STOCK PRICE PREDICTION BASED ON NEURAL NETWORKS

Vellaichamy U[1a], Meenakshi C[1b*]

1a - Research Scholar, Department of Computer Applications, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai 600 117.

1b - Assistant Professor, Department of Computer Applications, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai 600 117.

*- Corresponding Author

ABSTRACT

One of the most challenging problems of opinion mining is model based opinion mining, the results confirm about the accuracy of the prediction performance of neural network. Prediction of stock market returns has been one of the most challenging problems. The proposed model applies a kind of artificial neural network (ANN) named Backpropagation (BP) algorithm onto the prediction of stock price patterns. The BP algorithm neural network for the transaction data. The models are evaluated using standard strategic indicators: RMSE and MAPE for forecasting the stock prices. The result of simulation shows that the proposed method achieves higher rate of accuracy in predicting the instances than the existing methods.

Keywords: *LSTM, Stock Price, Prediction and Back Propagation*

1. INTRODUCTION

People have been debating in a passionate manner for decades on the topic of whether it is feasible to anticipate stock values. The topic at hand is whether it is possible to predict stock values [1]. According to the Efficient Market Hypothesis, the value of a company shares should be able to accurately reflect all the relevant information at any moment in time. There is no way to identify low-cost stocks or foresee changes in market patterns using fundamental or technical research if it is not possible to forecast stock prices by using past information. If this is the case, then it is impossible to find low-cost stocks. Another significant body of information suggests going in the complete opposite direction [2]. Individual and institutional investors continued to place a large degree of significance on fundamental and technical analysis. This was true regardless of the type of investment being conducted. In instance, both the theory and the facts imply that those who have a stronger ability to comprehend information have an edge. This is particularly relevant to the stock market. Processing and analysing data of this kind have historically been among the most important aspects of a financial analyst job obligations. On the other hand, there is a mountain of evidence demonstrating that financial analysts face conflicts of interest, which can result in recommendations that are substantially skewed. This can be a problem because it can lead to suggestions that are not followed [3]. A practise that was once exclusive to highly skilled professionals, individual investors are increasingly conducting and disseminating their own financial research. This is a practise that was previously only available to specialists. Academics and industry professionals are focusing their attention on the growth of online financial forums to determine whether new information can be gathered from investor talks [4]. This will allow them to improve their ability to predict changes in stock prices. There is a multitude of data to support the concept that the feelings of investors play a large role in explaining variations in stock prices. This can be seen in the fact that this theory is supported by a wealth of evidence [5].

To produce predictions, several pieces of study have integrated the past performance of stock prices with the textual data obtained from social networks. Stock price forecasting has been an important area of research, and most researchers have approached this subject as though it were a classification problem with two binary classes. On the other hand, only a small number of them have ever attempted to propose models for directly projecting prices based on time series analysis [6]. This is because such models are notoriously difficult to develop. When doing this research to establish the attitude of investors, the news is the only source that is ever taken into consideration as a potential indicator of investor sentiment. Several cognitive biases, like overconfidence, and loss aversion, all play a role in the way different investors take in and make sense of the same piece of news [7]. This research was carried out with the intention of developing a hybrid model for analysing the stock market and forecasting prices. This model will combine a deep learning strategy with a model for sentiment analysis to circumvent the limitations that are inherent to more conventional methods. To arrive at the sentiment factor and determine the investor covert feelings, we first classify the investor covert sentiments using a back propagation neural network (BPNN). In addition, the technical indicators that are derived from stock quotations as well as the investor sentiment component that is present in a well-known stock forum are input into an LSTM neural network technique to forecast the closing price for the following trading day one day in advance. This is done to maximise the accuracy with which the closing price can be predicted.

More than 80% of the amount of trading activity in a stock market is driven by individual investor purchases and sales of shares. As a direct result of this, we anticipate discovering evidence of the predictive value of technical indicators and investor sentiment for shares that are traded on the Singapore Stock Exchange (SSE). We can make accurate projections on the movement of stock prices if, in addition to technical indicators, we also take into account investor sentiment. We also demonstrate that our classification of sentiment is superior to the use of purely technical indicators in terms of its capacity to generate accurate predictions.

The model utilises a technique of artificial neural networks (ANNs) known as the backpropagation (BP) approach to create predictions regarding the patterns of stock prices. BP approach to computing with neural networks, which was created specifically for use with financial data Standard strategic criteria for predicting stock prices, such as RMSE and MAPE, are used in the evaluation of the models so that conclusions can be drawn about their accuracy.

2. PROBLEM DEFINITION

The time series of prices for most financial instruments can be understood to be continuously moving data series that fall within the realm of what can be categorised as regular trade. This interpretation is possible because most of these instruments are traded on a regular basis. This is the case since these pricing shifts are consistent with the norm. However, due to the discrete periods that are included within the transaction range, not all financial time series can be completely continuous data. This is because of the transaction range. In most instances, this is how the time series has played out. Additionally, the patterns that are seen for each series at each of the various time intervals tend to be somewhat different from one another in comparison to the patterns that are seen for the other time intervals. To better understand this topic, we will use the information that is available on various market indexes. Over the course of a cyclical amount of time, a pattern that is favourable to expansion has been noticed over the course of the longer term. The intermediate and short-term trends that are exhibited at daily or hourly intervals may shift or even be declining at many points throughout the whole of this extended timeframe. These trends are displayed at regular intervals of either 24 hours or 24 minutes. The primary focus of the investigations that were carried out to complete this work was placed on the analysis of data pertaining to stock prices. Trading days were employed as observation intervals to limit the number of confounding factors, assure consistency in the comparison studies, and take into consideration $T+1$ trading system. Trading days are defined as the period between two consecutive trading days. The following prices for the stock index have been recorded in order:

$$y_1, y_2, \ldots, y_t, \ldots, y_n.$$

Therefore, at time t, all the data that is currently available is utilised to search for a nonlinear process that replicates price fluctuations while simultaneously producing forecasts for a financial time series. The goal of this search is to find a process that can do both things. The formula is used as a guide in the construction of a prediction model, which is subsequently utilised in the process of developing projections pertaining to y_t. The objective is to develop an approximation for the process by picking an appropriate model with parameters that are optimal and making use of data from a time series. This will allow for the development of an approximation f.

$$y_t\text{-}s(t) = f(y_{t-d}, y_{t-d-1}, \ldots, y_{t-d-n+1}) + w(t)$$

where

d - delay,

n - time span,

$w(t)$ - noise and

t - time.

Figure 2 illustrates a simplified flowchart of the three steps that are involved in constructing a model for predicting the price of a stock index. These processes are shown in order from the most basic to the most complex. These are the steps that need to be taken:

(1) the processing and analysis of the collected data;

(2) the creation of a predictive model; and

(3) the forecasting and evaluation of the outcomes of experimental research.

In the first stage, the preprocessing operations, phase-space reconstruction, and data structuring procedures are used to build input and output datasets that are helpful to the prediction model. Other preprocessing operations include phase-space reconstruction.

All these steps are included in the first stage of the process (such as label definition and test set dividing). In a later step, a variety of machine learning methods are utilised to construct a prediction model. In addition, the model goes through the process of fine-tuning and refining, with parameters being chosen in the most rational way that is practically possible.

In addition to this, the model is used to produce a forecast that is founded on the data that was input into it and gained from it. This is done by using the data that was obtained from it. The study needs to conduct a multi-dimensional analysis on the data that is produced by the prediction model and develop assessment measures for the results of the prediction in the third step of the process.

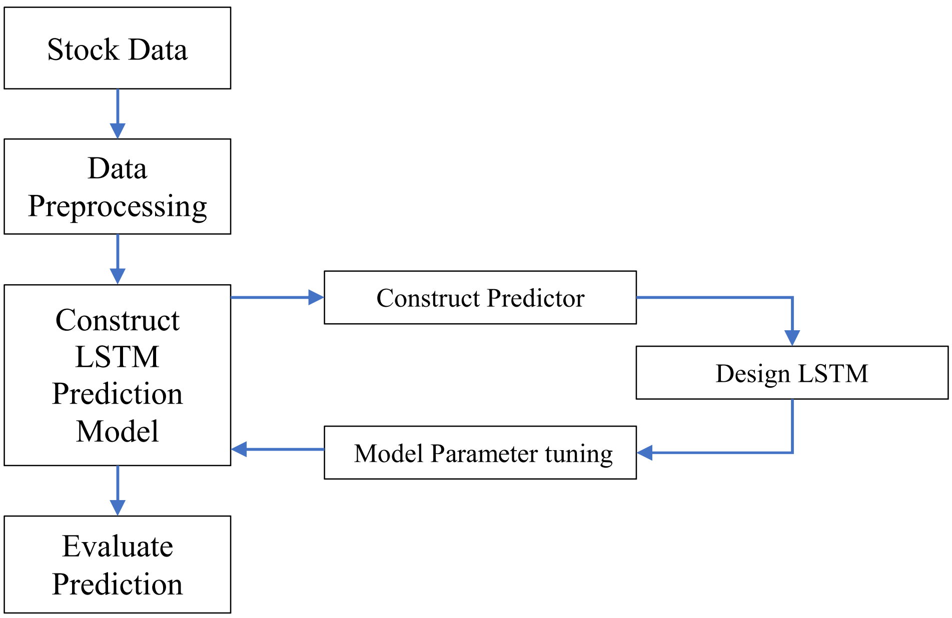

Figure 1: Framework of Prediction process

PSR of price data

Figure 2 provides an illustration of the process for forecasting stock index price series using price-sales ratio analysis. The great majority of financial products that are based on stock indexes have only been around for the past few decades. This is very recent in the history of finance. When compared to the number of data points in other machine learning datasets, the number of data points in a time series generated by using trading days as observation intervals is relatively low and frequently falls somewhere in the several thousand range. This is because trading days are used as the interval between observations in the time series.

It is usual for models trained with machine learning algorithms to overfit the data when the dataset in question contains an insufficient number of observations. Because the amount of data is so crucial to the findings of this investigation, we employ a tried-and-true method known as sliding window analysis to fold and replicate the data a great number of times. Anyone who has even a passing interest in dimensions ought to know PSR provides a dimensional extension for the data that pertains to price. The PSR model formula is written in terms of a time series [$y_t = x(t)$, t=1,…,n] that is known at some point in time t.

$$X(n) = x(n),$$

$$x(n+\tau),\ldots,x(n+(m\text{-}1)\cdot\tau)$$

where

τ - delay interval between $x(t)$ and $x(t+\tau)$

The embedding dimension of the generated series is denoted by the symbol m, and the delay interval is denoted by the symbol. If the embedding dimension is increased, the analytical potential of a system structure can be improved; however, this expansion must be carried out in a suitable manner in order for this improvement to take place.

To determining and quantifying, respectively, and m, we make use of the mutual information technique. After the values of a and m have been successfully computed, PSR can then be carried out. $X(t)$ is the label that has been given to the information that was recovered (t).

The supervised learning model that we have constructed will contain the following inputs and outputs, and a description of each may be found below: $X(i)$, $i=t_k$, t_{k+1},... is the input, and the output is the reconstructed data at time t, which is indicated by X. and the output is the reconstructed data at time t.

The number of reconstructed data points immediately preceding time t serves as the input, while the data that has been reconstructed at time t serves as the output. As the anticipated price series, the final data dimension $x(t\text{-}(m\text{-}1)\tau)$ of the predicted reconstructed data $X(t)$ is employed, and the correctness of the anticipated price series is evaluated by comparing it to the actual price data.

3. METHODOLOGY

The best way to interpret financial data, and stock index data, is as a time series in which the functions of several independent variables are entangled. This is the case because the stock market is a time series. This is the case for a wide range of different applications. Macroscopic factors are those that influence the market over a longer period of time, whereas microscopic variables are those that have an effect on the market more quickly.

These elements fall into two primary categories: the first is microscopic variables, whereas the second is macroscopic factors. Even though the financial market is a complex system that is comprised of elements that are always subject to change, it is still feasible to make educated choices by conducting research on both the macro and the micro levels.

On the other hand, due to the dynamic character of these systems, it is reasonable to suppose that they are governed by some process that cannot be directly viewed. This belief is supported by the fact that it is rational to believe that they are governed. It is not possible to compile a comprehensive list of all the conceivable elements, regardless of how large or tiny they may be, together with the levels of importance that are associated with each of those aspects. This is a direct result of the previous point.

Any sort of financial time series analysis must begin with the observation of data taken from a financial time series. This is the foundation upon which the analysis is built. Trading experiments typically make use of the rate of return as a metric because it does not depend on the size of the investment and because, as a stable series, it exhibits great statistical properties. This is one of the reasons why the rate of return is so widely employed. In addition, the rate of return is widely utilised as a metric since it is unaffected by the total amount of money invested.

If we assume that the price of the asset at time t is denoted by the symbol Pt, then the only way to obtain a simple gross return over a single period is to hold onto that asset between the times t-1 and t, while ignoring dividends and dividend returns along the way. This is the only way to obtain a simple gross return over a single period.

$$P_t = P_t(1+R_t)$$

The following must be used to calculate the same kind of simple return throughout the course of a single period:

$$R_t = (P_t/(P_{t-1}))-1$$

Financial time series display several statistical characteristics that are distinguishable from those of other forms of time series that are more conventional. These characteristics include:

Outliers:

The information that is obtained because of a financial time series having been subjected to a significant impact by macroscale factors, such as those that affect people and policies enacted by governments, may be significantly different from the initial data. This is because macroscale factors have the potential to have an outsized influence on the information that is obtained.

Trend:

Any financial time series, regardless of its dimension, can display a definite upward or decreasing trend during intermediate and extended time periods. This is true even when looking at smaller time periods.

Mean reversion:

The tendency of financial time series to return to their mean level is referred to as the mean reversion phenomenon. This phenomenon reflects the tendency of financial time series.

Fluctuation focus:

Variants on a larger scale are usually always followed by variations on an even larger scale, but variations on a smaller scale are frequently followed by variations on a smaller scale.

LSTM classification

The LSTM model can be found within the RNN family of models, which also includes several other model types. The key innovation that it presents is a self-loop architecture, which it introduces. This design creates a path for a gradient that can flow continuously for a significant amount of time. By making an additional adjustment to the weight of the self-loop at each iteration, it is feasible to circumvent the issue that arises as a natural consequence of the RNN model updating the weights, which is the disappearance of the gradient.

The modelling of time series places a significant emphasis on the nonlinear fitting of parameters as its core component. The LSTM model achieves outstanding results in the pursuit of the goal of stock prediction. It does this by shedding light on the correlation of a nonlinear time series within the delay state space. Utilizing the delay state space is one method for accomplishing this goal. Through the process of analysing historical stock and forex data, the LSTM-based trend prediction system gained the ability to recognise key patterns, which allowed it to become more accurate.

The recurrent issues of gradient explosion and gradient disappearance in RNN inspired the development of the LSTM network model, which was created as a solution to these issues. The most basic form of RNN has a straightforward design and is composed of a single module that is iterated repeatedly. This is going to be a brown coating most of the time. On the other hand, the operation of four of the LSTM modules is extremely like the operation of the standard RNN modules, and the way in which they interact with one another is also highly specific. The output gate, the input gate, and the forget gate are the three components that make up the LSTM memory cell.

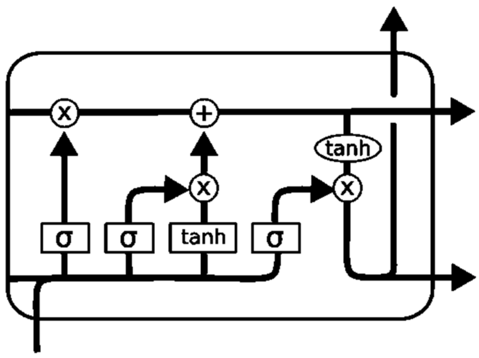

Figure 2: LSTM Architecture

$C_{t\text{-}1}$ – past cell state,

$h_{t\text{-}1}$ - final output,

x_t – present input,

σ - activation function,

f_t - forget gate output,

C_t - status of candidate cell,

o_t - output gate value,

C_t - cell state,

h_t - output.

The LSTM process is given below:

(1)

To build an LSTM, the following stages need to be completed in the correct order: (1) The forget gate requires both the current time and the newly calculated value to be entered into it to function properly. The following formula can be used to figure out the output of the forget gate:

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

where,

W_f - forget gate weight,

b_f - forget gate bias,

x_t – current input value,

h_{t-1} – previous output value.

The data for the input gate is comprised of both the currently read value from the input gate and the most recent value read from the output gate. After you have finished a series of mathematical computations, you will have both your output value and the potential state of the input gate cell.

The following is applied to the cell, the current state of the cell is changed:

$$C_t' = f_t * C_{t-1} + i_t * C_t'$$

where the value range of

Ct is 0 to 1.

Both the most recent value to come out of the output gate as well as the value that is currently being received at the input gate are both given into the output gate. We can compute the output value of the output gate, which is determined to be as follows:

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

where

W_o - output gate weight, and

b_o - output gate bias.

We can determine the output value of the LSTM, which requires that we make use of the output of the output gate in conjunction with the current state of the cell.

$$h_t = o_t * \tanh(C_t)$$

4. RESULTS AND DISCUSSIONS

In this section, various stock price index datasets are utilized to predict the performance of the proposed model. This includes S&P500, DJIA, Nikkei225, HIS, CSI300 and ChiNext. The simulation is conducted in python, and it is carried out in a high-end computing system. The proposed method is compared in terms of prediction accuracy, MSE and RMSE. The proposed method is compared with existing methods that includes ANN, ARIMA, LSTM and GRU.

Table 1: Prediction Accuracy of proposed LSTM and BP

Index	Using PSR	Accuracy	MSE	RMSE
S&P500	Yes	98.175	0.084	1.131
	No	91.031	0.044	0.819
DJIA	Yes	95.222	0.089	0.637
	No	93.143	0.068	0.871
Nikkei225	Yes	92.268	0.167	2.535
	No	88.028	0.085	1.053
HSI	Yes	94.644	0.063	0.923
	No	93.308	0.056	0.767
CSI300	Yes	93.506	0.045	1.040
	No	90.651	0.045	0.975
ChiNext	Yes	94.430	0.277	5.174
	No	89.793	0.145	3.276

Table 2: Prediction Accuracy of proposed LSTM and BP w.r.t hidden nodes

Hidden Node	Accuracy	MSE	RMSE
15	87.00	0.051	0.001
18	90.00	0.053	0.003
20	92.85	0.052	0.004
24	91.80	0.050	0.002
28	86.10	0.052	0.002
32	95.70	0.055	0.004
36	89.10	0.051	0.003
40	87.15	0.055	0.005
48	93.90	0.055	0.005
56	92.85	0.062	0.007

Table 3: Prediction Accuracy of proposed LSTM and BP w.r.t to number of layers

Layer	Accuracy	MSE	RMSE
1	91.84	0.061	0.004
2	96.00	0.053	0.003
3	92.08	0.057	0.003
4	94.24	0.097	0.008
5	90.72	0.119	0.013
6	86.72	0.140	0.011
7	85.12	0.016	0.011
8	83.52	0.022	0.006

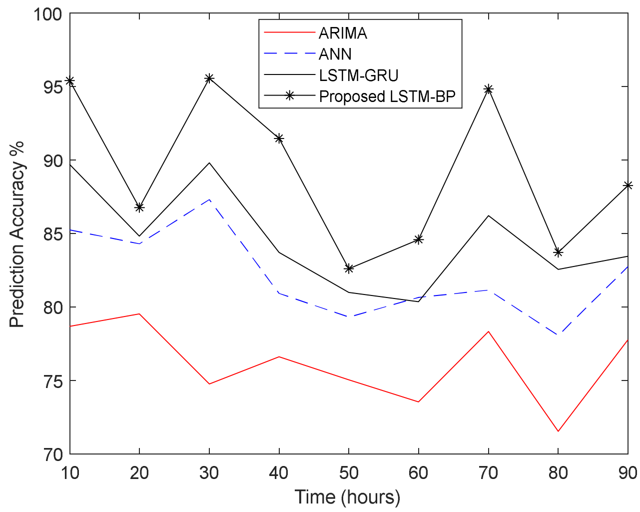

Figure 3: Accuracy

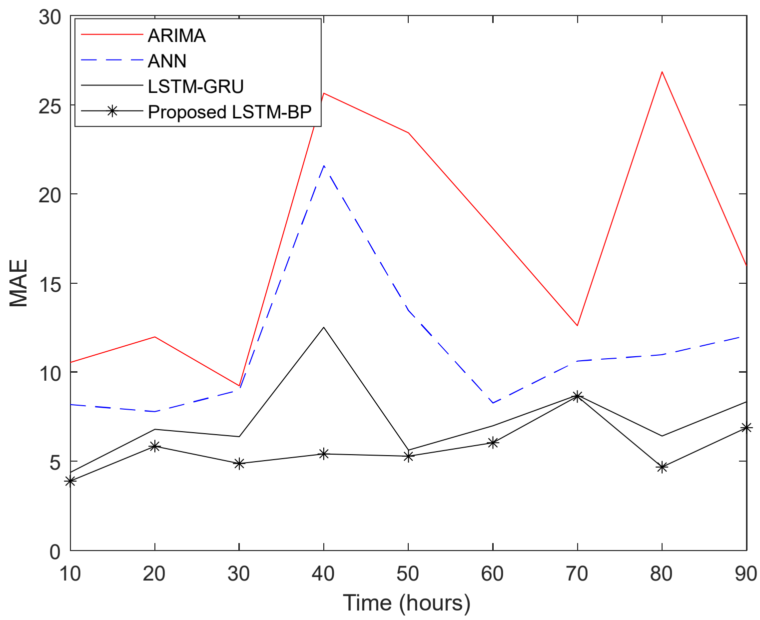

Figure 4: MAE

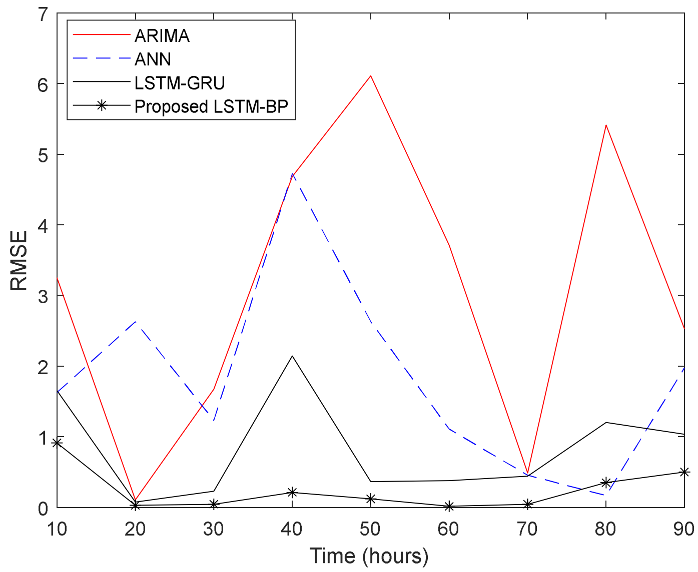

Figure 5: RMSE

The experimental design entails several significant processes, some of the most important of which are the categorization of markets and the selection of qualities. In addition, the accuracy of predictions made by six distinct categorization algorithms is compared in this study. The data for the study comes from ten distinct financial markets, and there are three characteristics associated with each stock as in Figure 3 - 5.

The suggested algorithm is superior to other methods of prediction, and its total accuracy is higher when compared to that of the historical data of a single forecast. In addition, there is only room for one forecast to be used in this comparison. The findings of the tests indicate that utilising leading indicators as experimental data is preferable to merely utilising historical data, and that utilising options is more accurate than utilising futures on their own. Additionally, the results of the tests indicate that utilising options is preferable to utilising futures on their own.

5. CONCLUSION

The proposed model utilises a technique of ANNs known as the BP approach to create predictions regarding the patterns of stock prices. BP approach to computing with neural networks, which was created specifically for use with financial data Standard strategic criteria for predicting stock prices, such as RMSE and MAPE, are used in the evaluation of the models so that conclusions can be drawn about their accuracy. The findings of the simulation indicate that the method that was provided is superior to the approaches that were utilised in the past in terms of the accuracy with which it forecasts the cases. It is found that the LSTM framework performed significantly better than the conventional statistical methods and the standard prediction tasks. This will be done because the primary objective of this piece of writing is to forecast the rise and fall of the stock market. Because predicting the rise and fall of the stock market is the primary purpose of this piece of writing, the next step will be the creation of an expert system for the management of investments after that step is complete.

REFERENCES

[1] Saha, S., Gao, J., & Gerlach, R. (2022). A survey of the application of graph-based approaches in stock market analysis and prediction. *International Journal of Data Science and Analytics*, *14*(1), 1-15.

[2] Soni, P., Tewari, Y., & Krishnan, D. (2022). Machine Learning approaches in stock price prediction: A systematic review. In *Journal of Physics: Conference Series* (Vol. 2161, No. 1, p. 012065). IOP Publishing.

[3] Liu, S., Zou, Y., Ji, W., Zhang, Q., Ahmed, A., Han, X., ... & Zhang, S. (2022). Energy-saving potential prediction models for large-scale building: A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, *156*, 111992.

[4] Wu, D., Wang, X., & Wu, S. (2022). Jointly modeling transfer learning of industrial chain information and deep learning for stock prediction. *Expert Systems with Applications*, *191*, 116257.

[5] Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. (2022). Predicting stock market index using LSTM. *Machine Learning with Applications*, *9*, 100320.

[6] Sisodia, P. S., Gupta, A., Kumar, Y., & Ameta, G. K. (2022, February). Stock market analysis and prediction for NIFTY50 using LSTM Deep Learning Approach. In *2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)* (Vol. 2, pp. 156-161). IEEE.

[7] Chen, Y., Wu, J., & Wu, Z. (2022). China's commercial bank stock price prediction using a novel K-means-LSTM hybrid approach. *Expert Systems with Applications*, *202*, 117370.

[8] Nikou, M., Mansourfar, G., & Bagherzadeh, J. (2019). Stock price prediction using DEEP learning algorithm and its comparison with machine learning

algorithms. *Intelligent Systems in Accounting, Finance and Management*, *26*(4), 164-174.

[9] Fazeli, A., & Houghten, S. (2019, December). Deep learning for the prediction of stock market trends. In *2019 IEEE International Conference on Big Data (Big Data)* (pp. 5513-5521). IEEE.

[10] Lakshminarayanan, S. K., & McCrae, J. P. (2019, December). A Comparative Study of SVM and LSTM Deep Learning Algorithms for Stock Market Prediction. In *AICS* (pp. 446-457).

[11] Xu, Y., & Keselj, V. (2019, December). Stock prediction using deep learning and sentiment analysis. In *2019 IEEE international conference on big data (big data)* (pp. 5573-5580). IEEE.

[12] Rana, M., Uddin, M. M., & Hoque, M. M. (2019, December). Effects of activation functions and optimizers on stock price prediction using LSTM recurrent networks. In *Proceedings of the 2019 3rd International Conference on Computer Science and Artificial Intelligence* (pp. 354-358).

[13] Nguyen, D. H. D., Tran, L. P., & Nguyen, V. (2019). Predicting stock prices using dynamic LSTM models. In *Applied Informatics: Second International Conference, ICAI 2019, Madrid, Spain, November 7–9, 2019, Proceedings 2* (pp. 199-212). Springer International Publishing.

[14] Naik, N., & Mohan, B. R. (2019). Study of stock return predictions using recurrent neural networks with LSTM. In *Engineering Applications of Neural Networks: 20th International Conference, EANN 2019, Xersonisos, Crete, Greece, May 24-26, 2019, Proceedings 20* (pp. 453-459). Springer International Publishing.

[15] Hossain, M. A., Karim, R., Thulasiram, R., Bruce, N. D., & Wang, Y. (2018, November). Hybrid deep learning model for stock price prediction. In *2018 ieee symposium series on computational intelligence (ssci)* (pp. 1837-1844). IEEE.

[16] Lai, C. Y., Chen, R. C., & Caraka, R. E. (2019, July). Prediction stock price based on different index factors using LSTM. In *2019 International conference on machine learning and cybernetics (ICMLC)* (pp. 1-6). IEEE.