Journal of Mineral Resources Engineering, 8(2): 83-97, (2023)

Imam Khomeini International University
Vol. 8, No. 2, Summer 2023

نشریه مهندسی منابع معدنی
Journal of Mineral Resources Engineering (JMRE)

Research Paper

Crude Oil Price Forecasting Using Text Mining and Big Data Modeling

Fattahi Sh.[1], Kianpoor S.[2,3*], Soheili K.[1]

1- Associate Professor, Faculty of Economics, Razi University, Kermanshah, Iran
2- Assistant Professor, Faculty of Economics, Payame Noor University, Tehran, Iran
3- Ph.D Student, Faculty of Economics, Razi University, Kermanshah, Iran

Received: 30 Jun. 2022 **Accepted: 04 Sep. 2022**

Abstract: This study uses data modeling and text mining techniques for oil price predictions. To improve the model's explanatory capability, text features from internet news articles on crude oil are automatically extracted using convolutional neural networks. Additionally, various time series models employ a state analysis approach called convolution. The years 2021 to 2011 saw the collection of almost 13000 news items, and it was discovered that text mining and data from large Internet-based apps perform better for prediction than other approaches. This means that it is pretty fair to say that there is a parallel link between news headlines, those headlines, and searches in the Google search engine. This relationship is highly appropriate for correctly forecasting the price of crude oil.

Keywords: Oil price, Google trends, Data mining, Deep learning, Convolution neural network.

How to cite this article
Fattahi, Sh., Kianpoor, S., and Soheili, K. (2023). *"Crude oil price forecasting using text mining and big data modeling"*. Journal of Mineral Resources Engineering, 8(2): 83-97.
DOI: 10.30479/JMRE.2022.17482.1593

*Corresponding Author Email: *s_kianpoor@pnu.ac.ir*

INTRODUCTION

Forecasting the price of crude oil has been an essential topic in the field of forecasting research [1]. A challenging question in modeling various factors is how to select and extract effective predictors, some of which are difficult to determine. The emergence of extensive data knowledge means that there is enough online data to reflect the factors that create oil markets [2]. Google search, the most effective tool for obtaining the latest relevant news on the Internet, ranks first among all search engines [3]. This research aims to provide a new method for predicting the price of crude oil, which focuses on improving the prediction accuracy according to online news information and Google Trends. It faces the primary hypothesis that combining complex neural networks with text mining techniques to extract Oil and gold news information can improve the accuracy of crude oil and gold price prediction. The purpose of this study is to investigate the impact of oil pricing through media influence and to combine qualitative information with statistical data to predict oil prices. This research uses convolutional neural networks to extract hidden patterns in online news media. Then, the analysis technique is used to identify the effects of online news by processing convolutional outputs. There are many kinds of research for extracting text. However, this is the first time that the hidden patterns in online news media have been extracted by combining artificial neural networks and analysis techniques. In this context, this study tries to answer this research question: To what extent can the everyday use of online media and Google Trends improve crude oil price forecasting?

METHODS

The research method is deep learning in terms of its purpose, application, and method of collecting and analyzing information. The statistical population of this research is the international indices of crude oil news, Google Trends, and the price of crude oil in the world. The details of each variable are as follows:

This study inputs three datasets: historical oil prices, news headlines, and Google trend datasets. A total of 12,842 news headlines published in the "Crude News" column of the popular energy news portal "Oilprice.com" were collected from June 1, 2011, to June 11, 2021. Every seven days, the news is collected as a sample and a total of 252 samples—weekly data from September 25, 2017, to August 25, 2019, covering 100 observations. Weekly oil price data cover the period from September 18, 2017, to August 18, 2019, with 100 observations. There are four specific Google trends, i.e., "crude oil", "oil inventory", "oil consumption" and "oil price" that come from Google search. The present study divides the data set into training and test sets. The convolution model's training period is from June 1, 2011, to November 17, 2015, which includes 9445 news and 228 weekly titles. The exam period is from November 18, 2015, to June 11, 2021, which includes 3397 papers and 298 weekly papers. The convolutional network is set to 70-30 for training and testing, so the amount of news in the training set is almost close to the test set. Considering that convolution models are used as input variables for crude oil price forecasting, the training and testing sets for the crude oil price forecasting model are solved using the convolution test period shown in Figure 1. The oil price forecasting model is from November 17, 2015, to October 28, 2020, and includes 268 weekly records. In contrast, the test set is from October 18, 2020, to June 11, 2021, which includes 30 weekly records.

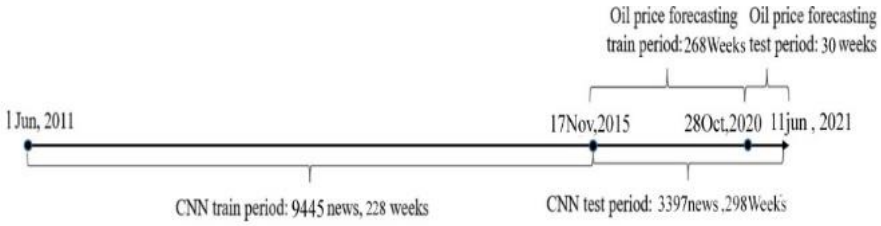

Figure 1. Training and testing sets of oil price forecasting models

FINDINGS AND ARGUMENT

The research findings indicate that all the time series are at the level or with one-time differentiation; also, the co-accumulation test was performed, and the results show the co-accumulation relationship between the global price of crude oil and Google trends. This study used a combination test and Granger causality analysis to investigate how Google Trends and crude oil prices interact. Since many time series

are unstable and different models are based on stationarity; the convergence test starts from the stationarity analysis of time series. The Engel-Granger test is used to test the effect of Google Trends on the price of crude oil. In this regard, it was concluded that the Google trend "Crude Oil" is the Granger causality of the price of crude oil in one and two orders, and the Google trend "Oil Price" is the Grangerian causality of crude oil price in one order at a significant level of 5%. However, Google's "Oil Inventories" trend shows no Granger causality in crude oil prices. Google's "Crude Oil" trend is more suitable to attract investors' attention compared to Google's "Crude Oil" and Google's "Oil Inventory" trends.

Also, in choosing the best Google Trends crude oil interval, the time interval of one was considered. While extracting the text of the online news, it was described by the cloud section of the best words. The 20 main words from which some of the factors affecting the price of crude oil can be extracted are as follows: "production", "crude oil", "United States of America", "output", and "pipeline". "Exports", "Reduction", "Iran", "Imports", "Gas", "Transaction", "Plan", "Increase", "LNG (Liquefied Natural Gas)", "OPEC (Organization of the Petroleum Exporting Countries)", "refinery", "BPD (barrels per day)", "energy" and "price". As shown, "crude oil", "gas", and "energy" show a close relationship with crude oil (Figure 2).

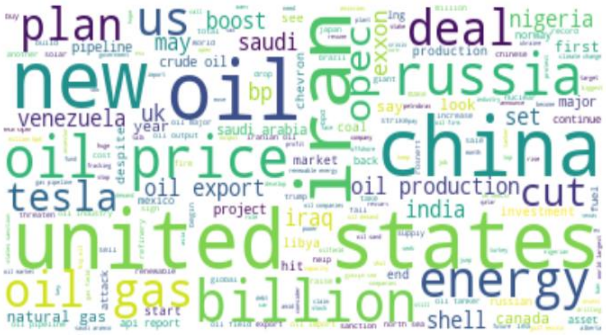

Figure 2. Word cloud of the top 100 words in the entire corpus

Meanwhile, "production", "increase", "liquefied natural gas", and "refinery", among others, may represent crude oil demand and supply. In addition, "The United States", "Iran", "Iranian", and "OPEC" reflect political events. The news headlines include various factors that affect the price of crude oil. Therefore, learning semantic relationships between crude oil news states enables researchers to make predictions. The results of IR were also stated; the results show that news headlines and Google trends complement each other in terms of improvement rates of mean square error and mean absolute error. It should also be noted that the results combining the text features and Google trends are improved more than the single results. The results of the average value of the influence of the historical oil price (-1 and 0) of WIT and the Google trend for "crude oil" (-1 and 0) show that it is an essential factor for predicting the price of crude oil. In general, it can be concluded that the combination of text features and Google trends can contribute to significant progress in crude oil price forecasting. Therefore, the proposed method with the text of oil news and Google trends as useful predictors can be considered an effective tool for crude oil price prediction. According to the obtained results, it is clear that the research hypothesis is accepted, and crude oil news can be compelling in its price prediction. This finding is in line with researchers' studies [1,2,4,5]. Many studies show this alignment [6-8]. Among the limitations of this research are the years of research because, before 2011, the collection of statistics was not available, and it is suggested that this model be carried out on a broader level or even its effects on other financial markets be investigated.

CONCLUSIONS

In order to select useful predictors, this study combines qualitative information with statistical data, and qualitative information and statistical data are used to predict the price of crude oil, which are complementary. It shows that News messages reflect different trends. Motivated by this topic, a combination of crude oil price forecasting methods and crude oil news using research, deep learning, and analysis techniques is suggested. A convolutional neural network can reveal hidden patterns, where online oil news headlines correspond to crude oil price developments. It considers that Google Trends and news text information can

promote each other. Google Trends can indicate investors' attention, while investors' attention and crude oil price fluctuations interact. Predictive information cannot be entirely obtained by using statistical data alone. The news text can reflect unexpected social or political events that play an essential role in the volatility of oil prices. Therefore, considering Google trends and news text is a scientific approach. In addition, this study helps the oil price prediction method.

Combining text features and Google trends can significantly progress crude oil price forecasting. Therefore, the proposed method with the text of oil news and Google Trends as useful predictors can be considered a more effective tool for crude oil price prediction. In the long term, the global price of oil has experienced one of the most volatile conditions. Therefore, it is necessary to combine the media information to judge whether this information is positive or negative for the oil market. In addition, the research results describe a significant positive relationship between online media news and market performance in the past period. In general, marketers can consider the joint effect of Google Trends and online media news in the oil market or other market fields

REFERENCES

[1] Abdollahi, H. (2020). *"A novel hybrid model for forecasting crude oil price based on time series decomposition"*. Applied Energy, 267: 115035.

[2] Yu, L., Zhao, Y., Tang, L., and Yang, Z. (2019). *"Online big data-driven oil consumption forecasting with Google Trends"*. International Journal of Forecasting, 35(1): 213-223.

[3] Li, X., Ma, J., Wang, S., and Zhang, X. (2015). *"How does Google search affect trader positions and crude oil price?"*. Economic Modelling, 49: 162-171.

[4] Zhao, L. T., Wang, Y., Guo, S. Q., and Zeng, G. R. (2018). *"A novel method based on numerical fitting for oil price trend forecasting"*. Applied Energy, 220: 154-163.

[5] Wu, B., Wang, L., Lv, Sh., and Zeng, Y. (2021). *"Effective crude oil price forecasting using new text-based and big-data-driven model"*. Elsevier, 168: 108468.

[6] Hu, H. L., Wang, L., Peng, L., and Zeng, Y. R. (2020). *"Effective energy consumption forecasting using enhanced bagged echo state network"*. Energy, 193: 116778.

[7] Bildirici, M., Bayazit, N., and Ucan, Y. (2020). *"Analyzing crude oil prices under the impact of COVID-19 by using LSTARGARCHLSTM"*. Energies, 13(11): 2980.

[8] Elshendy, M., Colladon, A. F., Battistoni, E., and Gloor, P. (2018). *"Using four different online media sources to forecast the crude oil price"*. Journal of Information Science, 44(3): 408- 421.