An Influential Stock Market Paradigm based on LSTM Modeling and Sequencing

Sanjay Krishnan
Data Science
School of Computing, Engineering and Intelligent Systems, Ulster University
Londonderry, UK
Krishnan-s1@ulster.ac.uk

Abstract—**Investors are paying close attention to the stock market. Shareholders and investment firms have long relied on it to comprehend the stock market's regularity of movement and forecast its trajectory. There are a variety of ways to anticipate stock prices at this time. Two broad prediction techniques exist; statistical and artificial intelligence (AI). This study has considered the stock market dataset based on time series data. The research explores this dataset using the sequencing function and LSTM model. After various experiments, the study computed that our proposed model's average Mean Absolute Error is just 0.0136. Thus, the researchers predicted new attributed "Close_Predicted" and "Open_Predicted" based on the previous time-series data. Our proposed approach also used exploratory analysis for a better understanding of the data. Consequently, presenting upcoming "open" and "close" prices for stock markets based on the available dataset.The Implications are that this time series analysis study is meaningful for stock markets, stock exchanges, regulatory authorities, and money exchangers.**

Keywords—stock market, time-series, LSTM, MAE

I. Introduction

The financial sector has received much attention lately, especially from investors. For investors and scholars alike, understanding the changing patterns of the financial markets and predicting the trajectory of financial markets has long been a popular topic of discussion. A variety of elements, including politics, the financial system, the community, and the stock exchange, impact the ups and downs of stocks in the market. The stock market's trend projection is closely tied to the accumulation of earnings for stock investors. The more precise a prediction is, the more efficiently it may be used to prevent hazards in the future. It is crucial to note that the stock price of publicly traded firms represents the company's operational circumstances and future development trends, but it is also a technical gauge used in the company's analysis and research. Stock forecasting research is especially significant in studying a country's economic growth since it helps predict future stock prices. As a result, research into the intrinsic value of stocks and the forecast of stock prices has significant theoretical relevance as well as broad practical potential.

To put it another way, in an intuitive sense, the time series does not change statistical parameters (like the mean and covariance) when it is shifted in time [1]. For the stationary succession, only recent results have a major impact on the current value, and the longer the period is, the less influence that the previous value has on the present value. Our ability to pick a more accurate prediction method depends on determining whether or not the data is stationary. However, non-stationary time series, the Auto Regressive Moving Average (ARMA) model is commonly utilised [2], [3].

Stock market prediction is the most prevalent projection in the financial market built on time-series data. Mostly, the stock price data is found as noisy, complicated, non-linear, and they can be influenced by various variables like government strategies, the economy, and mindset. Because of this, data pre-treatment and analysis must be performed prior to the forecasting (for example, stationarity detection). Conventional forecasting models are not capable of adequately capturing historical insights, relying on the LR (Linear Regression) model and statistical inference. Hence, non-linear models, such as support vector machines (SVM) and deep neural networks are used to make most financial predictions.

Machine Learning is a broad domain and deep learning is the subfield of this major domain that evolved from the basic neural network in order to improve performance [4]. In recent decades, many academics have attempted to address time-series problems using deep learning approaches to great success. The RNN (Recurrent Neural Networks) is a kind of neural network that incorporates the idea of time into its network topology, creating it well suited for time-series processing and data analysis [5]. In practice, the impact is ineffective. The vanishing gradient issue may arise during the optimisation process if the series is too lengthy. In 1997, Sepp Hochreiter and Jürgen Schmidhuber presented an LSTM model to resolve this challenge [6]. An RNN network with an LSTM structure is better able for long-term activities and is less susceptible to waning gradient issues when using an LSTM structure. A notable example is Graves' successful use of the Bidirectional RNN (BRNN) which is an efficient model for handwriting recognition [7] as well as voice recognition [8], as well as the application of many variations of the RNN network to other NLP (Natural Language Processing) issues. Similarly, Google's large-scale speech recognition system employed two layers of deep LSTM, where the built model performed significant outcomes [9]. In recent years, several researchers have used LSTM to predict stock prices with promising results [10]. For instance, a time-weighted LSTM framework for redefining stock tendency predictions is presented in the literature [11]. A comparison is made between the bidirectional as well as stacked LSTMs using basic LSTM. The findings revealed that the bidirectional LSTM outperformed the other two in stock prediction in the study [12]. An additional model was developed using K-means clustering and was shown to be more accurate in short-term stock prediction than the single branch LSTM stock model. Researchers have established the effectiveness of the LSTM forecasting model over conventional estimation techniques, which can be used for a variety of non-linear time-series data. However, they have not investigated the impact of time-series stationary variation on forecasting outcomes in these models.

The primary objective of this study is to deploy an LSTM model that can forecast the "opening amount" as well as the "closing amount" of the stock for the coming day based on the company's historical amount. Other technical parameter information. A deep recurrent neural network model built on the LSTM is suggested to predict the two values that are associated with each LSTM.

The rest of the study covers briefly the whole approach that the study used for analysing the stock market time series data using the LSTM model. In section II researchers have described the related work regarding the same approach that has been proposed. In section III .experiments are discussed that have been conducted and the outcome and results of the study. The final section IV elaborated the conclusion of this time series of stock market research.

II. Related Work

There have been numerous relevant studies conducted on stock market forecasting. SVMs are used to create a regression model from past stock data and forecast stock prices' direction and their movements [13]. The SVMs' parameters are optimised using the particle swarm optimisation approach, which has ability to forecast stock value with high accuracy [14]. The SVM approach is improved due to this research; however, the particle swarm optimisation methodology takes a long time to compute. The LSTM approach is integrated using the NB (Naïve Bayesian) technique for uncovering market emotion components in order to enhance forecasting accuracy [15]. This strategy may be used to anticipate financial markets across a wide range of time intervals, as well as in conjunction based on other factors. The behavioural analysis framework is merged with the LSTM time series architecture to provide an efficient time series forecasting model for the “opening amount” of shares, findings depict that the prediction accuracy could be improved by including this model in the process [16]. Jia examined the usefulness of LSTM for forecasting share value, and the research revealed that LSTM is a reliable strategy for predicting stock profits as well as stock prices [17]. It was possible to forecast the east Asian stock market index with the integration of the LSTM network with the real-time wavelet denoising, rectifying several logic faults found in prior experiments [18]. This hybrid model performs much better when contrasted with the real LSTM, with strong predictive accuracy and a minimal regression error in the predictions. The bagging strategy is utilised to integrate various neural network tactics in order to predict the Chinese stock index [19]; each neural network has been trained using the backpropagation technique using the Adam optimisation algorithm; outcomes show that the approach has varying model efficiency score for the forecasting of various stock indexes, however the forecasting on “close amount” is found unsatisfactory; The evolutionary technique is used to anticipate the recent trends of stock prices, and the results are promising [20]. The prediction of the company's stock time series is accomplished via applying a deep network with intrinsic flexibility [21]. Using a CNN (Convolutional Neural Network), it is possible to forecast the direction of the stock amount [22]. A forward multi-layer model based on Neural Networks is employed for forecasting stock market prices is developed utilising a hybrid technique that combines methodological analysis parameters with elementary analysis attributes of stock market characteristics, as well as the BP algorithm, to forecast future stock prices [23]. According to the study's findings, this strategy outperforms the scientific analysis method to forecast daily stock prices accurately. The Dhaka Stock Exchange (DSE) has developed a powerful soft computing system that can anticipate the closing price of the stock exchange [24]. Comparative research has shown that this strategy is more successful than ANN and adaptive neural fuzzy reasoning systems.

A. *Time-Series Analysis*

Stationarity is a precondition for most time-series forecasts, so time-series forecasting is important. The record and unit root test techniques are common stationary detection approaches. There exist various forms to observe a pattern, however, the common is to look at the original image of the series and see whether there is a clear trend or periodicity using partial autocorrelation function (PACF) [25], [26]. ADF (Augmented Dickey-Fuller Test) [27] is widely employed unit root test techniques, which he first introduced in 1979. Other categorisation approaches include the automated time series clustering in stationary as well as non-stationary groups. There is no in-depth research on them, although they are used as a technique to determine the presence of stationary conditions in studies.

B. *The Long Term Short Term Memory (LSTM)*

The LSTM algorithm works as an improved version of the RNN algorithm. RNN differs from regular neural networks because it has connections between its hidden layers, as opposed to the standard neural network. So, the hidden layer receives input from neurons as well as the outcome of the hidden layers at an earlier period, which is called intake of the hidden layer. Figure 1 depicts the RNN's extended structure in more detail.

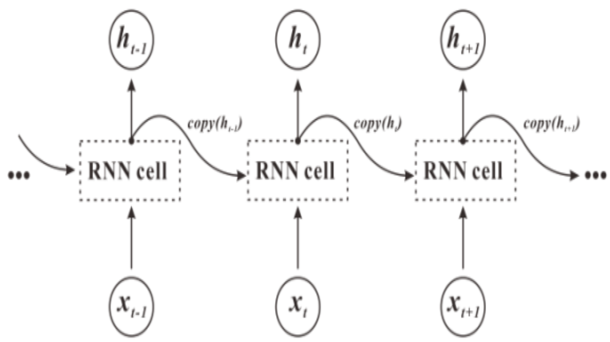

Fig. 1. RNN Model Expanded Structure

While the LSTM, as well as RNN, contains a unique extended structure, that is known as memory cell structure which varies in the place of the hidden layer. Hidden layers is based on gates that are

- Output Gate
- Forget Gate
- Output Gate

The hidden layer's memory cell of LSTM successfully handles the vanishing gradient problem, making it ideal for trading with long-term dependence issues. The structure of the cell for hidden layers of the LSTM model is displayed here in Figure 2.

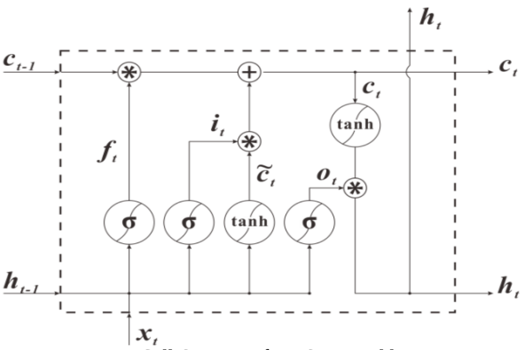

Fig. 2. Cell Structure for LSTM Hidden Layers

III. Experiments and Results

A. Dataset

The analysis is performed on an open-source, available stock market dataset. The dataset is based on 1565 records. There exist six attributes. These attributes include “Date”, “Open”, “Close”, “High”, “Low”, and “Volume”. Rather than the "Date" attribute, all the parameters are integers and float values; however, the "Date" parameter is the object. For conducting the study, researchers downloaded the dataset from an open-source Kaggle platform. The link on the Kaggle platform dataset is https://www.kaggle.com/datasets/borismarjanovic/price-volume-data-for-all-us-stocks-etfs.

B. Supporting Environment

The analysts have used Google Colab to evaluate this time series data of the stock market. Python 3 is used for computing the analysis results. GPU is used in computing the outcome in Google Colab. RAM of 1.11/12.69 GB and disk space of 39.90/78.19 GB is consumed in this part of the analysis.

C. Analysis

Researchers adjusted the "Date" attribute into a required pattern so that it can be easily readable by the machine learning algorithm. Researchers first computed the "Open" and "Close" prices in a time-series manner as the data was from 2010 to 2018. The stock market "open" and "close" values are illustrated in Fig. 3.

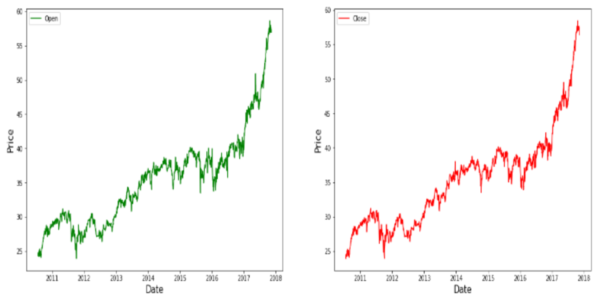

Fig. 3. Time series representation of Open and Close Prices

D. Data Preprocessing

For data preprocessing, analysts adopted the min-max normalisation technique. This technique converges the data between 0 and 1. The major goal of this approach is to make the whole data uniform. Hence Min Max Scaler of sklearn library of Python framework is used to normalise the data.

E. Model Building

In model building, first, researchers split the data into train and test parts. For test train splitting, 80% of the data is selected for model traning, i.e., 1252 rows of records, for the training purpose of the model. However, 20% of the data, i.e., 313 records, are selected for testing the model. Then used the sequencing function for the below-listed purposes.

1. Create state vectors from the input sequence.
2. Begin with a small target sequence.
3. The decoder will use the state vectors and the 1-char target sequence to anticipate the next character.
4. Using assumptions to know the relevant character.
5. Finally, paste in the sampled character.
6. On reaching the character limit or when creating the end-of-the-sequence character.

Analysts then called the LSTM model with the stock market data for model building. Based on the previous "Open" and "Close" attributes, new features are computed using the LSTM model. The two new features are named "Open_Predicted" and “Close_Predicted”. Both the new parameters present the forecasted prices against each date daily, as presented in Fig. 4 and Fig.5 respectively.

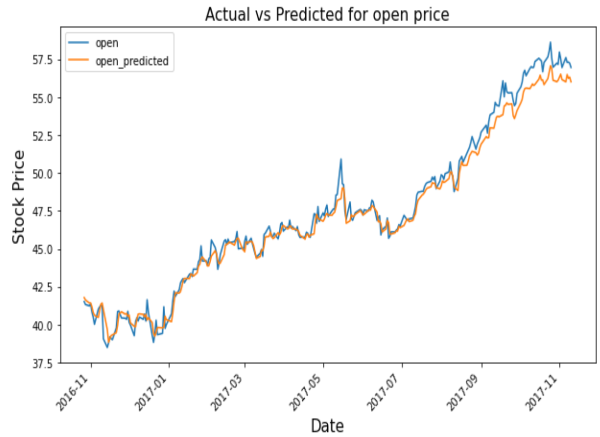

Fig. 4. Actual and Forecasting Amount Comparison for Open Price

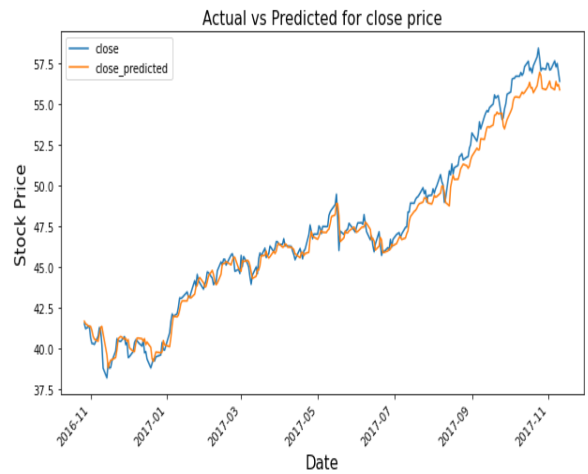

Fig. 5. Actual and Forecasting Amount Comparison for Close Price

Using this technique of stock market forecasting, the analysts predicted various records for both open and close values. These values are also given in the arrangement of time series because each predicted value represented in front of a specific date. Thus, it considers the previous time-series data to forecast new values for opening and closing the stock market.

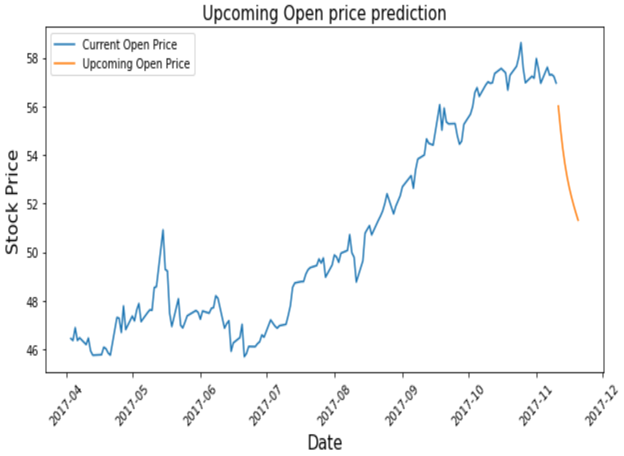

Fig. 6. Upcoming Open Price Forecasting

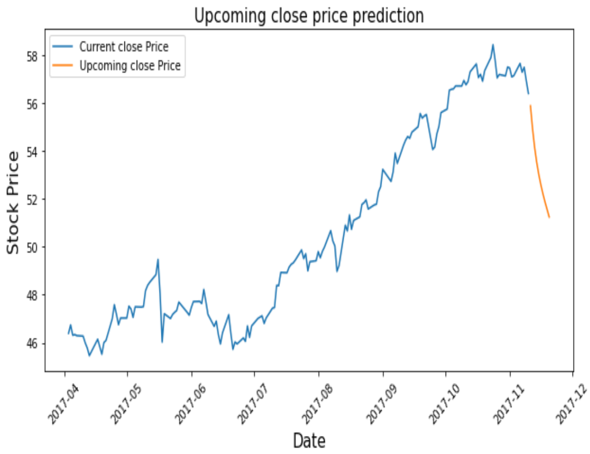

Fig. 7. Upcoming Close Price Forecasting

After successfully computing the "Open_Predicted" and "Close_Predicted" time series values for the stock market, then calculated the upcoming open price prediction as well as close price predictions for those dates that are not available in our dataset. Fig. 6 and Fig. 7 are clear evidence of upcoming prices.

F. Evaluation Metric

The researchers have computed the Mean Absolute Error after running all Epochs and found that the average of all Epochs MAE is just 0.0136. Thus, this means absolute error will be considered as an evaluation metric of the time series analysis.

IV. Conclusion

In this study, the researchers used a stock market dataset for analysing time-series data. Stock market data has a great significance that retians a key role in the economic development of a state. Hence, predicting the accurate stock values can help the regulatory authorities bring forward necessary actions if there is a risk of a decrease. However, they can boost their stock market by focusing on influential features to get a higher outcome from the data.

In the research case, analysts employed the LSTM model for predicting the stock market data. This study depicted the time series data in an exploratory form to get insights from the data. The exploratory data helps us decide which year the stock market "open", and "close" prices were lower or higher. Thus, the analyst can judge the seasonal inflation or growth in the stock data.

After successful exploratory analysis, employed a few preprocessing techniques to make the data more meaningful and easily understandable by the computer systems. Afterward, split the dataset into train and test data built a sequencing function, and created an LSTM model for further predictions. As an evaluation metric, analysts computed that the Mean Absolute Error for the built model is just 0.0136.

Consequently, achieved the new predicted attributes "Close_Predicted" and "Open_Predicted". These forecasted values are much closer to the actual "close" and "open" price values in the stock market dataset. At the end of the study, analysts evaluated the upcoming "open" as well as "close" values based on the previous time-series data.

Implications of this study can be used for different kinds of currency exchange datasets in future work. Furthermore, because bitcoin currency is becoming increasingly popular across the world, a modified version of this study may be able to produce more statistically significant results for the bitcoin dataset.

References

[1] J.-F. Zhao, "Modeling of Software Aging Based on Non-stationary Time Series," in 2016 International Conference on Information System and Artificial Intelligence (ISAI), 2016, pp. 176-180: IEEE.

[2] G. E. Box and G. M. J. C. H.-D. Jenkins, "Time series analysis: Forecasting and control San Francisco," 1976.

[3] F. Qian and X. Chen, "Stock prediction based on LSTM under different stability," in 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 2019, pp. 483-486: IEEE.

[4] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT Press, 2016.

[5] A. Gustavsson, A. Magnuson, B. Blomberg, M. Andersson, J. Halfvarson, and C. J. C. s. Tysk, "On the difficulty of training recurrent neural networks," vol. 52, no. 3, pp. 337-345, 2013.

[6] S. Hochreiter and J. J. N. c. Schmidhuber, "Long short-term memory," vol. 9, no. 8, pp. 1735-1780, 1997.

[7] A. Graves and J. J. A. i. n. i. p. s. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," vol. 21, 2008.

[8] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in 2013 IEEE international conference on acoustics, speech and signal processing, 2013, pp. 6645-6649: Ieee.

[9] T. Zia and U. J. I. J. o. S. T. Zahid, "Long short-term memory recurrent neural network architectures for Urdu acoustic modeling," vol. 22, no. 1, pp. 21-30, 2019.

[10] Z. Zhao, R. Rao, S. Tu, and J. Shi, "Time-weighted LSTM model with redefined labeling for stock trend prediction," in 2017 IEEE 29th international conference on tools with artificial intelligence (ICTAI), 2017, pp. 1210-1217: IEEE.

[11] K. A. Althelaya, E.-S. M. El-Alfy, and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction," in 2018 9th international conference on information and communication systems (ICICS), 2018, pp. 151-156: IEEE.

[12] X. Shao, D. Ma, Y. Liu, and Q. Yin, "Short-term forecast of stock price of multi-branch LSTM based on K-means," in 2017 4th International Conference on Systems and Informatics (ICSAI), 2017, pp. 1546-1551: IEEE.

[13] M. K. Alshaikh, F. T. Filippidis, H. A. Al-Omar, S. Rawaf, A. Majeed, and A.-M. J. B. p. h. Salmasi, "The ticking time bomb in lifestyle-related diseases among women in the Gulf Cooperation Council countries; review of systematic reviews," vol. 17, no. 1, pp. 1-17, 2017.

[14] T. M. Sands, D. Tayal, M. E. Morris, and S. T. Monteiro, "Robust stock value prediction using support vector machines with particle swarm optimisation," in 2015 IEEE Congress on Evolutionary Computation (CEC), 2015, pp. 3327-3331: IEEE.

[15] J. Li, H. Bu, and J. Wu, "Sentiment-aware stock market prediction: A deep learning method," in 2017 international conference on service systems and service management, 2017, pp. 1-6: IEEE.

[16] Q. Zhuge, L. Xu, and G. J. E. l. Zhang, "LSTM Neural Network with Emotional Analysis for prediction of stock price," vol. 25, no. 2, 2017.

[17] H. J. a. p. a. Jia, "Investigation into the effectiveness of long short term memory networks for stock price prediction," 2016.

[18] Z. Li and V. Tam, "Combining the real-time wavelet denoising and long-short-term-memory neural network for predicting stock indexes," in 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 2017, pp. 1-8: IEEE.

[19] B. Yang, Z.-J. Gong, and W. Yang, "Stock market index prediction using deep neural network ensemble," in 2017 36th Chinese control conference (ccc), 2017, pp. 3882-3887: IEEE.

[20] Y.-C. Tsai and C.-Y. Hong, "The application of evolutionary approach for stock trend awareness," in 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST), 2017, pp. 306-311: IEEE.

[21] X. Li, L. Yang, F. Xue, and H. Zhou, "Time series prediction of stock price using deep belief networks with intrinsic plasticity," in 2017 29th Chinese Control And Decision Conference (CCDC), 2017, pp. 1237-1242: IEEE.

[22] M. U. Gudelek, S. A. Boluk, and A. M. Ozbayoglu, "A deep learning based stock trading model with 2-D CNN trend detection," in 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 2017, pp. 1-8: IEEE.

[23] A. A. Adebiyi, C. K. Ayo, M. Adebiyi, S. O. J. J. o. E. T. i. C. Otokiti, and I. Sciences, "Stock price prediction using neural network with hybridised market indicators," vol. 3, no. 1, 2012.

[24] M. Billah, S. Waheed, and A. J. I. J. o. C. A. Hanifa, "Predicting closing stock price using artificial neural network and adaptive neuro fuzzy inference system (anfis): the case of the dhaka stock exchange," vol. 129, no. 11, pp. 1-5, 2015.

[25] P. Galeano and D. Peña, "Multivariate analysis in vector time series," 2001.

[26] W. S. J. T. Cleveland, "The inverse autocorrelations of a time series and their applications," vol. 14, no. 2, pp. 277-293, 1972.

[27] D. A. Dickey and W. A. J. J. o. t. A. s. a. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," vol. 74, no. 366a, pp. 427-431, 1979.