**TITLE:**

**PREDICTING BRENT OIL PRICES USING LSTM AND LINEAR REGRESSION**

**MODULE NAME:**

**MODULE CODE:**

**COURSE CODE:**

**TABLE OF CONTENTS**

[1. INTRODUCTION 1](#_Toc8257)

[2. METHODS AND DATA SOURCES 4](#_Toc31076)

2.1 Long Short-Term Memory 4

2.2 Comparison Model 4

2.3 Linear Regression 5

2.3.1 Multiple Linear Regression 5

2.4 Evaluation Metrics 6

2.5 Training and Test data Descriptive Statistics 6

[3. FITTING EFFECT ANALYSIS OF LSTM 9](#_Toc12711)

[4. FORECASTING EFFECT ANALYSIS 11](#_Toc29452)

[5. CONCLUSION 12](#_Toc4291)

[6. RECOMMENDATION 13](#_Toc28822)

[REFERENCES 13](#_Toc18980)

**LIST OF FIGURES**

[Figure 1 : The training and test set for brent oil prices............................................................7](#_Toc10095)

Figure 2 : Brent oil Prices for the last 20 years………………………………………………8

[Figure 3 : Predicted values using LSTM and actual values...... 10](#_Toc13431)

[Figure 4 : Predicted values using multiple linear regression and actual values 10](#_Toc9418)

**LIST OF TABLES**

[Table 1: LSTM Parameters 8](#_Toc124239942)

[Table 2: Linear Regression Parameters 9](#_Toc124239943)

[Table 3: Cross-validation Parameters 9](#_Toc124239944)

[Table 4: Forecasting techniques and evaluation metrics 12](#_Toc124239945)

# INTRODUCTION

According to Zhong et al. (2019), crude oil is a vital component of industrial production. Globally, Brent oil is frequently used as a benchmark for oil prices. It comes from the North Sea and serves as a standard for oil exported to the West from Europe, Africa, and the Middle East (Trading Economics, 2023).

It is challenging to forecast how supply and demand affect oil prices. Forecasting oi price is crucial for several reasons. Variations in oil prices directly affect the costs of related goods like petroleum fuels and industrial raw materials, impacting the costs of finished industrial goods. As a result, central banks and governments frequently consider crude oil prices when determining macroeconomic risks and developing projections. Second, fluctuations in oil prices may impact firms throughout the value chain. For instance, a sharp increase in oil prices may result in increased transportation expenses and lower earnings for businesses involved in the oil industry. A prolonged reduction in oil prices may stifle production in the chemical and oil refineries industries, resulting in a decline in profitability. Accurate oil price forecasting can assist companies in lowering cost risks and maintaining steady profit growth (Zhang & Hong, 2022).

Forecasting crude oil prices is crucial for investors, corporations, and policymakers (Zhong & Hong, 2022). Businesses can utilise price projections to make educated production decisions and effectively manage costs, while policymakers can use them to set reasonable budgets (Murat & Tokat, 2009; Zhang et al., 2019 & Zhao et al., 2017). On the other side, investors can utilize precise projections to efficiently deploy their resources and possibly profit from price unpredictability (Zhang et al., 2019). Consequently, it is crucial to focus on increasing the precision of crude oil price forecasts (Zhong & Hong, 2022).

Various models exists and have been deployed for forecasting the prices of oil (Abdollahi, 2020), including traditional econometric strategies and machine learning techniques (Chiroma et al., 2015 & Lu et al., 2020). Crude oil price predictions have been made using traditional econometric techniques, including Generalized Auto-Regressive Conditional Heteroskedasticity (GARCH) models and Auto-Regressive integrated moving average (ARIMA) random walk models (Lammerding et al., 2013). In their investigation of crack spread futures' ability to forecast adjustments to oil price, Murat and Tokat (2009) discovered that the crack spread model outperformed the random walk model, their comparison model. In another analysis of United Kingdom North Sea Brent (Brent) prices and WTI’s (West Texas Intermediate) volatility, Wei et al. (2010) employed another variety of GARCH-type models. It was discovered that nonlinear GARCH-based classical models outperformed linear models in the area of obtaining the long memory of the two prices as well as unsymmetrical volatility.

In addition to traditional econometric methods, various machine learning models have recently been used to anticipate oil prices (Butler et al., 2021 & Li et al., 2019). The two most popular techniques of these are neural networks (Ramyar and Kianfar, 2019 & Mostafa and El-Masry, 2016) and SVM (support vector machines) (Fan et al., 2021 & Zhao et al., 2017). They can simulate the intricate features of oil prices, such as their nonlinearity and volatility. An adaptive neural fuzzy inference system model was used by Gori et al. (2007) to predict medium-term oil prices. Zhao et al. (2017) made a forecast the prices of crude oil by employing stacked de-noising auto-encoders and bootstrap aggregation (bagging), ultimately displaying higher forecasting capabilities.

The goal is to develop a model or models that can forecast future Brent oil prices with precision and a minimum level of error. We will utilise regression and deep learning methods, including artificial neural network and linear regression models. One type of neural network that has proven particularly effective in this context is the long short-term memory (LSTM) network, which is a kind of recurrent neural network that may overcome the "vanishing gradient" and "exploding gradient" problems that often plague other types of RNNs.

To predict the prices of crude oil, we will first create a multiple linear regression and an LSTM model. Our findings indicate that the LSTM model is capable of broad generalisation. Additionally, we will show that the LSTM model is preferable for forecasting the prices of crude oil by contrasting the precision and stability of the LSTM model with a linear regression model.

# 

# METHODS AND DATA SOURCES

## Long Short-Term Memory

A deep learning neural network built on recurrent neural networks (RNN) is the long short-term memory (LSTM) (Zhang & Hong, 2022). It was first created by Hochreiter and Schmidhuber (1997), then Graves (2012) expanded it with a deep learning approach. RNNs are very helpful for time-series data analysis and prediction because they can use prior knowledge to complete the current task (Pabuçcu et al., 2020). In other words, they can recall and apply prior knowledge to generate the current output (Zhang & Hong, 2022). Some of the problems with RNNs, such as the vanishing and expanding gradient problems, are addressed by LSTM (Fischer & Krauss, 2018). It accomplishes this by incorporating its memory and a lengthy delay between the input, output, and gradient. As a result, the LSTM can anticipate crude oil prices with accuracy.

Additionally, LSTM models retain prior state data about crude oil prices because of their long- and short-term memory capacities. Additionally, they can fully extradite past data on crude oil prices and include current data characteristics in the price projection. The automatic discovery of nonlinear characteristics and intricate patterns in crude oil prices is made possible by LSTM models, which are particularly good at spotting long-term dependencies in crude oil price sequence data. They have proven extraordinary forecasting abilities in predicting oil prices and have found extensive use in prediction-related sectors (Zhong & Hong, 2022).

The LSTM cell structure comprises the input, output and forget gates. An essential component of the unit structure of LSTM, the cell state of an LSTM model uses memory storage to store the state of neurons. However, the control gate, which selectively communicates crude oil price data and can add or delete information from the cell state to manage it, determines whether it is retained or discarded. It comprises point-wise multiplication and a sigmoid layer (Zhang & Hong, 2022).

We have added batch-normalisation (BN) and dropout layers to our LSTM model to improve the neural network's structure. The dropout technique can lessen the chance of over-fitting, while BN can successfully solve the vanishing gradients issue, making model convergence challenging (Ouyang et al., 2021).

Three LSTM neural layers and two closely connected LSTM layers, each having 30 nodes, make up the LSTM model used in this work. Each LSTM neural layer is preceded by a BN layer and a dropout layer with 0.2 dropout probability. Mean squared error was used as the loss function and has been trained using a mini-batch technique. We employ the Adam optimisation method for optimisation training, renowned for its quick convergence time and good learning performance.

## Comparison Model

We used a linear regression model as the reference model to compare the predicted accuracy of various methods.

2.3 Linear Regression

### 2.3.1 Multiple Linear Regression (MLR)

A modelling method examines the relationship between a response variable and several explanatory variables. The model can be written as follows in an MLR model with p explanatory variables and n observations:

Y = ω0 + ω1x1 + ω2x2 + … + ωpxp + ɛt

Here, y is the response variable, x is the explanatory variable, ɛ is the fitting error, and ω is the regression coefficient (Shabri & Samsudin, 2014). The model coefficients are frequently estimated using the least squares approach. However, if the model is flawed or improperly stated, the least squares fit outcomes occasionally may not be acceptable (Bozdogan & Howe, 2012).

The traditional MLR model solution minimises the squared errors between the projected and actual values (Shabri & Samsudin, 2014).

The following are the assumptions of the multiple linear regression model:

1. The response and explanatory variables have a linear relationship.
2. The explanatory variables have a minimum level of correlation.
3. From the population, observations were selected randomly and separately.
4. The residuals have a mean of 0 and are normally distributed (Adam, 2022).

## 2.4 Evaluation Metrics

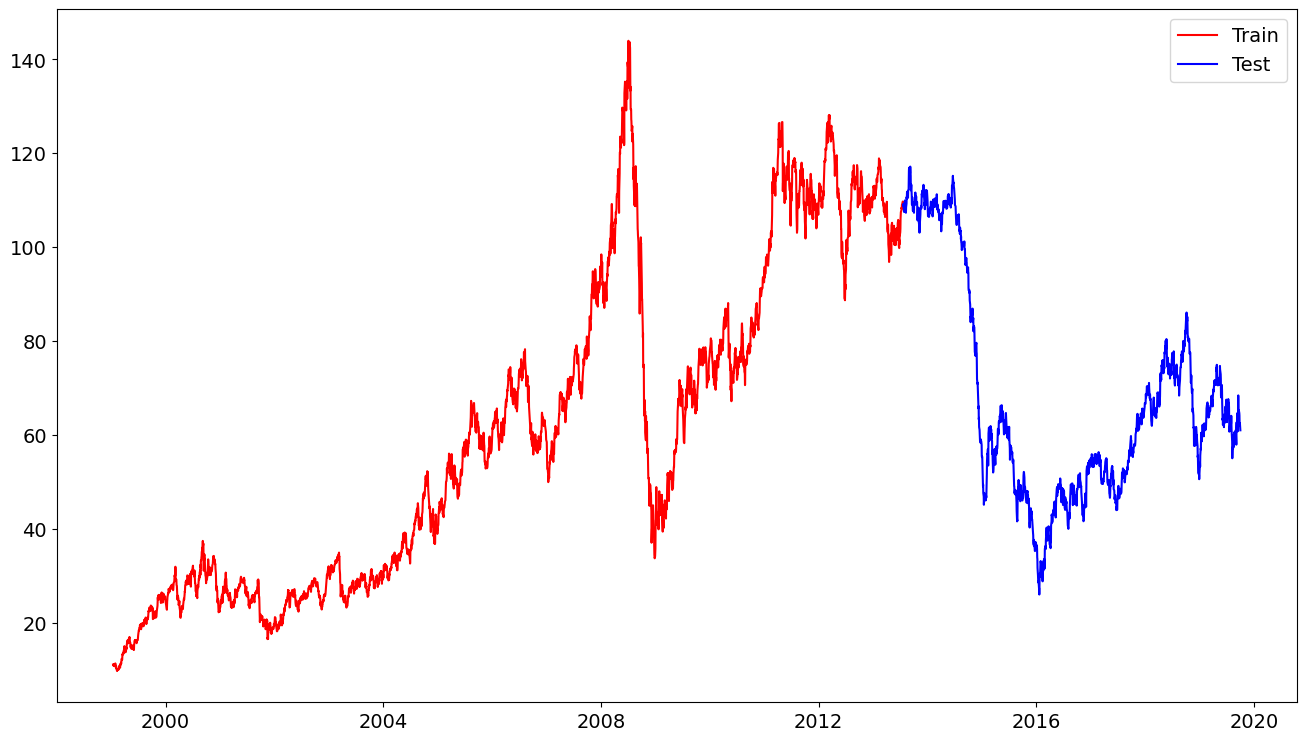
The key metrics used to measure predicting stability and accuracy were mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE).

## 2.5 Training and Test Data Descriptive Statistics

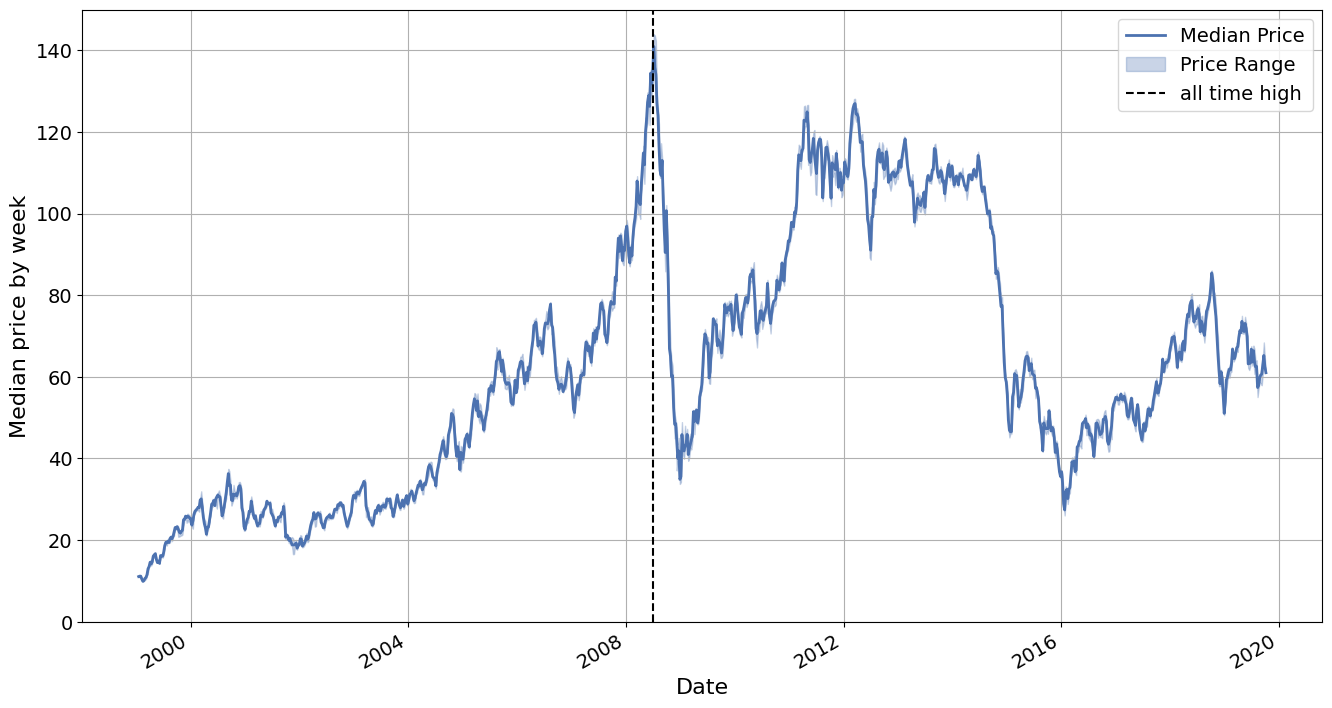
Our goal was to make price predictions for crude oil. We used daily prices of Brent crude oil price as the sample for thorough assessment of the usefulness of the LSTM model for an oil price forecast. The primary justification for choosing it is that Brent oil serves as the pricing benchmark for more than 75% of the world's exporting crude oil, either directly or indirectly (market pulse ICE). The sample ranges from May 20, 1987, to September 30, 2019, but for our research, only prices between January 4, 1999, and September 30, 2019, were taken into account.

We built an LSTM model to analyse forecasts. We separated the entire sample into the training and test sets following the standard data set partitioning rules for LSTM models. 70% of the data were used as the training set to fit our model, taking the volatility of crude oil spot prices into account. In other words, the training set's time frame covered January 14, 1999, to July 17, 2013. The test set consisted of 30% of the samples. The test set time frame covered the period from July 22, 2013, to September 30, 2019.

Figure 1 shows the training and test sets for Brent oil price data, and figure 2 shows the line chart of Brent oil price changes for the last 20 years.



**Figure 1**: The training and test set for brent oil prices



**Figure 2**: Brent oil Prices for the last 20 years

As can be seen in Figures 1 and 2, Since about 2007, significant fluctuations are found in the prices of Brent oil. The prices remained relatively steady during the 1990s, but the increase in global oil demand in early 1999 ended this stable period (Ajmi et al., 2021). In addition, commodity prices typically increased in the 2000s, and Brent oil prices went up even further. The price of Brent oil attained an all-time peak of 147.50 in 2008 (Trading Economics, 2023). The commodity bubble was then ended as crude oil prices plummeted. The cost fluctuated significantly after 2008. New price bubbles were noticed for both prices at the start of 2015. The principal factors determining how oil prices changes are demand and supply circumstances (Zhang & Hong, 2022).

Table 1: LSTM Parameters

|  |  |
| --- | --- |
| Parameters | Values |
| Number of layers | 5 (3 hidden, 1 input and 1 output layer) |
| Number of units | 30 |
| Normalization | Batch Normalization |
| Optimizer | Adam |
| Loss function | Mean squared error |
| Dropout probability | 0.2 |
| Scaler | MinMaxScaler |
| Epoch | 50 |
| Batch size | 32 |

Table 2: Linear Regression Parameters

|  |  |
| --- | --- |
| Parameters | values |
| Loss function | Mean squared error |
| Target Variable | Price |
| Features | 3-day and 9-day moving average of price |

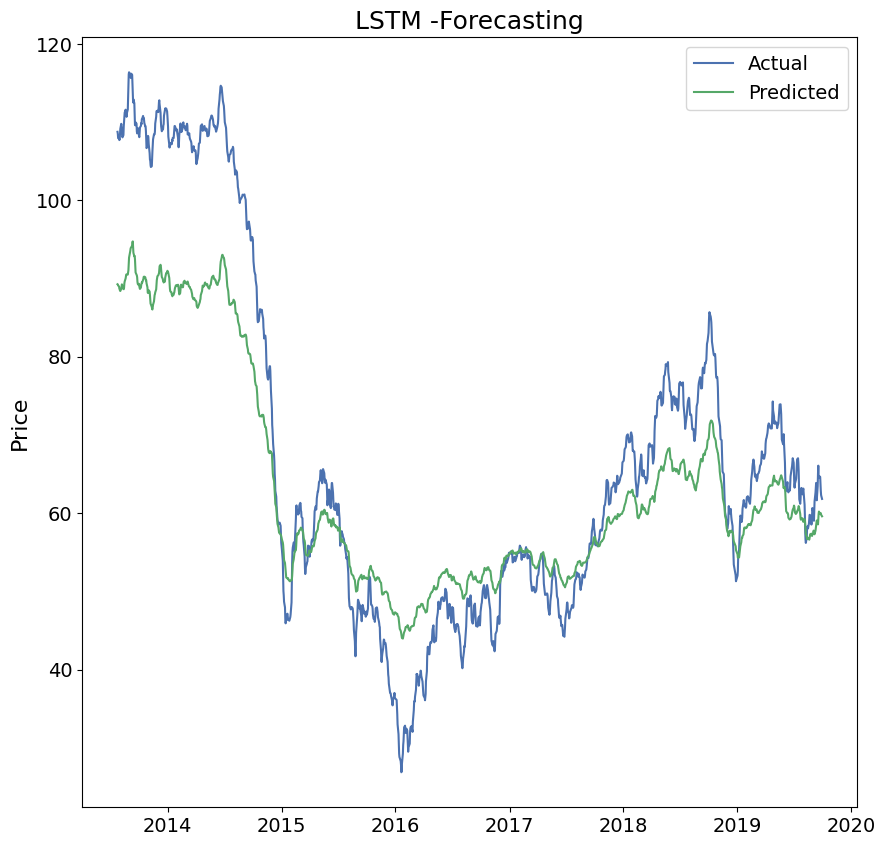
Table 3: Cross Validation Parameters

|  |  |
| --- | --- |
| Parameters | values |
| CV type | KFold |
| Number of split (n\_split) | 5 |
| shuffle | False |
| Scoring method | cross-val\_score |
| model | Linear Regression |
| Evaluation metric | neg\_mean\_squared\_error |
| n\_jobs | -1 |

**3. FITTING EFFECT ANALYSIS OF LSTM**

The LSTM model outperformed MLR significantly when the short-term effects of the current price's 9-day and 3-day moving averages were considered. The fitting values show exact trend as the actual values for the MLR and LSTM models. The fitting price decreases as the actual price does.

Secondly, it is essential to consider the variations between the actual and fitting values. Although the MLR model fits better, the Multiple Linear Regression model performs best when dealing with linear sequences due to its linear properties. However, it cannot demonstrate the nonlinear aspect of oil price fluctuations. Third, the LSTM model successfully fit the shifting crude oil price trend when a significant turning point in the oil prices occurred. There have been several instances in which significant events or factors have caused shifts in crude oil prices (Zhang & Hong, 2022). These include the Gulf War in 1990 and 1991 (Ajmi et al., 2021), production cuts by OPEC and recovery from the financial crisis in Asia in the late 1990s, the September 2011 terrorist attack in the US, storms in the Gulf of Mexico and ongoing supply issues in Nigeria and Iraq in the mid-2000s, and the global financial crisis in the late 2000s. The LSTM model may effectively predict changes in crude oil prices during these significant turning points (Zhang & Hong, 2022).



**Figure 3**: Predicted values using LSTM and actual values



**Figure 4**: Predicted values using multiple linear regression and actual values

# 

# 4. FORECASTING EFFECT ANALYSIS

The techniques used to forecast prices are displayed in the second column of Table 1. The MSE, MAE, and RMSE findings are displayed in columns 3, 4, and 5, respectively. Additionally, the RMSE value was used to predict stability, while the MAE and MSE values offer a reflection of the accuracy of the forecast. Lower values indicate higher accuracy.

Table 4: Forecasting techniques and evaluation metrics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Forecasting method | MSE | MAE | RMSE |
| Brent price | LSTM | 105.61 | 7.95 | 10.27 |
|  | MLR | 0.69 | 0.63 | 0.83 |

It was observed that the three metrics were higher for LSTM than MLR. This does not mean MLR performed better because there is some latent information that LSTM considers that MLR would not, and in the real world scenario, MLR would predict wrongly.

The forecasting accuracy of LSTM can further be improved by performing hyperparameter tuning, increasing the number of nodes and also using other features not available in the data provided.

**5. CONCLUSION**

The LSTM model produces strong generalization capabilities and reliable fitting outcomes at various timescales. It outperforms the multiple linear regression model in terms of fitting oil prices.

The LSTM model provides superior predicting accuracy and stability. It performs better than the multiple linear regression model, in particular, in terms of forecasting stability and accuracy.

The predicting accuracy of the LSTM model decreases as the forecast timeline for crude oil prices increases. Therefore, for more extended forecast periods, extra factors should be incorporated into the model (Zhang & Hong, 2022). It shows how well the LSTM model manages erratic and nonlinear changes in oil prices. Accurate forecasts from the LSTM model can also help to understand better the mechanisms underlying changes in the price of oil as well as the production practices and investment choices made by businesses, which can enhance domestic economic growth and security.

**6. RECOMMENDATION**

Brent oil is a widely traded type of crude oil shipped from various locations, including Russia, Nigeria, Angola, and Middle Eastern countries. It is highly liquid and is used as a benchmark for oil pricing. Various contracts are available for Brent oil, including ones for different types of oil such as Mars, Dubai and Caspian Sea oil. These contracts allow physical players to manage risk and financial traders to take positions and express their views on the market. (Wittner, 2020). Based on the data presented, the following potential recommendations are made:

1. The potential effects of changes in oil price on the company and the sector in which it works must be considered. It may have an effect on customer demand, shipping costs, and manufacturing costs.

1. Machine learning techniques can be one way to try and forecast future oil prices. However, it is crucial to carefully assess the performance of these models and consider other factors that may impact oil prices.
2. It may be helpful to monitor economic conditions, political events, and changes in global demand for oil, as these can all significantly affect oil prices.
3. It is advantageous to consider diversifying the company's energy sources or putting in place energy-saving measures.
4. It is also helpful to consider developing marketing strategies that are flexible and adaptable to changing market conditions in order to respond quickly to changes in oil prices.

Overall, it is critical to consider the potential effects of oil price fluctuations on the company and take the initiative to adjust to market developments. It may be possible to lessen the impacts of oil price variations and preserve a competitive edge in the market by carefully monitoring market developments and considering a variety of alternative tactics.

**REFERENCES**

Abdollahi, H. (2020). A novel hybrid model for forecasting crude oil price based on time series decomposition. *Applied energy*, 267, 115035. [https://doi.org/10.1016/j.apenergy.2020.115035](https://10.1016/j.apenergy.2020.115035)

Ajmi, A. N., Hammoudeh, S., & Mokni, K. (2021). Detection of bubbles in WTI, brent, and Dubai oil prices: A novel double recursive algorithm. *Resources Policy*, *70*, 101956. <https://doi.org/10.1016/j.resourpol.2020.101956>

Bozdogan, H., & Howe, J. A. (2012). Misspecified multivariate regression models using the genetic algorithm and information complexity as the fitness function. *European Journal of Pure and Applied Mathematics*, *5*(2), 211-249.

Butler, S., Kokoszka, P., Miao, H., & Shang, H. L. (2021). Neural network prediction of crude oil futures using B-splines. *Energy Economics*, 94, 105080. <https://doi.org/10.1016/j.eneco.2020.105080>

Chiroma, H., Abdulkareem, S., & Herawan, T. (2015). Evolutionary Neural Network model for West Texas Intermediate crude oil price prediction. *Applied Energy*, 142, 266-273. [http://doi.org/10.1016/j.apenergy.2014.12.045](https://10.1016/j.apenergy.2014.12.045)

El-Masry, A. A., & Mostafa, M. (2016). Oil price forecasting using gene expression programming and artificial neural networks. <http://dx.doi.org/10.1016/j.econmod.2015.12.014>

Fan, D., Sun, H., Yao, J., Zhang, K., Yan, X., & Sun, Z. (2021). Well production forecasting based on ARIMA-LSTM model considering manual operations. *Energy*, 220, 119708. <https://doi.org/10.1016/j.energy.2020.119708>

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654-669. <https://doi.org/10.1016/j.ejor.2017.11.054>

Gori, F., Ludovisi, D., & Cerritelli, P. F. (2007). Forecast of oil price and consumption in the short term under three scenarios: Parabolic, linear and chaotic behaviour. *Applied Energy*, 32(7), 1291-1296. [https://doi.org/10.1016/j.energy.2006.07.005](https://10.1016/j.energy.2006.07.005)

Graves, A. (2012). Long Short-Term Memory. In: Supervised Sequence Labelling with Recurrent Neural Networks. Studies in Computational Intelligence, vol 385. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-24797-2_4>

Hayes, A. (2022, June 23). Multiple Linear Regression (MLR) Definition, Formula, and Example. <https://www.investopedia.com/terms/m/mlr.asp>

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>

Lammerding, M., Stephan, P., Trede, M., & Wilfling, B. (2013). Speculative bubbles in recent oil price dynamics: Evidence from a Bayesian Markov-switching state-space approach. *Energy Economics*, 36( C ), 491-502. <https://doi.org/>[10.1016/j.eneco.2012.10.006](https://econpapers.repec.org/scripts/redir.pf?u=https://doi.org/10.1016%2Fj.eneco.2012.10.006;h=repec:eee:eneeco:v:36:y:2013:i:c:p:491-502)

Li, Z., Dong, H., Floros, C., Charemis, A., & Failler, P. (2022). Re-examining bitcoin volatility: a CAViaR-based approach. *Emerging Markets Finance and Trade*, 58(5), 1320-1338. <https://doi.org/10.1080/1540496X.2021.1873127>

Lu, Q., Li, Y., Chai, J., & Wang, S. (2020). Crude oil price analysis and forecasting: A perspective of “new triangle”. *Energy Economics*, 87, 104721.

Murat, A., & Tokat, E. (2009). Forecasting oil price movements with crack spread futures. *Energy Economics*, 31(1), 85-90. <https://doi.org/10.1016/j.eneco.2008.07.008>

Ouyang, Z. S., & Lai, Y. (2021). Systemic financial risk early warning of financial market in China using Attention-LSTM model. *The North American Journal of Economics and Finance*, *56*, 101383. <https://doi.org/10.1016/j.najef.2021.101383>

Pabuçcu, H., Ongan, S., & Ongan, A. (2020). Forecasting the movements of Bitcoin prices: an application of machine learning algorithms. *Quantitative Finance and Economics*, 4(4), 679-692. <https://doi.org/10.3934/QFE.2020031>

Ramyar, S., & Kianfar, F. (2019). Forecasting crude oil prices: A comparison between artificial neural networks and vector autoregressive models. *Computational Economics*, 53(2), 743-761. <https://doi.org/10.1007/s10614-017-9764-7>

Shabri, A., & Samsudin, R. (2014). Crude oil price forecasting based on hybridizing wavelet multiple linear regression model, particle swarm optimization techniques, and principal component analysis. *The Scientific World Journal,* 3, 854520 . <http://dx.doi.org/10.1155/2014/854520>

Trading Economics (2023). Brent crude oil 2023 Data - 1970-2022 Historical - 2024 Forecast - Price. <https://tradingeconomics.com/commodity/brent-crude-oil>

Wei, Y., Wang, Y., & Huang, D. (2010). Forecasting crude oil market volatility: Further evidence using GARCH-class models. *Energy Economics*, 32(6), 1477-1484. <https://doi.org/10.1016/j.eneco.2010.07.009>

Wittner, M. (2020, September). Brent: the world’s crude benchmark. *The ICE,* https://www.theice.com/insights/market-pulse/brent-the-worlds-crude-benchmark

Zhang, K., & Hong, M. (2022). Forecasting crude oil price using LSTM neural networks. *Data Science in Finance and Economics*, 2(3), 163-180. <https://doi.org/10.3934/DSFE.2022008>

Zhang, Y., Ma, F., & Wang, Y. (2019). Forecasting crude oil prices with a large set of predictors: Can LASSO select powerful predictors? *Journal of Empirical Finance*, 54, 97-117. <https://doi.org/10.1016/j.jempfin.2019.08.007>

Zhao, Y., Li, J., & Yu, L. (2017). A deep learning ensemble approach for crude oil price forecasting. *Energy Economics*, 66, 9-16. [https://doi.org/10.1016/j.eneco.2017.05.023](https://10.1016/j.eneco.2017.05.023)

Zhong, J., Wang, M., M Drakeford, B., & Li, T. (2019). Spillover effects between oil and natural gas prices: Evidence from emerging and developed markets. *Green Finance*, *1*(1), 30–45. <https://doi.org/10.3934/gf.2019.1.30>