while XGBoost is recommended for ensemble-based methods.

Accurate forecasting of crude oil prices is essential for the U.S. economy, informing energy policy, trade strategy, and business decision-making across key sectors. Governments can utilize these predictions to manage resources, mitigate price volatility, and strengthen energy security. Businesses and investors can apply these insights to optimize operations, guide investment decisions, and enhance market resilience.

Future work should include expanding datasets with additional variables, applying the developed models to other commodities to test cross-domain applicability, and exploring advanced hybrid approaches. In particular, time-series hybrid models and novel combinations of Neural Networks with machine-learning algorithms hold promise for further enhancing predictive accuracy.

References

Chen, W., Ma, F., Wei, Y., & Liu, J. (2020). Forecasting oil price volatility using high-frequency data: New evidence. International Review of Economics and Finance, 66, 1-12.

Zhou, W., Cheng, Y., Ding, S., Chen, L., & Li, R. (2021). A grey seasonal least square support vector regression model for time series forecasting. ISA Transactions, 114, 82-98.

Gupta, V., & Pandey, A. G. (2018). Crude oil price prediction using LSTM networks. World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 12, 226–230.

Chai, J., Guo, J. E., Meng, L., & Wang, S. Y. (2011). Exploring the core factors and its dynamic effects on oil price: An application on path analysis and BVAR-TVP model. Energy Policy, *39*(12), 8022-8036.

Bajpai, P. (2022). The world's top 5 oil consumers. *Nasdaq*. https://www.nasdaq.com/articles/the-worlds-top-5-oil-consumers/.

U.S. Energy Information Administration. (2022). *Frequently asked questions: What countries are the top producers and consumers of oil?* U.S. Department of Energy. https://www.eia.gov/tools/faqs/faq.php?id=709&t=6.

Kilian, L., & Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics, 29*(3), 454-478. https://doi.org/10.1002/jae.2322.

Baumeister, C., & Hamilton, J. D. (2019). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *American Economic Review, 109*(5), 1873-1910.

Mhalla, M. (2020). The impact of novel coronavirus (COVID-19) on the global oil and aviation markets. *Journal of Asian Scientific Research, 10*(2), 96-104.

Keerthan, J. S., Nagasai, Y., & Shaik, S. (2019). Machine learning algorithms for oil price prediction. *International Journal of Innovative Technology and Exploring Engineering, 8*(8).

Safari, A., & Davallou, M. (2018). Oil price forecasting using a hybrid model. *Energy*, *148*, 49-58.

Ding, Y. (2018). A novel decompose-ensemble methodology with AIC-ANN approach for crude oil forecasting. *Energy, 154*, 328-336.

Chen, Y., Zhang, C., He, K., & Zheng, A. (2018). Multi-step-ahead crude oil price forecasting using a hybrid grey wave model. *Physica A: Statistical Mechanics and Its Applications, 501*, 98-110.

Huang, L., & Wang, J. (2018). Global crude oil price prediction and synchronization based accuracy evaluation using random wavelet neural network. *Energy, 151*, 875-888.

Wang, J., Niu, T., Du, P., & Yang, W. (2020). Ensemble probabilistic prediction approach for modeling uncertainty in crude oil price. *Applied Soft Computing, 95*.

Busari, G. A., & Lim, D. H. (2021). Crude oil price prediction: A comparison between AdaBoost-LSTM and AdaBoost-GRU for improving forecasting performance. *Computers & Chemical Engineering, 155*.

Zhou, W., Cheng, Y., Ding, S., Chen, L., & Li, R. (2021). A grey seasonal least square support vector regression model for time series forecasting. *ISA Transactions, 114*, 82-98.

Gallo, A., Mason, P., Shapiro, S., & Fabritius, M. (2010). What is behind the increase in oil prices? Analyzing oil consumption and supply relationship with oil price. *Energy, 35*(10), 4126-4141.

Mensi, W., Hammoudeh, S., & Kang, S. H. (2015). Precious metals, cereal, oil and stock market linkages and portfolio risk management: Evidence from Saudi Arabia. *Economic Modelling, 51*, 340-358.

Wang, M., Zhao, L., Du, R., Wang, C., Chen, L., Tian, L., & Stanley, H. E. (2018). A novel hybrid method of forecasting crude oil prices using complex network science and artificial intelligence algorithms. *Applied Energy, 220*, 480-495.

Jha, N., Tanneru, H. K., Palla, S., & Mafat, I. H. (2024). Multivariate analysis and forecasting of the crude oil prices: Part I-Classical machine learning approaches. *Energy, 296*.

Bildirici, M. E., & Turkmen, C. (2015). Nonlinear causality between oil and precious metals. *Journal of Economic Behavior & Organization, 46*, 202-211.