JOURNAL OF PETROLEUM GEOMECHANICS
(JPG)

Original article

Presenting a model for predicting global crude oil prices based on artificial intelligence algorithm

Atefeh Heidari [1], Alireza Daghighiasli [1*], Ebrahim Abbassi [1], Marjan Damankeshideh [1]

1- Department of Economics, Central Tehran Branch, Islamic Azad university, Tehran, Iran

Received: 4 November 2023; Accepted: 30 December 2023

DOI: 10.22107/JPG.2024. 423782.1216

Keywords

Oil price prediction, Machine learning, Artificial intelligence, Prediction algorithm, Economic growth

Abstract

Predicting the price trend of crude oil and its fluctuations has always been one of the challenges facing traders in oil markets. Crude oil is the primary source of energy supply worldwide, serving as a fundamental pillar of the global economy. Furthermore, it plays a crucial role in financial markets and the development of the global economy. As a result, this study aims to review and evaluate algorithms, especially machine learning algorithms that are comparable based on real data (such as Bayesian networks, artificial neural networks, support vector machines, k-nearest neighbors), for crude oil price forecasting. . For this research, variables such as EU oil prices and US oil prices from 1990 to 2020 have been considered. Due to important events such as the outbreak of Covid-19, data from 1999 to 2020 have been selected. EU economic growth variables, US economic growth, gold price, EU death rate, US death rate, US oil reserves, US oil production, EU oil production, Corona disease, to predict the price of natural gas with West oil price Texas Global were selected. MAE, MSE, RMSE, R2, RMSLE, MAPE, TT (Sec) indicators have been used to check the accuracy and compare each algorithm. Python software was used for analysis.The results of this study have shown that, based on the machine learning algorithms used in the first model (Brent crude oil price), algorithms like Linear Regression, Least Angle Regression, Orthogonal Matching Pursuit, and Bayesian Ridge are the best algorithms for predicting the Brent crude oil price. Additionally, the results of machine learning algorithms in the second model (global Brent oil price and West Texas oil price) indicate that Linear Regression, Least Angle Regression, Orthogonal Matching Pursuit, and Bayesian Ridge are the best algorithms for predicting West Texas oil prices. Due to the complexity of the financial and economic markets, no model can predict with absolute accuracy, but using the available information and data, the proposed model can come very close to predicting the price of natural gas.

1. Introduction

Crude oil is the primary source of energy supply worldwide, serving as a fundamental pillar of the global economy. Furthermore, it plays a crucial role in financial markets and the development of the global economy [1]. This is because fluctuations in crude oil prices impact a country's economic activities, social stability, and national security [2].

In this regard, recent research reviews by the International Monetary Fund (IMF) indicate that when demand-based crude oil prices increase by 10% (with an initial price of approximately $50 to $55 per barrel), global production decreases by 0.1% to 0.15%. This means that a decline in crude oil prices can have repercussions on oil-exporting countries, potentially leading to budget deficits. Therefore, in the coming years, crude oil prices are expected to remain a key determinant governing investments [3].

Furthermore, crude oil serves as a critical variable in evaluating economic development, energy policy decisions, and stock markets. Despite the considerable instability in international fuel prices, their impact on domestic prices varies significantly from one country to another. Additionally, in some countries, international fuel price changes are promptly and entirely reflected in retail prices [4].

Therefore, being aware of future fluctuations in

*** Corresponding Author:** daghighiasli@gmail.com

crude oil prices can be instrumental in making better managerial decisions [5]. Economists and researchers have been grappling with the impact of crude oil price fluctuations on economic performance for years. Thus, making more accurate predictions of crude oil prices has been a concern and a priority for organizations, institutions, and research communities. In fact, there is no doubt that forecasting crude oil prices is highly beneficial for industries, governments, and individuals. Moreover, this matter holds profound scientific significance [2]. However, due to the conditions and complexities of the international crude oil market, predicting oil prices has become a challenging and yet pressing issue in the energy economy [6,7].

In light of the crucial role crude oil prices play in influencing key factors, including the performance of international financial markets, forecasting these prices becomes essential for shaping government policies and optimizing long-term production strategies [8]. Given the global significance of oil price prediction, previous studies and investigations have shed light on a notable gap and insufficient exploration of this subject, highlighting the necessity for a fresh perspective, which is the primary focus of this research. The urgency of addressing this issue from this particular standpoint is now considered imperative [1,4].

Hence, the central objective of this research is to unravel how machine learning and artificial intelligence techniques are harnessed to construct predictive models for crude oil prices. By delving into the intricate patterns and relationships within crude oil price data, this study aims to provide valuable insights that can aid in more accurate and informed decision-making across various sectors and industries. As we delve deeper into the complexities of predicting oil prices, we endeavor to enhance our understanding of this dynamic and multifaceted aspect of the global economy.

In addressing the urgency of this issue, this research aims to delve into the construction of a predictive pattern for crude oil using machine learning and artificial intelligence techniques. By harnessing the power of these advanced methodologies, this study endeavors to bridge the gap in current research by exploring how these techniques can be employed effectively for oil price prediction.

Machine learning algorithms, such as neural networks, support vector machines, random forests, and gradient boosting models, offer promising avenues for analyzing historical oil price data and discerning patterns that contribute to accurate predictions. Additionally, artificial intelligence methodologies, including natural language processing and deep learning frameworks, can be instrumental in extracting valuable insights from unstructured data sources, such as news articles, geopolitical events, and market sentiments, which significantly impact oil prices.

This research seeks to comprehensively analyze and compare various machine learning and artificial intelligence models in their effectiveness for crude oil price prediction. It aims to assess the suitability of these models in capturing the complexities of the oil market, considering factors like geopolitical tensions, economic indicators, supply and demand dynamics, and unexpected events like pandemics or natural disasters.

The research methodology involves collecting extensive historical data related to crude oil prices and relevant influencing variables, designing and implementing predictive models, and rigorously evaluating their performance based on accuracy, robustness, and adaptability to dynamic market conditions.

Ultimately, the findings of this study aspire to contribute significantly to the body of knowledge surrounding oil price prediction, providing insights and methodologies that can aid policymakers, industry stakeholders, and financial markets in making informed decisions and formulating effective strategies in response to the volatile nature of the oil market.

Therefore, since crude oil prices are among the most critical variables that significantly impact important factors, including the performance of international financial markets, forecasting oil prices can play a pivotal role in governmental policies and long-term production optimization. Consequently, given the global significance of oil price prediction, studies and investigations have been conducted, indicating a glaring oversight and insufficient examination of this matter, as well as the need for a new study, as is the focus of this research. The urgency of addressing this issue from this perspective is now deemed imperative. Therefore, this research seeks to provide an answer to the question of how the pattern of predicting crude oil using machine learning and artificial intelligence techniques is constructed?

2. A literature review and research background

Crude oil is a global commodity traded worldwide in both spot markets and through derivative contracts. The price of oil, in essence, refers to the current price per barrel of crude oil [9]. This product has been characterized by significant price fluctuations since the 1970s due to various global events and the emergence of major oil companies, which led to the establishment of large oil exchanges to transparently price and hedge the risks arising from price fluctuations [10].

Oil prices are highly volatile and adapt their price path according to the economic situation. Extreme fluctuations in oil prices are primarily a result of demand fluctuations, extraction costs, and its reserves. Supply and demand are the key factors determining oil prices [11]. Specifically, oil demand is linked to its consumption in developed and developing countries, while its supply is influenced by geopolitical events such as conflicts between Venezuela and the United States, Turkey and the Kurdish region of Iraq, or decisions made by OPEC member countries regarding production levels [12]. Nonetheless, crude oil prices have always been one of the most complex and challenging subjects to predict due to their relatively irregular, non-linear, and non-constant fluctuations. For this reason, many researchers have devoted significant efforts to developing various models for predicting crude oil prices. In fact, forecasting the trends and fluctuations in crude oil prices has always been one of the challenges facing traders in the oil markets [13].

Since World War II, scientists, sociologists, and practitioners in the field of operations research, among others who consider themselves futurists, have engaged in the establishment and advancement of quantitative and qualitative methods to make rational future predictions [14]. Consequently, forecasting methods have consistently been crucial tools in the hands of futurists [15]. The importance of predicting oil prices stems from the fact that it is essential not only for stakeholders such as oil-dependent industries, investors, financial companies, and risk managers but also for central banks in measuring financial and economic stability [16].

To delve further into this construct, the factors affecting oil prices and the forecasting tools must be examined in the light of theories surrounding this concept to substantiate the issue.

Many researchers have conducted studies in the field of crude oil prices, price prediction, and forecasting tools and algorithms. For instance, Elyakov et al. (2018) [16] found that factors such as changes in a country's macroeconomic conditions, global economic development, geopolitical changes, oil reserves, exchange rate fluctuations, and more are influential factors that significantly impact crude oil prices. Similarly, Peng et al. (2020) [17] consider economic and financial factors such as trade cycles and market sentiment, which increase the volatility of oil markets, as impactful elements on oil prices. Additionally, Liu et al. (2018) [18] discovered that factors affecting oil prices can be multidimensional financial factors. Sehgal and Pandey (2015) [19] conducted research using artificial intelligence methods to examine and predict oil prices. They believe that artificial intelligence methods are widely used as an alternative approach to conventional techniques for oil price prediction. They also argue that artificial intelligence methods encompass a wide range of techniques that can be employed to overcome the complexities and irregularities in oil price series. GAO and Lei (2017) [20] utilized machine learning paradigms for oil price prediction. They state that the advantage of this algorithm is its ability to capture changes in oil prices, as the model continuously updates whenever new oil price data becomes available.

In Iran, studies have also been conducted on the influential factors and prediction of crude oil prices. For example, Takroosta, Mohajeri, Mohammadi, Shakeri, and Ghasemi (2020) [21] analyzed the factors affecting oil prices. They found that factors such as risks arising from sub-indices of government stability, socio-economic conditions, domestic conflicts, external conflicts, corruption, religious tensions, law and order, and others can be influential factors. Furthermore, in the field of algorithms and price forecasting software, some studies have been mentioned. Razavi, SalimiFar, Mostafavi, and Baki Heskooyi (2014) [22] considered financial markets as influential on crude oil prices in Iran. In another article, Shahbazi and Salimian (2014) [23] conducted an evaluation and assessment of oil price prediction using the meta-analysis method. The new meta-analysis method, which is a combination of weighted least squares methods, was used for estimating oil prices, and they believe that the meta-analysis method has the best predictive power. Firouzi Jahantigh and Dehghani (2015) [24] showed that the genetic algorithm is effective in optimizing the architecture of the neural network and predicting the price of oil, and it performs better than the existing models. Rostami et al. (2017) [25] showed that based on the findings of financial econometric researchers, oil price as one of the most important macroeconomic variables affects the financial market and the economy of oil exporting countries. The results of oil price fluctuations using single and multi-regime MS-GARCH model indicate that the three-regime model is suitable for explaining variable behavior in the period under review.

Although many studies have been conducted in the field of oil price prediction, these models have both advantages and many disadvantages and limitations. Member countries of OPEC are often developing countries, and the economic indicators used in prediction models in these countries are subject to significant fluctuations even in the long term due to the unstable economic structure of these nations. This economic instability reduces the accuracy of oil price predictions by these models and can lead to long-term instability in the economic policies of these countries. Therefore, accurate short-term oil price prediction can mitigate the undesirable effects of political and economic events at the international level. To use these mentioned models and make accurate oil price predictions, it is essential to understand the international oil pricing system's structure. However, from the perspective that a study solely focuses on examining global oil price predictions and is a new study with fresh investigations, there is a noticeable research gap. We are seeking to address this gap and search for new and up-to-date information to better understand this subject.

Comparing the results of the importance of linear or near-linear methods has shown the desirability of using artificial neural networks and machine learning over other models. In the approach proposed in this article, comparing the results of different neural networks, all of which are linear in nature, has been used in predicting the oil price, so that the comparison of the results is closer to the reality of the oil price changes. Finally, while reviewing the conducted researches, it was found that until now genetic algorithm has only been used to determine the optimal weights of the neural network for predicting the price of crude oil; But the genetic algorithm has not been used to find the optimal values of weights, the number of hidden layers, the number of neural network nodes at the same time. In addition, the fitness function that simultaneously calculates the error and architecture of the neural network in crude oil price prediction and determines the best architecture in the output has not been used.

3. Methodology

The objective of the current research is to predict global oil prices using machine learning algorithms such as Bayesian, artificial neural networks, support vector machines, nearest neighbors, and others, utilizing historical data and variables derived from the existing research literature. Subsequently, these algorithms will be compared based on performance metrics like MAE, MAPE, and more, to determine the most suitable algorithm for forecasting global crude oil prices. The research model and variables are as follows:

$$\begin{aligned}\text{Oil Price} = {} & \beta_0 + \beta_1\text{Mortality} + \beta_2\text{Gas price} \\ & + \beta_3\text{Gold price} + \beta_4\text{Growth} \\ & + \beta_5\text{BCovid} + \beta_6\text{Oil reserves} \\ & + \beta_7\text{Oil Production} + \varepsilon\end{aligned}$$

in which the research variables are as follows:

Mortality

Gas price. The price of natural gas

Gold price

Growth Economic growth (Europe GDP, European economic growth, USGDP, American economic growth)

COVID is a livestock variable of zero and one caused by corona disease

Oil reserves

Oil production

Oil Price Global crude oil prices

Table 1. Operational definition of research variables

Variable name	symbpls	Calculation method
Oil price (Brent oil and Texas oil)	Oil price	Crude oil prices are numerically available in the World Bank for different years. Also, to measure the price of oil that has already been measured and the data on the site
Mortality rate	Mortality	https://www.macrotrends.net/charts/energy are available.
gas price	Gas price	A unit of measure for the number of deaths (in general, or due to a specific cause) divided by the total population per unit of time.
gold price	Gold price	In this research, in order to measure the price of gas, which has already been measured, from the statistics of the site
Economic Growth	Growth	https://www.macrotrends.net/charts/energy used.
Corona virus	Covid	In this research, to measure the price of gold from the statistics of the site
oil reserves	Oil reserves	https://www.macrotrends.net/charts/energy used.
Oil production	Oil production	Economic growth rates of countries are usually compared using the ratio of gross domestic product to population (income per capita). Data on economic growth rates are available from the World Bank and the statistical agencies of the countries that estimate them.

In this context, the scope of this research extends to encompass European Union member countries and the United States. Therefore, variables of interest in this study, including European Union oil prices and U.S. oil prices, will be considered from the years 1990 to 2020. The sampling method, aligned with philosophical assumptions, follows both a deductive and comparative inductive approach, in line with the quantitative research paradigm. Consequently, adhering to sampling theories in quantitative research, a probabilistic sampling method is employed.

4. Finding

Table 2 displays the results related to the descriptive statistics of the research variables. According to Table (1), since the COVID variable is a binary variable with values of zero and one, it has not been included in the descriptive statistics. Furthermore, due to differences in the measurement units of variables, some are very small, while others are very large. Therefore, in the next section, the data will be normalized using MinMaxScaler, and then the correlation between variables and regression and machine learning algorithms for predicting global oil prices will be investigated.

Table 2. Descriptive Statistics

	EuropeGDP	USGDP	GoldP	EurMort	USMort	USOilRs	USOILProd	EuroOilRP	covid	Gasprice	WTOP	BrentOP
	Economic Growth of the European Union	Economic Growth of the United States	Gold price	The mortality rate of the European Union	The mortality rate in the United States.	Reserves of oil in the United States.	Production of oil in the United States	Reserves of oil in the European Union	Coronavirus disease	Natural Gas Price	West Texas Intermediate (WTI) Crude Oil Price	Brent Crude Oil Price
Mean	1.81	1.96	591.26	245.91	257.70	27674.62	10779.71	1.39	0.06	4.25	46.98	49.28
Standard Deviation	1.80	2.10	455.41	47.41	90.26	4706.68	5041.94	0.56	0.25	1.19	21.67	24.03
Minimum	-5.75	-3.70	35.96	148.54	0.00	19121.00	4955.33	0.01	0.00	2.03	14.42	12.76
25th Percentile	1.58	1.09	313.75	210.14	221.64	22817.00	6436.88	1.15	0.00	4.25	30.73	33.56
50th Percentile (Median)	1.81	2.00	423.71	255.35	261.26	27674.62	8757.17	1.39	0.00	4.25	46.98	49.28
75th Percentile	2.53	3.24	609.55	286.19	321.20	31171.50	14514.00	1.72	0.00	4.25	47.82	51.16
Maximum	5.51	6.31	1798.89	311.47	371.81	39001.00	21582.00	2.20	1.00	8.86	99.67	111.57

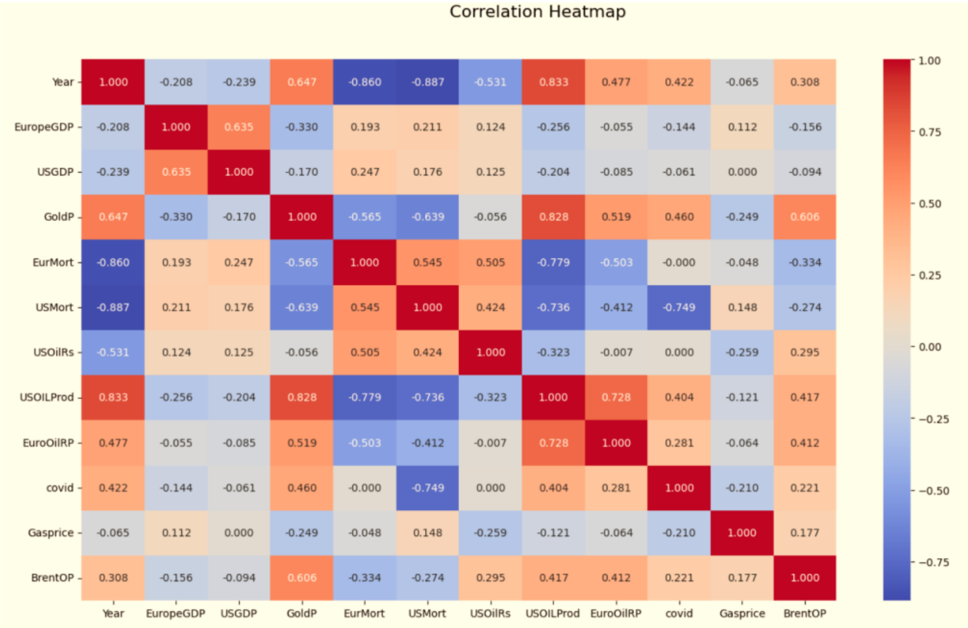

Fig. 1. Correlation between Independent Variables and Brent Crude Oil Price

To examine the correlation between research variables, a correlation heatmap was utilized in Python software. As observable in the above chart, there are significant relationships among the variables, including the economic growth of the European Union, the economic growth of the United States, the price of gold, the mortality rate of the European Union, the mortality rate of the United States, US oil reserves, US oil production, European Union oil production, COVID-19, and natural gas prices with the global Brent crude oil price. The highest correlation is related to the price of gold, which is approximately 0.60, indicating a strong positive and significant relationship (between 0.6 and 1 signifies a strong relationship) between the price of gold and the Brent crude oil price. Furthermore, the lowest level of correlation pertains to the variable "economic growth of the United States," which is approximately 0.094, indicating a weak negative and significant relationship (between 0.0 and 0.3 indicates a weak relationship) between the economic growth of the United States and the Brent crude oil price.

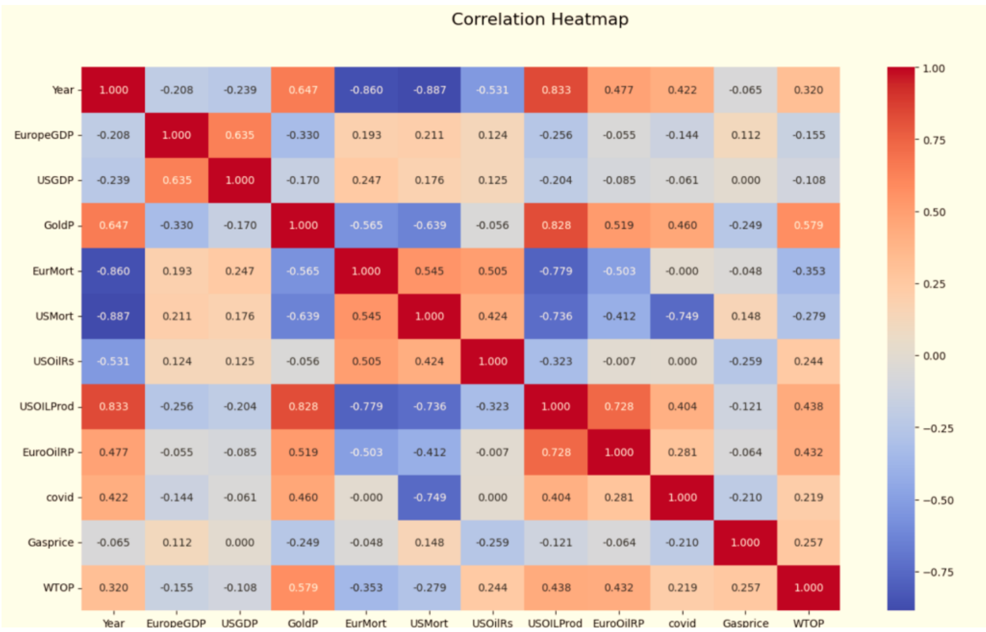

Fig. 2. Correlation Between Independent Variables and Texas Crude Oil Price

As shown in Figure (2), there is a meaningful correlation between the independent variables and the West Texas Global Crude Oil Price. In this diagram, the highest correlation is related to the gold price, which is 0.579, indicating a positive and statistically significant relationship and a moderate one (between 0.3 and 0.6 indicates a moderate relationship) between the gold price and the West Texas Global Crude Oil Price. Additionally, the lowest correlation is associated with the variable of U.S. economic growth, which is -0.108, indicating a negative and statistically significant weak relationship (between 0.0 and 0.3 indicates a weak relationship) between U.S. economic growth and the West Texas Global Crude Oil Price.

After preprocessing the data, before predicting the dependent variable, the data is divided into two parts: training and testing. In this section, 20% of the data is considered test data, and 80% of the data is considered training data.The foregoing data splitting methods can be implemented once we specify a splitting ratio. A commonly used ratio is 80:20, which means 80% of the data is for training and 20% for testing. Other ratios such as 70:30, 60:40, and even 50:50 are also used in practice. There does not seem to be a clear guidance on what ratio is best or optimal for a given dataset. The 80:20 split draws its justification from the well-known Pareto principle, but that is again just a thumb-rule used by practitioners. The training data is sent to the model as input data for model training (algorithm), and the model is designed or, in other words, fitted based on the training data. The general idea is that in each iteration, we randomly select samples from the entire data set as the test set. For example, if we decide that 20% of the data set will be our test set, 20% of the samples will be randomly selected and the remaining 80% will become the training set.

Furthermore, to select the best parameters and achieve the highest score for each algorithm, the evaluation criterion, which includes accuracy measures such as MAE, MSE, RMSE, R SQUARE, RMSLE and MAPE, is taken as an input parameter here. Additionally, for the selection and calculation of the accuracy of each model, the training data is also divided into two parts: training and testing. This partitioning is performed based on the k-fold algorithm, which operates as follows:

Data classification using the k-fold cross-validation method:

Assuming k is equal to 10, in this stage, one-tenth of the data is set aside for testing, and the remaining nine-tenths of the data, or the so-called training data, are used to train the model. This algorithm, for predicting parameters, employs nine sets of input data, and ultimately, it places the predicted variable under the model's control to provide accuracy based on the trained data. Now, the trained model is tested with the data that has been set aside for testing, and the prediction accuracy is obtained. The researcher repeats this step 10 times, with the difference being that the data will change in each iteration. In other words, one set of data that was set aside for testing is input into the model, while another set of data is removed from the model for testing, and the learning and prediction process is started again for these data points. This process is repeated enough times so that all data sets are used once as test data. Given that we have 10 sets of data, it is evident that the researcher must perform the explained cycle 10 times. Ultimately, the accuracy obtained from each set of data is averaged, and this average represents the final result for the entire dataset.

Next, to make predictions using the model, the test data is fed into the fitted model, and the target variable is predicted using the model. To calculate the percentage error between predictions and actual data, the target variable (global crude oil price) is compared with the data, and the deviation between predictions and actual values of the dependent variable (y) in the test data is calculated using MAE, MSE, RMSE, RMSLE, MAPE.

All of the above steps are performed for the following 20 models, and the accuracy index is calculated for each model.

Linear Regression, Least Angle Regression, Orthogonal Matching Pursuit, Bayesian Ridge, Huber Regressor, Extra Trees Regressor, Gradient Boosting Regressor, Extreme Gradient Boosting, AdaBoost Regressor, Random Forest Regressor, Decision Tree Regressor, Ridge Regression, Passive Aggressive Regressor, K Neighbors Regressor, Light Gradient Boosting Machine, Lasso Regression, Elastic Net, Lasso Least Angle Regression, Dummy Regressor, MLP Regressor

Results of the first model (Global Brent Crude Oil Price):

$$\mathrm{BrentOP} = \mathrm{F}(\mathrm{Mortality}, \mathrm{Gas\ price}, \mathrm{Gold\ price}, \mathrm{Growth}, \mathrm{BCovid}, \mathrm{Oil\ reserves}, \mathrm{Oil\ Production})$$

In the first model, the dependent variable is the Global Brent Crude Oil Price.

According to the above model, algorithms such as Linear Regression, Least Angle Regression, Orthogonal Matching Pursuit, Bayesian Ridge, Huber Regressor, Extra Trees Regressor, Gradient Boosting Regressor, Extreme Gradient Boosting, AdaBoost Regressor, Random Forest Regressor, MLP Regressor, Decision Tree Regressor, Ridge Regression, and Passive Aggressive Regressor have been able to accurately predict the Global Brent Crude Oil Price using independent variables such as mortality rate, gas price, gold price, economic growth, COVID, oil reserves, and oil production. As observed in the table above, the coefficient of determination (R-squared), which indicates the percentage of variance explained by the independent variables, is high, while MAE, MSE, RMSE, RMSLE, and MAPE, which represent the percentage of error between predicted and actual values of the dependent variable, are low. TT represents the execution time of the algorithm, and the lower this number is, the better the algorithm performs. Linear Regression, Least Angle Regression, Orthogonal

Matching Pursuit, and Bayesian Ridge algorithms are the best algorithms for predicting the Brent Crude Oil Price, with an R-squared value of 0.98 and zero error percentage, accurately forecasting the Brent Crude Oil Price.

Table 3. Comparison of Algorithm Accuracy in Predicting the Global Brent Crude Oil Price

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
lr	Linear Regression	0.000	0.000	0.000	0.989	0.000	0.000	0.030
lar	Least Angle Regression	0.000	0.000	0.000	0.989	0.000	0.000	0.047
omp	Orthogonal Matching Pursuit	0.000	0.000	0.000	0.989	0.000	0.000	0.051
br	Bayesian Ridge	0.000	0.000	0.000	0.989	0.000	0.000	0.045
huber	Huber Regressor	0.000	0.000	0.000	0.988	0.000	0.001	0.083
et	Extra Trees Regressor	0.012	0.000	0.018	0.988	0.012	0.059	0.204
gbr	Gradient Boosting Regressor	0.015	0.001	0.024	0.976	0.016	0.069	0.101
xgboost	Extreme Gradient Boosting	0.021	0.002	0.034	0.951	0.022	0.081	0.057
ada	AdaBoost Regressor	0.031	0.002	0.039	0.931	0.027	0.127	0.098
rf	Random Forest Regressor	0.028	0.003	0.041	0.914	0.028	0.158	0.232
mlp	MLP Regressor	0.055	0.008	0.074	0.864	0.051	0.306	0.285
dt	Decision Tree Regressor	0.031	0.005	0.049	0.848	0.033	0.120	0.029
ridge	Ridge Regression	0.062	0.007	0.078	0.818	0.055	0.430	0.031
par	Passive Aggressive Regressor	0.059	0.008	0.075	0.794	0.054	0.396	0.047
knn	K Neighbors Regressor	0.090	0.023	0.141	0.377	0.095	0.478	0.033
lightgbm	Light Gradient Boosting Machine	0.159	0.040	0.188	0.034	0.134	1.102	0.155
lasso	Lasso Regression	0.170	0.062	0.243	0.608	0.175	1.480	0.030
en	Elastic Net	0.170	0.062	0.243	0.608	0.175	1.480	0.030
llar	Lasso Least Angle Regression	0.170	0.062	0.243	0.608	0.175	1.480	0.047
dummy	Dummy Regressor	0.170	0.062	0.243	0.608	0.175	1.480	0.047

Table 4. Results of Algorithm Accuracy Comparison in Predicting Global West Texas Crude Oil Price

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
lr	Linear Regression	0.000	0.000	0.000	0.987	0.000	0.000	0.032
lar	Least Angle Regression	0.000	0.000	0.000	0.987	0.000	0.000	0.027
omp	Orthogonal Matching Pursuit	0.000	0.000	0.000	0.987	0.000	0.000	0.027
br	Bayesian Ridge	0.000	0.000	0.000	0.987	0.000	0.000	0.033
huber	Huber Regressor	0.000	0.000	0.000	0.987	0.000	0.001	0.080
et	Extra Trees Regressor	0.012	0.001	0.020	0.985	0.012	0.050	0.206
gbr	Gradient Boosting Regressor	0.023	0.002	0.040	0.944	0.026	0.074	0.096
rf	Random Forest Regressor	0.031	0.004	0.049	0.927	0.031	0.130	0.281
ada	AdaBoost Regressor	0.032	0.004	0.051	0.916	0.033	0.149	0.094
xgboost	Extreme Gradient Boosting	0.028	0.005	0.046	0.905	0.029	0.101	0.054
dt	Decision Tree Regressor	0.027	0.003	0.047	0.882	0.031	0.110	0.044
ridge	Ridge Regression	0.060	0.006	0.074	0.840	0.052	0.423	0.027
par	Passive Aggressive Regressor	0.061	0.006	0.074	0.832	0.052	0.375	0.045
knn	K Neighbors Regressor	0.078	0.017	0.115	0.602	0.076	0.410	0.048
lightgbm	Light Gradient Boosting Machine	0.157	0.036	0.182	0.122	0.129	0.877	0.140
lasso	Lasso Regression	0.188	0.072	0.252	0.692	0.182	1.733	0.028
en	Elastic Net	0.188	0.072	0.252	0.692	0.182	1.733	0.028
llar	Lasso Least Angle Regression	0.188	0.072	0.252	0.692	0.182	1.733	0.028
dummy	Dummy Regressor	0.188	0.072	0.252	0.692	0.182	1.733	0.026
mlp	MLP Regressor	885.80	814565.33	898.38	0.347	6.48	5322.18	0.3020

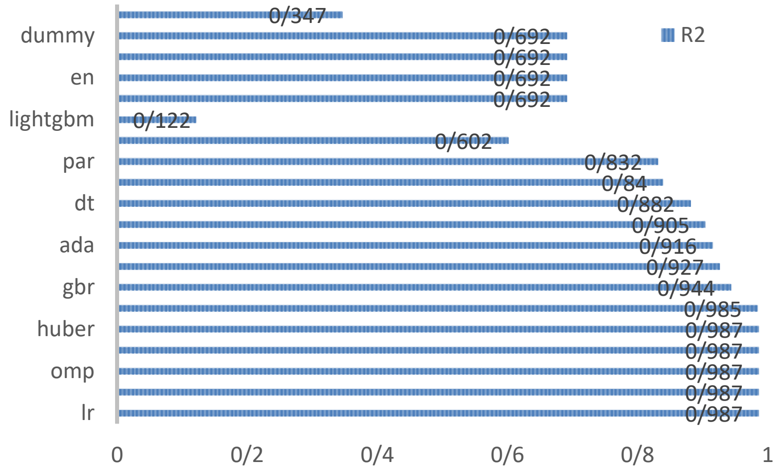

Fig. 3. The result of the coefficient of determination for each prediction

Based on the information and data of the research, the researcher has discussed several estimates in different ways. First, he calculated the data of the European Union and the United States at the same time, and some examples of different estimates can be seen in the regression appendix number one, and the final estimate can be seen in the following table:

Table 5. dependent variable of Brent oil price (BrentOP)

Variable	symbol	Coefficient	Probability statistic	t statistic
Year	Year	0.0441	0.719	0.362
European economic growth	EuropeGDP	0.2322	0.147	1.473
America's economic growth	USGDP	-0.0020	0.985	-0.018
gold price	GoldP	0.3222	0.087	1.747
Death rate and Mirarupa	EurMort	-0.1373	0.372	-0.902
America's death rate	USMort	-0.0997	0.580	-0.557
America's oil reserves	USOilRs	0.127	0.320	1.005
American oil production	USOILProd	-0.2544	0.239	-1.191
Europe's oil reserves	EuroOilRP	0.0872	0.485	0.703
Corona virus	covid	-0.0016	0.993	-0.009
Natural gas price	Gas price		–	–
The coefficient of determination	0.92	Probability statisticsF	2.87e-24	
Adjusted coefficient of determination	0.90	durbin watson test	1.466	

Based on the table above, it is clear that in the final model, by removing the gas price variable, Watson's camera has increased and reached 1.46. It also shows the coefficient of determination of 92%, which indicates that the independent variables have been able to explain 92% of the changes in the dependent variable. The F probability statistic is less than 5%, which indicates the existence of a regression line. Based on the t-statistics, none of the variables had a significant effect on the dependent variable, i.e. Brent oil price.

The above table is the regression estimate for the European Union and the United States with the dependent variable of West Texas oil. After estimating several models (can be seen in the appendix) and removing the natural price, the researcher reached the final model, so that the coefficient of determining the number is 91% and the F probability statistic is less than 5%. Watson's camera shows the number 1.467. Based on the t-statistics, none of the variables had a significant effect on the dependent variable, i.e. the price of West Texas oil. Now, by separating the two statistical communities and separating the European Union from the United States, the researcher intends to estimate new estimates. As in the past, different estimates have been estimated to achieve an ideal estimate, which can be seen in the regression appendix. The forecast of Brent and West Texas Intermediate (WTI) oil prices depends on several reasons, some of which include the following:

Reference basket: Brent oil pricing is based on condensates that are taken from different regions of the world in online information markets. Meanwhile, West Texas oil is pricing based on oil that is extracted in a specific region in the United States. This difference in reference sources may cause differences in price predictions.

Table 6. dependent variable of West Texas oil price(WTOP)

Variable	symbol	Coefficient	Probability statistics	t statistic
Year	Year	0.0278	0.833	0.212
European economic growth	EuropeGDP	0.2632	0.127	1.551
America's economic growth	USGDP	0.0098	0.935	0.082
gold price	GoldP	0.1731	0.375	0.895
Death rate and Mirarupa	EurMort	-0.1066	0.518	-0.652
America's death rate	USMort	-0.1408	0.475	-0.720
America's oil reserves	USOilRs	0.0741	0.568	0.575
American oil production	USOILProd	-0.1562	0.492	-0.692
Europe's oil reserves	EuroOilRP	0.0595	0.656	0.448
Corona virus	covid	0.0107	0.955	0.057
Natural gas price	Gas price	-	-	-
The coefficient of determination	0.91	Probability statisticsF	2.39e-23	
Adjusted coefficient of determination	0.89	durbin watson test	1.467	

Different Transactions and Markets: Brent and WTI oil markets may be affected by various variables. It is possible that different markets have different changes in supply and demand, government policies, sanctions and other things that cause differences in price forecasts.

Transportation and Logistics Issues: Costs associated with transporting Brent and WTI may vary. For example, the cost of transporting oil from the production area to international markets for Brent oil may be different than the cost of transporting WTI oil. This issue can also affect the price difference.

Temporal Matters: Changes in the demand for oil as well as the amount available in the market may change over time. These changes may have a slight impact on price forecasts.

Financial and trading activities: Financial and trading market participants may also make their forecasts based on different criteria for Brent and WTI oil, which can lead to differences in forecast performance.

Basically, oil price forecast analysis is based on different cases and these differences may affect the forecast results.

For Brent oil price (BrentOP) and West Texas Intermediate oil price (WTOP) in EU and US, the results showed that independent variables such as year, European economic growth, US economic growth, gold price, European mortality rate, US mortality rate, reserve American oil, American oil production, European oil reserves, and the corona virus had no significant effect on oil prices. The coefficient of determination showed that the independent variables were able to explain the changes in Brent and West Texas oil prices by 92% and 91%. After removing several Variable in the separate analysis for each country (EU and USA), the Watson camera improved and the coefficient of determination reached 0.90. However, none of the variables had a significant effect on the price of oil, and a significant effect was observed only on some variables.

Considering the results of separate analysis for each country (European Union and USA) and improving the coefficient of determination to 0.90, it seems that Watson's camera is included as a significant determining variable for predicting oil prices in these analyses. This result shows that separating the data for the EU and the US improves the model and this is the best possible variable to explain oil price changes in these analyses.

5. Conclusion

According to the above model, algorithms such as Linear Regression, Least Angle Regression, Orthogonal Matching Pursuit, Bayesian Ridge, Huber Regressor, Extra Trees Regressor, Gradient Boosting Regressor, Random Forest Regressor, AdaBoost Regressor, Extreme Gradient Boosting, Decision Tree Regressor, Ridge Regression, Passive Aggressive Regressor, and K Neighbors Regressor have successfully and accurately predicted the Global West Texas Crude Oil Price using independent variables such as mortality rate, gas price, gold price, economic growth, COVID, oil reserves, and oil production. As seen in the table above, the coefficient of determination (R-squared), which indicates the percentage of variance explained by independent variables, is high, while MAE, MSE, RMSE, RMSLE, MAPE, which represent the percentage of error between predicted and actual values of the dependent variable, are low. The "TT" indicates the execution time of the algorithm, and a lower value indicates better algorithm performance. Linear Regression, Least Angle Regression, Orthogonal Matching Pursuit, and Bayesian Ridge are the best-performing algorithms for predicting West Texas Crude Oil Price, achieving an R-squared of 0.98 and zero error percentage, indicating accurate prediction

Discussion

In this study, a model for predicting crude oil prices using machine learning was presented. Based on the machine learning algorithms employed in the first model (Global Brent Oil Price), it was determined that Linear Regression, Least Angle Regression, Orthogonal Matching Pursuit, and Bayesian Ridge are the best-performing algorithms for accurately predicting Brent oil prices, with an R-squared of 0.98 and zero error percentage.

The results of machine learning algorithms in the second model (Global West Texas Oil Price) also indicate that Linear Regression, Least Angle Regression, Orthogonal Matching Pursuit, and Bayesian Ridge are the best-performing algorithms for predicting West Texas oil prices, achieving an R-squared of 0.98 and zero error percentage.

Hasan et al. (2022) [1] demonstrated, similar to this research, that the proposed ensemble learning model outperforms existing prediction models in terms of prediction errors. The proposed model shows good predictive performance for both short-term and long-term forecasts, which is beneficial for stakeholders and industries dependent on this energy source.Therefore, in line with this research, the consideration of multi-modal data was emphasized. Behmiri and Manso (2013) [24] achieved high accuracy in predicting oil prices using machine learning techniques. He et al. (2023) [3] demonstrated that artificial intelligence and machine learning, along with decision trees, are suitable tools for price prediction in the market.The lower error rate and high accuracy of the proposed model are promising and potentially valuable for policy makers and shareholders in making future decisions. The oil market is highly volatile in the short term, but the proposed model allowed us to capture the nature of these short-term fluctuations. The diversity of the base prediction models used in our ensemble approach considered various fluctuation patterns in short-term crude oil price segments. These findings are in line with those provided by Gao et al. (2017) [25] and Li et al. (2021) [26] and Razavi et al. (2014) [22]. The observed increase in prediction performance for the daily crude oil price could be interpreted as confirmation of the feasibility of modeling these fluctuations using the blending approach. Interestingly, there are also differences in the prediction performance for the two crude oil prices, Brent and WTI. The results in terms of the error criteria showed that more accurate forecasts were obtained for the WTI crude oil prices. Similar trends have been reported by Nademi and Nademi (2018) [27] in their work on predicting crude oil prices using semiparametric Markov models. Similarly, Karasu and Altan (2022) [28] achieved better LSTM performance for the WTI crude oil prices . Therefore, this research demonstrates that explaining oil price changes requires addressing more complex issues, and the interactive effects of variables need to be carefully examined. Furthermore, the use of WTI Cushing as a determining variable shows improvements in the models, indicating that this variable may be crucial in explaining significant oil price variations

It can be concluded that linear models such as Linear Regression and other regression models like Least Angle Regression, Orthogonal Matching Pursuit, and Bayesian Ridge delivered superior performance in predicting Brent crude oil prices with remarkable accuracy and minimal error values (MAE, MSE, RMSE, RMSLE, MAPE). Not only did these models demonstrate excellent predictive capabilities, but they also exhibited efficient execution times, positioning them as the top choices for forecasting this variable. These findings hold significant relevance for decision-making processes within the oil market and investments associated with Brent crude oil prices. Moreover, the top-performing models stand as valuable predictive tools for informing future decisions in the realm of Brent crude oil price forecasting.

In light of these outcomes, it is apparent that the selected linear models, characterized by their precision and efficiency, can empower stakeholders to make more informed and data-driven choices, thus enhancing their capacity to navigate the complexities of the Brent crude oil market and optimize their investments.Based on the obtained results, the following results are suggested:

Increase in production: The development of advanced technologies in oil drilling and extraction can lead to an increase in oil production. Improving extraction methods and increasing the productivity of oil reservoirs can help countries to increase crude oil production and be more present in the oil market.

Cost reduction: New research and technology can help reduce the costs associated with oil production and extraction. This cost reduction can help oil companies to be profitable in a situation where oil prices are low.

Risk management: The development of advanced technologies in oil-related risk management, including the development of decision-making systems and more accurate oil price forecasting, will help companies and countries to successfully manage changes in the oil market.

Reducing dependence on oil prices: By developing alternative energy sources and increasing diversity in energy supply, countries can reduce dependence on oil prices. For example, the development of renewable energy sources can help reduce dependence on oil.

Environmental protection: Research in the field of sustainable petroleum technologies can help protect the environment. The use of low-consumption and clean technologies helps to reduce pollution and the negative effects of the oil industry on the environment.

6. Reference

[1] Hasan, M. Abedin, M. and hajek, p. Sultan, N. and Lucey, BM. (2022).A Blending Ensemble Learning Model for Crude Oil Price Prediction. Available at SSRN: https://ssrn.com/abstract=4153206 or http://dx.doi.org/10.2139/ssrn.4153206

[2] Sen, A, Dutta Choudhury, K. Kumar Datta, T. (2023) An analysis of crude oil prices in the last decade (2011-2020): With deep learning approach. *PLoS ONE* 18(3): e0268996. https://doi.org/10.1371/journal.pone.0268996

[3] He, H. Sun, M. Li, X. & Mensah, I, (2022). A novel crude oil price trend prediction method: Machine learning classification algorithm based on multi-modal data features. Energy, Elsevier, vol. 244. https://doi.org/10.1016/j.energy.2021.122706

[4] Karathanasopoulos, A. Zaremba, A. Osman, M. and Mikutowski, M. (2019). Oil Forecasting Using Artificial Intelligence. *Theoretical Economics Letters*, 9, 2283-2290. https://doi.org/10.4236/tel.2019.97144

[5] Manickavasagam J, Visalakshmi S, Apergis N. (2020). A novel hybrid approach to forecast crude oil futures using intraday data. *Technol Forecast Soc Change*. 158:120126. https://doi.org/10.1016/j.techfore.2020.120126

[6] Jovanovic, L. DejanJovanovic, N. Jovancai Stakic, A. Milos, A. Magd, H. AND ET AL. (2022). "Multi-Step Crude Oil Price Prediction Based on LSTM Approach Tuned by Salp Swarm Algorithm with Disputation Operator" *Sustainability* 14 (21), 14616. https://doi.org/10.3390/su142114616

[7] Jahanshahi, H. Süleyman, U. Sezgin, K. Qijia, Y. and Madini O. Alassafi. (2022). "Artificial Intelligence-Based Prediction of Crude Oil Prices Using Multiple Features under the Effect of Russia–Ukraine War and COVID-19 Pandemic" *Mathematics* 10(22), 4361. https://doi.org/10.3390/math10224361

[8] Lu, Q., Sun, S., Duan, H. et al. (2021).Analysis and forecasting of crude oil price based on the variable selection-LSTM integrated model. *Energy Inform* 4 (Suppl 2), 47. https://doi.org/10.1186/s42162-021-00166-4

[9] Zhao, L. Wang, Z. Wang, S & He, Y. (2021) Predicting Oil Prices: An Analysis of Oil Price Volatility Cycle and Financial Markets, *Emerging Markets Finance and Trade*, 57:4, 1068-1087, https://doi.org/10.1080/1540496X.2019.1706045

[10] Kpodar, K. & Imam, P.(2020). To Pass (or Not to Pass) Through International Fuel Price Changes to Domestic Fuel Prices in Developing Countries: What Are the Drivers?, *International Monetary Fund*, 20, 1-35. https://doi.org/10.1016/j.enpol.2020.111999

[11] Huang, L., Wang, J. (2018). Global crude oil price prediction and synchronization-based accuracy evaluation using random wavelet neural network. *Energy*, 151, 875–888. https://doi.org/10.1016/j.energy.2018.03.099

[12] Mohammadi Almoti, M., Hadadi, M., & Nadimi, Y. (2020). Investigating the Weak Efficiency Hypothesis in Two Low-Volatility and High-Volatility Regimes of the OPEC Crude Oil Market. *Investment Knowledge*, 9(33), 109-127.

[13] Bristone, M. Prasad, R. & Abubakr, A. .(2020). CPPCNDL: Crude oil price prediction using complex

network and deep learning algorithms, Petroleum, https://doi.org/10.1016/j.petlm.2019.11.009.

[14] Shabri, A. and Samsudin, R. (2014). Daily Crude Oil Price Forecasting Using Hybridizing Wavelet and Artificial Neural Network Model, Hindawi Publishing Corporation Mathematical Problems in Engineering. https://doi.org/10.1155/2014/201402

[15] Asghari, H., Omid, M., Malekinejad, Z., & Omid, A. (2018). Forecasting global crude oil demand by using vector autoregressive, autoregressive distributed lag, and gravitational search algorithms. *Future Studies in Management*, 29(4), 101-118**.**

[16] Elyakov, A. and et al.(2018). World oil prices: a retrospective analysis and elevation of factors, Corpus ID: 54783239.

[17] Peng, J. and et al. (2020). Dynamic Characteristics of Crude Oil Price Fluctuation—From the Perspective of Crude Oil Price Influence Mechanism, *Energies*, 13, 4465; https://doi.org/10.3390/en13174465

[18] Liu, Z.and et al .(2018). Financial factors affecting oil price change and oil-stock interactions: a review and future perspectives, *Natural Hazards*, https://doi.org/10.1007/s11069-018-3473-y.

[19] Sehgal, N. & Pandey, K.K. (2015). Artificial intelligence methods for oil price forecasting: a review and evaluation, *Energy Syst*, https://doi.org/10.1007/s12667-015-0151-y

[20] Gao, S. & Lei, Y.(2017). A new approach for crude oil price prediction based on stream Learning, *Geoscience Frontiers* 8, 183-187. https://doi.org/10.1016/j.gsf.2016.08.002

[21] Takroosta A, Mohajeri P, Mohammadi T, Shakeri A, & Ghasemi A. (2019). An Analysis of Oil Prices Considering the Political Risk of OPEC. *Jemr*, 10 (37), 105-138. http://jemr.khu.ac.ir/article-1-1832-en.html

[22] Razavi, A., Salimifar, M., Mostafavi, M., & Baki Heskooyi, M. (2014). Short-term impact of financial markets on the behavior of heavy Iranian crude oil prices. *Quantitative Economics Quarterly*, 11(2), 115-143. http://dx.doi.org/10.5539/mas.v10n3p45

[23] Shahbazi, K., & Salimian, S. (2015). Predicting oil prices using meta-analysis method. *Journal of Energy Economics Studies*, 11, 67-93**.** http://iiesj.ir/article-1-301-en.html

[24] Firouzi Jahantigh, F., & Dehghani, S. (2015). Application of genetic algorithm in optimization of neural network architecture and oil price prediction (GADDN). *Applied Economic Studies of Iran (Applied Economic Studies)*, 5(20), 122-97.

https://dorl.net/dor/20.1001.1.22519165.1397.9.35.8.1

[25] Rostami, M., Naqvipour, M., Holy Bayat, M. (2017). Oil market volatility model based on regime choice approach. *Financial Engineering and Securities Management*, 9(35), 179-196.

https://dorl.net/dor/20.1001.1.22519165.1397.9.35.8.1D

[26] Behmiri, N. B., Manso, J. R. P. (2013). How crude oil consumption impacts on economic growth of Sub-Saharan Africa? *Energy*, 54, 74–83. https://doi.org/10.1016/j.energy.2013.02.052

[27] Gao, X., Fang, W., An, F., Wang, Y. (2017). Detecting method for crude oil price fluctuation mechanism under different periodic time series. Applied Energy, 192, 201–212. https://doi.org/10.1016/j.apenergy.2017.02.014

[28] Li, R., Hu, Y., Heng, J., Chen, X. (2021). A novel multiscale forecasting model for crude oil price time series. *Technological Forecasting and Social Change*, 173, 121181. https://doi.org/10.1016/j.techfore.2021.121181

[29] Nademi, A., Nademi, Y. (2018). Forecasting crude oil prices by a semiparametric Markov switching model: OPEC, WTI, and Brent cases. *Energy Economics*, 74, 757–766. https://doi.org/10.1016/j.eneco.2018.06.020

[30] Karasu, S. and Altan, A. (2022). Crude oil time series prediction model based on LSTM network with chaotic Henry gas solubility optimization. Energy, 242, 122964. https://doi.org/10.1016/j.energy.2021.122964