ISSN: 0067-2904

Crude Oil Price Forecasts Using Support Vector Regression and Technical Indicators

Redha A. Redha*, Suhad A. Yousif
Computer Science Department, College of Science, Al-Nahrain University, Baghdad, Iraq

Received: 28/4/2022 Accepted: 31/8/2022 Published: 30/5/2023

Abstract

Oil price forecasting has captured the attention of both researchers and academics because of the unique characteristics of crude oil prices and how they have a big impact on a lot of different parts of the economic value of the product. As a result, most academics use a lot of different ways to predict the future. On the other hand, researchers have a hard time because crude oil prices are very unpredictable and can be affected by many different things. This study uses support vector regression (SVR) with technical indicators as a feature to improve the prediction of the monthly West Texas Intermediate (WTI) price of crude oil. The root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) measure how well the model is working. The RMSE was 1.5456, the MAE was 1.3219, and the MAPE was 1.9173 in the experiment. The results show that WTI crude oil prices are affected by technical indicators and get good performance that outperforms most other models that can be found.

Keywords: Support vector regression (SVR) model, RMSE, MAE, WTI Crude oil price.

تأثير المؤشرات الفنية على التنبؤ بأسعار النفط الخام

رضا علي رضا*, سهاد عبد الرحمن يوسف
علوم الحاسوب, كلية العلوم, جامعة النهرين, بغداد, العراق

الخلاصة

استحوذت توقعات أسعار النفط على اهتمام الباحثين والأكاديميين على حد سواء بسبب الخصائص الفريدة لأسعار النفط الخام ولتأثيرها الكبير على الكثير من الجوانب الاقتصادية المهمة. نتيجة لذلك ، استعمل معظم الأكاديميين الكثير من الطرق المختلفة للتنبؤ بالمستقبل. من ناحية أخرى ، يواجه الباحثون أوقاتًا صعبة لأن أسعار النفط الخام لا يمكن التنبؤ بها ويمكن أن تتأثر بالعديد من الأشياء المختلفة. في هذه الدراسة استعمل دعم الانحدار الاتجاهي SVR مع المؤشرات الفنية كخوارزمية في الخصائص لتحسين النموذج الخاص للتنبؤ بسعر النفط الخام لغرب تكساس WTI الشهري. تم استعمال RMSE و MAE و MAPE لقياس مدى جودة أداء النموذج. حيث تظهر النتائج أن أسعار خام غرب تكساس الوسيط تتأثر بالمؤشرات الفنية وتحصل على أداء جيد يفوق أداء معظم النماذج الأخرى التي تستعمل خصائص اخرى.

*Email: redha.ali.cs2020@ced.nahrainuniv.edu.iq

1. Introduction

According to pricing theories, changes and fluctuations in crude oil prices directly affect stock prices, the returns of some companies, and the total global economic activity [1]. Furthermore, the crude oil price is frequently influenced by a variety of factors, including supply and demand, stock levels, weather, political factors, and other unanticipated events that cause problems for business leaders and policymakers; thus, forecasting crude oil prices has captured the attention of both researchers and academics and has become an intriguing field [2]. Therefore, further research on crude oil time series forecasting has been done [3-7]. These studies assume that future data values are proportional to present and historical data values. Many real-world time series data sets, on the other hand, contain complex nonlinear, seasonal, and non-stationary patterns that traditional machine learning techniques struggle to capture. An example of traditional forecasting techniques is autoregressive integrated moving average (ARIMA) models, which can be accurately forecasted if time series data is steady and stationary [8, 9]. In addition, time-series data, such as stock market statistics, is not seasonal.

Furthermore, time-series data may exhibit varying levels of volatility. For example, the price of crude oil has been quite volatile during the COVID-19 period[10] . This study can capture the volatility of crude oil prices by using SVR with technical indicators. There are many markets for crude oil prices in the world. The essential crude oil is the West Texas Intermediate (WTI). This research demonstrates that WTI crude oil prices are influenced by technical indicators and have excellent performance, outperforming the majority of other models that have been developed to this point.

The following is how the paper is structured: the first section introduces the subject. Then, the literature reviews are discussed in Section 2. Next, Section 3 explains time series forecasting and technical indicators, whereas Section 3.1 discusses datasets. Section 3.2 explains technical indicators. Section 3.3 shows feature selection, and Section 3.4 shows learning models. Finally, Section 4 gives SVR experimental results, and Section 5 concludes the research.

2. Literature Review

Crude oil price time-series forecasts have become more critical in the last few decades. Many people have used artificial intelligence to figure out and predict the movements of the Brent and WTI benchmarks based on historical time-series data. Previous authors have researched machine learning algorithms that can help them predict.

In [11], the authors utilize an ARIMAX model to anticipate future Brent Crude prices using a dummy variable from the US-China trade war. Because ARIMAX(1,1,0) met all the fit statistical requirements while the other five did not, it was chosen as the most effective model parameter among the six possibilities. The MAPE value of ARIMAX (1,1,0) is 13.6733 percent. The study's drawback is that it is impossible to create a dummy variable for the long term, leaving the model ineffective for forecasting in this situation.

To forecast the Brent crude oil price, the authors [6] utilized the Box–Jenkins time series technique and created an ARIMA model. AIC and BIC information criteria were used to determine the best model. After attempting many steps, ARIMA (1, 1, 1) was picked. The ARIMA model was created to operate with linear issues. The ARIMA model was designed to function with a linear problem because numerous variables impact crude oil pricing, such as the economic cycle, international relations, and geopolitics. However, Brent crude oil data is a nonlinear problem.

In [13], the authors compare Artificial Neural Networks (MLP) with Vector Autoregressive Models. The MLP neural network can forecast crude oil prices more reliably than a VAR model according to the matching R-squared for each model. For example, the MLP model had an MSE of 0.0897 and an R2 of 0.99, whereas the VAR model had an MSE of 0.4816 and an R2 of 0.96. However, the MLP has a short memory, which is a drawback. As a result, it cannot be used to foresee the future.

In [5], the researchers wanted to see how RW, ARMA, DBN, RW-DBN, ARMA-DBN, LSTM, RW-LSTM, and ARMA-LSTM compared. With an MSE of 5.3811, the DBN model delivers the most significant outcomes when the MSEs are the smallest across all models in the experiment. The study's flaw is that DBN's accuracy for this dataset is relatively high. The study's shortcoming is that the accuracy may decline if the time or dataset changes.

In [14], the authors applied ARIMAX to the WTI crude oil time series with the Google Index as an exogenous variable. According to the empirical findings, the Google Index negatively impacts crude oil prices, but it cannot be used to predict crude oil prices where ARIMAX's RMSE, MAE, and MAPE are 6.072, 5.072, and 138.751, respectively.

In [15], on WTI crude oil time series forecasts, the authors compare Long Short Term Memory (LSTM) against Moving Average (MA), Linear Regression (LR), and Autoregressive Integrated Moving Average (ARIMA) When measured by RMSE and R-Square, the findings show that LSTM outperforms the other three approaches, with an RMSE of 1.18 and an R-Square of 0.97. In addition, the deep learning model (LSTM) was shown to be the best at gathering nonlinear data points to forecast crude oil prices in the future in this study. On the other hand, tuning hyperparameters in LSTM is difficult and time-consuming.

In[16], the authors employ Artificial Neural Networks (ANN), Hybrid models, and Random Forest models to forecast the price of WTI crude oil and other energy sources. According to the study, random forests can produce more accurate findings than hybrids, ANNs, and RFs of the WTI time series, with RMSEs of 23.3935, 14.7208, and 7.8539 for the hybrid, ANN, and RF, respectively.

Many problems and limitations were found in the studies done before this one: capturing unknown complicated nonlinear (volatility) properties in a time series dataset is one of them. Most researchers approach solving the problem of nonlinear prediction by using deep learning, but in some cases, deep learning is not helpful. Deep learning requires high hardware specifications such as CPU, RAM, disk space, disk I/O, network I/O, and so on. These are all part of this task. Deep learning is also challenging work since it requires a large quantity of data to obtain the best model. It is challenging to decide the ideal number of layers. It may take a long time. Thus, deep learning has several downsides and limits [17]. Therefore, this study solves this significant effort or challenge by using a simple machine learning algorithm with generated technical indicators that will give us the best results.

3. Methodology

The suggested model is shown in Figure 1:

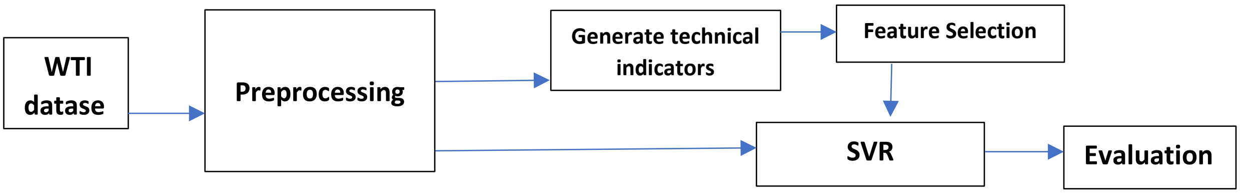


Figure 1: Abstract view of the proposed methodology

As shown in the figure above, our proposed research has many steps, which we will clear up one by one.

3.1 Data Sources

There are several crude oil markets located across the world. The NASDAQ, which stands for National Association of Securities Dealers Automated Quotations, provided the data used in this study, which is a prominent global electronic marketplace for buying and selling stocks. From January 1, 1988, to December 31, 2017, monthly historical WTI crude oil prices are illustrated in Table 1.

Table 1: Monthly WTI crude oil prices

Date	Price
01-February-17	53.47
01-March-17	49.33
01-April-17	51.06
01-June-17	48.48

3.2 Technical indicators

In the stock market, technical analysis involves appraising equities to predict future price patterns [6, 7]. Techniques such as technical indicators and specific mathematical formulations that use previous stock price data are employed as tools in technical analysis. From the previous study [4,7], a total of eight technical indicators that were employed in this article, as well as their unique formulation, are displayed in Table 2 [18] [19].

Table 2: Technical indicators used in this article

Technical Indicators	Formulas
(SMA)	$\mathrm{SMA}_n(t) = \frac{\sum_{k=t-n+1}^{t} \mathrm{P}(k)}{n}$ (1)
Standard Deviation (SD)	$SD(t) = \sigma = \sqrt{\frac{\sum_{k=t-n}^{t} \left(A_r - SMA_{A_x}(n)\right)^2}{n}}$ (2)
Exponential Moving Average (EMA)	$\mathrm{EMA}_n(t) = P(k) \times a + (1-a) \times \mathrm{EMA}_n(t-1)$ (3)
Moving Average Convergence Divergence MACD	MACD = EMA (12) – EMA (26) (4) MACD Signal=EMA (MACD, 9) (5)

Where P(k), and A_k refers to the price at time k. $\mathrm{EMA}_n(t)$ represents the EMA of today's price. $\mathrm{EMA}_n(t-1)$ represents the EMA of yesterday's t, and a = 2/ (n + 1). a is the weight coefficient. Finally, the default values of the MACD parameters (12, 26, and 9) can be adjusted to meet the needs of individual traders. This indicator is used as a feature in the SVR.

3.3 Features selection methods

Feature selection picks the most essential and relevant qualities from many features in a dataset. The feature selection methodologies are the filter, wrapper, and embedding methods [20]. In this study, the wrapper method (forward selection) is employed. The feature selection process in wrapper methods is based on a particular machine learning algorithm that we attempt to fit to a given dataset[21]. In forward selection, starting with one feature in the model, after fitting the model with each unique feature, one selects the best subset of features based on R-squared[3].

3.4 Metric

There are many different ways to measure how well a model works. The most commonly used metrics for prediction accuracy are MAE and RMSE, defined in Equations (6, 7, and 8), [22-24].

$$MAE = \frac{1}{n-m}\sum_{t=m+1}^{n} |y_t - \hat{y}_t|. \tag{6}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n-m}\sum_{t=m+1}^{n} (y_t - \hat{y}_t)^2} \tag{7}$$

$$MAPE = 100 \cdot \frac{1}{N}\sum_{i=1}^{N} \frac{|y_i - \hat{y}|}{y_i} \tag{8}$$

R^2 is another metric used to measure the performance of a forecasted model. It usually falls between 0 and 1. A model's fit is better if R^2 is close to 1 or -1. Equations 13 and 14 are used to compute it[3].

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y})^2}{\sum_{i=1}^{N} (y_i - \hat{y})^2} \tag{9}$$

Where y_i is the actual value and $\hat{y}$ is represented by the outcome value from the forecasting model.

3.5 support vector regression (SVR)

The supervised learning technique, Support Vector Regression, is used to predict discrete values. SVMs and Support Vector Regression are both based on the same theory. SVR's core principle is to locate the best-fitting line [25]. The best fit line in SVR is the hyperplane with the most significant number of points. The SVR, unlike other regression models, aims to fit the best line inside a threshold value rather than minimize the error between the real and projected value. The distance between the hyperplane and the boundary line is the threshold value. The SVM regression function is written as in Equation (10) [26, 27].

$$y(x) = w\phi(x) + b \tag{10}$$

Where ϕ(x) is named the feature nonlinearly planned from the input space x, the *w* and *b* are coefficients.

4. Results and Discussion

In this section, we look at whether the technical indicator obtained from WTI crude oil impacts WTI crude oil price forecasting performance or not. In the past, we used data from the WTI crude oil market. The dataset for this experiment was collected from January 1, 1988, to December 31, 2017, and it was divided into two parts: 80% was used as a training set, and the other 20% was used as a test set to evaluate the models' performance. In this experiment, we utilize the SVR model.

4.1 Pre-processing stage

After importing datasets into the system, the date column type must be converted from string to date via casting and assigned to the index. After that, sort the time series data by date in descending order, from oldest to newest. Finally, the data was normalized to a scalable range to be entered into SVR later.

4.2 Generate technical indicators

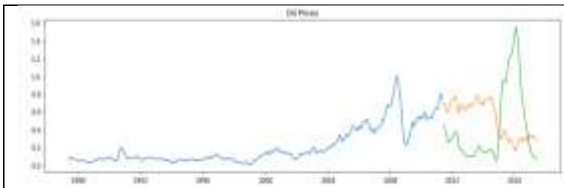
Construct a technical indicator from the WTI crude oil time series, using the earlier equations defined in the methodology section with many timeframes (lags), and then use it as a feature (the x variable) in SVR.

Table 3: The periods of monthly time of technical indicators

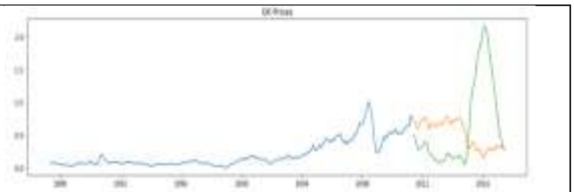
No.	Technical indicators	Selected period of lags
1	Simple Moving Average (SMA)	3 lags – 12 lags
2	Standard Deviation (SD)	3 lags – 12 lags
3	Exponential Moving Average (EMA)	3 lags –12 lags
4	Moving Average Convergence Divergence (MACD)	12 lags -26 lags -9 lags

4.3 Forecasting Model

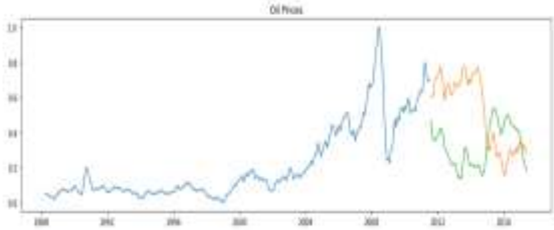
Not every feature in the model is significant. Thus, as seen in Figure 2, this experiment employs the SVR model and sees the technical indicator's impact on it. Forward selection begins with a null model, which is then fitted with each indicator one at a time.



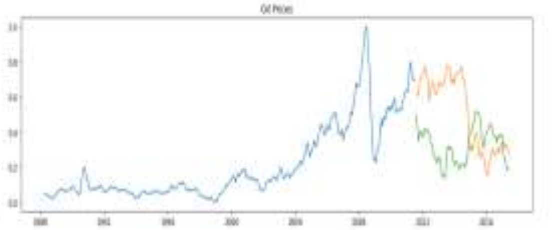
(a)SVR with MACD



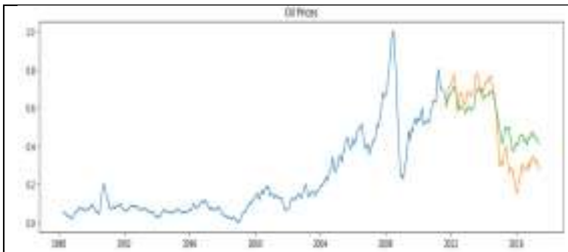
(b)SVR with MACD and MACD_SIGNALS



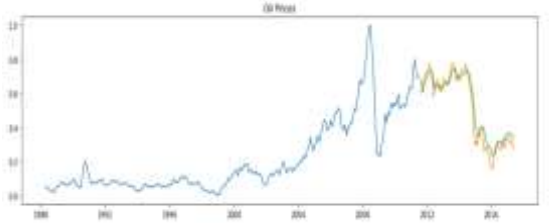
(c)SVR with MACD ,MACD_SIGNALS , and wti_std_12



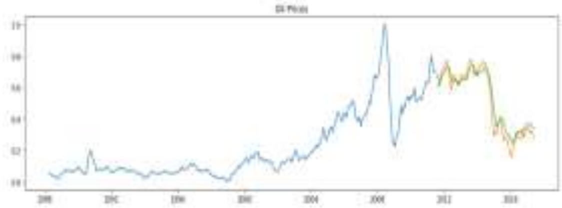
(d) SVR with MACD ,MACD_SIGNALS , wti_std_12,and wti_std_3



(e) SVR with MACD ,MACD_SIGNALS , wti_std_12, wti_std_3,and wti_sma_12



(f) SVR with MACD ,MACD_SIGNALS , wti_std_12, wti_std_3,wti_sma_12,and wti_sma_3



(g) SVR with MACD ,MACD_SIGNALS , wti_std_12, wti_std_3,wti_sma_12,wti_sma_3,and wti_ema_12

(h) SVR with MACD ,MACD_SIGNALS , wti_std_12, wti_std_3,wti_sma_12,wti_sma_3,wti_ema_12,and wti_ema_3

Figure 3: SVR Model improvement performance stages.

By looking at the figure above, it is clear that the technical indicators have an influence on the accuracy of the model, which is used to predict the price of WTI oil. Therefore, we can see gradual progress at each level of the experiment, where the green line represents the predicted model and the orange one represents actual data. In order to support our concepts of the experiment above, we calculated errors (MAE, RMSE, and MAPE) and saw improvement in each stage, as shown in Table 4.

Table 4: Accuracy measures of the SVR model's attempt

Stage	MAE	RMSE	MAPE
(a)	63.8288	72.2657	142.7909
(b)	84.1465	103.45	136.4665
(c)	37.2152	42.2396	86.8734
(d)	34.0042	39.7175	77.2621
(e)	11.7384	13.7885	16.4291
(f)	4.7477	6.0381	7.0927
(g)	5.1405	6.3845	7.6813
(h)	1.3219	1.5456	1.9173

As shown in the table above, we ensure that the technical indicator of WTI is its influence on the model when making gradual progress at each level of the experiment and its ability to provide good results.

4.4 **Comparison of accuracy between this study and related works**

In Table 5, a comparison will be made between the forecasting capabilities of the proposed model for WTI crude oil and the capabilities of one related work that worked with the same type of crude oil and utilized deep learning techniques and random forests.

Table 5: Comparison accuracy with other models

Author	Year	Title	Period	Model	RMSE	MAPE
Herrera, Gabriel [16]	2019	"Long-term forecast of energy commodities price using machine learning"	From 1980-01-01 To 2017-06-01	ANN	14.7208	16.5554
				Random forests	7.8539	8.9603
				Hybrid	23.3935	33.7099
		Proposed Models		**SVR with a technical indicator**	6.7907	8.0118

5. Conclusion

Forecasting crude oil prices is a hot topic among governments and large companies, who want a detailed profile to comprehend future demands and the evolution of the existing system. This paper demonstrates the impact of technical indicators that are generated from WTI such as MACD, SMA, EMA, and SD and utilizes them as features in the SVR model. The experimental findings reveal that the technical indicator has high influence and is accurate on the forecasting model. As shown in the results of the comparison experiment above, the proposed model was more accurate than those of ANN and Random Forests. In the future, we will try to generate technical indicators from other datasets, such as gold prices, to see how they affect the forecasting model.

References

[1] R. Bollapragada, A. Mankude, and V. Udayabhanu, "Forecasting the price of crude oil," *Decision*, vol. 48, no. 2, pp. 207-231, 2021, doi: 10.1007/s40622-021-00279-5.

[2] A. Malik, "The Impact of Oil Price Changes on Inflation in Pakistan," International Journal of Energy Economics and Policy, 6(4), 727-737 2016.

[3] J. Guo, "Oil price forecast using deep learning and ARIMA," in *Proceedings - 2019 International Conference on Machine Learning, Big Data and Business Intelligence, MLBDBI 2019*, 2019/11// 2019: Institute of Electrical and Electronics Engineers Inc., pp. 241-247, doi: 10.1109/MLBDBI48998.2019.00054.

[4] B. J. Ayeni and R. Pilat, "Crude oil reserve estimation: An application of the autoregressive integrated moving average (ARIMA) model," Journal of Petroleum Science and Engineering, vol. 8, 1992.

[5] Y. Chen, K. He, and G. K. F. Tso, "Forecasting Crude Oil Prices: A Deep Learning based Model," *Procedia Computer Science*, vol. 122: Elsevier B.V., pp. 300-307, 2017. doi: 10.1016/j.procs.2017.11.373.

[6] Mensah, Emmanuel Kwasi, "Box-Jenkins modelling and forecasting of Brent crude oil price," MPRA Paper 67748, University Library of Munich, Germany, 2015.

[7] C. l. Zhao and B. Wang, "Forecasting crude oil price with an autoregressive integrated moving average (ARIMA) model," in *Advances in Intelligent Systems and Computing*, Springer Verlag, vol. 211: pp. 275-286, 2014, doi: 10.1007/978-3-642-38667-1_27.

[8] Ü. Ç. Büyükşahin and Ş. Ertekin, "Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition," *Neurocomputing*, vol. 361, pp. 151-163, 2019/10// 2019, doi: 10.1016/j.neucom.2019.05.099.

[9] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," in *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, 2019/1// 2019: Institute of Electrical and Electronics Engineers Inc., pp. 1394-1401, doi: 10.1109/ICMLA.2018.00227.

[10] J. Sadefo Kamdem, R. Bandolo Essomba, and J. Njong Berinyuy, "Deep learning models for forecasting and analyzing the implications of COVID-19 spread on some commodities markets volatilities," *Chaos Solitons Fractals,* vol. 140, p. 110215, Nov 2020, doi: 10.1016/j.chaos.2020.110215.

[11] I. A. Rahmayanti, C. Andreas, and S. M. Ulyah, "Does US-China trade war affect the Brent crude oil price? An ARIMAX forecasting approach," presented at the International Conference on Mathematics, Computational Sciences and Statistics 2020, 2021.

[12] I. N. Mahmood and H. S. Abdullah, "Telecom Churn Prediction based on Deep Learning Approach", Iraqi Journal of Science, vol. 63, no. 6, pp. 2667–2675, Jun. 2022.

[13] S. Ramyar and F. Kianfar, "Forecasting Crude Oil Prices: A Comparison Between Artificial Neural Networks and Vector Autoregressive Models," *Computational Economics,* vol. 53, no. 2, pp. 743-761, 2017, doi: 10.1007/s10614-017-9764-7.

[14] T. Yao and Y.-J. Zhang, "Forecasting Crude Oil Prices with the Google Index," *Energy Procedia,* vol. 105, pp. 3772-3776, 2017, doi: 10.1016/j.egypro.2017.03.880.

[15] D. Makala and Z. Li, "Economic Forecasting with Deep Learning: Crude Oil," *MATTER: International Journal of Science and Technology,* vol. 5, no. 2, pp. 213-228, 2019, doi: 10.20319/mijst.2019.52.213228.

[16] G. P. Herrera et al., "Long-term forecast of energy commodities price using machine learning," *Energy,* vol. 179, pp. 214-221, 2019, doi: 10.1016/j.energy.2019.04.077.

[17] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data,* vol. 8, no. 1, p. 53, 2021, doi: 10.1186/s40537-021-00444-8.

[18] Y. Shynkevich et al., "Forecasting price movements using technical indicators: Investigating the impact of varying input window length," *Neurocomputing,* vol. 264, pp. 71-88, 2017, doi: 10.1016/j.neucom.2016.11.095.

[19] D. C. Yıldırım, I. H. Toroslu, and U. Fiore, "Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators," *Financial Innovation,* vol. 7, no. 1, 2021, doi: 10.1186/s40854-020-00220-2.

[20] A. Zheng and A. Casari, *Feature engineering for machine learning: principles and techniques for data scientists*. " O'Reilly Media, Inc.", 2018.

[21] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering,* vol. 40, no. 1, pp. 16-28, 2014, doi: 10.1016/j.compeleceng.2013.11.024.

[22] D. Banerjee, "Forecasting of Indian stock market using time-series ARIMA model," 2014 2nd International Conference on Business and Information Management (ICBIM), pp. 131-135, 2014. doi: 10.1109/ICBIM.2014.6970973.

[23] P. Areekul, T. Senjyu, H. Toyama, and A. Yona, "A hybrid ARIMA and neural network model for short-term price forecasting in deregulated market," *IEEE Transactions on Power Systems,* vol. 25, no. 1, pp. 524-530, 2010, doi: 10.1109/TPWRS.2009.2036488.

[24] G. Ciulla and A. D'Amico, "Building energy performance forecasting: A multiple linear regression approach," *Applied Energy,* vol. 253, 2019, doi: 10.1016 /j.apenergy. 2019. 113500.

[25] N. Sehgal and K. K. Pandey, "Artificial intelligence methods for oil price forecasting: a review and evaluation," *Energy Systems,* vol. 6, no. 4, pp. 479-506, 2015, doi: 10.1007 /s12667- 015-0151-y.

[26] Ahmed, Rana & Shabri, Ani., "Daily crude oil price forecasting model using arima, generalized autoregressive conditional heteroscedastic and Support Vector Machines," American Journal of Applied Sciences, 11, 425-432, 2014. 10.3844/ajassp.2014.425.432.

[27] H. S. Dhiman, D. Deb, and J. M. Guerrero, "Hybrid machine intelligent SVR variants for wind forecasting and ramp events," *Renewable and Sustainable Energy Reviews,* vol. 108, pp. 369-379, 2019, doi: 10.1016/j.rser.2019.04.002.