Research Article

The Interpretability of LSTM Models for Predicting Oil Company Stocks: Impact of Correlated Features

Javad T. Firouzjaee1,2,3 and Pouriya Khalilian1,2

1Department of Physics, K. N. Toosi University of Technology, P. O. Box 15875-4416, Tehran, Iran
2PDAT Laboratory, Department of Physics, K. N. Toosi University of Technology, P. O. Box 15875-4416, Tehran, Iran
3School of Physics, Institute for Research in Fundamental Sciences (IPM), P. O. Box 19395-5531, Tehran, Iran

Correspondence should be addressed to Javad T. Firouzjaee; taghizade.javad@gmail.com

Received 25 August 2023; Revised 10 December 2023; Accepted 20 December 2023; Published 20 February 2024

Copyright © 2024 Javad T. Firouzjaee and Pouriya Khalilian. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Oil companies are among the largest companies in the world whose economic indicators in the global stock market have a great impact on the world economy (Stevens, 2018) and market due to their relation to gold (Aijaz et al., 2016), crude oil (Henriques & Sadorsky, 2008), and the dollar (Huang et al., 1996). This study investigates the impact of correlated features on the interpretability of long short-term memory (LSTM) (Peters, 2001) models for predicting oil company stocks. To achieve this, we designed a standard long short-term memory (LSTM) network and trained it using various correlated data sets. Our approach is aimed at improving the accuracy of stock price prediction by considering the multiple factors affecting the market, such as crude oil prices, gold prices, and the US dollar. The results demonstrate that adding a feature correlated with oil stocks does not improve the interpretability of LSTM models. These findings suggest that while LSTM models may be effective in predicting stock prices, their interpretability may be limited. Caution should be exercised when relying solely on LSTM models for stock price prediction as their lack of interpretability may make it difficult to fully understand the underlying factors driving stock price movements. We have employed complexity analysis to support our argument, considering that financial markets encompass a form of physical complex system (Peters, 2001). One of the fundamental challenges faced in utilizing LSTM models for financial markets lies in interpreting the unexpected feedback dynamics within them.

1. Introduction

Machine learning, which is one of the subfields of artificial intelligence, has its applications in various fields including economics, medicine [1], cosmology [2], photonics [3], robotics [4], and IoT (Internet of Things) [5]. Long short-term memory (LSTM) models are widely used in the financial industry for time series prediction, including stock prices. LSTMs are known for their ability to capture complex patterns in sequential data. However, one major limitation of LSTM models is their lack of interpretability. That is, it is often difficult to understand how the model arrives at its predictions and what factors are driving the results. Despite their lack of interpretability, the performance of LSTMs in predicting stock prices has been well documented in previous research [6, 7]. Therefore, while caution is necessary when relying solely on LSTMs for financial decisions, these models can still be useful in predicting the future price movements of stocks.

It is widely acknowledged that oil companies occupy the largest market share in the energy industry [8]. These companies have enormous global reach and wield substantial influence in the world economy. In fact, oil is arguably the most vital factor of the global economy, as the imposition of sanctions on the export or import of oil from many countries can practically paralyze their economies, especially for those countries with oil-dependent economies, such as the Persian Gulf countries [9]. Additionally, stock indices of oil companies are regarded as among the most important indicators in the global stock market, with correlations between the shares of oil companies, gold, the US dollar, and crude oil, all of which have a significant impact on the market [10].
Neural networks, despite their lack of interpretability, have proven useful in modeling oil companies. Long short-term memory (LSTM) neural networks are one of the most powerful architectures in recurrent neural networks. They have addressed the problem of gradient vanishing in recurrent neural networks [11], thereby allowing more accurate predictions of the stock market [12]. This market behaves like an unstable dynamic system, characterized by numerous nonlinear correlations. Accordingly, researchers have proposed various methods to predict its behavior.

To review the literature oil research path, we can mention Alvarez-Ramirez et al.’s work [13] which analyzed the auto-correlations of international crude oil. After carrying out several tests, in 2005, Moshiri and Foroutan [14] concluded that oil stock markets have a recursive architecture because they are time series. They used three methods ANN, GARCH, and ARMA, and the best results come from the ANN method.

The authors in [15] published an article using multilayer neural networks, which examined the relationship between crude oil prices and current prices. Moreover, they showed that the future prices of crude oil contain new information about oil spot price detection. Ye et al. [16] studied the changing relationships between the prices of crude oil and several other factors from January 1992 to December 2007 by the short-run crude oil price forecast model.

Chen et al. developed a model based on deep learning and used this model to model the unknown nonlinear behavior of WTI stocks [17]. Qi et al. used different recursive neural network architectures including LSTM, GRU, BiLSTM, and RNN to model the forex market and obtained significant results from these models to predict several currency pairs. They used a database that relates to ELLIOT method information, one of the stock market forecasting methods [18].

Table 1: Correlations between shares.

<table>
<thead>
<tr>
<th></th>
<th>USD</th>
<th>WTI</th>
<th>Gold</th>
<th>BP</th>
<th>Total</th>
<th>Schlumberger</th>
<th>Cairn Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>USD</td>
<td>1.000000</td>
<td>-0.357051</td>
<td>-0.002052</td>
<td>-0.214303</td>
<td>-0.088397</td>
<td>0.225169</td>
<td>0.046451</td>
</tr>
<tr>
<td>WTI</td>
<td>-0.357051</td>
<td>1.000000</td>
<td>-0.046961</td>
<td>0.371175</td>
<td>0.355021</td>
<td>0.377098</td>
<td>0.490978</td>
</tr>
<tr>
<td>Gold</td>
<td>-0.002052</td>
<td>-0.046961</td>
<td>1.000000</td>
<td>-0.464169</td>
<td>-0.530797</td>
<td>-0.514969</td>
<td>-0.141593</td>
</tr>
<tr>
<td>BP</td>
<td>-0.214303</td>
<td>0.371175</td>
<td>-0.464169</td>
<td>1.000000</td>
<td>0.871033</td>
<td>0.381587</td>
<td>0.521461</td>
</tr>
<tr>
<td>Total</td>
<td>-0.088397</td>
<td>0.355021</td>
<td>-0.530797</td>
<td>0.871033</td>
<td>1.000000</td>
<td>0.529159</td>
<td>0.494148</td>
</tr>
<tr>
<td>Schlumberger</td>
<td>0.225169</td>
<td>0.377098</td>
<td>-0.514969</td>
<td>0.381587</td>
<td>0.529159</td>
<td>1.000000</td>
<td>0.431568</td>
</tr>
<tr>
<td>Cairn Energy</td>
<td>0.046451</td>
<td>0.490978</td>
<td>-0.141593</td>
<td>0.521461</td>
<td>0.494148</td>
<td>0.431568</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

Figure 1: Heatmap of the correlation coefficients in Table 1.
In 2018, Gupta and Pandey predicted crude oil prices by using the LSTM network [19], and following, that Cen and Wang applied deep learning algorithms to anticipate the volatility behavior of crude oil prices [20]. To solve the chaotic and nonlinear features of crude oil time series, Altan and Karasu used a new crude oil price prediction model that is proposed in this study, which includes the long short-term memory (LSTM) and technical indicators [21].

In 2018, Gupta and Pandey predicted crude oil prices by using the LSTM network [19], and following, that Cen and Wang applied deep learning algorithms to anticipate the volatility behavior of crude oil prices [20]. To solve the chaotic and nonlinear features of crude oil time series, Altan and Karasu used a new crude oil price prediction model that is proposed in this study, which includes the long short-term memory (LSTM) and technical indicators [21].
Figure 4: BP discrete correlation statistical parameter analysis heatmap.

Figure 5: Total discrete correlation analysis.

Figure 6: Total discrete correlation histograms.
In 2017, Arfaoui and Rejeb published an article examining the effects and relationships of stock markets, oil, dollars, and gold based on global market documentation. They concluded that oil prices are significantly affected by stock markets, gold, and the dollar and that there are always indirect effects, which also confirms the presence of correlations in the global market [22]. In this paper, we compare this correlation feature and the relationships between stocks with the dollar, crude oil, gold, and major oil company stock indices, and we create data sets and compare the results of forecasts with real data.

This [23] article dives deep into the challenges posed by LSTM models in financial market forecasting. Smith et al. discuss the concept of “black box” models—LSTM architectures that lack transparency in their decision-making process. By examining the weights, biases, and internal mechanisms of LSTMs, the authors ascertain the difficulties in understanding and interpreting these models, especially in financial contexts. They emphasize the need for interpretability, highlighting potential risks and the consequences of solely relying on uninterpretable LSTM models.

Johnson et al. tackle the enigma of LSTM interpretability within the financial market domain. They discuss the challenges of interpreting LSTM models specifically for financial forecasting. The authors emphasize the importance of explaining predictions and making sense of underlying

<table>
<thead>
<tr>
<th>Table 4: Correlation amount of Total.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corr (-1, -0.5) Corr (-0.5, 0) Corr (0, 0.5) Corr (0.5, 1)</td>
</tr>
<tr>
<td>Count of corr (Total-WTI) 3.0 17.0 14.0 15.0</td>
</tr>
<tr>
<td>Percent of corr (Total-WTI) 6.0 34.0 30.0 30.0</td>
</tr>
<tr>
<td>Count of corr (Total-USD) 7.0 15.0 15.0 12.0</td>
</tr>
<tr>
<td>Percent of corr (Total-USD) 14.0 30.0 30.0 26.0</td>
</tr>
<tr>
<td>Count of corr (Total-gold) 10.0 15.0 15.0 9.0</td>
</tr>
<tr>
<td>Percent of corr (Total-gold) 22.0 30.0 30.0 18.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5: Total discrete correlation statistical parameters.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
</tr>
<tr>
<td>-----------------------------------------------------------</td>
</tr>
<tr>
<td>USD-Total</td>
</tr>
<tr>
<td>Gold-Total</td>
</tr>
<tr>
<td>WTI-Total</td>
</tr>
</tbody>
</table>

*Figure 7: Total discrete correlation statistical parameter analysis heatmap.*
Figure 8: Schlumberger discrete correlation analysis.

Figure 9: Schlumberger discrete correlation histograms.

Table 6: Correlation amount of Schlumberger.

<table>
<thead>
<tr>
<th></th>
<th>Corr (-1, -0.5)</th>
<th>Corr (-0.5, 0)</th>
<th>Corr (0, 0.5)</th>
<th>Corr (0.5, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count of corr (Schlumberger-WTI)</td>
<td>3.0</td>
<td>12.0</td>
<td>21.0</td>
<td>13.0</td>
</tr>
<tr>
<td>Percent of corr (Schlumberger-WTI)</td>
<td>6.0</td>
<td>24.0</td>
<td>42.0</td>
<td>28.0</td>
</tr>
<tr>
<td>Count of corr (Schlumberger-USD)</td>
<td>3.0</td>
<td>19.0</td>
<td>12.0</td>
<td>15.0</td>
</tr>
<tr>
<td>Percent of corr (Schlumberger-USD)</td>
<td>6.0</td>
<td>38.0</td>
<td>26.0</td>
<td>30.0</td>
</tr>
<tr>
<td>Count of corr (Schlumberger-Gold)</td>
<td>11.0</td>
<td>13.0</td>
<td>12.0</td>
<td>13.0</td>
</tr>
<tr>
<td>Percent of corr (Schlumberger-Gold)</td>
<td>22.0</td>
<td>26.0</td>
<td>26.0</td>
<td>26.0</td>
</tr>
</tbody>
</table>

Table 7: Schlumberger discrete correlation statistical parameters.

<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>Mean</th>
<th>Variance</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>USD-Schlumberger</td>
<td>0.043825</td>
<td>0.100825</td>
<td>0.224941</td>
<td>0.474280</td>
</tr>
<tr>
<td>Gold-Schlumberger</td>
<td>0.049048</td>
<td>0.015501</td>
<td>0.309095</td>
<td>0.555963</td>
</tr>
<tr>
<td>WTI-Schlumberger</td>
<td>0.299134</td>
<td>0.193702</td>
<td>0.175441</td>
<td>0.418857</td>
</tr>
</tbody>
</table>
patterns and features to build trust in financial decision-making. The article provides a holistic overview of conceptual and technical difficulties, articulating the limitations of LSTM models in this context [24]. Brown et al. delve into the intricate aspect of feature importance within LSTM models. They argue that understanding which features are most influential in driving predictions is vital for financial market forecasting. The article explores various techniques and approaches that attempt to unravel the complexity of feature transformations within LSTMs. By illuminating the challenges of achieving feature importance analysis, the authors call for further research to address the uninterpretability of LSTM models [25].}

LSTM is a dynamic network that is used to predict these markets due to the dynamic nature of energy financial markets. The transparency of this algorithm in detecting surprising based on stochastic distortion [26] is a fundamental challenge for investors. Considering the importance of oil company stocks for investors, checking the transparency of this algorithm can be significant. Considering that there are various causes for fluctuations and disruptions in the markets of oil companies, the investigation of the main

Figure 10: Schlumberger discrete correlation statistical parameter analysis heatmap.

Figure 11: Cairn Energy discrete correlation analysis.
economic factors in the comparative impact of these stocks can be of great help in modeling. The purpose of this article is to investigate the effect of these main economic factors that have correlation with the stocks of major oil companies and to investigate the effect on the interpretability for regression with LSTM. These articles collectively establish the significance of interpretability in LSTM models for financial market forecasting. They consistently highlight the limitations and potential risks of relying solely on uninterpretable LSTM models and advocate for the need to strike a balance between accuracy and interpretability.

This article is structured as follows: First, we investigate the complexity of the relationships between stocks through discrete correlation analysis. Subsequently, we design a standard LSTM network, optimize its hyperparameters, and then perform modeling. Finally, we review the analysis and modeling results and draw conclusions.

2. Oil Shares and Complexity of Stock Relations

The relation between oil companies’ stocks on the one hand and the prices of other assets like gold, crude oil, and the price of the US dollar, on the other hand, is intrinsically interesting for many economical and political reasons. Before calculating the complexity, it is worth having a short introduction for oil companies.

2.1. Oil Companies. In the global economy, large companies operating in the field of energy and oil are among the most authoritative companies in the global economy. Companies such as Total, BP, Cairn Energy, and Schlumberger are among the most authoritative companies. These companies are trading in energy and oil, and their investors and shareholders are among the largest investors and shareholders in the world, for whom the trend of changing the indicators of these companies in the stock market is very worthwhile. Moreover, the strong dependence of industry [27] in various fields on oil has made oil a major factor influencing the policies of each country. In the following, we introduce some of these companies which we analyze their shares in this paper [28].

(a) WTI: one of the known light sweet crude oil is West Texas Intermediate (WTI) which refers as one of the main global oil benchmarks. WTI has high quality, is easy to refine, and is sourced primarily from inland Texas

(b) BP.L: BP plc is a British multinational oil and gas company based in London, England. It is one of the world’s seven oil and gas companies

(c) FP.PA: TotalEnergies SE is a French multinational unified oil and gas company which is established in 1924 and is one of the seven major oil companies

(d) SLB.PA: Schlumberger Limited is an oil field service company whose members have more than 140 nationalities working in more than 120 countries. Schlumberger executive offices are located in Paris, Houston, London, and The Hague

2.2. Correlation between Shares and Economic Stocks. The crude oil index is the most important characteristic in determining the index of oil companies. The sale of crude oil in the world market is done in dollars. Commonly, the dollar can change the value of the indexes of oil companies. Moreover, the gold value can cause changes and fluctuations in currency indices. Their effects on each other can be obtained...
by calculating the correlation coefficient between their data [22]. In this article, in the value of the stock indices that we have examined, from 08/03/2009 to 07/01/2021, the results of the correlation calculation are shown in Table 1. This table is represented as an array where its principal diameter is equal to one, and the other component represents the correlation coefficient between the value of stock indices introduced in the global market.

To have a better presentation, we draw the heatmap diagram which is a type of information visualization (data visualization) in which the value of each cell of the matrix input is displayed in one color. The correlation coefficient obtained in the previous part is shown in the diagram of the thermal map (Figure 1). The lighter the color, the more direct the relationship and the higher the direct correlation, and the darker color, the higher the inverse relationship and correlation.

2.3. Discrete Correlation Analysis for Oil Shares. To conduct a detailed analysis of the relationships between WTI, USD, and gold with shares of Total, BP, Schlumberger, and Cairn Energy oil companies, we discretized the data and then examined the Pearson correlations between them. Between 2013/6/24 and 2021/4/7, we discretized around 2000 data points for each stock into 50 40-day data sets.

Table 9: Cairn Energy discrete correlation statistical parameters.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>Median</th>
<th>Mean</th>
<th>Variance</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>USD-Cairn Energy</td>
<td>0.248828</td>
<td>0.198628</td>
<td>0.217601</td>
<td>0.46647</td>
</tr>
<tr>
<td>Gold-Cairn Energy</td>
<td>0.016285</td>
<td>-0.008477</td>
<td>0.290874</td>
<td>0.539328</td>
</tr>
<tr>
<td>WTI-Cairn Energy</td>
<td>0.321247</td>
<td>0.261865</td>
<td>0.214425</td>
<td>0.463061</td>
</tr>
</tbody>
</table>

Figure 13: Cairn Energy discrete correlation statistical parameter analysis heatmap.

Figure 2 presents the discrete correlations of the BP company data set with WTI, USD, and gold data sets. The correlation histograms of the BP stock index with the WTI, USD, and gold stock indices are displayed in Figure 3. The histograms are divided into four columns based on the results of the correlation analysis, as shown in Table 2.

The statistical parameters, including the average, median, variance, and standard deviation, for the correlations between the BP stock index and the WTI, USD, and gold are presented in Table 3 and illustrated in Figure 4.

Figure 14: LSTM architecture.
The discrete correlations between the Total company data set and the WTI, USD, and gold data sets are displayed in Figure 5.

The correlation histograms of the Total stock index with the WTI, USD, and gold stock indices are presented in Figure 6. The histograms are divided into four columns based on the results of the correlation analysis, as shown in Table 4.

The statistical parameters, such as the average, median, variance, and standard deviation, for the correlations between the Total stock index and the WTI, USD, and gold are displayed in Table 5 and Figure 7.

The discrete correlations between the Schlumberger company data set and the WTI, USD, and gold data sets are displayed in Figure 9. The histograms are categorized into four columns based on the results of the correlation analysis, as presented in Table 6.

The statistical parameters, namely, the average, median, variance, and standard deviation, for the correlations of the Schlumberger stock index with the WTI, USD, and gold are exhibited in Table 7 and Figure 10.
The discrete correlations between the Cairn Energy company data set and the WTI, USD, and gold data sets are presented in Figure 11.

The correlation histograms indicating the relationship between the Cairn Energy stock index and the WTI, USD, and gold stock indices are illustrated in Figure 12. The histograms are arranged into four columns based on the results of the correlation analysis presented in Table 8.

The statistical parameters, namely, the average, median, variance, and standard deviation, for the correlations of the Cairn Energy stock index with the WTI, USD, and gold are displayed in Table 9 and Figure 13.

3. LSTM Architecture for Oil Prices

Long short-term memory (LSTM) is an architecture for RNNs to avoid the gradient vanishing problem in RNNs in memory building. LSTM architecture has three gates, forget gate, input gate, and output gate, as shown in Figure 14. LSTM can read, write, and delete information from its memory. The LSTM structure has the capability to identify which cells are stimulated and compressed based on the previous state, available memory, and current input.

Forget gate determines what information data we are going to throw away from the cell state. In this gate,
information is deleted or stored based on the output of the sigmoid function. This gate cell is formulated as follows:

\[ f_t : \text{forget gate}, \sigma : \text{sigmoid}, W : \text{weight}, h_{t-1} : \text{hidden state}, \]

\[ x_t : \text{input}, b : \text{bias}, i_t : \text{input gate}, c_t : \text{cell state}, o_t : \text{output gate}, \]

\[ f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f), \]

\[ i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i), \]

\[ c \sim t = \tanh(W_c \times [h_{t-1}, x_t] + b_c). \]

Cell state drops the information about past subjects and adds new information as follows:

\[ c_t = f_t \times c_{t-1} + c \sim t 	imes i_t, \]

Output gate decides what we are going to output and reading from memory which is formulated as

\[ o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o), \]

\[ h_t = o_t \times \tanh(c_t). \]

3.1. Data and Features. In this paper, we use four major (open, high, low, and closing prices) oil stock pairings FP.PA (Total), CNE.L (Cairn Energy), BP.L (BP), and SLB.PA (Schlumberger) with their daily interval data and use for comparison stock pairings WTI, gold, and dollar with their daily interval data from 2009/8/14 to 2020/7/19 for training and data from 2020/7/20 to 2021/7/15 for testing. Data was extracted from Alpha Vantage API [29].

This section provides a summary of the findings obtained through experiments conducted using our proposed methodology. The neural network model utilized was the LSTM network, implemented using the Python Neural Networks library and Keras, running within the TensorFlow 2.0 Python development environment to train the data.

3.2. Hyperparameter Selection and Network Topology. The fundamental determination of hyperparameters for the neural network has a very important effect on the modeling results. Appropriate number of time lags is one of the most important hyperparameters that has a very important effect on the results in modeling with recurrent neural networks. Autocorrelation is a statistical method used to measure the relationship between a time series data point and its lagged values. It indicates whether there is a correlation between data points that are a certain distance apart in time. Autocorrelation can be used in the context of long short-term memory (LSTM) models to find the appropriate number of time lags to use for prediction. In LSTM models, the previous time steps are used as features for predicting the next step. One of the key challenges is to determine the appropriate number of time lags to use for prediction. If too few lags are used, important features may be missed, leading to poor performance. If too many lags are used, the performance may also suffer due to overfitting. Autocorrelation can be used to address this challenge by measuring the correlation between the time series data point and various lagged values. By plotting the autocorrelation function, we can determine the optimal number of lags to use for prediction. In general, the autocorrelation function declines as the lag increases, with the most significant lags occurring at the beginning of the time series. Basic determination of this value prevents. The results of the autocorrelations of the close price of each stock from 2013/6/23 to 2017/6/12 with each of the dates in Figures 15–18 are as follows.

Due to the fact that in Section 2.3, we have discretized the data as 40-day packages and then performed the autocorrelation analysis, and according to Figures 15–18, we can see that the data has a strong direct relationship with the future 40 days. According to the obtained results, we set the value of the appropriate number of time lags equal to 40, so that we can check the interpretability of LSTM for predicting the data of oil stocks based on the results obtained from the modeling and the results of Section 2.3.

Adam’s [30] optimization function is used with 0.0005 learning rate, 10 percent data is validation data, epoch value is equal to 50, and batch size is equal to 25. For all simulations with LSTM, these values are the same.

<table>
<thead>
<tr>
<th>Layers</th>
<th>Neurons</th>
<th>Computational parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inputs</td>
<td>(None, 50, 6)</td>
<td>0</td>
</tr>
<tr>
<td>LSTM</td>
<td>(None, 50)</td>
<td>11400</td>
</tr>
<tr>
<td>Dropout</td>
<td>(None, 50)</td>
<td>0</td>
</tr>
<tr>
<td>Dense</td>
<td>(None, 64)</td>
<td>3264</td>
</tr>
<tr>
<td>Activation (sigmoid)</td>
<td>(None, 64)</td>
<td>0</td>
</tr>
<tr>
<td>Dense</td>
<td>(None, 1)</td>
<td>65</td>
</tr>
<tr>
<td>Activation (linear)</td>
<td>(None, 1)</td>
<td>0</td>
</tr>
<tr>
<td>Total parameters</td>
<td>14729</td>
<td></td>
</tr>
<tr>
<td>Trainable parameters</td>
<td>14729</td>
<td></td>
</tr>
<tr>
<td>Nontrainable parameters</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

| Table 10: The network topology which includes each layer along with neurons and computational parameters. |
For modeling, we have used a simple standard LSTM topology, which is displayed in Table 10.

3.3. Evaluation Measures. To measure the badness of the model, people use mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), as defined below. For measuring the average of the squares of the errors, the MSE measure can be used which is defined in the following formula:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\text{true} - \text{prediction})^2.$$  \hspace{1cm} (8)

To measure the average root of the squares of the errors, we use RMSE which is defined as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\text{true} - \text{prediction})^2},$$ \hspace{1cm} (9)

which describes how spread out these residuals are by the standard deviation of prediction errors. To measure the average of absolute error, we use MAE which calculates errors between pair observations. It is defined as

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\text{true} - \text{prediction}|.$$ \hspace{1cm} (10)

Finally, to measure the size of the error in percentage terms, we use MAPE which is calculated for the mean of the absolute percentage errors of prediction. It is defined in the following formula:

$$MAPE = \frac{\sum_{i=1}^{n} |\text{true} - \text{prediction}|/\text{true}}{n} \times 100.$$ \hspace{1cm} (11)

Now, we are in the position where we can apply these evaluation measures as error metrics in our prediction.

4. Discussion

In this study, we have demonstrated the potential of deep learning as a tool for forecasting stock prices in the oil industry, either as a replacement or a complement to traditional forecasting methods.

We began by investigating the correlation coefficients between the stocks of various oil companies, as well as the USD, WTI crude oil, and gold indices. We then discretized the stock data and performed the discrete correlation analysis.

In Section 2.2, we observed that there is a correlation between the stocks of the oil companies in question and the gold, dollar, and crude oil indices. Based on the findings presented in Table 1, we obtained the following results.

Overall, our analysis suggests that there are significant correlations and relationships among the stock indices of the oil companies, as well as with the USD, WTI, and gold.
indices. Specifically, BP and Cairn Energy stock indices show strong direct relationships with each other, while Total appears to be more inversely correlated with gold. Schlumberger has weaker correlations with the other indices but exhibits a strong direct relationship with Total and weak direct relationships with BP and Cairn Energy. Our findings suggest that deep learning algorithms may be particularly useful in predicting stock prices within the oil industry given these complex interrelationships and correlations.

Secondly, in order to verify the accuracy and stability of the results obtained from Section 2.2, and prior to modeling and checking the results of the LSTM model, we will examine the results of Section 2.3 which presents a discrete analysis of the correlations.

(i) According to Figure 2, BP shares display fluctuations in their correlations with the index of gold, USD, and WTI stocks throughout abnormal correlation intervals, which are not confined to specific correlation intervals. In Figure 3 and Table 2, around 30 percent of the correlations exhibit weak direct relationships with WTI, which contradicts the results of Table 1, whereas approximately 40 percent show a strong direct association. About 36 percent of the correlations indicate an inverse relationship with the USD, which is contrary to Table 1, but around 64 percent of the correlations demonstrate a strong direct relationship with the USD. Approximately 46 percent of the correlations suggest an inverse relationship with gold, while about 54 percent of the correlations demonstrate a direct relationship with gold, which contradicts the findings of Table 1. The average correlation of BP shares with the WTI stock index is 0.27, indicating a weak direct association, which is consistent with the results of Table 1. However, this average has a standard deviation of 0.45, representing high dispersion. The median value of 0.32 is noteworthy, acting as the intermediary between the outcome obtained from Table 1 and the average of the correlations, thereby indicating that the correlations with an inverse relationship carry more weight. This may suggest the influence of indirect factors such as political events. The average correlation of BP’s shares with the gold stock index almost indicates a lack of relationship and negative correlation, which contradicts the findings of Table 1. The median value is smaller than the average, indicating that the correlations outweigh the inverse relationship between the two stocks. The average correlation of BP shares with the dollar suggests a weak direct relationship, contradicting the results of Table 1. Moreover, the average of these correlations highlights the weight of the correlations with a direct relationship.

(ii) Similar to BP’s correlations shown in Figure 2, shares of Total, Schlumberger, and CNE companies display high fluctuations in the correlation ranges based on Figures 5, 8, and 11. According to Figure 6 and Table 4, roughly 60 percent of the correlations demonstrate a direct relationship between

| Table 13: Experimental result for Schlumberger shares. |
|----------------|----------------|----------------|----------------|
|                | MSE        | RMSE       | MAE       | MAPE      |
| WTI            | 0.00964    | 0.09818    | 0.00621   | 0.02830   |
| Main data set  | 0.00102    | 0.03203    | 0.00202   | 0.02569   |
| Dollar         | 0.00112    | 0.03361    | 0.03361   | 0.025234  |
| Gold           | 0.00108    | 0.03296    | 0.00208   | 0.025690  |

Figure 20: Cairn Energy real prices vs. predicted prices.
Total shares and the stock index and WTI, which contradicts the results of Table 1. About 56 percent of the correlations exhibit a direct relationship between Total shares and the USD, which contradicts the findings of Table 1, but in terms of gold, it confirms the results of Table 1. The median correlations of Total shares with the USD and gold are approximately zero, which aligns with the results of Table 1 in the case of USD shares, but in the case of gold, it contradicts the outcomes of Table 1. This is an essential point. Based on the standard deviation value, there is a high dispersion among the correlation values. The average and median correlations of this stock with WTI are consistent with the results of Table 1.

(iii) Based on Figure 9 and Table 6, approximately 70 percent of the correlations imply a direct relationship between Schlumberger company shares and WTI shares, matching the results of Table 1. About 56 percent of the correlations suggest a direct relationship between Schlumberger shares and the USD, while roughly 52 percent of the correlations contradict the outcomes of Table 1 regarding their relationship with gold stocks. The mean correlations in Table 9 for Schlumberger stocks with WTI show a weak direct relationship. The mean and average of these shares with gold stocks are both zero, which contradicts the results of Table 1. Furthermore, concerning the mean and median of these shares with the USD, it indicates a greater weight of correlations that demonstrate a direct relationship.

(iv) Based on Figure 12 and Table 8, around 72 percent of the correlations indicate a direct relationship between Cairn Energy shares and WTI, consistent with the outcomes of Table 1. About 70 percent of the correlations demonstrate a direct relationship between these stocks and the USD, contradicting the results of Table 1, which suggested an absence of a relationship. Roughly 52 percent of the correlations suggest a direct relationship between Cairn Energy shares and gold shares, contradicting the outcomes of Table 1. The median and average correlations of Cairn Energy shares with the USD imply a weak direct relationship, differing from the results of Table 1. The mean and average of these shares with gold show a contradiction with the outcome of Table 1 by indicating a lack of a relationship. Also, the mean and average correlations of these stocks with WTI demonstrate a direct relationship, aligning with the results of Table 1.

Based on the results obtained from the discrete analysis of correlations, we can observe that the correlation outcomes are not always consistent and do not necessarily increase within a specific correlation interval. Instead, they fluctuate across different correlation intervals. In the oil markets, various factors like political decisions, shifts in politics, and conflicts impact the price, which changes over time. Third, to forecast stock prices of different companies, we employed recurrent neural networks with LSTM architecture, given that these stocks exhibit changes in time series. We conducted empirical experiments and...
performed on the stock index data set to evaluate the predictive performance in terms of several common error metrics, namely, MSE, MAE, RMSE, and MAPE. Here are the summary results based on these metrics:

(i) In Table 11, it was observed that including the WTI (crude oil), gold, and dollar indices did not enhance the model’s prediction and did not lessen its errors (as seen in the error metrics). Furthermore, the same four primary features had the lowest cost function. The empirical analysis results for BP are illustrated in Figure 19.

(ii) In Table 12, it was observed that incorporating the crude oil and dollar indices did not decrease the cost function. However, during the learning process, when the machine analyzed the gold index, it significantly improved the measurement and resulted in reduced costs, thereby helping us better our modeling, as depicted in the Cairn Energy diagrams in Figure 20.

(iii) Table 13 indicates that the WTI index resulted in an increase in the cost function and the addition of the gold and dollar indices made no significant impact on the learning process. The diagrams for the Schlumberger company are illustrated in Figure 21.

(iv) Table 14 regarding Total shares reveals that none of the WTI, dollar, or gold indicators enhanced the learning process, as evidenced by Figure 22.

5. Conclusion
In general, the presented results show that the range of errors for each stock varies among different models, though these differences are consistent across all stocks. Based on our analysis, it can be concluded that the use of correlated features did not improve the performance of the LSTM model for the oil market.

Our findings revealed that although LSTM models could achieve high accuracy in predicting stock prices, their interpretability was limited. Specifically, their internal states and weight parameters were difficult to interpret and could not provide clear insights into the underlying factors driving stock price movements. Overall, caution should be exercised when relying solely on LSTM models for stock price prediction, as their lack of interpretability may make it difficult to fully understand the factors driving stock price movements. Further research is needed to improve the interpretability of LSTM models for predicting stock prices.

Data Availability
The stock markets data that support the findings of this study are publicly available at the Alpha Vantage API and can be found at https://www.alphavantage.co.

Disclosure
An earlier version of this study has been presented as an arXiv preprint according to the following link: https://arxiv.org/abs/2201.00350.

Conflicts of Interest
The authors declare that they have no conflicts of interest.
References


[28] An earlier version of it has been presented as a arxiv preprint according to the following link: https://arxiv.org/abs/2201.00350.

[29] https://www.alphavantage.co/.


