



# A Deep Learning-Based Method for Forecasting Gold Price with Respect to Pandemics

Mahtab Mohtasham Khani<sup>1</sup> · Sahand Vahidnia<sup>2</sup> · Alireza Abbasi<sup>2</sup>

Received: 6 March 2021 / Accepted: 22 May 2021 / Published online: 12 June 2021  
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2021

## Abstract

The spread of COVID-19 has had a devastating impact on the world economy, international trade relations, and globalization. As this pandemic advances and new potential pandemics are on the horizon, a precise analysis of recent fluctuations of trade becomes necessary for international decisions and controlling the world in a similar crisis. The COVID-19 pandemic made a new pattern of trade in the world and affected how businesses work and trade with each other. It means that every potential pandemic or any unprecedented event in the world can change the market rules. This research develops a novel model to have a proper estimation of the stock market values with respect to the COVID-19 dataset using long short-term memory networks (LSTM). The goal of this study is to establish a model that can predict near future regarding the variable set of features. The nature of the features in each pandemic is completely different; therefore, prediction results for a pandemic by a specific model cannot be applied to other pandemics. Hence, recognizing and extracting the features which affect the pandemic is pivotal. In this study, we develop a framework that provides a better understanding of the features and feature selection process. Although the global impacts of COVID-19 are complicated, we are trying to show how additional features like COVID-19 cases can help to forecast in a real-world scenario, rather than relying solely on the history of tickers, which is used conventionally for prediction. This study is based on a preliminary analysis of features such as COVID-19 cases and other market tickers for enhancing forecasting models' performance against fluctuations in the market. Our predictors are based on the market value data and COVID-19 pandemic daily time-series data (i.e. the number of new cases). In this study, we selected Gold price as a base for our forecasting task which can be replaced by any other markets. We have applied Convolutional Neural Networks (CNN) LSTM, vector sequence output LSTM, Bidirectional LSTM, and encoder–decoder LSTM on the dataset. The results of the vector sequence output LSTM achieved an MSE of  $6.0e - 4$ ,  $8.0e - 4$ , and  $2.0e - 3$  on the validation set respectfully for 1 day, 2 days, and 30 days predictions in advance which are outperforming other proposed method in the literature.

**Keywords** COVID-19 · Economy · Deep learning · Time series

---

Mahtab Mohtasham Khani and Sahand Vahidnia contributed equally to this work.

✉ Mahtab Mohtasham Khani  
khani17@itu.edu.tr

Sahand Vahidnia  
s.vahidnia@unsw.edu.au

Alireza Abbasi  
a.abbasi@unsw.edu.au

<sup>1</sup> Istanbul Technical University, Istanbul, Turkey

<sup>2</sup> School of Engineering and IT, University of New South Wales, Canberra, Australia

## Introduction

The novel Coronavirus (SARS-CoV-2) disease identified as COVID-19 has been initiated in Wuhan, China, and had a quick global spread. This disease first led to big human health problems in all countries and many cascading financial problems resulted from social distancing and travel restrictions. Governments proceeded with specific procedures to control the speed of spreading the disease such as canceling flights, locking down their national and state borders, preventing most exports and imports, and shutting down some businesses which lead to a huge economic shock to the world. In this situation, trading (including business travel) which is essential and crucial part of today's life can

facilitate spreading the virus in many ways [39]. On the other hand, without trading economy will collapse and the effects of such economic devastation will remain for a long time. The pandemic has damaged the global economy by creating problems in the world supply chain [4]. For example, equity markets in the EU and US dropped by as much as 30% [19]. Generally, there was a dramatic shock to global trading activities during the COVID-19 pandemic such as increasing the demand for essential goods such as medical products and food, as well as a sharp decrease in the prices of some products like oil, or the collapse of some airlines declaring bankruptcy with hopes to resume operations after the end of the outbreak [34]. According to a recent review, [5] trade implications of the COVID-19 pandemic that China and the rest of the world follow a new pattern which leads some economies to win and some to lose.

Understanding the economic implications and impact of the pandemic on the economy is a way to minimize the impact and support the economic decision-makers on their trade choices. Conventional methods such as tacking economic factors of companies and the markets, like the statement of cash flow and balance sheet, or technical analysis to understand the market trend based on the volume and price, have been in place for many years and predictions have been made for decades using these methods. With the emergence and advancements in machine learning and deep learning, new methods have emerged to better understand the patterns and make more accurate predictions. These data-driven methods require access to large datasets, which have been available during recent years, and has paved the ground for the research in this direction [29]. The nature of the features of each pandemic is different, hence, one result cannot be applied to other cases. Hence, in this study, we develop a framework facilitating the overall procedure of time series analysis and predictive robust models in various cases.

In this work, we use LSTM for time series analysis in order to better understand the patterns in the sequences and improve the prediction performance. LSTMs are basically artificial recurrent neural networks (RNNs) that can process the entire sequence of data. LSTM was first investigated by Hochreiter [22] to deal with the problem of vanishing or exploding the gradients of RNNs. LSTM Networks are a solution for extracting the pattern of the input data-set which spans over a sequence of time which can be responsible for COVID-19 as a nonlinear feature over a sequence of time.

The main objective of this research is to analyze and explore the implications of COVID-19 on the economy considering short-term implications and patterns of pandemics on the world trade and market sector, by studying the case of Gold price on the stock market, based on the COVID-19 time series and 11 sectors of the market during the recent years and months. We study the correlation between different sectors and COVID-19 new cases for distinguishing the

effective features in market prediction. This paper contributes to market prediction accuracy by developing an accurate forecasting model on the stock market values with respect to the COVID-19 dataset, using LSTM. Our forecasts are based on the market value data extracted from Yahoo! Finance and COVID-19 pandemic daily time-series data (i.e. the number of new cases). We applied our method for 1, 2, and 30 days in advance predictions, concerning COVID-19 data. The performance of our proposed method has been examined and validated using Mean Square Error (MSE) on the validation dataset, by comparing it to the most recent developments in the field.

## Literature Review

### Economy and Pandemics

Human beings have faced many pandemics throughout the history such as SARS Epidemic [6], Middle East respiratory syndrome (MERS) [11], West African Ebola epidemic [24], and H1N1 Swine Flu pandemic [17]. The pandemics are known to have significant effects on the economy, some of which may persist for decades after the pandemic [23]. There are many studies trying to understand the pandemic and its effects on trade. Some of these studies are purely economic and are not relying on machine learning for predictions. Machine learning is a branch of analytic predictions in which has a lower reliance on human expert supervision for understanding and analyzing the reasons and causes [31]. Machine learning algorithms do not count out the need for human expert analyzes, meanwhile they are one of the best solutions for extracting the underlying patterns [33]. The studies in this section analyze the economical, social, and even political aspects of the markets to understand and explain the market. These studies provide insights into the economic point of view of the pandemics and help to better understand the variables. In a review paper by Barua [5], likely trade implications of the COVID-19 pandemic have been found to provide a better understanding of the impacts of COVID-19 on coming days. This study investigates a standard trade framework and then proposes a theoretical mapping that depicts the progress of trade implications. The study then reviews some real-life evidence for testing his map and concluded with clues for controlling pandemic situations. The study divides the effects of a pandemic into different short and long-term stages, and it suggests reducing reliance on specific countries to reduce the economic impact. In another survey about macroeconomic implications of COVID-19, the effects of negative supply shocks on the shortage of demands are discussed [21]. They argue the future of economic shocks of COVID-19 as well as presenting the theory of Keynesian supply shock, which is about

larger changes in aggregate demand than the shocks, in comparison to standard supply shocks which concluded in a good understanding about availability or lack of some goods in the pandemic situation. The Center of Economic Policy Research (CERF), which is a network of economists mostly from European universities, has asserted that “The virus is likely to be as ‘contagious’ economically as it is medically” [3]. In the past recessions, global trade has slowed faster than global growth. This study [3] also discussed the demand shock on trade as well as the supply-sides of this virus. This study is an analysis across the COVID-19 and trade which emphasized the danger of permanent collapse of the trading system. Also, the optimal lock-down policies for minimizing the output costs of the lock-down was researched by Alvarez et al. [1]. The study uses a linear economy model to formalize the planner’s dynamic control problem and identifies the features to measure the optimal intensity, shape, and time that this lockdown will last. Gormsen et al. [19] study the impacts of Coronavirus on stock prices and the growth expectation. They show how effective are news, events, and the data on dividend futures on forecasting and analyzing the drop and growth of the economy over time. They have studied the expected growth in the US and EU S&P 500 and Euro Stoxx 50 and estimated that Expected growth over the next year is down by 2.6% both in the US and the EU.

## Stock Market Prediction

The studies on market prediction and the surrounding literature can be generally divided into two categories: (1) the studies which predict the market using social networks analysis; and (2) the studies of time series analysis. The former category comprises methods like the diffusion models which are useful for minding lots of latent information and market predicting. Li et al. [26] investigate the research methods and techniques on diffusion models that could facilitate the prediction of social network influences or predicting the trade future leads. In another study, [2] proposed a system for detecting influenza epidemics using Twitter data. They extracted the tweets that mention actual influenza patients using the support vector machine (SVM)-based classifier for separating the negative and positive tweets.

The second category, which is more in terms of quantity, mainly focus on methodologies to predict various financial time series. LSTM networks are always one of the best solutions for a sequence of data and single data points [16]. Hence, Fischer and Krauss propose and use the use of LSTMs in financial time series. The feature space in their study is the standardized daily stock market return of specific markets and days. They provide a comparison of LSTM, random forest, a standard deep net, and logistic regression and conclude that LSTM outperforms all other methods. In a more recent study, Livieris et al. [27] implement a CNN–LSTM model to predict

gold price time-series and its fluctuations. The paper asserts that the combination of LSTM layers with some other convolutional layers increases the forecasting performance. They do not consider other variables and the pandemic, which could be a limitation of this study.

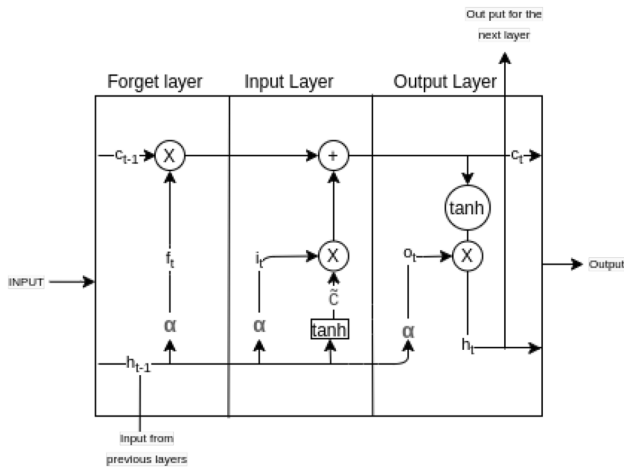
In a review paper on time series classification [15], it has been emphasized that LSTM networks are the most efficient way for time series classification. Among the usages of LSTMs, time series forecasting problems are considered to reduce the complication of study in comparison to the other traditional ones [18]. This study [18] suggest using traditional methods in non-complicated tasks and to use LSTMs as substitute methodology. They conclude that LSTM has a great promise and opportunity for applying to the problem of time series forecasting. In another study [7], LSTM is used in order to predict China’s stock market return and has reported the power of LSTM in stock market prediction. Besides, Persio and Honchar [12] investigated a similar problem by applying artificial neural network architectures to predict trend movement on the stock market based on past returns. According to their results, in a comparison of multi-layer perceptron (MLP), convolutional neural networks, LSTMs for feature extractions, and the combination of wavelet transform and CNN, the latter one has achieved the best results.

In a study surveying time series forecasting of the transmission of COVID-19 in Canada [8], LSTMs are used to predict the possible stopping time of COVID-19 in Canada and around the world. They propose a multi-step LSTM method and predicted 2, 4, 6, 8, 10, 12, and 14th day for two successive days. The study proposes the use of bidirectional LSTM for the forecasting model. LSTM is also used as a data-driven estimation method in India for predicting the spread of COVID-19 and the effects of preventing protocols [36]. In this study, they predicted the number of COVID-19 cases in India 30 days ahead.

The prediction of the number of cases in a pandemic has been demonstrated to be somehow predictable based on the historical data. However, the same is not true for market data, as it has far more fluctuations. Additionally, as the studies suggest, the world pandemic situations would make the market movements more chaotic and unpredictable. As a result, we aim to adopt LSTM, as one of the best forecasting methods, to develop a predictive model for stock market prediction using LSTM-series by considering the COVID-19 data-set for market prediction for the first time. This will establish a guideline for future works of this category.

## Methodology

Our method comprises four steps. First, our data in which the overall market sectors and tickers are analyzed, compiled, and extracted. Then the relevant features are processed



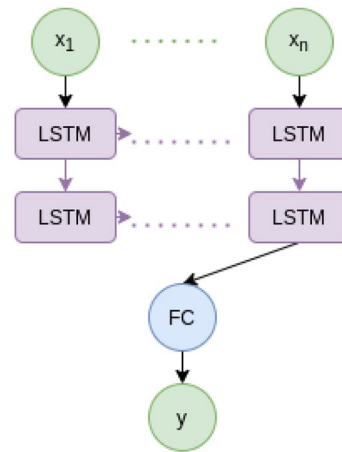
**Fig. 1** LSTM internal architecture which consist of:  $h_{t-1}$  for the output from last LSTM unit,  $c_{t-1}$  for memory of the last LSTM unit and  $c'_t$  for candidates of cell state at time  $t$  Two  $\alpha$ s to represent non-linearity as a sigmoid layers,  $\tan h$ : represent  $\tan h$  layer Vector operations:  $\times$  expresses scaling the information, and  $+$  expresses adding information

and extracted. Later, LSTM models are trained and tuned on the aforementioned features to create forecasting models with a range of hyperparameters, steps, and historical periods. Finally, the method and the results are evaluated and validated to be used for future forecasting tasks.

**LSTM Models**

Originally, LSTM was investigated in 1991 by Hochreiter [22] to deal with the problem of vanishing or exploding gradients, which are very common in RNNs. Generally, RNNs are good at handling the sequence dependencies. LSTMs are a type of RNNs which are better suited for larger architectures and more capable of extracting patterns from large sequences of datasets. LSTMs are also known to have a better response for non-linearity [28]. The COVID-19 and market time-series data show a non-linear behavior that motivates the application of LSTM in this research.

As illustrated in Fig. 1, each LSTM unit consists of three gates: a forget gate which remembers the values over arbitrary time intervals, and two other gates to regulate information into and out of the cell, which is called input and output gates. Each LSTM cell maintains a cell state vector and at each time step, the next LSTM can choose to read from it, write to it or reset the cell. These gates give the ability to control the process of memorizing to LSTM, and therefore, it can avoid long-term dependency [25] which is a key factor of solving problems related to COVID-19 with a short historical dataset. The parameters of the gates are expressed in Eq. 1, where  $\alpha$  expresses the sigmoid function,  $w_x$  expresses the weights for the neurons of gate  $x$ ,  $h_{t-1}$  expresses the



**Fig. 2** LSTM network architecture.  $X$  is input,  $Y$  is output, with stacked LSTM cells

output from last LSTM unit,  $x_t$  expresses current input, and  $b_x$  expresses the biases for the gate  $x$ .

$$\begin{aligned}
 i_t &= \alpha(w_i[h_{t-1}, x_t] + b_i) && \text{for input gate} \\
 f_t &= \alpha(w_f[h_{t-1}, x_t] + b_f) && \text{for forget gate.} \\
 o_t &= \alpha(w_o[h_{t-1}, x_t] + b_o) && \text{for output gate}
 \end{aligned}
 \tag{1}$$

In this study, we conduct both single-step and multi-step analysis, predict gold prices at least a day ahead. To have a better sensitivity analysis, we also employ both multi-variate and uni-variate approaches to demonstrate the effectiveness of other variables. There are numerous methods and approaches involving LSTMs to tackle time series forecasting analysis which depends on the dataset and the task. In the following sections, these approaches will be discussed and established to pave the ground for comparison and model selection.

**Single-Step LSTM**

To predict the gold price a day in advance, single-step feed-forward stacked LSTM networks are used. As mentioned earlier, a series of hyper-parameters and input variables are tested to better understand the effect of feature space on the prediction error. The overall structure of the LSTM networks employed in this section is depicted in Fig. 2. Generally, the higher number of LSTM cells within a layer would allow us to have a longer memory. This means that for longer historic days, we can grow the width of the LSTM network and vice versa to have an optimal fit. The activation layers in all architectures are Rectified Linear Unit (ReLU) (Eq. 2) and the optimizer of choice for LSTM networks are usually ADAM, as opposed to Stochastic Gradient Descent (SGD) which are usually known for robust optimization. After tuning

numerous hyper-parameters, the selection of top models has been presented at Table 3.

$$\text{ReLU}(x) = \max(0.0, x). \quad (2)$$

### Multi-step LSTM

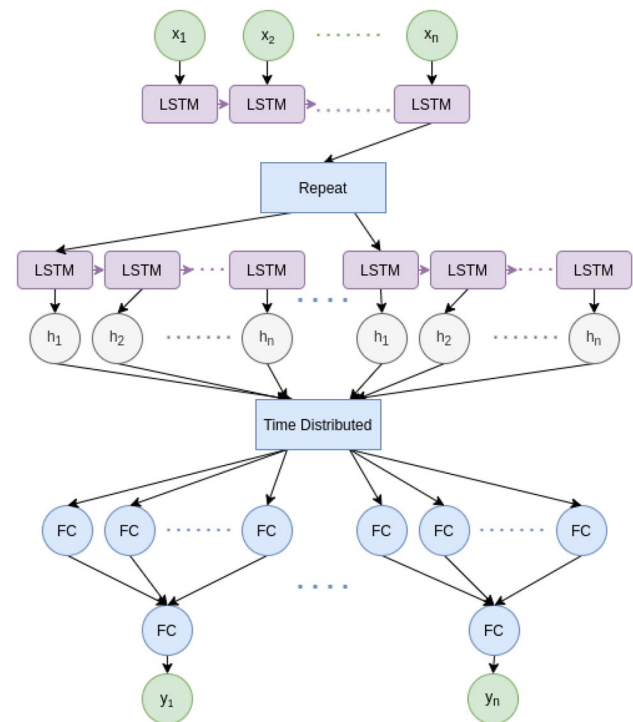
Forecasting for more than one day or step makes it a multi-step forecasting problem. The methods for addressing the multi-step forecasting problem can be categorized under the vector-output sequence prediction approach and encoder–decoder approach. The encoder–decoder approach in addition to the vector output sequence prediction methods is the main focus of this study. To validate the results and have a comparison to other suggested methods in the literature, the results will also be compared to Bidirectional LSTM and CNN–LSTM.

The output for a multi-step forecasting LSTM can be a vector sequence. This can be achieved by simply adding  $n$ -output neurons to a simple vanilla LSTM network. Hence, the overall architecture of the multi-step vector output approach is almost identical to Fig. 2.

Encoder–decoder as it is explained by its name, predicts by encoding the inputs and then decoding the output. This approach is used for multi-step time series forecasting [10]. The model was designed to solve sequence to sequence problems like natural language processing [35], text translation, and answering textual questions. Encoder–decoder is also known to yield good results for image classification, image to text, movement classification, and describing images by text tasks [38]. The encoder–decoder approach in LSTM can have many different implementations, suiting different workloads. In general, the overall architecture of the experimented encoder–decoder models in this study can be illustrated as Fig. 3. As illustrated, the model takes and encodes the inputs, then repeats the final state of the encoding layer for all time steps. The decoder comprises at least an LSTM layer and time distributed dense layer to provide the output of the desired shape and structure.

CNN–LSTMs are the combination of CNNs and LSTMs [29] that are common to be utilized in computer vision problems [13]. CNN–LSTMs are also encoder–decoder-based approaches, where the encoding happens in the CNN section. They have been utilized in various tasks in the literature like caption generation [38] and prediction of gold prices [27]. Thus, CNN–LSTM models have also been experimented in this study, and the results are discussed in the following section.

Bidirectional LSTM is inspired by Bidirectional Recurrent Neural Networks [32]. The network learns the sequences both from forward and backward and then concatenates all



**Fig. 3** Encoder–decoder LSTM network architecture.  $X$  is input,  $Y$  is output, and  $h$  is hidden state of the LSTM cells

the data for prediction. Bidirectional LSTM networks can be more beneficial than unidirectional ones in terms of results [20].

Similar to the single-step LSTM approaches, ReLU activation and ADAM optimizer have been used in all architectures. In the following section, we discuss the evaluation metrics, the data required for this analysis and forecasting models, and the result of the analysis.

### Validation and Evaluation

For validation purposes, we separate a 90 recent days period of our dataset as a validation set and validated the model on the aforementioned period. In order to accommodate the randomness factor, the tests have been repeated a number of times and only the top-performing seeds and training instances have been referred to.

There are various metrics to calculate the loss in regression and prediction tasks. Root Mean Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Error (MAE), Mean Squared Logarithmic Error (MSLE) are all techniques to find the difference between the predicted value and the actual value. However, in this study the models are optimized on the MSE values. Hence, the comparison results will favor this metric. This affects both best-model and training checkpoint selection, and the comparison of different LSTM methodologies and models based on validation

errors. However, to better understand the results, we take advantage of all of them which are respectfully defined by Eq. 3, where  $n$  is for the number of predictions and  $y_i$  is the ground truth of  $i$  instance and the  $\hat{y}_i$  is the predicted results of them.

$$\begin{aligned}
 \text{RMSE} &= \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \\
 \text{MSE} &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\
 \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \\
 \text{MSLE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n \log(y_i - \hat{y}_i)^2}
 \end{aligned} \tag{3}$$

## The Data

There are thousands of publicly traded stocks around the world and every one of them can be categorized as a member of the 11 major market sectors [14], including Financial, Utilities, Consumer Discretionary, Consumer Staples, Energy, Healthcare, Industrial, Technology, Telecom, Materials, and Real Estate. These 11 sectors are responding to the key areas of the economy and all the companies in each sector share the same broad focus. The list of corresponding tickers of the 11 market sectors has been collected from ETFdb.<sup>1</sup> As the list of the market sectors is long and the market values and volume of the tickers vary drastically, only the top 10 tickers with the highest values for each sector have been selected. The market data in this study is gathered for this selected top 10 sector tickers for the past five years from 07/2015 to 07/2020 from Yahoo! Finance, which also covers the recent global COVID-19 pandemic period. On the other hand, we also need COVID-19 pandemic data (including newly infected and total infections) to incorporate in our model and it has been collected from the ‘‘JHU CSSE COVID-19 Data’’ daily time series. As the market data has been collected from 30-07-2015 to 30-07-2020 and COVID-19 data starts from 22-01-2020, the COVID-19 values prior to 22-01-2020 have been set to zero to match the dimensionality of market dates.

## Understanding the Feature Space

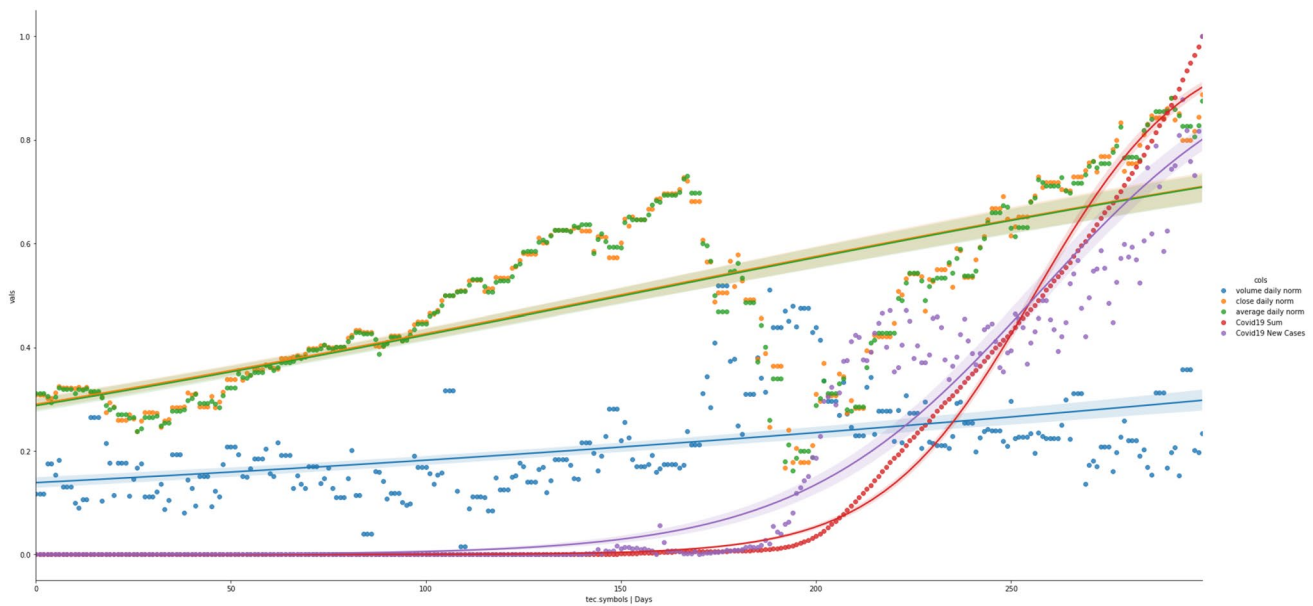
The COVID-19 time-series data comprises both the world and USA data separately. To best utilize this data,

an aggregation of these cases have also been recorded and added to the feature space. Then the new cases have been calculated as the difference of the daily cases, yielding six features as follows: US-new, US-all, World-wide-new (except the US), World-wide-all (except the US), Total-new, Total-all. Finally, the case numbers are normalized before feeding to the neural networks, making the feature space even better suited for the task. The stock market data has many missing rows as a result of market closures for holidays and weekends. Hence, data was padded to interpolate the missing data. The reason to prefer padding to other interpolation techniques is that it’s intuitive to refer to the final exchange rates and values as the current one. To obtain sector data, the selected ticker data for market close rate, volume, and daily average rates are acquired and calculated. Later, the mean of the corresponding values is taken as the overall sector values and then normalized to better fit the neural networks. To better understand the feature space and the relationships among the features, Fig. 4 illustrates the technology sector symbol average vs. the COVID-19 cases, which hints at a correlation in the data. Hence, a correlation analysis is carried out to better understand these relationships among COVID-19 and the market. This analysis provides further insights into the strength of relationships among the variables and parameters to be used in our model, helping us to confidently include important variables in the feature space and eliminate less important ones.

To prepare the data for analysis, the volume, and the average of daily sector values are calculated. Later, the results are normalized and then the logistic regression model of the values is calculated. As the prediction goal and case of this study are to forecast gold price, the feature space should be prepared. To better understand the feature space, the correlation coefficients are calculated against the daily gold price (as the possible dependent variable) within the last 300 days period. A strong correlation can be subjective and vary from one study to other [37], but in this study,  $r = \pm 0.4$  has been selected as the correlation coefficient threshold to eliminate weak correlations. As shown in Table 1, only some sectors including total new COVID-19 cases in the world, Consumer-staples (closing price), and Technology (closing price) have statistically significant correlations (significance of over 95% and  $r > 0.4$  or  $r < -0.4$ ) with gold price. Hence, all the remaining sectors values with non-significant correlations are eliminated to finalize the feature space. Table 1 shows the correlation coefficients across the sectors. In Table 1 the correlations for both ‘Close-gold’ and ‘Average-gold’ are presented to illustrate the difference between the correlations of the closing price of the market and the daily average prices. Overall, we can see similar results on both.

As shown in Table 1, new COVID-19 cases have stronger correlation ( $r$  value) with the market data. On the other hand, it is also observable that the daily market average value

<sup>1</sup> ETFdb.



**Fig. 4** Technology symbols and COVID-19 time series in 300 days until 03 Jul 2020

(normalized) has a stronger correlation with the COVID-19 pandemic than the market volume (normalized). It should also be reminded that correlation does not necessarily result in causation, yet the strong correlations can bear latent underlying connections, relationships, or meanings. For instance, the energy sector has the strongest (negative) correlation coefficient to the new cases. As the cases rise, the energy sector market value falls. This is very sensible, as can be observed from the recent drop in fuel prices, which hints at causality. The same also applies to the industrial sector, which comes second after the energy sector. The financial sector also has a very strong correlation coefficient, with  $-0.954$ , which comes third in this table. This indicates that the financial sector has also been hit hard by the pandemic at similar levels.

## Results

The experiments in this study have been implemented in Python 3.8. Keras [9] has been used for deep learning implementations and Scikit-learn [30] has been used for some loss evaluations.

Time series analysis has 1, 2 and 30 days of future data (single and multi-step) predictions. The history and the range of the data to infer from has also been a variable tuned in this study for both single and multi-step predictions. The number of historical days to tune is 5, 9, 15, 22, and 30 for the 1 and 2-day predictions, and 45 and 60 historical days for 30 days prediction models. All models have been trained for 400 epochs, with the best model checkpoints.

## The Feature Space

An objective of this study is to see the effect of COVID-19 pandemic data in the market future prediction error. Hence, the models are trained with varying features to see the influence of COVID-19. In this study as presented in the results Table 3, the inclusion of COVID-19 data in feature space would increase the model's performance for 1 and 2 days in advance predictions. Taking into account that the variation in feature space might require architectural modifications and tuning, the tests and tuning were carried out on three different feature spaces of uni-variate gold data, multi-variate data with COVID-19, and multi-variate data without COVID-19. We also notice that as opposed to many other studies that implement uni-variate methods, implementation of the multi-variate approach and taking advantage of related variables in the market has improved the performance of the models by reducing their errors. As discussed in *Understanding the Feature Space* section, the features have been selected based on their correlation coefficients with the target market value.

The features in this study can only capture so much of the real-world influence of variables on the stock market. Many other variables, including the sentiment of the news related to the pandemic does influence the markets. The use of COVID-19 case data in addition to the historic market data for the models in this study has been proposed to demonstrate this effect and its influence on the market data.

**Table 1** Correlation analysis of all sectors vs. gold

| Variables                    | $r$ (Close-gold) | $r$ (Average-gold) |
|------------------------------|------------------|--------------------|
| Consumer-staples (Close)     | 0.817            | 0.817              |
| Consumer-staple (Avg.)       | 0.817            | 0.818              |
| Volume energy                | 0.438            | 0.440              |
| Financial (Close)            | 0.468            | 0.469              |
| Financial (Avg.)             | 0.466            | 0.467              |
| Healthcare (Close)           | 0.721            | 0.722              |
| Healthcare (Avg.)            | 0.722            | 0.722              |
| Materials (Close)            | 0.600            | 0.600              |
| Materials (Avg.)             | 0.598            | 0.599              |
| Real-estate (Close)          | 0.492            | 0.493              |
| Real-estate (Avg.)           | 0.495            | 0.496              |
| Tec-symbols (Close)          | 0.777            | 0.778              |
| Tec-symbols (Avg.)           | 0.775            | 0.777              |
| Telecom (Close)              | 0.675            | 0.675              |
| Telecom (Avg.)               | 0.675            | 0.675              |
| Utilities (Close)            | 0.710            | 0.711              |
| Utilities (Avg.)             | 0.720            | 0.711              |
| Gold (Close)                 | 1.0              | 0.999              |
| Gold (Avg.)                  | 0.999            | 1.0                |
| Volume-food                  | 0.4665           | -0.682             |
| Covid-all-total-cases (norm) | 0.877            | 0.885              |
| Covid-all-new-cases (norm)   | 0.881            | 0.886              |
| Covid-US-total-cases (norm)  | 0.889            | 0.898              |
| Covid-US-new-cases (norm)    | 0.858            | 0.860              |
| Covid-ww-total-cases (norm)  | 0.873            | 0.881              |
| Covid-ww-new-cases (norm)    | 0.879            | 0.884              |
| Covid-all-total-cases        | 0.877            | 0.885              |
| Covid-all-new-cases          | 0.881            | 0.886              |
| <b>Covid-US-total-cases</b>  | <b>0.889</b>     | <b>0.898</b>       |
| Covid-US-new-cases           | 0.858            | 0.860              |
| Covid-ww-total-cases         | 0.873            | 0.881              |
| Covid-ww-new-cases           | 0.879            | 0.884              |

The  $p$  values for the corresponding correlations in this table, which already have been filtered, are below  $1.59e - 7$

ww notion in COVID-19 cases indicates world-wide cases except the United States, US indicates the United States cases, norm indicates normalized data, all denotes the aggregated cases of the United States and the rest of the world, Close indicates the mean closing price of sector symbols, Avg. indicates the mean daily-average price of sector symbols

## Single Step

The best results were achieved with 30 days of historical data as sequence inputs for the 1-day predictions, and the LSTM network with the architecture of the winner model is mentioned in Table 3. It is worth mentioning that the output of the stacked LSTM layers was taken and the hidden states were ignored, as it was noticed that having the hidden states does not improve the performance. A sample

of this experiment has been recorded in Table 3. The statement regarding the improved predictions using COVID-19 data stands true in single-step models as illustrated in Fig. 5. We can see that the predictions can follow the trend much faster when we incorporate and enrich the market data features with COVID-19 data, visible in Fig. 5b. Error values are also self explanatory in this comparison, where the model on Fig. 5b has MSE of 0.00053 and the model on Fig. 5a stops at 0.00197.

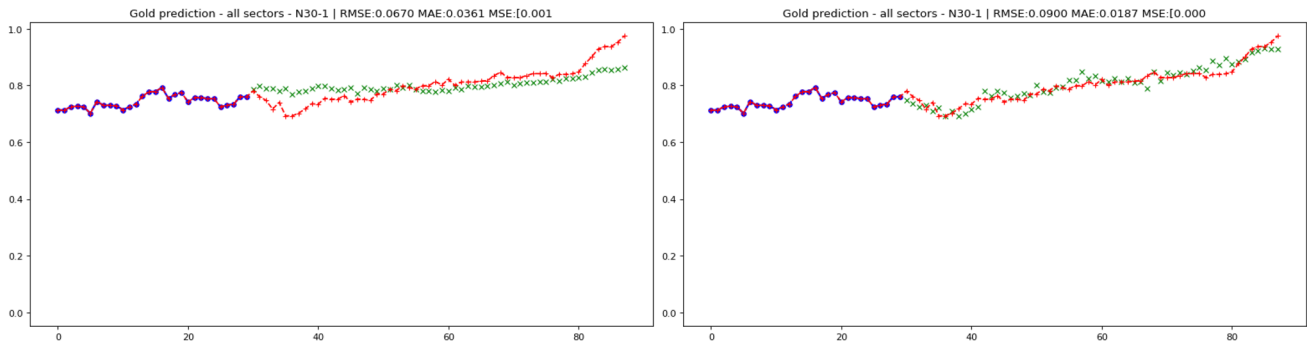
## Multiple Steps

As discussed previously in *Multi-Step LSTM* section, various multi-step models are tested in this study, including vector-output sequence prediction and encoder–decoder. According to our experiments and after many different runs due to the stochastic and random behavior, we determined that the winner method is vector output sequence prediction with 22 days of history, predicting two days in advance. The MSE of the model on validation data has been recorded as 0.00080. The encoder–decoder approach also performed very well in comparison to the rest of the approaches. Using the encoder–decoder approach, the MSE of 0.00091 was hit on the validation set. We noticed that simple stacked LSTM networks would outperform other methods for 1 and 2 days of prediction in advance, as probably it requires a less complicated approach for this kind of prediction task, and over-complicating the approach does not necessarily increase the performance.

As it can be seen in Table 3, COVID-19 data have a noticeable effect on the results. Generally, the results of the experiments without COVID-19 as demonstrated in Table 3 do not show any improvement. Besides, to calculate the effects of uni-variate data-sets, we predict some of the experiments with only the Gold-close data-set and our winner architecture and the results did not show any improvement in the performance. As it can be seen in Table 3 the error rate with uni-variate data-set is increasing totally.

Finally, 30 days prediction of the stock market value sequence is carried out, which is presented in Table 2. The historical periods of 45 days and 60 days are compared as the best resulted in experiments in this category. The MSE of the best model is 0.0021 which indicates the model has converged and reached a low error value. As illustrated, the vector output model has achieved the best results in comparison to encoder–decoder, bidirectional, and CNN–LSTM models. These results are in line with 2-day prediction models, hinting at the capability of the vector output model, given correct architecture and hyper-parameters.





**Fig. 5** Single-step time series prediction. Figure on the left **a** has been trained on the market data without COVID-19 time series. Figure on the right **b** has been trained on all features including COVID-19 time series data. Red markings are ground truth validation points

**Table 2** MSE error rates for 30 days prediction in advance with some of the best architectures

| Type      | History   | Network                  | MSE           |
|-----------|-----------|--------------------------|---------------|
| E-D       | 60        | L500-L100-d100           | 0.0217        |
| BD        | 45        | BDL300-L100-d100         | 0.0024        |
| CNN       | 45        | C100-C200-pool-L100-d100 | 0.0061        |
| <b>SE</b> | <b>60</b> | <b>L250-L250-d30</b>     | <b>0.0021</b> |

History column shows the number of days that predictions are based on.

*L* denotes LSTM layers, *d* denotes Dense layers, *C* denotes Convolution layers, *BDL* denotes bidirectional LSTM layers, *pool* denotes Max-pooling layers, *E-D* denotes encoder–decoder model, *BD* denotes bidirectional model, *CNN* denotes CNN–LSTM model, *SE* denotes vector output sequence prediction model

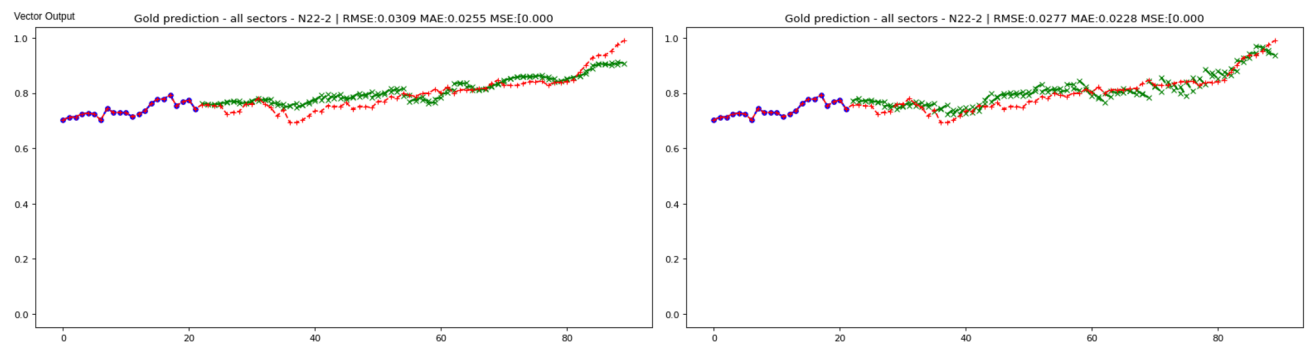
### Comparison of the Proposed Method

Here we discuss and compare our proposed methods’ results with two other proposed models. The first study proposes a time-series forecasting prediction for the price of gold, by

and green marking are the predictions. The model makes prediction based on 30 days of historical data. The architecture is: LSTM(300)-LSTM(300)-Dense(100)-Dense(1)

a CNN–LSTM model [27] that using uni-variate data and single-step prediction (single-day forecast) which is shown in Table 3 (marked by \* on row 26). Changing the uni-variate to multivariate model using the same overall architecture improves the results (specifically MSE), as illustrated in the next row. Also, comparing these results with our best architectures (i.e. vector-output sequence) with the same features, we notice that regarding the prediction of Gold price with this much fluctuations, our models yield better results in terms of prediction error (lower MSE) (Fig. 6).

The second study [8] surveyed the prediction of COVID-19 in Canada Using LSTM Networks for 14 days in advance and used uni-variate dataset and Bidirectional LSTM (model ‘28, td14’ in Table 3). It was noticed that the method could not be fitted to this problem efficiently and the errors are higher than other methods. For a more accurate comparison, we also predicted 14 days with encoder–decoder architecture and uni-variate dataset (model ‘L200-r-L200-td100-td1’), and as the results reflect, our proposed model outperforms the approach. Besides, we further modified their model by changing the dataset from uni-variate to multi-variate



**Fig. 6** Multi-step time series prediction. Figure on the left **a** has been trained on the market data without COVID-19 time series. Figure on the right **b** has been trained on all features including COVID-19 time series data. Red markings are ground truth validation points

and green marking are the predictions. The model makes prediction based on 22 days of historical data. The architecture is: LSTM(200)-LSTM(200)-Dropout-Dense(2)

**Table 3** Table of results

| #  | Steps  | Variable   | Method | History | Model                           | RMSE    | MSE     | MAE     | MSLE    | R2       | ACC     |
|----|--------|------------|--------|---------|---------------------------------|---------|---------|---------|---------|----------|---------|
| 01 | 1      | Multi      | SE     | 5       | L100-L100-d100                  | 0.0251  | 0.00063 | 0.01774 | 0.00018 | 0.85759  | 0.97803 |
| 02 | 1      | Multi      | SE     | 22      | L100-L100-d100                  | 0.0290  | 0.00084 | 0.02168 | 0.00024 | 0.76198  | 0.97438 |
| 03 | 1      | Multi      | SE     | 30      | L100-L100-d100                  | 0.0246  | 0.00060 | 0.02072 | 0.00017 | 0.79249  | 0.97592 |
| 04 | 1      | Multi      | SE     | 9       | L100-L100-drop-d100-d1          | 0.0781  | 0.00067 | 0.02066 | 0.00020 | 0.82069  | 0.94616 |
| 05 | 1      | Multi      | SE     | 30      | L300-L300-d100-d1               | 0.0900  | 0.00053 | 0.01870 | 0.00016 | 0.80063  | 0.91020 |
| 06 | 1      | COVID less | SE     | 30      | L300-L300-d100-d1               | 0.0020  | 0.00197 | 0.03611 | 0.00059 | 0.50048  | 0.93532 |
| 07 | 2      | Multi      | SE     | 5       | L200-L200-drop-d2               | 0.0439  | 0.00192 | 0.03294 | 0.00057 | 0.54020  | 0.95891 |
| 08 | 2      | Multi      | SE     | 15      | L200-L200-drop-d2               | 0.0426  | 0.00174 | 0.03738 | 0.00052 | 0.58592  | 0.95340 |
| 09 | 2      | Multi      | SE     | 22      | L200-L200-drop-d2               | 0.0311  | 0.00096 | 0.02539 | 0.00028 | 0.70735  | 0.97065 |
| 10 | 2      | Multi      | SE     | 22      | L200-L200-drop-d2               | 0.0283  | 0.00080 | 0.02334 | 0.00024 | 0.75998  | 0.97118 |
| 11 | 2      | Uni        | SE     | 22      | L200-L200-drop-d2               | 0.0983  | 0.00479 | 0.09456 | 0.00274 | 0.51678  | 0.87320 |
| 12 | 2      | COVID less | SE     | 22      | L200-L200-drop-d2               | 0.0287  | 0.00082 | 0.02410 | 0.00024 | 0.75448  | 0.97081 |
| 13 | 2      | Multi      | SE     | 30      | L200-L200-drop-d2               | 0.0414  | 0.00171 | 0.03342 | 0.00048 | 0.36443  | 0.96104 |
| 14 | 2      | Multi      | SE     | 30      | L200-L300-drop-d2               | 0.0423  | 0.00178 | 0.02961 | 0.00050 | 0.34163  | 0.96603 |
| 15 | 2      | Multi      | E-D    | 5       | L200-r-L200-td100-td1           | 0.0525  | 0.00275 | 0.04250 | 0.00082 | 0.32775  | 0.94633 |
| 16 | 2      | Multi      | E-D    | 15      | L200-r-L200-td100-td1           | 0.0303  | 0.00091 | 0.02290 | 0.00027 | 0.78431  | 0.97180 |
| 17 | 2      | COVID less | E-D    | 15      | L200-r-L200-td100-td1           | 0.0317  | 0.00100 | 0.02456 | 0.00030 | 0.76589  | 0.96948 |
| 18 | 2      | Uni        | E-D    | 15      | L200-r-L200-td100-td1           | 0.0745  | 0.00554 | 0.05769 | 0.00165 | -0.30800 | 0.83155 |
| 19 | 2      | Multi      | E-D    | 15      | L200-r-L200-drop-td100-td1      | 0.0386  | 0.00148 | 0.03276 | 0.00045 | 0.64927  | 0.95866 |
| 20 | 2      | Multi      | E-D    | 22      | L200-r-L200-td100-td1           | 0.0424  | 0.00180 | 0.03723 | 0.00052 | 0.10212  | 0.96339 |
| 21 | 2      | Multi      | E-D    | 30      | L100-r-L100-drop-td200-td1      | 0.0500  | 0.00250 | 0.03244 | 0.00070 | 0.42893  | 0.95437 |
| 22 | 2      | Multi      | E-D    | 30      | L200-r-L200-drop-td100-td1      | 0.0348  | 0.00121 | 0.02520 | 0.00033 | 0.55437  | 0.97100 |
| 23 | 14     | Multi      | E-D    | 14      | L200-r-L200-td100-td1           | 0.0938  | 0.00280 | 0.04374 | 0.00088 | -0.10390 | 0.94558 |
| 24 | 14     | Multi      | E-D    | 15      | L200-r-L200-td100-td1           | 0.0523  | 0.00270 | 0.04312 | 0.00084 | -0.09150 | 0.94591 |
| 25 | 14     | Uni        | E-D    | 14      | L200-r-L200-td100-td1           | 0.0513  | 0.00263 | 0.04029 | 0.00080 | 0.07561  | 0.95002 |
| 26 | * 1    | Uni        | CNN    | 9       | C64-C128-Max128-L200-td32-td1   | 0.0762  | 0.00581 | 0.05752 | 0.00172 | -0.31150 | 0.93195 |
| 27 | 1      | Multi      | CNN    | 9       | C64-C128-Max128-r-L200-td32-td1 | 0.0768  | 0.00278 | 0.03868 | 0.00080 | 0.37183  | 0.93054 |
| 28 | 2      | Uni        | CNN    | 9       | C64-C128-Max128-L200-td32-td1   | 0.0532  | 0.00283 | 0.03821 | 0.00083 | 0.32742  | 0.95390 |
| 29 | 2      | Multi      | CNN    | 9       | C64-C128-Max128-L200-td32-td1   | 0.0663  | 0.00430 | 0.04930 | 0.00130 | -0.03590 | 0.94067 |
| 30 | * * 14 | Uni        | BD     | 14      | BDL28, td1                      | 0.46173 | 0.54144 | 0.4371  | 0.09122 | -87.826  | 0.45193 |
| 31 | 14     | Multi      | BD     | 14      | BDL28, td1                      | 0.0337  | 0.25308 | 0.02388 | 0.00034 | 0.57346  | 0.96969 |
| 32 | 14     | Multi      | BD     | 14      | BDL100,td1                      | 0.0436  | 0.25430 | 0.03513 | 0.00057 | 0.33846  | 0.95614 |
| 33 | 14     | Multi      | BD     | 14      | BDL100,td28,td1                 | 0.0459  | 0.25460 | 0.03723 | 0.00064 | 0.26752  | 0.95343 |

The column “Step” indicates the number of steps ahead (days) predicted that are 1, 2, and 14 days in this table.

The column “Variable” shows different variables used in feature space (i.e Uni indicates the dataset which only includes Gold, Multi includes all the dataset, and COVID less includes all the financial, variables without COVID-19 data)

History column shows the number of days that predictions are based on

*L* denotes LSTM layers, *d* denotes Dense layers, *C* denotes Convolution layers, *BDL* denotes bidirectional LSTM layers, *pool* denotes Max-pooling layers, *td* denotes time distributed dense, *pool* denotes Max-pooling layers, *E-D* denotes encoder–decoder model, *BD* denotes bidirectional model, *CNN* denotes CNN–LSTM model, *SE* denotes vector sequence output prediction model

\*Based on [27]

\* \*Based on [8]

(‘BDL28, td1’ on rows 30 and 31), and as it can be seen the results improved with reducing the errors (considering all four errors). Changing the architecture, yet preserving the bidirectional LSTM approach improves the errors, but fails to outperform the rest of the approaches. We can see in Table 3 on rows 23–25 that using the encoder–decoder

approach reduces the errors drastically. Hence, we can conclude that tuning and adapting bidirectional and CNN–LSTM approaches for this task would be much harder and even impossible. Finally, we can assert that the multi-variate dataset affects the final results positively in almost every architecture, yielding lower errors.

## Conclusion and Future Work

In this study, we developed a set of prediction models for financial markets based on various features (including COVID-19 and different analyses of the market). We further investigated the effects of these features, specifically the COVID-19 pandemic, on the predictive models. The paper focuses on three aspects: (1) days to predict, (2) machine learning approaches, and (3) the feature space.

The results and experiments show that although vanilla stacked LSTMs might seem simple, yet they are very powerful at prediction compared to encoder–decoder, Bidirectional, and CNN–LSTM approaches in time series prediction problems. In this study, we achieve the MSE of  $5e - 4$  for single-step and  $8e - 4$  for multi-step forecasting for two days of prediction in advance as illustrated in Table 3. These results can be achieved using at least 5 days of historic data for single-step predictions, 22 days of historic data for two-step prediction, and 45 days of historic data for 30 days prediction in advance. The models expect normalized data, and should be trained on a sufficient amount of historic time series data.

We also show the power of other financial indicators (i.e. market data like oil, technology symbols, etc.) and non-financial parameters (e.g. COVID-19 new cases), in the prediction of a specific market. Correlation analyses are utilized in this study to establish the connection among markets and the features in general. According to the results, we can assert that having a dataset of other pandemics and the variations of the economy in that situation in history can boost the results. Although the models can achieve very low errors, it should also be acknowledged that markets rely on many variables like geopolitical decisions, global news, and escalations that can result in sudden movements. This is a limitation of machine learning at market forecasting, as designing an AI to comprehend every variable is not easily achievable. A reliable predictive model for real-world use may rely on numerous data points and features. It should be acknowledged that this study is a simplified proof of concept showing the effectiveness of non-economic features for short term predictions. As a future work, understanding the semantics and context of social networks can also be considered as a potential feature in predictive models. Further economic and market calculations as pre-processing stages of the features may also be explored in future works to increase the accuracy and making the real-world implementations more reliable.

### Declaration

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Alvarez FE, Argente D, Lippi F. A simple planning problem for covid-19 lockdown. Working Paper 26981, National Bureau of Economic Research. 2020. <https://doi.org/10.3386/w26981>. <http://www.nber.org/papers/w26981>.
2. Aramaki E, Maskawa S, Morita M. Twitter catches the flu: detecting influenza epidemics using twitter. In: Proceedings of the conference on empirical methods in natural language processing, pp. 1568–76. Association for Computational Linguistics; 2011.
3. Baldwin R, Di Mauro BW. Economics in the time of COVID-19: a new E-book. VOX CEPR Policy Portal; 2020.
4. Baldwin R. The greater trade collapse of 2020: learnings from the 2008–2009 great trade collapse. 2020. <https://voxeu.org/article/greater-trade-collapse-2020>. Accessed 6 2020.
5. Barua S. Covid-19 pandemic and world trade: some analytical notes. Available at SSRN. 2020;3577627.
6. Chen Q, Liu Y, Wang F. A chronicle on the sars epidemic. Chin Law Government. 2003;36(4):12–5.
7. Chen K, Zhou Y, Dai F. A LSTM-based method for stock returns prediction: a case study of china stock market. In: 2015 IEEE International Conference on Big Data (Big Data); 2015. pp. 2823–4.
8. Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos Solitons Fractals. 2020;135:109864
9. Chollet F et al. Keras. 2015. <https://keras.io>. Accessed 5 2020.
10. Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint. 2014; [arXiv:1406.1078](https://arxiv.org/abs/1406.1078).
11. de Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, Fouchier RA, Galiano M, Gorbalenya AE, Memish ZA, et al. Commentary: middle east respiratory syndrome coronavirus (mers-cov): announcement of the coronavirus study group. J Virol. 2013;87(14):7790–2.
12. Di Persio L, Honchar O. Artificial neural networks approach to the forecast of stock market price movements. International Journal of Economics and Management Systems. 2016;1, 158–62
13. Donahue J, Anne Hendricks L, Guadarrama S, Rohrbach M, Venugopalan S, Saenko K, Darrell T. Long-term recurrent convolutional networks for visual recognition and description. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. pp. 2625–34
14. etfdb: The 11 sectors of the stock market. <https://etfdb.com/etf-education/what-is-an-in-kind-redemption-for-etfs/>. Accessed: 04-2020.
15. Fawaz HI, Forestier G, Weber J, Idoumghar L, Muller PA. Deep learning for time series classification: a review. Data Min Knowl Disc. 2019;33(4):917–63.
16. Fischer T, Krauss C. Deep learning with long short-term memory networks for financial market predictions. Eur J Oper Res. 2018;270(2):654–69.
17. for Disease Control C, Prevention et al. The 2009 h1n1 pandemic: summary highlights, april 2009–april 2010. Official Online Article Published by the Centers for Disease Control, vol. 4. 2010.
18. Gers FA, Eck D, Schmidhuber J. Applying lstm to time series predictable through time-window approaches. In: Neural Nets WIRN Vietri-01. Springer; 2002. pp. 193–200.
19. Gormsen NJ, Koijen RS. Coronavirus: impact on stock prices and growth expectations. Rev Asset Pric Stud. 2020;10(4):574–97. <https://doi.org/10.1093/rapstu/raaa013>.

20. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* 2005;18(5–6):602–10.
21. Guerrieri V, Lorenzoni G, Straub L, Werning I. Macroeconomic implications of COVID-19: Can negative supply shocks cause demand shortages? National Bureau of Economic Research: Tech. rep; 2020.
22. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* 1997;9(8):1735–80.
23. Jordà Ò, Singh SR, Taylor AM. Longer-run economic consequences of pandemics. National Bureau of Economic Research: Tech. rep; 2020.
24. Kaner J, Schaack S. Understanding ebola: the 2014 epidemic. *Glob Health.* 2016;12(1):1–7.
25. Le XH, Ho HV, Lee G, Jung S. Application of long short-term memory (LSTM) neural network for flood forecasting. *Water (Switzerland).* 2019. <https://doi.org/10.3390/w11071387>.
26. Li M, Wang X, Gao K, Zhang S. A survey on information diffusion in online social networks: models and methods. *Information (Switzerland).* 2017. <https://doi.org/10.3390/info8040118>.
27. Livieris IE, Pintelas E, Pintelas P. A CNN–LSTM model for gold price time-series forecasting. *Neural Comput Appl.* 2020;32(23):1–10.
28. Marchi E, Vesperini F, Weninger F, Eyben F, Squartini S, Schuller B. Non-linear prediction with LSTM recurrent neural networks for acoustic novelty detection. In: 2015 International Joint Conference on Neural Networks (IJCNN), pp. 1–7. IEEE; 2015.
29. Mohtasham Khani M, Vahidnia S, Ghasemzadeh L, Ozturk YE, Yuvalaklioglu M, Akin S, Ure NK. Deep-learning-based crack detection with applications for the structural health monitoring of gas turbines. *Struct Health Monit.* 2020;19(5):1440–52.
30. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011;12:2825–30.
31. Pettie C. Modeling syndromic surveillance and outbreaks in sub-populations. Ph.D. thesis, University of Maryland; 2020.
32. Schuster M, Paliwal KK. Bidirectional recurrent neural networks. *IEEE Trans Signal Process.* 1997;45(11):2673–81.
33. Shouval R, Bondi O, Mishan H, Shimoni A, Unger R, Nagler A. Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in sct. *Bone Marrow Transplant.* 2014;49(3):332–7.
34. Slotnick D. Many of the world’s airlines could be bankrupt by may because of the COVID-19 crisis. 2020. <https://www.businesinsider.com.au/coronavirus-airlines-that-failed-bankrupt-covid-19-pandemic-2020-3>. Accessed 6 2020.
35. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: *Advances in neural information processing systems*; 2014. pp. 3104–12.
36. Tomar A, Gupta N. Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. *Sci Total Environ.* 2020;728:138762.
37. Vahidnia S, Tanrıöver ÖÖ, Askerzade I. An early phase software project risk assessment support method for emergent software organizations. *Int J Advan Comp Sci Appl.* 2017;8(5):105–18.
38. Vinyals O, Toshev A, Bengio S, Erhan D. Show and tell: a neural image caption generator. In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2015. pp. 3156–64.
39. Wilson ME. Travel and the emergence of infectious diseases. *Emerg Infect Dis.* 1995;1(2):39.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.