TAS Journal Journal of Technical and Applied Sciences

Journal of Technical and Applied Sciences
01(02), 14-24, 2021
DOİ: http://dx.doi.org/10.53296/tasjournal.27
e-ISSN: 2791-6715

Research Article

Artificial Intelligence in Crude Oil Price Forecasting

Fetih KAYA[1] and Can BALKAYA[2]

Geliş / Received: 15/07/2021 ***Revize / Revised: 18/08/2021*** ***Kabul / Accepted: 09/09/2021***

ABSTRACT

The crude oil is a common energy source for nearly all commercial sectors, its price forecasting activities have always been an important issue for both governments and commercial firms to make better decisions and investments. In this research, both the history of the crude oil price forecasting and used artificial intelligence methods on forecasting were investigated. In early stages of crude oil price forecasting, traditional statistical and mathematical models were used, while afterwards computer based artificial intelligence models became more popular. These models were more appropriate to the non-linear, volatile and complex structure of oil prices. Artificial intelligence gave chance to evaluate the situation in many aspects at the same time with the help of the computers' power. Evaluation of these produced outcomes together with other variables such as historical prices, weather condition, political situations etc. gave much better forecasting results for crude oil. Fuzzy logic, Artificial Neural Network (ANN), Genetic Algorithms (GA), Support Vector Machine (SVM), expert systems, text-mining algorithms and their sub-versions were the frequently used AI based algorithms in the crude oil forecasting models. Among them for forecasting of crude oil prices, ANN algorithms, with its layered structure which makes it possible to relate many parameters with target variable a detailed way, have the most appropriate working principle for forecasting the complex and sensitive structure of crude oil prices. Hybrid models usually give better results, its combination with other algorithms such as text-mining or the most used one ANN, could improve the prediction results.

KEYWORDS: *- Artificial intelligence, crude oil price, price forecasting, computer based models.*

[1]*Corresponding author: fetih.kaya@bahcesehir.edu.tr Department of Big Data Analytics and Management, Faculty of Engineering and Natural Sciences, Bahcesehir University, Istanbul, Turkey
[2]Communication: canbalkaya@aydin.edu.tr Department of Civil Engineering, Faculty of Engineering, Istanbul Aydın University, Istanbul, Turkey

1. INTRODUCTİON

Crude oil is a key component of World's energy supply by being a common energy source for almost every commercial sectors. This means, any movement in the oil market can easily affect all other markets. So, prediction of crude oil prices is an important issue for both governments and commercial firms to make better decisions and investments by knowing its future movements [1].

The forecasting activities of crude oil prices dates back to 1970s, the Yom Kippur War between Arab countries and Israel. Arab oil producers decided to boycott and impose an embargo to America and threat the west in response to support Israel in the War. As a result, crude oil prices began to increase rapidly and economic activities were interrupted. So, the significance of crude oil price prediction was understood and professional prediction activities were triggered to start.

In early stages of oil price forecasting, econometric, statistical and mathematical models were popular due to the absence of machine-made algorithms [2]. Historical price movements, new oil field openings, production rates and economic situations of oil producing countries were the used independent variables of forecasting. In short, time lagged supply and demand balance based approximations were the underlying decision parameters of forecasting for traditional methods. However, these methods were able to solve only linear or near-linear problems with the limited human skills and abilities. In case of the non-linear, volatile and complex structure of oil prices which is sensitive to many exogenous variables such as political decisions, weather conditions etc., the traditional models become ineffective and cannot achieve accurate forecasts [1,3,4].

Development of computer based forecasting methods in early 1990s, helped for generating more complex and accurate prediction models. The forecasting performance of these methods was enough, even more, to compete with traditional methods. In addition, the algorithms perform better in discovering the relationship between every given individual parameter on target variable than traditional methods by conducting many iterations and learning the pattern that cannot be achieved by traditional methods. Computer based algorithms began to substitute traditional algorithms to produce more improved results and "Artificial Intelligence" (AI) term became popular to express these complex machine-made algorithms in early 2000s. The forecasting accuracies of these computer based artificial intelligence models were much better than traditional models because these models can conduct more complex algorithms for forecasting. In this study, development of artificial intelligence models on crude oil price forecasting and possible future works were discussed.

2. ARTİFİCİAL INTELLİGENCE METHODS USED İN PRİCE FORECASTİNG

Artificial intelligence methods used in the scope of crude oil price forecasting are discussed as Artificial Neutral Networks, Genetic Algorithms, Expert Systems, Text Mining and Support Vector Machine in general.

2.1 Artificial Neural Networks

The structure of artificial neural network (ANN) algorithm is a product of inspiration of human brains neural system. The algorithm basically consists input, hidden and output layers in order respectively, as shown in Fig.1. The input and output layers of the algorithm are filled with known data from a given data set and algorithm generates hidden layer(s) that models the problem with the best way. In short, algorithm tries to learn the followed path according to given input and output data and builds a model for future predictions. The algorithm conducts a

deep process to search relations between every given individual input and this makes it successful on forecasting. The number of hidden layers are assigned by the algorithm itself according to the complexity level of the data. In addition, there are many types of neural network such as feed-forward neural networks (FFNN) and generalized-regression neural network (GRNN) which are differentiated from each other based on their processing systems.

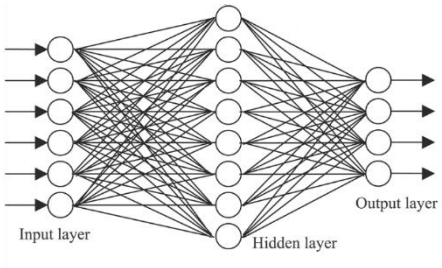

Fig. 1. Neural network structure

2.2 Genetic Algorithms

Genetic algorithms (GA) are searching methods which are based on Darwin's theory of natural selection. The process begins with an initial random population that is thought it has the optimal solutions. After that selection, crossover and mutation steps are applied to this initial population to get a better describing model. During the selection stage, the best performing chromosomes are kept and the others are eliminated from the population. Then, the remaining chromosomes are paired on the crossover stage and adapted on mutation stage. This loop is repeated until the best representing population of the data set is reached. The working principle of the GA is similar to the ANN algorithm. They both go over each input to discover the best explanation of the relationships. The whole process of the GA is represented in Fig.2.

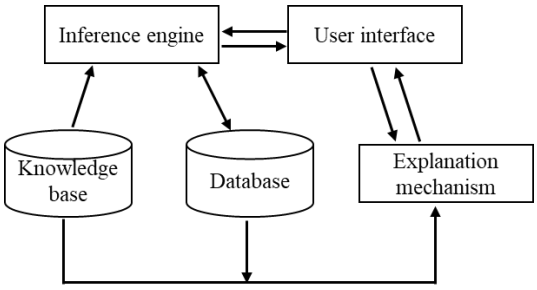

Fig. 2. Genetic algorithm structure

2.3 Expert Systems

Expert systems are a group of systems that are database, user interface, explanation mechanism, inference engine and knowledge base, as shown in Fig.3. The base knowledge which is crude oil prices for our case, is the most significant part of problem solving. The knowledge is represented in a set of rules which determine conditions and actions to perform according to defined rules. The database holds the facts that paired with the defined rules consisted in knowledge base. The inference engine serves as an interface between the database and knowledge base to provide communication and make intelligently produced results. The explanation mechanism produces expressions for produced results. Expert systems provide a better understanding of data sets by acknowledging people about explanations of taken steps. In short, expert systems produce a set of rules by using knowledge base

which is filled with approved information by experts and perform modelling on the given data set to reach the best explaining model of the data set. In addition, this process is conducted interactively with people.

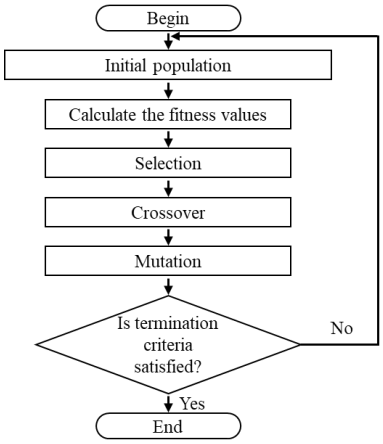

Fig. 3. Expert system structure

2.4 Text Mining

Text mining is a process of reinterpreting any given text such as news, books etc. The text, first of all, cleaned from any punctuations and divided into words which is called tokenization. Then, the stop words which do not carry important meanings such as “the”, “a” etc. are removed from the text. After that, lemmatization which is decreasing the words to its original phase such as “goodness” and “best” turn to “good”, is conducted over the text. Finally, these words are vectorised and put in a table for comparison or searching purposes. A representative flow of text mining is illustrated in Fig.4.

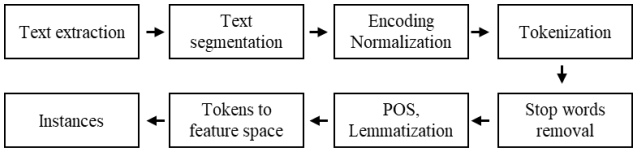

Fig. 4. Text mining process

2.5 Support Vector Machine

Support vector machine (SVM) is a machine learning algorithm which is used for both classification and regression types of analysis. The algorithm trains a given set of examples and then constructs a prediction model for assigning new coming examples to their categories. The SVM draws lines to differentiate each group from each other by trying to achieve assigning maximum distances between given group examples in space. The working system of SVM is shown and it tries to draw a line which is equally distant from each groups element in Fig.5.

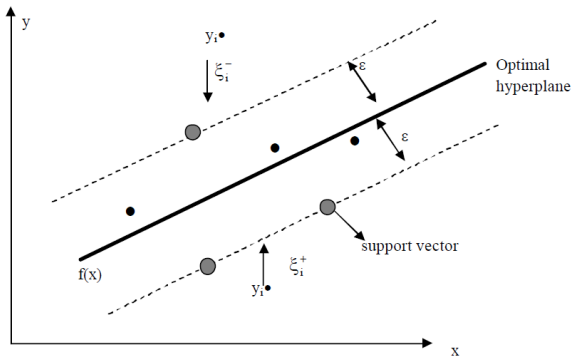

Fig. 5. Support vector machine system [5].

3. ARTIFICIAL INTELLIGENCE APPLICATIONS ON CRUDE OIL FORECASTING

Application of artificial intelligence algorithms on crude oil price forecasting dates back to early 2000s. In the beginning, individual algorithms were implemented into the models to forecast oil prices. Afterwards, hybrid models, two or more algorithms applied together, were constructed to achieve better predictions. The results of the models showed that; in general, computer based algorithms outperformed over traditional ones for individual algorithm applied models, while hybrid models were the best for forecasting of crude oil prices. Thus, hybrid models have become more and more popular over years and many studies have been conducted with hybrid models.

In 2001, Kaboudan used two artificial intelligence (AI) algorithms individually, genetic programming (GP) and artificial neural network (ANN), to predict crude oil prices for a given period of time. The performance of these two algorithms were compared with random walk algorithm's (RW) results which is a conventional mathematical based algorithm that has been used for a long time. The GP algorithm outperformed over other algorithms while the ANN algorithm had the worst results. The study showed two things actually; AI based algorithms get better result in appropriate conditions, but not always [2]. On the other hand, the price data of crude oil over the period of Gulf War was excluded from the input dataset of the study, because the price volatility was very high and thus hard to integrate a price pattern over that period [2]. This means, AI based algorithms can outperform over conventional methods, but they still suffer from being not able to integrate effects of unexpected events into forecasts.

In the same year with Kaboudan, Rast conducted a hybrid model on oil price forecasting with fuzzy logic and ANN algorithms. The integrated model works in a way that; ANN makes predictions but priorly, fuzzy logic supports the ANN on modelling by giving some predetermined criteria to plot a route such as if weak demand and high inventory situations come together price goes down [3]. In the study, hybrid model was compared with individual

Table 1. AI based studies on crude oil price forecasting with their used AI based methods

	Researchers	Year	Used Artificial Based Methods
1	Kaboudan	2001	Genetic programming and FFNN
2	Rast	2001	Fuzzy logic and FFNN
3	Wang et al.	2004	FFNN, expert systems
4	Yu et al.	2005	Knowledge based systems
5	Fernández	2006	FFNN and support vector machine
6	Yu et al.	2008	FFNN and adaptive linear neural network
7	Mehdi	2009	Fuzzy logic, expert systems and FFNN
8	Khashman and Nwulu	2011	Support vector machine and back-propagation FFNN
9	He et al.	2012	Wavelet analysis
10	Ahmed and Shabri	2014	Support vector machine
11	Yu et al.	2016	Least-squares support vector regression and GA
12	Bon and Isah	2019	Hidden Markov
13	Herrera et al.	2019	ANN and Random forest
14	Wang et al.	2019	Text mining and convolutional neural network
15	Lu et al.	2020	Dynamic-bayesian time series

ANN algorithm. The result; hybrid model outperformed over individual ANN and, the result forms a base for performance enhancement with hybrid models [6].

An important study was conducted by Wang et.al.(2004) to build a model that considers the effects of unexpected events on crude oil prices. The study puts together web-based text mining (WTM), rule-based expert system (RES) and ANN algorithms to generate a hybrid model. The logic behind the study is; searching the crude oil relevant texts in the internet such as news, articles etc. and digging them into to conclude with words that match with oil price movements. The study revealed that the word “war”, for example, affects prices in an upward direction within a range of 5 - 45% and, conversely “OPEC raise production” phrase affects prices downward direction within a range of 5 – 6%. After defining the words or phrases with their effects on oil prices, this information is loaded into RES for processing as a rule over predictions. In final stage, ANN algorithm is applied to historical crude oil dataset by considering determined rules. As it seen in Table 2, the root mean squared error (RMSE) values of hybrid model for each case less than only ANN applied model and also direction change statistics (Dstat) are higher. So that, the hybrid system gave better results rather than individual ANN algorithm, especially in the period of many important events which have effect on prices, happened. The text mining algorithm played a significant role to integrate the effects of unexpected events on oil price forecasting [7].

Fernandez (2006) applied support vector machine (SVM) algorithm which is an AI based classification technique, to forecast crude oil prices. She built both individual and hybrid models with SVM, ANN and traditional methods. The comparison of results showed that; apparent difference between applied techniques was originated from time interval. Applied traditional method gave the best prediction result in short term (couple days) while SVM and ANN gave better in long term (up to 20 days). Moreover, the forecasts of linear combination of SVM and ANN were more accurate than individuals. The study shows that, hybrid models perform better, in addition, short term forecasts of traditional models are better than AI based models [8].

Furthermore, Yu et al. (2008) applied a neural network ensemble learning paradigm which is based empirical mode decomposition (EMD), to conclude with a more accurate crude oil price forecast over worldwide [9]. Khashman

and Nwulu (2011) conducted a study that compares the back-propagation neural network and SVM for crude oil price forecasting [10]. He et al. (2012) performed a wavelet analysis study to get a better explanation of dynamic movement of oil prices and make precise price forecasting [11]. Ahmed and Shabri (2014) suggested a SVM-based model to predict crude oil prices [12]. Yu et al. (2015) proposed a four-step model; decomposition, reconstruction, prediction and ensemble, on the basis divide and conquer with data-characteristic-driven reconstruction to predict oil prices [13]. Again Yu et al. (2016) proposed a new hybrid model which combines least-squares support vector regression (LSSVR) and a hybrid optimized parameter searching approach with grid method and genetic algorithm (GA) to forecast oil prices [14]. Yu et al. (2017) applied a different approach to its early study in 2016, the model works many times with randomly selected parameters and their average was accepted as actual price forecast [15]. Bon and Isah (2019) developed a hidden Markov-based model to forecast monthly crude oil prices and declared that performance of the model is better than the regular Markov model [16]. Herrera et al. (2019) carried a study by using neural network and random forest algorithms to compare with econometric methods, the result showed that AI based algorithms outperform over traditional ones and they are better at predicting price turning points [17].

Since statistical data such as oil price, supply, demand etc. cannot achieve a perfect prediction accuracy by alone on forecasting of crude oil prices, in 2019 a text-based prediction method was again become a current issue for better predictions. After the study of Wang et al. in 2004, Das and Chan in 2007 and Nguyen, Shirai and Velcin in 2015 have conducted text-based prediction studies too, however, this new one which is conducted by again Wang et al., works on an unaddressed way [18]. The study works on the effects of articles topics on oil prices because they consist more compact information and less words and repetition according to whole article [19]. The model can be summarized like; the headings of articles, news or blogs in online sources which are related with crude oil, are collected and unnecessary parts of them like ‘the’, ‘in’, ‘on’, ‘at’ etc. are cleaned to get core words. Then, the next day’s oil price was compared with today’s for assigning these words’ effects and coefficient values. For example, if the oil prices decrease by 1.5% after the used words on heading “Iran”, “supply” a day before, it is coded that these words together decrease oil prices by 1.5%. This step, of course, is made many times to conclude with a word-effect pair. Latent Dirichlet Allocation (LDA) method is used for grouping the words in the topics according to their effects, increasing or decreasing. After generating a whole set of words and their effects on crude oil prices, convolutional neural network (CNN) algorithm which is a type of ANN algorithm, is used to predict oil prices. The model performance is much better than the conventional methods and also many ANN algorithms used models. In Table 3, comparison of random forest, support vector regression.

Table 2. Comparison of ANN method and used hybrid AI method by Wang et. al. in 2004 [7].

Evaluation	Full period (2000-2002)	Sub-period I (2000)	Sub-period II (2001)	Sub-period III (2002)
ANN method				
RMSE	3.413	3.405	3.020	3.324
D_{stat} (%)	61.11	50.00	66.67	66.67
Hybrid AI method				
RMSE	2.369	3.000	2.040	1.916
D_{stat} (%)	80.56	75.00	83.33	91.67

*RMSE: root mean squared error; Dstat: direction change statistics.

Table 3. Forecasting performances of the random forest, SVR and linear regression models by used hybrid system [18].

	Text features (1)	Financial features (2)	Combination: (1) + (2)	Percentage improvement from (2) to (1) + (2)
MAE				
Random forest	0.0785	0.0082	0.0073	12.32%
SVR	0.0252	0.0032	0.0030	6.67%
Linear regression	0.0854	0.0035	0.0045	-22.22%
RMSE				
Random forest	0.0883	0.0092	0.0088	4.55%
SVR	0.0325	0.0041	0.0040	2.50%
Linear regression	0.0953	0.0044	0.0056	-21.42%

*MAE: mean absolute error; RMSE:root mean squared error; SVR: support vector regression.

(SVR) and linear regression algorithms were made according to used parameters such as only text parameter, only financial parameters and their combination. The study proves that again, texts that consists news about oil market carry a meaningful information which statistical data do not, about forecasting oil prices [18].

Additionally, Tang et al. (2020) applied a multi-scale model with the data collected from search engines by matching oil price driven factors in search engines [20]. Lu et al. (2020) conducted a dynamic-bayesian structural time series model by using 415 exploratory variables commonly from Google trend search data [21].

4. CONCLUSİONS

Many models have been developed on the crude oil price forecasting by using both artificial intelligence-based and traditional algorithms since 2000s Among them, in general, artificial intelligence-based models outperform over traditional ones, and hybrid models were the best for forecasting of crude oil prices. Fuzzy logic, ANN, GA, SVM, expert systems, text-mining algorithms and their sub-versions were the frequently used AI based algorithms in the models. However, a remarkable amount of models was constructed by using ANN algorithm and its sub-versions like feed-forward neural network (FFNN), recurrent neural network (RNN). Because, the most appropriate algorithm for the non-linear and complex structure of oil prices is ANN with its layered structure which makes it possible to relate many parameters with target variable a detailed way. On the other hand, the sensitive structure of oil prices to exogenous variables such as political, geographical, weather conditions etc. needs a model that can involve wide range parameters in many aspects. To achieve that, text-mining approach on articles, news, reports etc. was commonly used on crude oil price forecasting throughout years.

ItANN algorithms has the most appropriate working principle for forecasting the complex and sensitive structure of crude oil prices, because ANN assigns a coefficient number to all parameters according to their predictive power and then combine them all to reach the best. In addition, because of the oil price is very sensitive, usage of text-mining algorithms in order to explore a massive amount of texts in the internet and determine the effects of each word or phrases on oil prices is the most helpful way to feed ANN algorithm. So, changing the type of applied ANN algorithm to its other sub-types or widening the text search extent may help to reach a more developed model. On the other hand, decision tree algorithm could be a possible appropriate way to construct a model for crude oil price forecasting, because its working principle is similar with ANN. For further works, artificial bee colony (ABC) algorithm can be explored, because it is seen that the ABC algorithm have not been taken into

account for crude oil price forecasting despite its promising performance. Also, because the hybrid models usually give better results, its combination with other algorithms such as text-mining or the most used one ANN, could improve the prediction results.

REFERENCES

[1] Amin-Naseri, M.R., Gharacheh, E.A.: A hybrid artificial intelligence approach to monthly forecasting of crude oil price time series. In: The Proceedings of the 10th International Conference on Engineering Applications of Neural Networks, CEUR-WS284, pp. 160–167 (2007).

[2] Kaboudan M. A. (2001). Compumetric forecasting of crude oil prices. Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546), Seoul, South Korea, pp. 283-287 vol. 1. Doi: 10.1109/CEC.2001.934402.

[3] Chiroma, H., Abdul-Kereem, S., Shukri, A., Mohd Noor, A.S., Abubakar, A., Safa, N., Shuib, L., Hamza, M., Gital, A., Herawan, T. (2015). A Review on Artificial Intelligence Methodologies for the Forecasting of Crude Oil Price. Intelligent Automation and Soft Computing.

[4] Sehgal, N., Pandey, K.K. (2015). Artificial intelligence methods for oil price forecasting: a review and evaluation. Energy Syst 6, 479–506. Doi: https://doi.org/10.1007/s12667-015-0151-y

[5] Fernández, V. (2006). Forecasting crude oil and natural gas spot prices by classification methods (No. 229).

[6] Rast, M. (2001, July). Fuzzy neural networks for modelling commodity markets. In IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th (Vol. 2, pp. 952–955). IEEE. Doi: 10.1109/NAFIPS.2001.944733.

[7] Wang S., Yu L., Lai K.K. (2005) A Novel Hybrid AI System Framework for Crude Oil Price Forecasting. In: Shi Y., Xu W., Chen Z. (eds) Data Mining and Knowledge Management. CASDMKM 2004. Lecture Notes in Computer Science, vol 3327. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-30537-8_26.

[8] Fernandez, V (2006). Forecasting commodity prices by classification methods: The cases of crude oil and natural gas spot prices. Documentos de Trabajo from Centro de Economía Aplicada.

[9] Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. Energy Economics, 30(5), 2623-2635.

[10] Khashman, A., & Nwulu, N. I. (2011, May). Support vector machines versus back propagation algorithm for oil price prediction. In International Symposium on Neural Networks (pp. 530-538). Springer, Berlin, Heidelberg.

[11] He, K., Yu, L., & Lai, K. K. (2012). Crude oil price analysis and forecasting using wavelet decomposed ensemble model. Energy, 46(1), 564-574.

[12] Ahmed, R. A., & Shabri, A. B. (2014). Daily crude oil price forecasting model using arima, generalized autoregressive conditional heteroscedastic and support vector machines. American Journal of Applied Sciences, 11(3), 425.

[13] Yu, L., Wang, Z., & Tang, L. (2015). A decomposition–ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. Applied Energy, 156, 251-267.

[14] Yu, L., Dai, W., Tang, L., & Wu, J. (2016). A hybrid grid-GA-based LSSVR learning paradigm for crude oil price forecasting. Neural computing and applications, 27(8), 2193-2215.

[15] Yu, L., Xu, H., & Tang, L. (2017). LSSVR ensemble learning with uncertain parameters for crude oil price forecasting. Applied Soft Computing, 56, 692-701.

[16] Bon, A. T., & Isah, N. Application of Markov Model and Artificial Neural Network in Crude Oil Price Forecasting for PETRONAS Malaysia.

[17] Herrera, G. P., Constantino, M., Tabak, B. M., Pistori, H., Su, J. J., & Naranpanawa, A. (2019). Long-term forecast of energy commodities price using machine learning. Energy, 179, 214-221.
[18] Li, X., Shang, W., & Wang, S. (2019). Text-based crude oil price forecasting: A deep learning approach. International Journal of Forecasting, 35(4), 1548-1560.
[19] Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2015). Text mining of news-headlines for FOREX market prediction: A multi-layer dimension reduction algorithm with semantics and sentiment. Expert Systems with Applications, 42(1), 306–324
[20] Tang, L., Zhang, C., Li, L., & Wang, S. (2020). A multi-scale method for forecasting oil price with multi-factor search engine data. Applied Energy, 257, 114033.
[21] Lu, Q., Li, Y., Chai, J., & Wang, S. (2020). Crude oil price analysis and forecasting: A perspective of "new triangle". Energy Economics, 104721.