



IMAM KHOMEINI
INTERNATIONAL UNIVERSITY
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](https://doi.org/10.30479/JMRE.2022.17482.1593)

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

COPYRIGHTS



©2023 by the authors. Published by Imam Khomeini International University.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/>)

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.

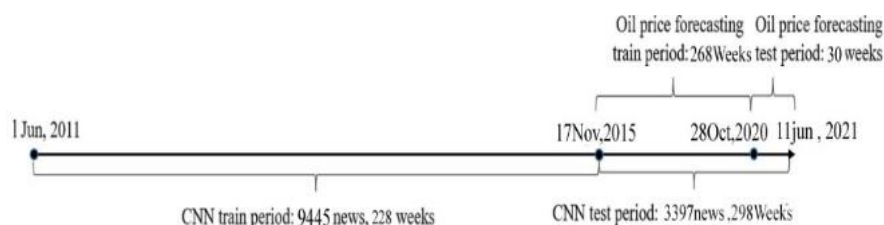


Figure 1. Training and testing sets of oil price forecasting models

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.



دوره هشتم، شماره ۲، تابستان ۱۴۰۲، صفحه ۸۷ تا ۹۷
Vol. 8, No. 2, Summer 2023, pp. 87-97



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

علمی-پژوهشی

پیش بینی قیمت نفت خام با استفاده از روش متن کاوی و مدلسازی داده های بزرگ

شهرام فتاحی^۱، سعید کیان پور^{۲*}، کیومرث سهیلی^۱

- ۱- دانشیار، دانشکده اقتصاد، دانشگاه رازی، کرمانشاه
- ۲- استادیار، دانشکده اقتصاد، دانشگاه پیام نور، تهران
- ۳- دانشجوی دکتری، دانشکده اقتصاد، دانشگاه رازی، کرمانشاه

پذیرش: ۱۴۰۱/۰۶/۱۳

دریافت: ۱۴۰۱/۰۴/۰۹

چکیده

هدف این تحقیق بررسی پیش بینی قیمت نفت با رهیافت متن کاوی و مدل داده های بزرگ است. برای استخراج خودکار ویژگی های متن از اخبار آنلاین نفت خام از روش شبکه عصبی کانولوشنال استفاده می شود و از این طریق قدرت توضیح دهنده مدل افزایش می یابد. همچنین حالت مختلف سری زمانی با استفاده از تجزیه حالت از روش کانولوشن مورد استفاده قرار می گیرد. نزدیک به ۱۳۰۰۰ عنوان خبری طی سال های ۲۰۲۱-۲۰۱۱ جمع آوری شد و در نتیجه مشخص شد روش های پیش بینی مبتنی بر متن کاوی و داده های بزرگ مبتنی بر اینترنت از روش های دیگر بهتر عمل می کند. از این رو می توان گفت ارتباط موازی عنوان های خبری و تیتراژها و جستجو در موتور جستجوی گوگل در پیش بینی دقیق قیمت نفت خام بسیار مناسب است.

کلمات کلیدی

قیمت نفت، گوگل ترندز، داده کاوی، یادگیری عمیق، شبکه عصبی کانولوشن.

استناد به این مقاله

فتاحی، ش.، کیان پور، س.، سهیلی، ک.؛ ۱۴۰۲؛ "پیش بینی قیمت نفت خام با استفاده از روش متن کاوی و مدلسازی داده های بزرگ". نشریه مهندسی منابع معدنی، دوره هشتم، شماره ۲، ص ۸۳-۹۷.

DOI: 10.30479/JMRE.2022.17482.1593



حق مؤلف © نویسندگان

ناشر: دانشگاه بین المللی امام خمینی (ره)

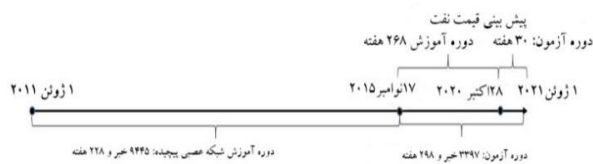
*نویسنده مسئول و عهده دار مکاتبات Email: s_kianpoor@pnu.ac.ir

۱- مقدمه

حلقه بازخورد شاخص‌های گوگل و قیمت نفت خام را تایید کردند. در واقع، برخی از مطالعات از گوگل ترندز به عنوان یکی از پیش‌بینی‌کننده‌های مفید برای پیش‌بینی بازار استفاده کردند [۳]. نویسندگان یک مطالعه نشان دادند که پیام‌های خبری آنلاین با تجزیه و تحلیل بیش از ۴۵ میلیون پیام خبری، ظرفیت پیش‌بینی قوی دارند. آنها استدلال کردند که بازار نفت عمیقاً تحت تاثیر رویدادهای شدید (مانند بی‌ثباتی‌های سیاسی و توسعه اقتصادی) قرار دارد و الگوریتم‌های متن‌کاوی می‌توانند اطلاعات قابل استفاده را از اخبار آنلاین نفت خام استخراج کنند [۱۲]. تحقیقات، تکنیک‌های در حال ظهور هوش مصنوعی را برای طبقه‌بندی، ارزیابی کردند و چارچوب مناسبی را برای استخراج متن پیشنهاد کردند. آنها شبکه‌های عصبی پیچشی را شایسته مطالعه و پتانسیل دانستند [۱۳]. بر اساس تحقیقات گوناگون، این مطالعه دیدگاه‌های نظری و کاربرد عملی برای پردازش اطلاعات را ارائه می‌دهد. این اولین بار است که اطلاعات مخفی، در رسانه‌های خبری آنلاین با ترکیب تکنیک‌های یادگیری عمیق و تکنیک‌های تجزیه استخراج می‌شود. علاوه بر این، دیگر نوآوری این مطالعه در ترکیب اخبار نفت خام و گوگل ترندز برای پیش‌بینی قیمت نفت خام با استفاده از الگوریتم‌های جدید است، بنابراین این مطالعه منابع خبری آنلاین را به عنوان منابع داده کیفی مرتبط و موثر در نظر می‌گیرد. پیام‌های خبری و گوگل ترندز نسبتاً مکمل یکدیگر هستند زیرا مورد اول نشان‌دهنده رویدادهای مهم اخیر در مورد بازار نفت است و دومی گرمای این رویدادها را نشان می‌دهد. در بسیاری از موارد، اخبار آنلاین یک منبع اطلاعاتی محوری است که تجزیه و تحلیل آن از پیش‌بینی‌های مالی، اقتصادی و سیاسی پشتیبانی می‌کند [۱۴، ۱۵]. با توجه به این موضوع، این تحقیق قصد دارد روش جدیدی برای پیش‌بینی قیمت نفت خام ارائه دهد که تمرکز آن بر بهبود دقت پیش‌بینی با توجه به اطلاعات خبری آنلاین و گوگل ترندز است و با این فرضیه اساسی مواجه است که ترکیب شبکه عصبی پیچیده با تکنیک متن‌کاوی برای استخراج اطلاعات اخبار نفت و طلا صحت پیش‌بینی قیمت نفت خام و طلا را بهبود می‌بخشد. هدف این مطالعه تاثیر قیمت‌گذاری نفت از طریق تاثیرپذیری از رسانه‌ها است و اطلاعات کیفی را با داده‌های آماری برای پیش‌بینی قیمت نفت ترکیب می‌کند. در این تحقیق شبکه‌های عصبی پیچشی برای استخراج الگوهای پنهان در رسانه‌های خبری آنلاین مورد استفاده قرار می‌گیرند،

نفت خام یکی از اجزای مهم اقتصاد جهانی است. پیش‌بینی قیمت نفت خام یک موضوع قابل توجه در زمینه پیش‌بینی تحقیقات بوده است [۱]. سوال چالش برانگیز در مدل‌سازی عوامل مختلف، نحوه انتخاب و استخراج پیش‌بینی‌کننده موثر است که تعیین برخی از آنها دشوار است. ظهور دانش داده‌های بزرگ به این معنی است که داده‌های آنلاین کافی برای انعکاس عوامل ایجادکننده بازارهای نفت وجود دارد [۲]. جستجوی گوگل که موثرترین ابزار برای به دست آوردن آخرین اخبار مرتبط در اینترنت است، در بین همه موتورهای جستجو در رتبه اول قرار دارد [۳]، بنابراین گوگل ترندز به عنوان یک مدل خاص از داده‌های بزرگ با انبوه اطلاعات در نظر گرفته می‌شود. بر این اساس، گوگل ترندز به ویژه یکی از پیش‌بینی‌کننده‌های مفید برای پیش‌بینی قیمت نفت خام در مطالعه حاضر است. با این حال، روند بازار نفت به مشکل در پیش‌بینی حرکت بازار منجر شده است. منابع خطا در پیش‌بینی قیمت نفت خام شامل ساختار پیچیده عرضه و تقاضا و بسیاری از عوامل غیرقابل پیش‌بینی است که تعادل بازار را مختل می‌کند. در بسیاری از موارد، اخبار آنلاین یک منبع اطلاعاتی محوری است که تجزیه و تحلیل آن از پیش‌بینی‌های مالی، اقتصادی و سیاسی پشتیبانی می‌کند [۴]. جامعه تحقیقاتی سعی کرده است از تکنیک‌های زیادی برای پیش‌بینی قیمت نفت خام و عوامل تعیین‌کننده آن استفاده کند که اکثریت مطالعات تجربی، پیش‌بینی قیمت نفت خام در اقتصاد مانند تکنیک‌های اقتصادسنجی و الگوریتم‌های هوشمند را پذیرفته‌اند [۷، ۶، ۵]. تکنیک‌های مختلف تجزیه مانند تجزیه حالت تجربی (برای به دست آوردن اطلاعات زمان-فرکانسی سیگنال باید روشی استفاده شود که بتواند مودهای ذاتی نهفته در سیگنال را استخراج کند، از این رو از تجزیه حالت تجربی یا EMD استفاده می‌شود)، تجزیه موجک، تجزیه و تحلیل طیفی منحصر به فرد می‌تواند استفاده شود [۸-۱۱]. به عنوان مثال تحقیقات پیش‌بینی قیمت نفت خام را با تجزیه‌های متعدد موجک تجزیه و تحلیل کردند. ضعف مشترک همه این رویکردها این است که روندهای آینده قیمت از داده‌های آماری تاریخی مشتق شده است. این مطالعات همچنین مشخص کرده‌اند که این تکنیک‌های تجزیه، زمانی که رویدادهای نامنظم تاثیر کوتاه‌مدت داشته باشند موثر نیستند، بنابراین در نظر گرفتن عوامل برون‌زا برای پیش‌بینی ضروری است [۹]. تحقیقات،

قیمت نفت خام با استفاده از دوره آزمون کانولوشن حل می‌شود. همان‌طور که در شکل ۱ نشان داده شده است. مدل پیش‌بینی قیمت نفت ۱۷ نوامبر ۲۰۱۵ تا ۲۸ اکتبر ۲۰۲۰ است و ۲۶۸ رکورد هفتگی دارد. در مقابل، مجموعه آزمایش ۱۸ اکتبر ۲۰۲۰ تا ۱۱ ژوئن ۲۰۲۱ است که ۳۰ رکورد هفتگی دارد.



شکل ۱: مجموعه آموزش و تست پیش‌بینی قیمت نفت

در این تحقیق نمونه بین سال‌های ۲۰۱۱-۲۰۲۱ انتخاب شده است. برای برآورد مدل از نظر اجرا و آماری از مدل کانولوشن تمام الگوریتم‌های پیشنهادی با استفاده از پایتون، ژوپیتِر نوت بوک و گوگل کولب کدگذاری شده‌اند. این مطالعه همچنین از شش مدل یادگیری عمیق شامل کی-نزدیکترین همسایه، درخت تصمیم، شبکه عصبی کانولوشن، ماشین بردار پشتیبانی، جنگل تصادفی، الگوریتم تقویت گرادیان به عنوان تکنیک‌های پیش‌بینی قیمت استفاده می‌کند. ضمناً نرم‌افزار متلب، ایوبیوز و اکسل نیز استفاده شده است. این محاسبه بر روی یک رایانه شخصی با CPU 2.7 گیگاهرتز، ۸ گیگابایت RAM و سیستم عامل حرفه‌ای ویندوز ۱۰ پیاده‌سازی می‌شود.

۲-۱- شبکه عصبی کانولوشن

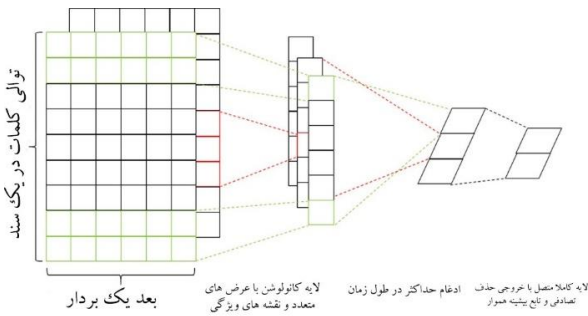
در ابتدا باید گفت شبکه عصبی مصنوعی از چندین پرسپترون در هر لایه تشکیل می‌شود و ورودی‌ها در جهت روبه جلو پردازش می‌شوند، اما شبکه عصبی پیچشی الگوریتم یادگیری عمیق است که تصویر را دریافت می‌کند و به هر یک از جنبه‌های تصویر وزن قابل یادگیری و بایاس می‌دهد تا آن‌ها را از هم متمایز کند و به پیش‌پردازش کمتری در مقایسه با دیگر الگوریتم‌ها نیازمند است از این رو دقت بالاتری دارند و تشخیص خودکار ویژگی‌های مهم بدون نظارت انسانی را دارند، اما به داده‌های آموزشی زیادی نسبت به شبکه عصبی مصنوعی نیاز دارند. شبکه عصبی پیچشی یکی از مهم‌ترین روش‌های یادگیری عمیق است که لایه‌های متعدد در آن به شیوه‌ای جدید و مستحکم آموزش می‌بینند. این روش، موثر واقع

سپس از تکنیک تجزیه برای تشخیص تاثیرات اخبار آنلاین با پردازش خروجی‌های کانولوشنال استفاده می‌شود. تحقیقات زیادی برای استخراج متن وجود دارد، اما این اولین بار است که الگوهای پنهان در رسانه‌های خبری آنلاین با ترکیب شبکه‌های عصبی مصنوعی و تکنیک‌های تجزیه استخراج می‌شود. در این شرایط، این مطالعه سعی دارد به این سوال تحقیق پاسخ دهد: تا چه حد، استفاده مشترک از رسانه‌های آنلاین و گوگل ترندز می‌تواند پیش‌بینی قیمت نفت خام را بهبود بخشد؟

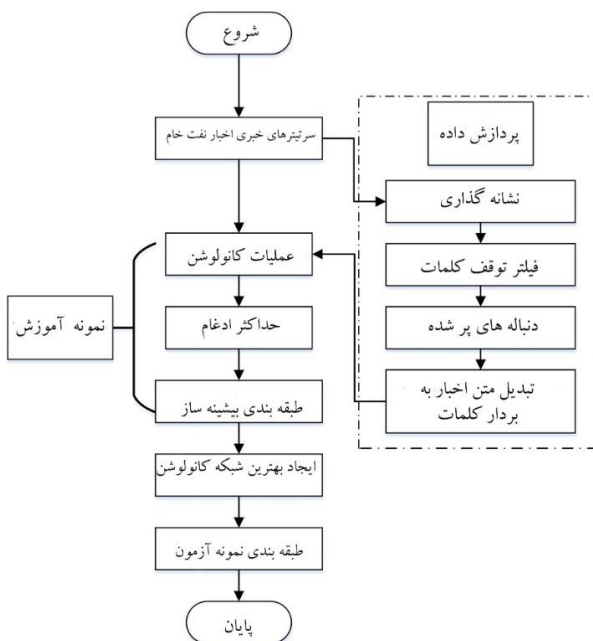
۲- روش روش کار

روش تحقیق از لحاظ هدف، کاربردی و از نظر شیوه گردآوری و تحلیل اطلاعات از نوع یادگیری عمیق است. جامعه آماری این پژوهش شاخص‌های بین‌المللی اخبار نفت خام، گوگل ترندز و قیمت نفت خام در جهان است. جزئیات مربوط به هر یک از متغیرها به شرح زیر است.

این مطالعه از سه مجموعه داده شامل قیمت نفت تاریخی، عنوان خبرها و مجموعه داده‌های روند گوگل به عنوان ورودی استفاده می‌کند. در مجموع ۱۲۸۴۲ سرفصل خبری منتشر شده در ستون "اخبار نفت خام" در پرتال خبری محبوب انرژی "Oilprice.com" از ۱ ژوئن ۲۰۱۱ تا ۱۱ ژوئن ۲۰۲۱ جمع‌آوری شده است. هر هفت روز، اخبار به صورت نمونه و در مجموع ۲۵۲ نمونه جمع‌آوری می‌شود. داده‌های هفتگی از ۲۵ سپتامبر ۲۰۱۷ تا ۲۵ آگوست ۲۰۱۹ و در مجموع ۱۰۰ مشاهده را پوشش می‌دهد. داده‌های هفتگی قیمت نفت در بازه زمانی ۱۸ سپتامبر ۲۰۱۷ تا ۱۸ آگوست ۲۰۱۹ با مجموع ۱۰۰ مشاهده را پوشش می‌دهد. چهار گرایش ویژه گوگل شامل "نفت خام"، "موجودی نفت"، "مصرف نفت" و همچنین "قیمت نفت" که از جستجوی گوگل به دست می‌آید، وجود دارد. مطالعه حاضر مجموعه داده‌ها را به مجموعه‌های آموزشی و آزمایش تقسیم می‌کند. در مدل کانولوشن، دوره آموزش ۱ ژوئن ۲۰۱۱ تا ۱۷ نوامبر ۲۰۱۵ که شامل ۹۴۴۵ عنوان خبری و ۲۲۸ عنوان هفتگی است. دوره آزمون ۱۸ نوامبر ۲۰۱۵ تا ۱۱ ژوئن ۲۰۲۱ است که شامل ۳۳۹۷ مقاله و ۲۹۸ پرونده هفتگی است. شبکه پیچشی برای آموزش و آزمایش ۷۰ تا ۳۰ تنظیم شده است تا حجم اخبار مجموعه آموزشی تقریباً نزدیک به مجموعه تست باشد. با توجه به این‌که مدل‌های کانولوشن به عنوان متغیرهای ورودی پیش‌بینی قیمت نفت خام استفاده می‌شوند، مجموعه آموزش و آزمایش برای مدل پیش‌بینی



شکل ۲: ساختار مدل پیچشی



شکل ۳: نمودار شبکه عصبی پیچشی

شده و به صورت رایج در کاربردهای بینایی ماشین استفاده می‌شود. این شبکه‌ها زیرمجموعه‌ای از شبکه‌های عصبی چند لایه هستند که برای داده‌های دو بعدی مانند تصویر، طراحی شده‌اند. بخش‌های مختلف تصویر به عنوان ورودی به لایه‌های سلسله مراتبی داده می‌شوند و در هر لایه با اعمال فیلترهایی دیجیتال، ویژگی‌های چشم‌گیری از تصویر استخراج می‌شود. به طور کلی کانولوشن از سه نوع لایه اصلی کانولوشن، ادغام و لایه تماماً متصل تشکیل می‌شود [۱۶]. با توجه به مزایای کانولوشن بر اساس شبکه یادگیری عمیق و ساختار تحول، این روش تجزیه و تحلیل احساسی متون کوتاه را مورد هدف قرار داده است. لایه کانولوشن، کانولوشن عملیات متقاطع را روی ماتریس ورودی انجام می‌دهد. در نتیجه آن را قادر می‌سازد تا قطعات معنایی جملات مختلف را ترکیب کرده و تعامل بین قطعات سنتز شده را بیاموزد. بر این اساس، روابط معنایی بین شیوه‌های اخبار نفت خام به طور کامل مورد استفاده قرار می‌گیرد. شکل ۳ ساختار مدل کانولوشن را نشان می‌دهد. ابتدا، مدل کانولوشن با یک ماتریس جمله توکن شده شروع می‌شود که در آن هر سطر یک کلمه را نشان می‌دهد. پس از آن، ماتریس جمله از طریق ضرب جمع می‌شود. فیلترهای خطی، مانند تقسیم سه، چهار یا پنج کلمه در هر بار انجام می‌شود. این طول جمله و اندازه ناحیه فیلتر ابعاد را تعیین می‌کند. خروجی تولید شده از نگاشت فیلتر را می‌توان به یک بردار ویژه "سطح بالا" با طول ثابت متصل کرد که پس از آن با استفاده از یک تابع پیشینه هموار برای تولید وارد می‌شود. یک نمودار مختصر و کامل در شکل ۳ نشان داده شده است تا درک ما از الگوریتم کانولوشن را بهبود بخشد. رویه‌ها در شکل ۳ از این قرار است:

مرحله ۱: پیش‌پردازش داده‌ها: مجموعه داده‌ها به داده‌های آموزشی و داده‌های آزمایشی تقسیم می‌شود. داده‌های اخبار نفت خام از چهار عملیات از جمله "نشانه‌گذاری"، "فیلتر کردن کلمات را متوقف کنید"، "دنباله پر شده"، "تبدیل متن خبر به بردار کلمه" عبور می‌کند.

مرحله ۲: آموزش مدل کانولوشن با استفاده از نمونه آموزشی. بردار کلمه وارد مدل کانولوشن می‌شود و فرآیند مدل کانولوشن شامل "عملیات متقابل"، "حداکثر جمع‌آوری" و "طبقه‌بندی تابع پیشینه هموار" است.

مرحله ۳: با استفاده از بهترین مدل آموزش دیده کانولوشن، نمونه آزمایش را طبقه‌بندی می‌کند. در این مطالعه، کانولوشن به وسیله پایتون کدگذاری شده است. همچنین از کتابخانه

پایتون به نام تنسورفلو و پاندا استفاده می‌شود که یک رابط برنامه‌نویسی اپلیکیشن ساده برای ایجاد شبکه‌های عمیق آرایه می‌دهد.

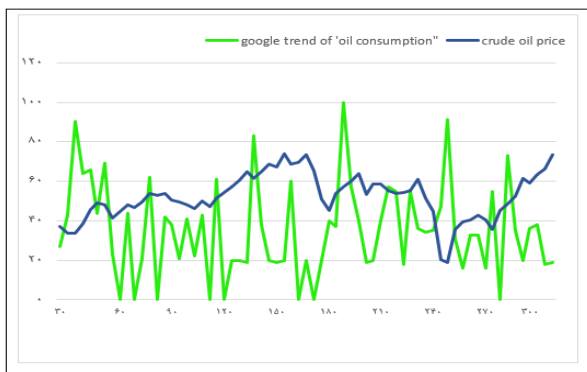
۲-۲- سایر الگوریتم‌ها

الگوریتم K نزدیک‌ترین همسایه (KNN) در بازشناخت الگوی یک متد آمار ناپارامتری است که برای طبقه‌بندی آماری و رگرسیون استفاده می‌شود. در هر دو حالت کی شامل نزدیک‌ترین مثال آموزشی در فضای داده‌ای و خروجی آن بسته به نوع مورد استفاده در طبقه‌بندی و رگرسیون متغیر است. در حالت طبقه‌بندی با توجه به مقدار مشخص شده برای کی، به محاسبه فاصله نقطه‌ای که می‌خواهد برچسب آن را مشخص

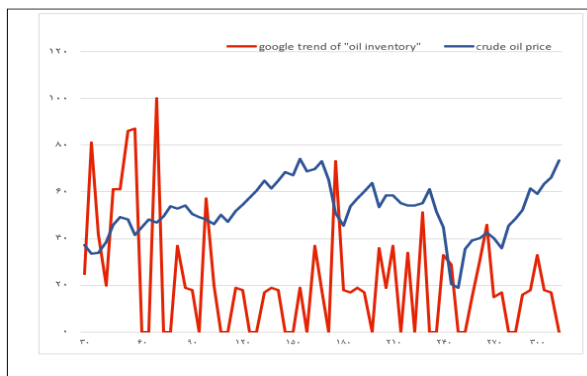
کنند. روش‌های تقویت‌گرادیان زیادی وجود دارد، اما تا اینجا محبوب‌ترین‌ها AdaBoost (مخفف Adaptive Boosting) و تقویت‌گرادیان هستند.

۳- نتایج و بحث

شکل‌های ۴، ۵ و ۶ داده‌های سری زمانی قیمت نفت خام و سه مورد روند گوگل ترندز را نشان می‌دهد. توجه داشته باشید که هیچ رابطه آشکاری بین قیمت نفت خام و برخی از گوگل ترندزها، مانند "موجودی نفت" و "مصرف نفت" وجود ندارد. شکل ۷ داده‌های سری زمانی قیمت نفت خام و روند "نفت خام" گوگل را نشان می‌دهد. روند "نفت خام" گوگل روندهای مشابهی را در مورد قیمت نفت چه با اندکی تاخیر و چه همزمان نشان می‌دهد، بنابراین روند "نفت خام" گوگل در پیش‌بینی قیمت نفت مفید است. بر این اساس، رابطه بین قیمت نفت خام و چهار روند گوگل بیشتر بررسی می‌شود.



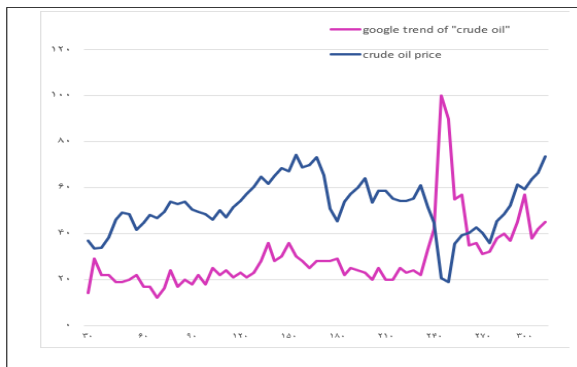
شکل ۴: سری زمانی قیمت نفت خام و روند گوگل مصرف نفت



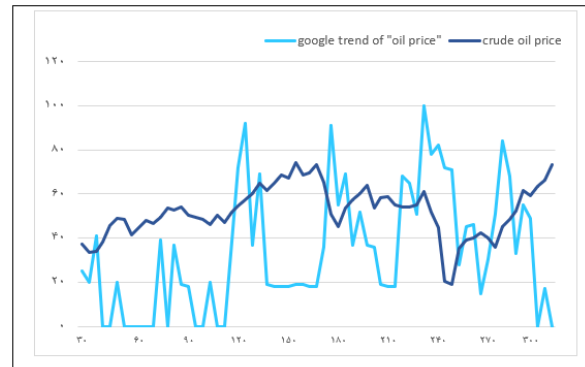
شکل ۵: سری زمانی قیمت نفت خام و روند گوگل موجودی نفت

کند با نزدیک‌ترین نقاط می‌پردازد و با توجه به تعداد رای حداکثری این نقاط همسایه، در رابطه با برچسب نقطه مورد نظر تصمیم‌گیری می‌کند. برای محاسبه این فاصله می‌توان از روش‌های مختلفی استفاده کرد که یکی از مطرح‌ترین این روش‌ها، فاصله اقلیدسی است. در حالت رگرسیون نیز میانگین مقادیر به دست آمده از کی خروجی آن است. از آنجا که محاسبات این الگوریتم بر اساس فاصله است.

الگوریتم درخت تصمیم (Decision Tree)، این الگوریتم جمعیت را برای چندین مجموعه بر اساس برخی خصوصیات انتخاب شده (متغیرهای مستقل) یک جمعیت طبقه‌بندی می‌کند. معمولاً این الگوریتم برای حل مشکلات طبقه‌بندی استفاده می‌شود. طبقه‌بندی با استفاده از برخی از تکنیک‌ها مانند آنترپوی، کای اسکور و جینی و نظایر آن انجام می‌شود. نرمال‌سازی داده‌ها به بهبود عملکرد آن کمک می‌کند. ماشین بردار پشتیبانی (SVM) یکی از روش‌های یادگیری با نظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. این روش از جمله روش‌های نسبتاً جدیدی است که در سال‌های اخیر کارایی خوبی نسبت به روش‌های قدیمی‌تر برای طبقه‌بندی نشان داده است. مبنای کاری دسته‌بندی کننده SVM دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌شود خطی انتخاب شود که حاشیه اطمینان بیشتری داشته باشد. حل معادله پیدا کردن خط بهینه برای داده‌ها به وسیله روش‌های QP که روش‌های شناخته شده‌ای در حل مسایل محدودیت‌دار هستند، صورت می‌گیرد. جنگل تصادفی یا جنگل‌های تصمیم تصادفی (Random forest) یک روش یادگیری ترکیبی برای دسته‌بندی، رگرسیون است که بر اساس ساختاری متشکل از شمار بسیاری درخت تصمیم، بر روی زمان آموزش و خروجی کلاس‌ها (کلاس‌بندی) یا برای پیش‌بینی‌های هر درخت به شکل مجزا، کار می‌کنند. جنگل‌های تصادفی برای درختان تصمیم که در مجموعه آموزشی دچار بیش‌برازش می‌شوند، مناسب‌اند. عملکرد جنگل تصادفی معمولاً بهتر از درخت تصمیم است، اما این بهبود عملکرد تا حدی به نوع داده هم بستگی دارد. الگوریتم تقویت‌گرادیان (GRADIENTBOOST) که در اصل Hypothesis Boosting نام دارد به هر روش گروهی گفته می‌شود که چند یادگیرنده ضعیف را با هم ترکیب می‌کند تا یک یادگیرنده قوی بسازد. ایده عمومی بیشتر روش‌های تقویت‌گرادیان این است که پیش‌بینی‌کننده‌ها به ترتیب آموزش ببینند که هر کدام سعی می‌کنند اشتباهات پیشین را درست



شکل ۷: سری زمانی قیمت نفت خام و روند گوگل نفت خام



شکل ۶: سری زمانی قیمت نفت خام و روند گوگل قیمت نفت خام

جدول ۱: نتایج حاصل از آزمون دیکی فولر تعمیم یافته (ADF)

بررسی مانایی	مقادیر بحرانی		آماره ADF	متغیرها
	%۵	%۱		
با ۱ بار تفاضل گیری مانا است .	-۲,۹۰۶۲	-۳,۵۳۳۲	-۱,۹۵۶۴	قیمت نفت خام
در سطح مانا است .	-۲,۹۰۶۲	-۳,۵۳۳۲	-۷,۱۶۵۴	روند گوگل قیمت نفت
در سطح مانا است .	-۲,۹۰۶۲	-۳,۵۳۳۲	-۳,۹۱۳۴	روند گوگل موجودی نفت
در سطح مانا است .	-۲,۹۰۶۲	-۳,۵۳۳۲	-۷,۸۷۶۶	روند گوگل مصرف نفت
با ۱ بار تفاضل گیری مانا است .	-۲,۹۰۶۲	-۳,۵۳۳۲	-۲,۹۲۹۸	روند گوگل نفت خام

جدول ۳: انتخاب وقفه بهینه گوگل ترند نفت خام

LR	FPE	AIC	HQ	
Na	۲,۲۵۸۸۵	۱۵,۸۳۷۱	۱۵,۸۶۴۳	۰
۱۳۴,۱۲۸۴*	۲۹۲۲,۶۱۵*	۱۳,۶۵۵۸*	۱۳,۷۳۷۲*	۱
۴,۴۳۲۲۹	۳۰۸,۶۹۱	۱۳,۷۰۷۸	۱۳,۸۴۳۴	۲
۱,۸۰۹۸۶	۳۴۰,۷۰۱	۱۳,۸۰۵۴	۱۳,۹۹۵۳	۳
۵,۰۶۸۲۳	۳۵۲۵,۳۲۱	۱۳,۸۳۹۱	۱۴,۰۸۳۲	۴
۰,۸۳۱۹۶	۳۹۶۷,۴۱۷	۱۳,۹۵۳۶	۱۴,۲۵۲۰	۵

جدول ۲: آمارهای توصیفی متغیرهای مربوطه

Google trend of "oil price"	Crude oil price	
۰,۳۳۰۱	۰,۳۹۲۴	میانگین
۰,۳۲۱۸	۰,۳۹۵۵	میانه
۰,۰۳۹۴	۰,۰۲۵۶	انحراف معیار
۰,۹۹۳۵	-۱,۴۷۳۸	چولگی
۴,۵۱۶۳	۶,۵۹۱۰	کشیدگی
۲۸,۱۷۴۴*	۶۰,۲۵۵۹*	نرمالیتی

**در سطح ۱ درصد

نتایج نشان‌دهنده روابط هم‌انباشتگی بین قیمت جهانی نفت خام و روندهای یاد شده گوگل است. سوم، هدف از تجزیه و تحلیل علیت گرنجری بررسی آماری این است که آیا "قیمت نفت"، "موجودی نفت" و "نفت خام" برای پیش بینی قیمت نفت خام مفید هستند یا خیر. مثلاً جدول ۴ نشان می‌دهد که آیا "نفت خام" گوگل علیت گرنجری قیمت نفت خام در یک و دو سفارش می‌شود؟ و آیا روند گوگل "قیمت نفت" علیت گرنجری قیمت نفت خام در یک سفارش در سطح معنی‌داری ۵ درصد است. با این حال، روند "موجودی نفت" گوگل نشان

ابتدا، سری‌های مختلف از نظر ایستایی آزمایش می‌شوند. جدول ۱ نتایج آزمایش را نشان می‌دهد. همچنین آمار توصیفی در جدول ۲ و انتخاب وقفه بهینه در جدول ۳ نمایش داده شده است. از این رو، شرایط لازم برای انباشتگی و علیت گرنجری به درستی برآورده می‌شود. با این حال، قیمت نفت خام و "مصرف نفت" به ترتیب متفاوت هستند. دوم، از آزمون همگرایی برای بررسی رابطه هم‌انباشتگی بین قیمت نفت خام و سه روند گوگل (یعنی "قیمت نفت"، "موجودی نفت" و "نفت خام") استفاده می‌شود. همچنین آزمون همگرایی انجام شد و

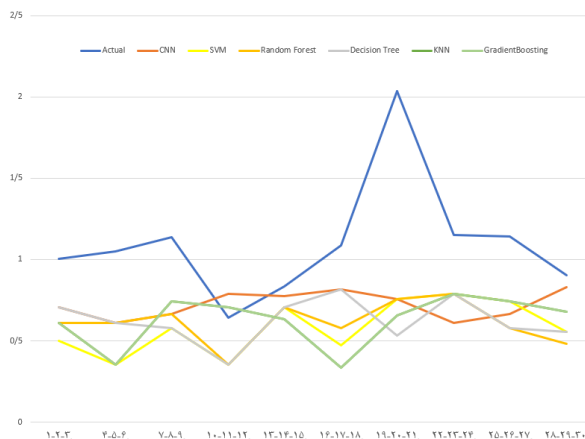
شکل ۹ روند ویژگی های طبقه بندی شبکه عصبی کانولوشن را در مقایسه با قیمت نفت در طول مجموعه داده های آزمایش توصیف می کند. طبقه بندی کانولوشن روندهای مشابه قیمت نفت را با اندکی تاخیر ارائه می دهد، کانولوشن روند مشابهی را برای قیمت نفت ارائه می دهد که همان وضعیت معاصر یا تاخیر جزئی با بخش واقعی (منحنی آبی رنگ) است که وجه تمایز آن ها را مشخص می کند و ما را قادر می سازد تا روند کلی نوسانات قیمت نفت تعیین شود. همچنین آزمون ویلکاکسون رتبه علامت دار برای انجام بررسی دو نمونه وابسته یا انطباق بین دو نمونه به کار گرفته می شود. این آزمون را می توان مشابه آزمون t برای گروه های وابسته برای میانگین جامعه غیرنرمال، در نظر گرفت (جدول ۶).

سیس دقت پیش بینی در طول دوره آزمون با استفاده از دو مورد ارزیابی می شود [۱۹].

ضمنا هنگام محاسبه این معیارها، روش پیش بینی مبتنی بر نورد (حالت متحرک) انتخاب می شود. نرخ بهبود (IR) برای تعیین قدرت توضیح بیشتر گوگل ترندز و ویژگی های متن معرفی می شود [۸]. نرخ بهبود از نظر فنی، نشان دهنده درصد بهبود قدرت توضیح بیشتر گوگل ترندز و عناوین خبری است (جدول ۷). این مطالعه از روش میانگین ارزش تاثیر (MIV)

جدول ۵: نتایج طبقه بندی مدل کانولوشن

معیار F	معیار پوشش	معیار صحت	دقت تست	دقت آموزش
۰٫۶۶	۰٫۵	۰٫۴۹	۰٫۵۰۹۸	۰٫۹۹۹۴



شکل ۹: نتایج مدل های متفاوت در طی ۳۰ هفته

می دهد که هیچ علیت گرنجری در قیمت نفت خام وجود ندارد. روند "نفت خام" گوگل در مقایسه با روند "نفت خام" گوگل و روند "موجودی نفت" گوگل گزینه مناسب تر برای جلب توجه سرمایه گذاران است.

جدول ۴: نتایج آزمایش تجزیه و تحلیل علیت گرنجر (سطح ۵ درصد)

		Lags	
		۱	۲
Panel A	H0: روند گوگل نفت خام علیت گرنجری قیمت نفت خام نیست.		
	F آماره	۰٫۴۱۵۷	۰٫۸۱۹۹۳
	احتمال	۰٫۰۵۲۱	۰٫۰۴۴۵
Panel B	H0: روند موجودی نفت گوگل علیت گرنجری قیمت نفت خام نیست.		
	F آماره	۱٫۰۱۹۰۸	۰٫۵۶۰۶۳
	احتمال	۰٫۳۱۶۶	۰٫۵۷۳۸
Panel C	H0: روند گوگل قیمت نفت باعث علیت گرنجری قیمت نفت خام نمی شود.		
	F آماره	۰٫۸۷۵۹۶	۰٫۱۸۵۲۴
	احتمال	۰٫۰۳۵۲	۰٫۰۸۳۱

شکل ۸ فراوانی وزنی ۱۰۰ کلمه کلیدی را با بیشترین وزن TF-IDF (فراوانی اصطلاح- معکوس فراوانی متن) در کل مجموعه توصیف می کند.



شکل ۸: فضای ابری کلمات از ۱۰۰ کلمه برتر در کل مجموعه

جدول ۵ نشان می دهد که CNN دقت تست خیلی بالایی ندارد. با این حال، چنین خطاهایی بیانگر ایراد در این رویدادها نیست [۱۸].

جدول ۶: آزمون ویلکاکسون رتبه علامت دار

مدل‌های مورد مقایسه	آزمون ویلکاکسون	احتمال	میانگین رتبه
کانولوشن در مقابل درخت تصمیم	۱,۷۷۶۴	۰,۰۷۵۷	۱۲,۹۰۰۰
کانولوشن در مقابل تقویت گرادیان	۱,۳۹۸۴	۰,۱۶۲۰	۱۲,۴۰۰۰
کانولوشن در مقابل K	۱,۳۹۸۴	۰,۱۶۲۰	۱۲,۴۰۰۰
نزدیک ترین همسایه			
کانولوشن در مقابل ماشین بردار پشتیبانی	۲,۰۴۱۰	۰,۴۱۳	۱۳,۲۵۰۰
کانولوشن در مقابل جنگل تصادفی	۱,۲۴۷۲	۰,۲۱۲	۱۲,۲۰۰۰

جدول ۷: بهبود پیش‌بینی مدل کانولوشن

میانگین مربع خطاها	میانگین خطا مطلق	
۰,۴۹۹۴	۰,۵۸۷۵	ویژگی‌های ترندهای گوگل (۱)
۰,۷۱۳۱	۰,۴۹۱۳	ویژگی‌های متن (۲)
۰,۹۶۸۴	۰,۹۱۴۱	ویژگی‌های تاریخی (۳)
۹۳,۹۱٪	۵۵,۵۹٪	IRR از (۳) به (۱)
۳۵,۸٪	۸۶,۰۵٪	IRR از (۳) به (۲)

جدول ۸: متغیرهای ورودی و رتبه‌بندی MIVها

متغیرهای ورودی	میانگین ارزش تاثیر	شماره رتبه بندی
قیمت تاریخی نفت (0) WIT	۰,۳۵۸۲۶۷۲۰	۱
قیمت تاریخی نفت (-1) WTI	۰,۳۵۸۲۵۸۳۱	۲
روندهای گوگل در مورد نفت خام (-1)	۰,۲۰۸۵۹۵	۳
گوگل ترندهای نفت خام (۰)	۰,۲۰۸۳۷۷	۴

(بشکه در روز)، "انرژی" و "قیمت". همان‌طور که نشان داده شده است، "نفت خام"، "گاز" و "انرژی" نشان‌دهنده ارتباط نزدیک با نفت خام هستند. در همین حال، "تولید"، "افزایش"، "گاز طبیعی مایع" و "پالایشگاه"، از جمله، ممکن است تقاضا و عرضه نفت خام را نشان دهند. علاوه بر این، "ایالات متحده"، "ایران"، "ایرانی" و "اوپک" منعکس‌کننده رویدادهای سیاسی هستند. عناوین خبری شامل عوامل مختلفی است که بر قیمت نفت خام تاثیر می‌گذارد، بنابراین یادگیری روابط معنایی

استفاده می‌کند که یکی از بهترین شاخص‌ها برای ارزیابی مقادیر ضریب تعیین شده در نظر گرفته می‌شود و از نظر کمی اهمیت نسبی هر یک از عوامل را در پیش‌بینی قیمت نفت خام بررسی می‌کند. در روش MIV، مجموعه آموزش با مدل کانولوشن آموزش داده می‌شود، سپس دو مجموعه آموزشی جدید با تغییر اندازه هر متغیر ورودی به ترتیب $\pm 10\%$ ، که برای شبیه‌سازی بر اساس مدل مناسب استفاده می‌شود، ایجاد می‌شود. میانگین مقادیر مختلف دو نتیجه شبیه‌سازی برای تولید MIV محاسبه می‌شود. میانگین ارزش تاثیر بزرگ در متغیر مستقل نشان‌دهنده قدرت پیش‌بینی بالای نورون ورودی نسبت به نورون خروجی است (جدول ۸).

یافته‌های پژوهش حاکی از آن است که تمامی سری‌های زمانی در سطح و یا با یکبار تفاضل‌گیری مانا هستند. همچنین آزمون هم‌انباشتگی انجام شد و نتایج نشان‌دهنده روابط هم‌انباشته بین قیمت جهانی نفت خام و روندهای گوگل است. در این مطالعه از آزمون تلفیق و تجزیه و تحلیل علیت گرنجر استفاده شده است تا بررسی شود که چگونه گوگل ترندز و قیمت نفت خام با یکدیگر تعامل دارند. با توجه به این که بسیاری از سری‌های زمانی غیرثابت و همچنین مدل‌های مختلف بر اساس ایستایی هستند، آزمون همگرایی از تجزیه و تحلیل ایستایی سری‌های زمانی شروع می‌شود. از آزمون انگل-گرنجر برای آزمایش تاثیر گوگل ترندز بر قیمت نفت خام استفاده می‌شود. در این رابطه به این نتیجه رسیده شد که روند گوگل "نفت خام" علیت گرنجری قیمت نفت خام در یک و دو سفارش است و روند گوگل "قیمت نفت" علیت گرنجری قیمت نفت خام در یک سفارش در سطح معناداری ۵ درصد است. با این حال، روند "موجودی نفت" گوگل نشان می‌دهد که هیچ علیت گرنجری در قیمت نفت خام وجود ندارد. روند "نفت خام" گوگل در مقایسه با روند "نفت خام" گوگل و روند "موجودی نفت" گوگل مناسب‌تر برای جلب توجه سرمایه‌گذاران است. همچنین در انتخاب وقفه بهینه گوگل ترندز نفت خام بهترین وقفه، وقفه زمانی یک در نظر گرفته شد. در خلال کار استخراج متن اخبار آنلاین به وسیله بخش ابری کلمات برتر توصیف شد. ۲۰ کلمه اصلی که برخی از عوامل تاثیرگذار بر قیمت نفت خام را می‌توان استخراج کرد، به شرح زیر است: "تولید"، "نفت خام"، "ایالات متحده آمریکا"، "خروجی"، "ستانده"، "خط لوله"، "صادرات"، "کاهش"، "ایران"، "واردات"، "گاز"، "معامله"، "طرح"، "افزایش"، "LNG (گاز طبیعی مایع)"، "اوپک (سازمان نفت کشورهای صادرکننده)"، "پالایشگاه"، "BPD"

گرفت، مطالعه یک چارچوب جامع پیش‌بینی را ارائه می‌دهد. نتایج تجربی ما نشان می‌دهد که ویژگی‌های متن و روند گوگل مکمل یکدیگر هستند. علاوه بر این، این روش برای پیش‌بینی قیمت نفت خام دقت رضایت بخشی دارد. مشارکت اصلی معرفی یک روش جدید مبتنی بر متن و مبتنی بر داده‌های بزرگ است که گوگل ترندز و اطلاعات متن اخبار را در نظر می‌گیرد. گوگل ترندز و اطلاعات متنی خبری می‌توانند یکدیگر را تبلیغ کنند. گوگل ترندز می‌تواند نشان‌دهنده توجه سرمایه‌گذاران باشد، در حالی که توجه سرمایه‌گذاران و نوسان قیمت نفت خام می‌توانند با یکدیگر تعامل داشته باشند. تنها با استفاده از داده‌های آمار نمی‌توان اطلاعات پیش‌بینی کننده را به طور کامل به دست آورد. متن اخبار می‌تواند رویدادهای اجتماعی یا سیاسی غیرمنتظره‌ای را بازتاب دهد که نقش مهمی در بی ثباتی قیمت نفت دارند، بنابراین در نظر گرفتن گوگل ترندز و متن اخبار یک رویکرد علمی است. علاوه بر این، این مطالعه به روش پیش‌بینی قیمت نفت کمک می‌کند، زیرا شبکه عصبی پیچشی می‌تواند برای تولید شاخص‌های سری زمانی اطلاعاتی بر اساس داده‌های متنی بدون ساختار استفاده شود. این مطالعه، روند گوگل و متون جدید را مورد مقایسه و بررسی قرار می‌دهد. مدل‌های آماری و هوش مصنوعی معمولی، یعنی الگوریتم تقویت گرادیان، الگوریتم ماشین بردار پشتیبانی، الگوریتم K نزدیک‌ترین همسایه، الگوریتم درخت تصمیم، الگوریتم جنگل تصادفی و شبکه عصبی پیچشی برای انتخاب مناسب‌ترین تکنیک برای پیش‌بینی قیمت نفت خام به کار می‌رود. نتایج نشان می‌دهد که مدل شبکه عصبی پیچشی عملکرد مطلوبی را نشان می‌دهد و بهترین عملکرد را از نظر معیارهای میانگین مربع خطا و میانگین خطا مطلق از همه مدل‌ها به دست می‌آورد. نتایج نرخ بهبود نشان می‌دهد که عناوین خبری و روند گوگل از نظر معیارهای نرخ بهبود میانگین مربع خطا و میانگین خطا مطلق مکمل یکدیگر هستند. مهم‌تر از همه، این یافته‌ها نشان می‌دهد که ویژگی‌های متنی یا ویژگی روند گوگل به تنهایی ممکن است اطلاعات پیش‌بینی کننده کافی برای پیش‌بینی قیمت نفت خام ارائه ندهد، اما ترکیب متن و ویژگی‌های روند گوگل می‌تواند تا حد زیادی دقت پیش‌بینی قیمت نفت خام را بهبود بخشد. این مطالعه از روش میانگین ارزش تاثیر (MIV) استفاده می‌کند که یکی از بهترین شاخص‌ها برای ارزیابی مقادیر ضریب تعیین شده در نظر گرفته می‌شود و از نظر کمی اهمیت نسبی هر یک از عوامل را در پیش‌بینی قیمت نفت خام بررسی می‌کند. در روش

بین حالت‌های خبری نفت خام، تحقیقات را قادر می‌سازد تا پیش‌بینی کنند. نتایج IR نیز در جدول ۷ بیان شد. نتایج نشان می‌دهد که عناوین خبری و روند گوگل از نظر معیارهای نرخ بهبود میانگین مربع خطا و میانگین خطا مطلق یکدیگر را تکمیل می‌کنند. همچنین باید توجه داشته باشید که نتایجی که ترکیبی از ویژگی‌های متن و روند گوگل است، در مقایسه با نتایج تکی بهبود یافته‌تر است. نتایج میانگین ارزش تاثیر قیمت تاریخی نفت (-۱ و ۰) WIT و روند گوگل برای "نفت خام" (-۱ و ۰) نشان می‌دهد عامل مهمی برای پیش‌بینی قیمت نفت خام است. به طور کلی می‌توان نتیجه گرفت، ترکیب ویژگی‌های متن و روند گوگل به پیشرفت چشمگیری در پیش‌بینی قیمت نفت خام کمک می‌کند، بنابراین روش پیشنهادی با متن اخبار نفت و گوگل ترندز به عنوان پیش‌بینی کننده مفید یک ابزار موثر برای پیش‌بینی قیمت نفت خام در نظر گرفته می‌شود. با توجه به نتایج به دست آمده مشخص می‌شود که فرضیه تحقیق پذیرفته شده است و اخبار نفت خام بر پیش‌بینی قیمت آن موثر است. این یافته با مطالعات پژوهشگران همسو است [۱۷،۵،۲،۱]. همچنین تحقیقات زیادی وجود دارد که این همسویی را نشان می‌دهد [۲۰-۲۲]. از جمله محدودیت‌های این تحقیق سال‌های تحقیق است چرا که قبل از سال ۲۰۱۱ جمع‌آوری آمار غیر قابل دسترس بوده و پیشنهاد می‌شود این مدل در سطح وسیع‌تری و یا حتی تاثیرات آن بر بازارهای مالی دیگر انجام شود.

۴- نتیجه گیری

این مطالعه با هدف انتخاب پیش‌بینی کننده‌های مفید، اطلاعات کیفی را با داده‌های آماری ترکیب می‌کند و اطلاعات کیفی با داده‌های آماری برای پیش‌بینی قیمت نفت خام که مکمل هستند به کار گرفته شده است و نشان می‌دهد پیام‌های خبری بازتاب گرایش‌های متفاوت است. رویدادهای مربوط به بازار نفت و گوگل ترندز بخش جذاب این رویدادها هستند، بنابراین پیش‌بینی قیمت نفت خام با استفاده از این عوامل ممکن است به عملکرد رضایت بخشی منجر شود. با انگیزه این موضوع، ترکیبی از روش‌های پیش‌بینی قیمت نفت خام و اخبار نفت خام با استفاده از تحقیقات، تکنیک‌های یادگیری عمیق و تکنیک‌های تجزیه پیشنهاد می‌شود. شبکه عصبی پیچشی می‌تواند الگوهای پنهان را فاش کند، که در آن عناوین اخبار آنلاین نفت با تحولات قیمت نفت خام مطابقت دارد. پس از اینکه ویژگی‌ها و اثرات تاخیر بیشتر مورد بررسی قرار

- 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] Fronzetti Colladon, A. (2020). "Forecasting election results by studying brand importance in online news". International Journal of Forecasting, 36(2): 414-427.
- [5] 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.
- [6] Bekiroglu, K., Duru, O., Gulay, E., Su, R., and Lagoa, C. (2018). "Predictive analytics of crude oil price by utilizing the intelligent model search engine". Applied Energy, 228: 2387-2397.
- [7] Atalla, T., Joutz, F., and Pierru, A. (2016). "Does disagreement among oil price forecasters reflect volatility? Evidence from the ECB surveys". International Journal of Forecasting, 32(4): 1178-1192.
- [8] Li, X., Shang, W., and Wang, S. (2019). "Text-based crude oil price forecasting: A deep learning approach". International Journal of Forecasting, 35(4): 1548-1560.
- [9] [9] Shahbaz, M., Tiwari, A. K., and Tahir, M. I. (2015). "Analyzing time-frequency relationship between oil price and exchange rate in Pakistan through wavelets". Journal of Applied Statistics, 42(4): 690-704.
- [10] Ma, X., Jin, Y., and Dong, Q. (2017). "A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting". Applied Soft Computing, 54: 296-312.
- [11] Wang, J., Li, X., Hong, T., and Wang, Sh. (2018). "A semi-heterogeneous approach to combining crude oil price forecasts". Information Sciences, 460: 279-292.
- [12] Wex, F., Widder, N., Liebmann, M., and Neumann, D. (2013). "Early warning of impending oil crises using the predictive power of online news stories". In: 2013 46th Hawaii International Conference on System Sciences, 1512-1521. DOI: 10.1109/HICSS.2013.186.
- [13] Hemmatian, F., and Sohrabi, M. K. (2019). "A survey on classification techniques for opinion mining and sentiment analysis". Artificial Intelligence Review, 52: 1495-1545.
- [14] Shaterkhabazi, M., Solh Mirzaei, Kh., and Kianpour, S. (2021). "The Role of the Convention on Intellectual
- میانگین ارزش تاثیر، مجموعه آموزش با مدل کانولوشن آموزش داده می‌شود، سپس دو مجموعه آموزشی جدید با تغییر اندازه هر متغیر ورودی به ترتیب $\pm 10\%$ ، که برای شبیه‌سازی بر اساس مدل مناسب استفاده می‌شود، ایجاد می‌شود. میانگین مقادیر مختلف دو نتیجه شبیه‌سازی برای تولید میانگین ارزش تاثیر محاسبه می‌شود. میانگین ارزش تاثیر بزرگ برای متغیر مستقل نشان‌دهنده قدرت پیش‌بینی بالای نورون ورودی نسبت به نورون خروجی است. با توجه به میانگین ارزش تاثیرهای مطلق محاسبه شده در دوره آموزشی، به ترتیب متغیرهای ورودی طبقه‌بندی شده است. به طور کلی، ترکیب ویژگی‌های متن و روند گوگل به پیشرفت چشمگیری در پیش‌بینی قیمت نفت خام کمک می‌کند، بنابراین روش پیشنهادی با متن اخبار نفت و گوگل ترندز به عنوان پیش‌بینی کننده مفید می‌تواند یک ابزار موثرتری برای پیش‌بینی قیمت نفت خام در نظر گرفته شود. قیمت جهانی نفت در بلندمدت یکی از پرنوسان‌ترین و بی‌ثبات‌ترین حال‌ها را تجربه کرده است، از این رو برای کاهش نوسان‌های شدید و بی‌ثبات در مقیاس جهانی به عنوان توصیه سیاستی پیشنهاد می‌شود که کاربران آنلاین در مورد تبدیل اطلاعات آنلاین به تصمیم‌گیری احتیاط بیشتری داشته باشند. البته به استثنای حوادث شدید (مانند بی‌ثباتی‌های سیاسی و توسعه اقتصادی) که می‌تواند نگرانی‌های زیادی را ایجاد کند، بنابراین باید اطلاعات رسانه‌ها را با هم ترکیب کرد تا قضاوت شود که آیا این اطلاعات برای بازار نفت مثبت هستند یا منفی. علاوه بر این، نتایج تحقیق یک رابطه مثبت حیاتی بین اخبار رسانه‌های آنلاین و عملکرد بازار در دوره گذشته را توصیف می‌کند. وقتی اخبار آنلاین وضعیت مثبتی برای بازار نفت نشان می‌دهد، عملکرد بازار بهتر می‌شود و بالعکس. اطلاعات اخبار آنلاین برای پیش‌بینی قیمت نفت خام کوتاه مدت است. با توجه به این دلیل برای دوره بلندمدت در سطح جهانی، در نظر گرفتن عملکرد قبلی بازارهای مالی به عنوان یک عامل هنگام اجرای فعالیت‌های بازاریابی مهم است. به طور کلی، بازاریابان می‌توانند تاثیر مشترک گوگل ترندز و اخبار رسانه‌های آنلاین را در بازار نفت یا سایر زمینه‌های بازار در نظر بگیرند.

۵- مراجع

- [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

- TSP.2013.2288675.
- [19] Hu, H. L., Wang, L., and Lv, S. X. (2020). "Forecasting energy consumption and wind power generation using deep echo state network". *Renewable Energy*, 154: 598-613.
- [20] 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.
- [21] 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.
- [22] Elshendy, M., Fronzetti Colladon, A., 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.
- Property Rights and Moral Corruption on Iran's Economic Growth*". *Journal of Ethics in Science and Technology*, 16(2): 187-189. (In Persian).
- [15] Shaterkhabazi, M., Kianpour, S., and Homayounfar, A. (2022). "The Relationship between Corruption and Business Environment with Economic Growth and Development". *Journal of Ethics in Science and Technology*, 16(4): 111-118. (In Persian).
- [16] Qayyum, A., Anwar, S. M., Awais, M., and Majid, M. (2017). "Medical image retrieval using deep convolutional neural network". *Neurocomputing*, 266: 8-20,
- [17] 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.
- [18] Dragomiretskiy, K., and Zosso, D. (2013). "Variational mode decomposition". *IEEE Transactions on Signal Processing*, 62(3): 531-544. DOI: 10.1109/