Communications in Mathematics and Applications
Vol. 15, No. 2, pp. 679–695, 2024
ISSN 0975-8607 (online); 0976-5905 (print)
Published by RGN Publications
DOI: 10.26713/cma.v15i2.2595

http://www.rgnpublications.com

Research Article

Estimation of Stock Price Using LSTM Algorithm and Sentiment Analysis

Varsha Bendre*[1], Deepti Khurge[1], P. K. Rajani[1], Jayashree Katti[2], M. T. Kolte[1] and Sayali Nemade[1]

[1] *Department of Electronics and Telecommunication Engineering, PCET's Pimpri Chinchwad, College of Engineering, Pune, India*
[2] *Department of Information Technology, PCET's Pimpri Chinchwad College of Engineering, Pune, India*
***Corresponding author:** varsha.bendre@pccoepune.org

Received: February 17, 2024 **Accepted:** April 9, 2024

Abstract. Predicting price of the stock is a complex job and it is almost impossible to predict the exact amount of the stock. However, it's possible to estimate future amount of the stock through searching at historical information. There are various machine learning models available with which stock price can be estimated. In this paper, initially a detailed survey is presented that is based on a comparison and examination of various deep learning algorithms for estimation of stock price. It includes Linear Regression, Support Vector Machine, Auto Regressive Integrated Moving Average, Convolutional Neural Network, Long Short-Term Memory, Artificial Neural network. It is observed that the LSTM algorithm has a better performance as compared to the others. In this work a univariate model is trained to estimate the price of the stock using LSTM model. To improve the results further a sentiment analysis based on twitter data is performed and a multivariate LSTM model is trained. The results presented in this work are based on three stocks Britannia, HDFC bank and Axis bank. The state-of-the-art natural language processing tool BERT are used to recognize the sentiments of text, and the long short term memory neural network (LSTM), which is good at analyzing time series data, is applied to forecast the stock price with stock historical transaction information and text sentiments The various performance parameters are also discussed in this paper. The analysis is performed based on changing number of epochs and it is observed that error percentage reduced with increasing number of epochs.

Keywords. Stock price prediction, Convolutional Neural Network, Training, LSTM, Sentiment analysis, Deep learning

Mathematics Subject Classification (2020). 68 Computer Science

1. Introduction

The stock market is a market where publicly traded corporations' shares are issued and traded. Stock investments are common investment activity and great source of secondary income for common people around the world. However, stock price is affected by various physical and psychological factors. The major factors affecting the stock prices are global market conditions, fundamentals and debt policy of the company (Adebiyi *et al*. [27]), government policies and bonds, news, twitter and other social media reviews. Because of this, it is challenging to determine the stock's price. Even identifying whether a given stock's price will climb or fall is challenging. Without the knowledge of future prediction of stock price, successful investments are difficult. However, a prediction model can be implemented and estimated price of stock can be predicted using historical data related to that stock and current market trends.

The proposed approach considers available historic data of a share and it provides prediction on a particular feature. The features of shares are opening price, day high, day low, previous day open price, previous day close price, date of trading, total trade quantity and turnover. The proposed model uses the time series analysis in order to predict a share price for a required time span. The National Stock Exchange (NSE) is the Indian stock exchange entity, the NSE was the first exchange in India to provide a modern, provides latest facility to the investors spread across the length and breadth of the country. It has thoroughly modern with all latest facilities, which provides investors with the facility to trade from anywhere in India. This has a decisive role in reforming the Indian equity market to add increased

Transparency, convergence and efficiency to the capital market. NSE's Common Index, The CNX NIFTY, is used prodigiously by the investor across India as well as globally. It provides accommodation for the exchange, settlement and clearing in equity and debt market and

Additionally in derivatives. This is one of India's most astronomically enormous mazuma, currency and index options trading exchanges worldwide.

Share price forecasting models can be created using a variety of deep learning methods. Many such deep learning models are available in literature and each one is implemented with either physical factors or psychological factors with different accuracy of prediction. In this paper, a performance comparison of various deep learning algorithms like Linear Regression (LR), Support Vector Machine (SVM), Auto Regressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM), Artificial Neural network (ANN), Convolutional neural network (CNN) [2] is done to find out the most accurate algorithm.

Nowadays, there has been a debate on the effectiveness of the sentiments conveyed via social media in forecasting the change in the stock market. Various researchers have revealed that sentiments might influence the stock market movement and act as potential predictors for trade-off outcomes.

2. Literature Review

Most of the previous work in stock price prediction use classical algorithms (Vijh *et al*. [29]) like linear regression (Seber and Lee [25]), Random Walk Theory (RWT) (Reichek and Devereux [4]), Moving Average Convergence/Divergence (MACD) (Chong and Ng [5]) and using some linear models like Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA) [31], for predicting stock prices. Recent research demonstrates that deep learning can improve share price prediction. A number of neural network-based techniques, including Support Vector Machine (SVM), Random Forest (RF) (Zhang [14]), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and deep neural networks like Long Short-Term Memory (LSTM), have also produced remarkable results (Li *et al*. [13], and Oyeyemi *et al*. [21]). In this section, survey on these techniques and various performance metrics that can be used for stock price prediction is explicated.

Wijesinghe *et al*. [30] examined the ANN and ARIMA models to determine which sectors and companies contribute the most to the All-Share Price Index (ASPI) value in the Central Stock Exchange (CSE). Twenty industries' ASPI values from 2008 to 2017 and nine banks' daily closing costs were gathered from the CSE database, along with other financial records. The outcomes showed that modern deep learning algorithms, like ANN, surpass older ones, like the ARIMA model. With publicly accessible market data from the NY Stock Exchange, Adebiyi *et al*. [1] examined the predicting effectiveness of ARIMA and artificial neural networks algorithms. As per the observation the ARIMA forecasting model is directional and ANN performance is superior that the ARIMA. Mehtab and Sen [17] predicted NIFTY 50 index values using five deep learning-based regression models, including two CNN models and three LSTM models. The models are trained and tested using data from December 29, 2008 to July 31, 2020. With the previous two weeks' data as input, it is observed that the univariate encoder-decoder convolutional LSTM was the most accurate model. The univariate CNN model with the prior week's data as input, on the other hand, was the quickest.

Rasheed *et al*. [24] proposed LSTM and CNN model. The dataset was chosen because of the steep decline and ramping behavior of stock values over time. Principal Component Analysis (PCA) is a technique for determining the most important feature. Data was gathered from Yahoo Finance from January 1, 2008, through September 15, 2020. Both LSTM and 1D-CNN were found to be effective in accurately predicting stock value. LSTM with PCA, on the other hand, generated slightly better outcomes. Siami-Namini *et al*. [26] looked at whether and how the newly designed LSTM outperformed the traditional ARIMA algorithm. Data was extracted from Yahoo Finance from January 1985 to August 2018, while monthly economics time series were acquired from the Federal Reserve Bank of St. Louis and the International Monetary Fund (IMF) Website for various time periods. When compared to ARIMA, the LSTM-based system improved prediction by 85% on average. For a comparative examination of performance, Lakshminarayanan *et al*. [12] developed four LSTM and four SVM models. The Dow Jones Index (DJI) stock price dataset was used, as well as a combination of this stock price dataset with externally contributed crude oil and gold price parameters. Data mining in Architecture is based on a cross-industry standard process. The results show that the LSTM outperforms the SVM in all circumstances. This is due to its superior capacity to remember and forget data compared to SVM.

Rajput and Kaular [23] proposed an ANN and SVM model for predicting SBI and L&T stock closing prices. Models for SBI were trained for 2012 and tested for 2013 data, while models for L&T were trained for 2015 and tested for 2016. When compared to ANN, the results showed that SVM produced better outcomes. Panwar *et al*. [22] compared SVM and LR models. They used web scrapping to collect the three months Amazon stock data. Support vector do not work well with the noisy and unlabeled data. The results of the comparison showed that linear regression is superior to SVM.

3. Deep Learning Algorithms

The deep learning algorithms are widely used for stock price prediction. In this section, various deep learning algorithms are discussed.

3.1 Support Vector Machine (SVM)

A supervised learning model is an SVM. Regression analysis and data categorization can both benefit from it. It works well for identifying trends and regression. SVM is frequently used to solve classification issues including classifying linear and nonlinear data. The fundamental working tenet of SVM is to construct a hyperplane that classifies the data. The information in Figure 1 is clearly separated by the Maximum Margin Hyperplane (MMH). For hyperplane selection, only the points closest to the border are crucial; all other points are irrelevant. These points are known as supporting vectors, and the hyperplane used to classify them is known as the Support Vector Classifier (SVC) [16].

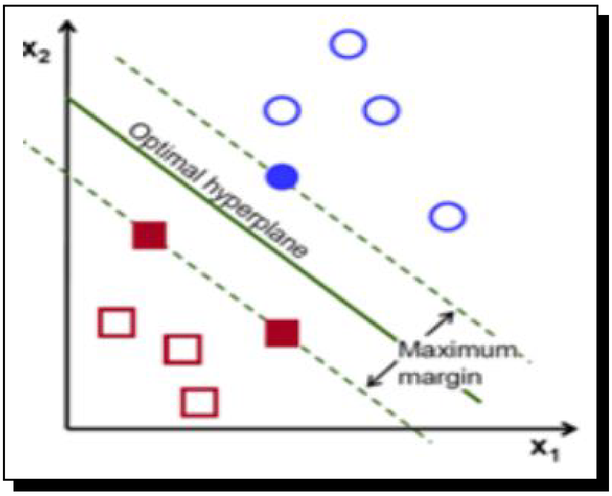

Figure 1. Support Vector Machine

3.2 Auto Regressive Integrated Moving Average (ARIMA)

ARIMA is a class of models that define a certain time series based on its past values. ARIMA is used for training the dataset and for generating the short-term forecasts. In short-term forecasts, it always exceeds the complex building models. The ARIMA model is a stand-alone series also called the ARMA model (p,q) (Ma [15], and Meyler *et al*. [18]). The upcoming values of a variable in the ARMA model is just a linear product of the previous value and the previous error, as shown in the following expression:

$$X_t = \mu + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}\,, \tag{3.1}$$

where

μ_t = actual value at t,	ϕ_t = parameters of autoregressive part of a model,
θ_t = parameters of moving average part,	ε_t = error term,
p = autoregressive integer,	q = moving average.

This is valid for stationary processes, but in order to create a new stationary sequence from a nonstationary sequence, specific changes are necessary. Differences can be created to convert nonstationary sequence with short term trends in to the stationary time series.

3.3 Long Short-Term Memory (LSTM)

Long-term connectivity learning is a specialty of the LSTM form of the recurrent neural network. It features a series-like structure with four neural network layers that communicate with one another. The cell state, shown by the top-to-bottom horizontal line in Figure 2, is the key to LSTMs. By deleting or introducing information, it can change a cell's state. Gates control how this happens. It has three gates to protect and control the condition of the cell.

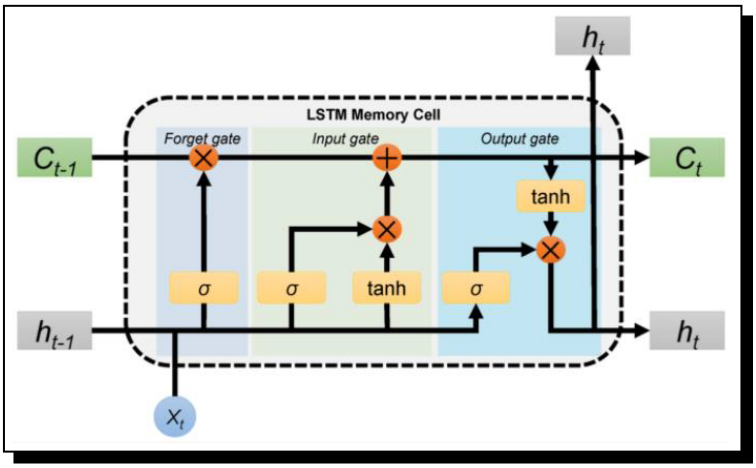

Figure 2. LSTM memory cell

Forget gate: Removing the unwanted information from the state of cell is the initial stage in LSTM. The sigmoid layer is responsible for taking this decision. Sigmoid layer is also called forget gate. For respective number in the state of cell C_{t-1}, sigmoid layer produces a value between 0 and 1 after checking at h_{t-1} and x_t.

This decision is taken by a sigmoid layer also termed as the "forget gate layer". It looks at h_{t-1} and x_t and outputs a number between 0 and 1 for each number in the cell state C_{t-1}. Value 1 stands for totally accept this information and 0 for totally clear this information.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f). \tag{3.2}$$

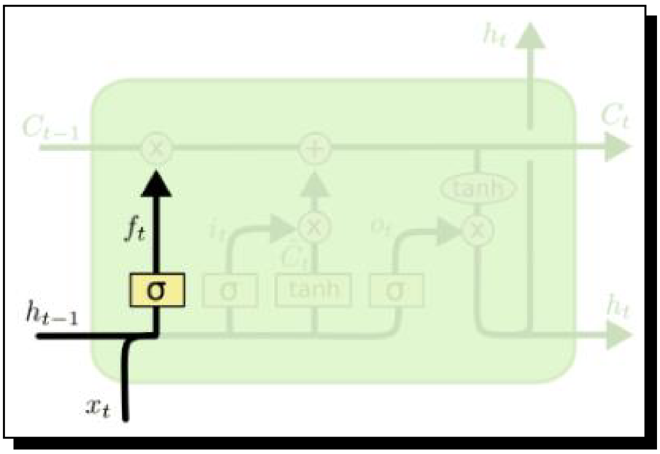

Figure 3. Forget gate

Input gate: Once the unwanted information is removed in first step, making a choice what new data to store in the state of cell. Input gate contains two layers. First layer called as input layer and second is tanh layer. Input gate layer determine value needs to be updated and tanh layer produces a vector of possible new values, $\widetilde{C}_t$, that the state might include. Combining the output of input gate layer and tanh layer results in a state update during the output stage.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{3.3}$$

$$\widetilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c). \tag{3.4}$$

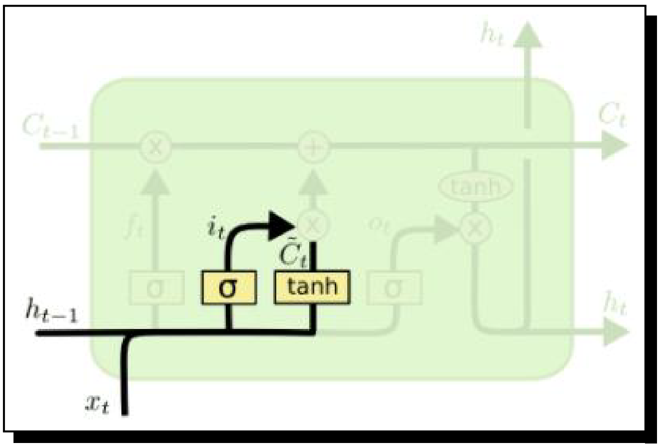

Figure 4. Input gate

Output gate: Last gate is the output gate which changes the old state of cell, C_{t-1}, into the new state of cell C_t. First step in the output gate layer is to execute sigmoid function. This sigmoid layer decides which portions of the state of cell will be the output. Later, combine the outcome of the sigmoid gate by the state of cell after passing it through tanh to force the values to be between -1 and 1 such that the output has the predetermined components (Taneja and Vaibhav [28]).

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t, \tag{3.5}$$

$$o_t = \sigma(W_0 \cdot [h_{t-1}, x_t] + b_0), \tag{3.6}$$

$$h_t = o_t * \tanh(C_t). \tag{3.7}$$

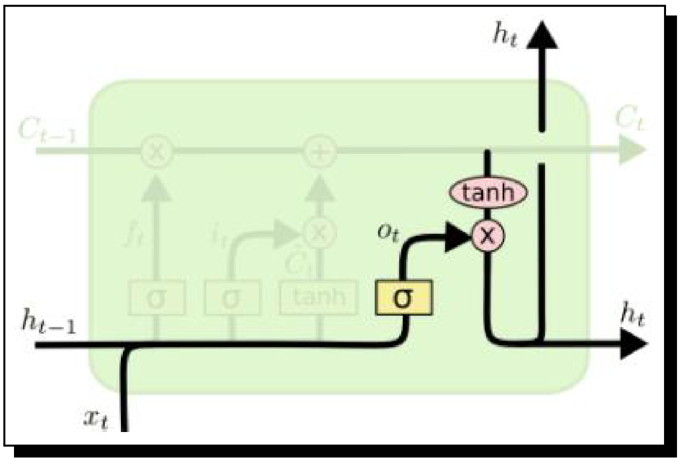

Figure 5. Output gate

3.4 Linear Regression (LR)

Linear regression is a statistical Method. It is used to forecast how the dependent variable will interact with independent variables or vice versa. It is represented as below:

$$P = S + Q * R \tag{3.8}$$

where

P = dependent variable,
R = independent variable,
S = constant,
Q = slope of the regression line (Taneja and Vaibhav [28]).

It performs a regression task using supervised learning as its foundation. Regression modelling forecasts the target value based on the independent variables to determine the link between the variables and forecasting (Aissaoui *et al*. [3]). Output of the regression model is degree equation. This equation establishes the correlation between the variables. Below is the equation for regression:

$$S = \theta_1 R_1 + \theta_2 R_2 + \cdots + \theta_n R_n, \tag{3.9}$$

where

$R_1, R_2, \ldots, R_n$ = freelance variables,
$\theta_1, \theta_2, \ldots, \theta_n$ = weights (Gupta and Nagalakshmi [8]).

3.5 Artificial Neural Network (ANN)

The strength of ANN models in comparison to other nonlinear models lie in their ability to accurately approximate a extensive range of functions (Khashei and Bijari [11]).

There are multiple ANN models out of which multi-layer perceptron is the most significant. MLP comes with the flexibility to map arbitrary inputs and outputs. Due to this inherited capability of MLP, it is mostly used for predictions.

The multi-layer perceptron consists of several layers. Each layer has nodes. Features or input data are received at the input layer. Prediction results are acquired at the output layer, which is the last layer. Between the input layer and the output layer, there exist hidden layers. The number of hidden layers depends on the issue description. Figure 6 shows an example of a fully linked MLP with one hidden layer operation that an ANN may perform through the connecting arcs (Zhang *et al*. [32]).

Below is the equation of correlation between the variables:

$$y_t = w_0 + \sum_{j=1}^{q} w_j g\left(w_{0,j} + \sum_{i=1}^{p} w_{ij} y_{t-i}\right) + \varepsilon_t, \tag{3.10}$$

where

w_i $(j = 0,1,2,3,\dots,q)$ and w_{ij} $(i = 0,1,2,3,\dots,p;\ j = 0,1,2,3,\dots,q)$ = connection weight,

p = number of input nodes,

q = hidden nodes,

y_t = output variable,

$y_{t-1}, y_{t-2}, \dots, y_{t-p}$ = input variables (Khashei and Bijari [11]).

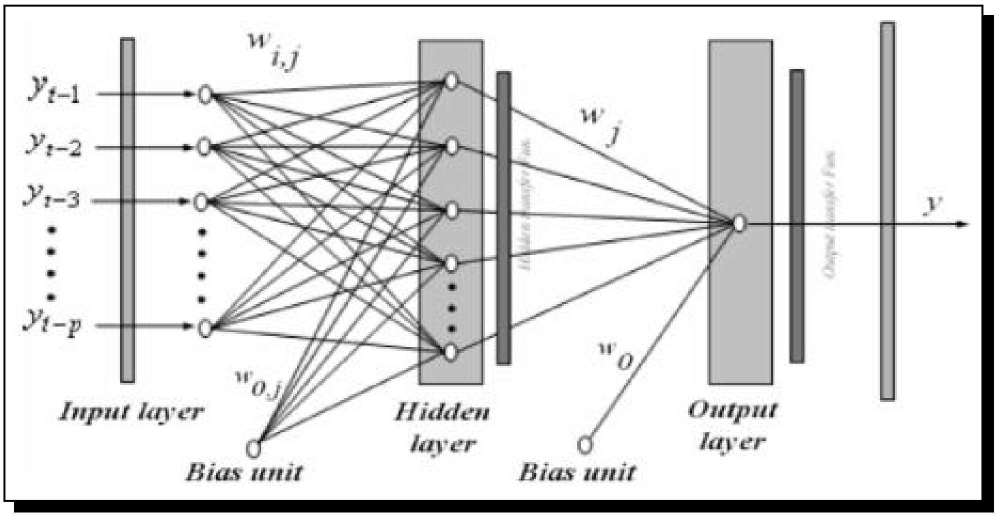

Figure 6. Neural Network Structure [29]

3.6 Convolutional Neural Network (CNN)

Convolution and pooling layers make up CNN's two layers. The placement of these two layers is alternated, one after the other. The objective of this layered sequence differs from conventional machine learning methods as it aims to generate high-level features without the need for human involvement in the feature engineering stage. The pooling layers reduce the size of the feature map that the convolutional layers output. More fully linked layers are joined at the network's conclusion to increase nonlinearity and produce the final output (Hoseinzade and Haratizadeh [10]).

In addition to the above algorithms, KNN, Prophet, Random Forest Regressor, bidirectional LSTM are also used for understanding the effect of sentiment analysis on stock price prediction for Indian markets [6].

4. Performance Metrics

In regard to several performance metrics, the efficacy of deep learning systems for stock price prediction may be examined. This part reviews the performance metrics and their values as they were obtained in several publications for various equities using the algorithms covered in Section 3.

4.1 Root-Mean-Square Error (RMSE)

The RMSE is a measurement of inaccuracy or correctness. It is the difference between the values predicted by a model and the values observed, which is computed by multiplying the sum of the squares of the discrepancies between the expected and actual values by the number of samples (Ahuja *et al*. [2], Gupta and Chen [7], Gururaj *et al*. [9], Nandakumar *et al*. [20]).

$$\text{RMSE} = \sqrt{\frac{\sum_{n=1}^{n} (Y_i - \widehat{Y}_i)^2}{n}}. \tag{4.1}$$

The RMSE values for different Deep Learning Models obtained by Darapaneni *et al*. [6] are summarized in Table 1.

Table 1. RMSE Values for different models

Models	RMSE
LSTM	38.19
Bidirectional LSTM	184.29
Linear Regression	1030.83
Arima	532.64
KNN	1273.05
Prophet	311.46
Random Forest Regressor	580.49

4.2 Mean Absolute Percentage Error (MAPE)

MAPE serves as a statistical measure for assessing the prediction accuracy of a forecasting method. Typically, it measures accuracy by representing it as a ratio defined by a specific formula:

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|. \tag{4.2}$$

Here, A_t represents the actual value, and F_t stands for the forecast value. The division of their difference by the actual value A_t is performed. The absolute values in this ratio are summed for each forecasted point in time and then divided by the number of fitted points, denoted as n.

The MAPE values obtained for Apple stock using different models is presented in Table 2.

Table 2. MAPE Values for different models [19]

Models	MAPE
ARIMA	7.39
Prophet	7.98
RNN-p	2.13
RNN-pp	2.03
RNN-pt	2.17
RNN-mv	10.43

4.3 Mean Absolute Error (MAE)

The average of absolute error levels is referred to as the MAE. The difference between the expected and actual values is divided by the number of samples is calculated in order to obtain the value of MAE as shown in equation (4.3).

$$\text{MAE} = \frac{\sum\limits_{i=1}^{n} |y_i - x_i|}{n}. \tag{4.3}$$

4.4 Mean Squared Error (MSE) or *R*-Squared

MSE measures the average of the squares of the errors that is, the average squared difference between the predicted values and the actual value:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2. \tag{4.4}$$

In addition to the above performance parameters. The F-score and Area Under Curve (AUC) is also used to check the stock price prediction potential of the model.

5. Methodology

From the tables above, the LSTM models are more accurate and have a lower error rate than the other algorithms. Consequently, LSTM univariate and multivariate models are developed and analyzed in this paper.

5.1 Univariate LSTM Model

In this paper, a univariate LSTM model for stock price prediction of Britannia, HDFC Bank and Axis bank is developed. The previous day close price of the stock is used as an input variable to the LSTM model. With the neuron size one and number of epochs 60, following results are obtained for MAPE.

Table 3. Univariate LSTM model results

Dataset	Variable	MAPE
Britannia	Close price	8.79
HDFC Bank	Close price	19.49
Axis Bank	Close price	25.57

To improve the results further, LSTM with sentiment analysis is used. Sentiment analysis is studying the emotions of the user. Customer feedback plays the vital role in any kind of business. Nowadays, social media is the common platform where users add comments daily. Somehow these comments are related to the flow of share market. To analyze this and to improve the univariate results, The multivariate LSTM model is implemented.

5.2 Multivariate LSTM Model

The block diagram of Multivariate LSTM Model with sentiment analysis is shown in Figure 8.

First layer is the LSTM layer or input layer. Then the 4 hidden layers are added and finally the output layer.

5.3 Data Collection

Financial data for the Britannia stock is collected from the yahoo finance with the help of ticker “BRITANNIA.NS”. This fetched data has Date, High, Low, Close, Open, Adj Close, Volume price of the stock. For the sentiment analysis tweets posted by the users with the twitter id like “@BritanniaIndLtd” are collected from the twitter with the help of twitter developer account. Both the financial data and twitter data fetched from 1st January 2012 till current date.

5.4 Pre-Processing

Data fetched from the twitter and yahoo finance needs processing before applying it to LSTM algorithm. In this section procedure followed for the processing is discussed. Initially, the null values are removed. Calculated the stock trend by taking the difference between the closing rates. Twitter data has the emojis, hash tags and words which do not contribute in sentiment analysis. Such tags and emojis are removed using the natural language processing library. Later, the sentiment score using the textblob library in python is calculated. Three emotions positive, negative and neutral and stock trend based on the sentiment score are considered for stock developing the model. The multiple variables like Volume, Adj Close, Positive, negative, neutral are used to build the multivariate LSTM model.

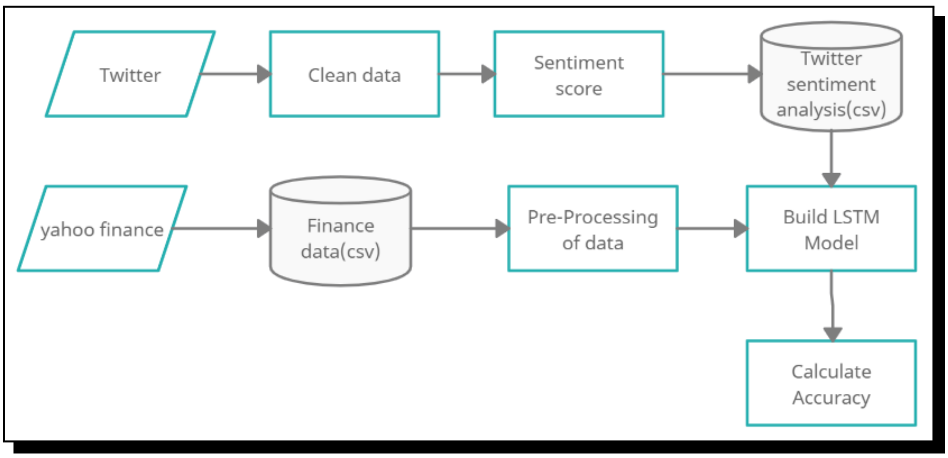

Figure 7. Block diagram

The raw data obtained from internet news articles and forum posts through a crawler cannot be directly fed into deep learning models. Subsections below describe the preprocessing of raw data. After collection of the news related to stock from Internet, for preprocessing, all news related to individual stocks are kept and the news are separate into sentences. If the the article is too long then the accuracy of SA will be less according to the experiment. In this work, sentiments in an article are recognized according to sentences in it.

The preprocessing of forum posts from PTT involves a rigorous approach. The board managers on the PTT stock board strictly permit posts falling within nine defined categories: NEWS, EXPERIENCE, SUBJECT, QUESTION, INVESTMENT ADVICE, TALK, ANNOUNCE, QUESTIONNAIRES, and OTHERS.

5.5 LSTM Model

Figure 8 shows the architecture of the designed multivariate LSTM model. Neurons r units selected for each layer are as shown in Table 4.

Table 4. Number of Neurons for Multivariate LSTM model

Layer	Neuron size
LSTM layer	6
Hidden layer-1	5
Hidden layer-2	4
Hidden layer-3	3
Hidden layer-4	2
Output Layer	1

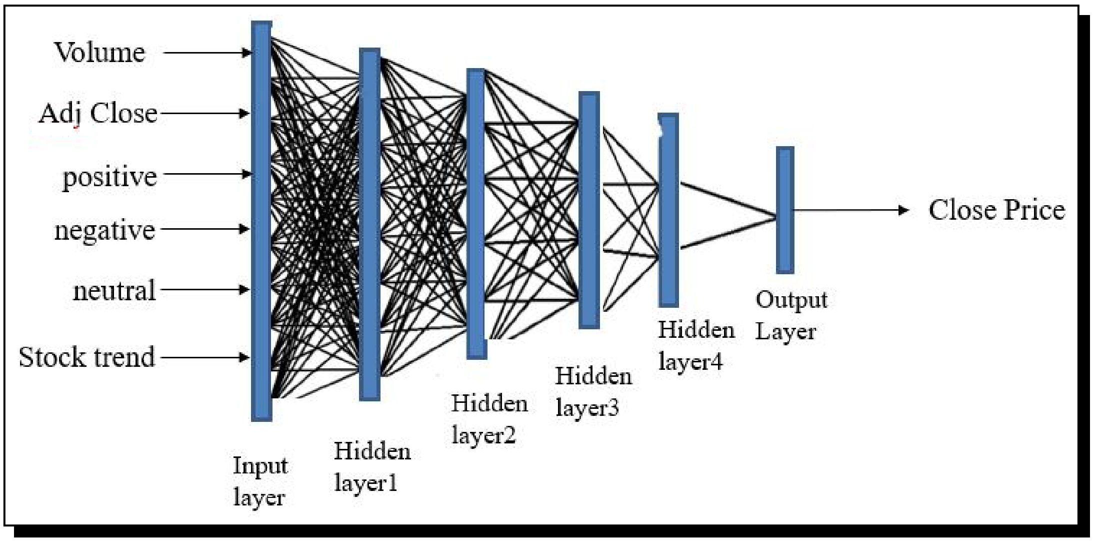

Figure 8. LSTM model

6. Experimental Results

Epochs are the iterations which denotes how many times the machine learning algorithm has gone over the full training dataset. In this paper, employed 10, 20, 30, 40, 50, and 60 epochs to investigate how epochs affected the outcomes. It is noticed that as the epoch increases, the results get better and better with the increase in epoch. After 60 epoch results were not changing.

Observations of the error percentage with the different epochs size are mentioned in Table 5.

Table 5. MAPE values for different number of epochs and different stocks

Number of Epochs	MAPE		
	Britannia	HDFC Bank	Axis Bank
10	17.38	53.53	26.38
20	14.22	18.477	13.87
30	6.57	8.61	4.65
40	3.79	6.46	4.41
50	3.18	6.64	3.46
60	3.03	2.61	2.96

The results obtained for MAPE, RMSE, MAE and MSE for 60 number of epochs is mentioned in Table 6.

Table 6. Performance parameters

Epochs	MAPE	RMSE	MAE	MSE
Axis Bank	2.96	3.46	18.57	767.39
HDFC Bank	2.61	36.10	20.87	770.00
Britannia	3.03	39.73	30.45	385068

The obtained results are then compared with the similar work on stock price prediction as shown in Table 7. With the comparison, it is evident that the proposed model with sentiment analysis gives better performance than other models without sentiment analysis.

Table 7. Model Comparison Results

Model	Reference	Stock	MAPE	RMSE	MSE	MAE
LSTM Base	[12]	DJI stock	1.41	399.39	159519.52	356.04
LSTM Base + MV			1.05	349.54	122183.47	275.73
LSTM Advanced			2.52	683.68	467426.27	603.93
LSTM with Sentiment analysis	Proposed model	HDFC Bank	2.61	36.10	20.87	770.00

7. Conclusion

This research provided a comprehensive assessment of works on estimating stock price using deep learning approaches, as well as a comparison of the various algorithms. This research review included an extensive variety of algorithms Linear regression, SVM, ANN, ARIMA, CNN, LSTM, and their performance comparison to understand the best suitable algorithm for stock price prediction. The major performance metrics of all models are discussed in this paper are RMSE, MAPE, MAE, MSE, accuracy, NMSE. Proposed method of LSTM with Sentiment analysis combines text features from social media, which helps to improve the performance of traditional methods. The stock index variables for finance may represent the expansion trend of the stock price, but the sentiments, thought process of investors can describe the possible trend of the stock price. Such contributions are usually neglected in traditional prediction methods. Public comments are important aspect to study the flow of stock some used sentiment analysis for the multivariate LSTM model. We conclude that multivariate as better accuracy as compared to univariate LSTM model as well as comments posted by the users have relation with the movement of stock. Stock Market working is affected by sentiments. People who know sentiments in market can get better market returns.

Proposed work calculates the daily sentiment probabilities collected with historical stock transaction data as input vector to LSTM neural network to predict next day stocks opening price.

The experimental results states, sentiment analysis features of news articles or PTT posts for the forecast model can reduce the RMSE. When sentimental information is combined to the forecast model, the RMSE reduces.

8. Future Studies

Results stated in the work proposed could be improved by fine-tuning the hyperparameters like SGX NIFTY index, increase the size of the training dataset, and considering various data sources such as previous day closing price of the US stock market. Other machine learning technologies such as Clustering SVM and Clustering Deep Learning improve results. Introduction of fuzzy deep learning into learning and prediction can be used since many news items are fuzzy in terms of their positive or negative impacts. Dynamically changing the data window size based on the type of news, as it is having long-lasting impact on many stock parameters can be considered. These psychological studies when combined with Deep learning problem can give better predictions.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

All the authors contributed significantly in writing this article. The authors read and approved the final manuscript.

References

[1] A. A. Adebiyi, A. O. Adewumi and C. K. Ayo, Comparison of ARIMA and artificial neural networks models for stock price prediction, *Journal of Applied Mathematics* **2014** (2014), Article ID 614342, 7 pages, DOI: 10.1155/2014/614342.

[2] H. Ahuja, S. Mehra, G. Manju, D. Aishwarya and Y. Bhavaraju, Stock price prediction using SVM and LSTM, *International Research Journal of Engineering and Technology* **7**(7) (2020), 4963 – 4968.

[3] O. El Aissaoui, Y. El Alami El Madani, L. Oughdir, A. Dakkak and Y. El Allioui, A multiple linear regression-based approach to predict student performance, in: *Advanced Intelligent Systems for Sustainable Development* (AI2SD'2019), M. Ezziyyani (ed.), *Advances in Intelligent Systems and Computing*, Vol. 1102. Springer, Cham., DOI: 10.1007/978-3-030-36653-7_2.

[4] T. T. Chitenderu, A. Maredza and K. Sibanda, The random walk theory and stock prices: evidence from Johannesburg stock exchange, International Business & Economics Research Journal 13 (6) (2014), 1241 – 1250, DOI: 10.19030/iber.v13i6.8918.

[5] T. T.-L. Chong and W.-K. Ng, Technical analysis and the London stock exchange: Testing the MACD and RSI rules using the FT30, *Applied Economics Letters* **15**(14) (2008), 1111 – 1114, DOI: 10.1080/13504850600993598.

[6] N. Darapaneni, A.R. Paduri, H. Sharma, M. Manjrekar, N. Hindlekar, P. Bhagat, U. Aiyer and Y. Agarwal, Stock price prediction using sentiment analysis and deep learning for Indian markets, *arXivpreprint* arXiv:2204.05783 (2022), DOI: 10.48550/arXiv.2204.05783.

[7] R. Gupta and M. Chen, Sentiment analysis for stock price prediction, in: *2020 IEEE Conference on Multimedia Information Processing and Retrieval* (MIPR), Shenzhen, China (2020), pp. 213 – 218, DOI: 10.1109/MIPR49039.2020.00051.

[8] A. Gupta and T. Nagalakshmi, Stock price prediction using linear regression in machine learning, *International Journal of Innovative Technology and Exploring Engineering* **8**(12) (2019), 1382 – 1385, DOI: 10.35940/ijitee.L3932.1081219.

[9] V. Gururaj, V. R. Shriya and K. Ashwini, Stock market prediction using linear regression and support vector machines, *International Journal of Applied Engineering Research* **14**(8) (2019), 1931 – 1934.

[10] E. Hoseinzade and S. Haratizadeh, CNNpred: CNN-based stock market prediction using a diverse set of variables, *Expert Systems with Applications* **129**(2019), 273 – 285, DOI: 10.1016/j.eswa.2019.03.029.

[11] M. Khashei and M. Bijari, An artificial neural network (p,d,q) model for timeseries forecasting, *Expert Systems with Applications* **37**(1) (2010), 479 – 489, DOI: 10.1016/j.eswa.2009.05.044.

[12] S. K. Lakshminarayanan and J. McCrae, A comparative study of SVM and LSTM deep learning algorithms for stock market prediction, in: *CEUR Workshop Proceedings – AICS* Volume 2563, pp. 446 – 457 (2019), URL: https://ceur-ws.org/Vol-2563/aics_41.pdf.

[13] L. Li, Y. Wu, Y. Ou, Q. Li, Y. Zhou and D. Chen, Research on machine learning algorithms and feature extraction for time series, in: *IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications* (PIMRC2017) Montreal, QC, Canada,2017, pp. 1 – 5 (2017), DOI: 10.1109/PIMRC.2017.8292668.

[14] A. Liaw and M. Wiener, Classification and regression by randomForest, *R News* **2-3** (2002), 18 – 22.

[15] Q. Ma, Comparison of ARIMA, ANN and LSTM for stock price prediction, in: *E3S Web of Conferences, 2020 International Symposium on Energy, Environmental Science and Engineering (ISEESE 2020)*, Vol. 218, (2020), Article Number: 01026, DOI: 10.1051/e3sconf/202021801026.

[16] S. Madge, Predicting stock price direction using support vector machines, *Independent Work Report Spring – 2015*, Department of Computer Science, Princeton University, Princeton, 14 pages (2015), URL: https://www.cs.princeton.edu/sites/default/files/uploads/saahil_madge.pdf.

[17] S. Mehtab and J. Sen, A robust predictive model for stock price prediction using deep learning and natural language processing, in: *Proceedings of the 2019 International Conference on Business Analytics and Intelligence* (ICBAI 2019), December 2019, Bangalore, India (2019).

[18] A. Meyler, G. Kenny and T. Quinn, Forecasting Irish inflation using ARIMA models, *Research Technical Papers 3/RT/98*, Central Bank and Financial Services Authority of Ireland, 46 pages (1998), URL: https://mpra.ub.uni-muenchen.de/11359/1/cbi_3RT98_inflationarima.pdf.

[19] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia and D. C. Anastasiu, Stock price prediction using news sentiment analysis, in: *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications* (BigDataService, Newark, CA, USA, 2019), pp. 205 – 208, (2019), DOI: 10.1109/BigDataService.2019.00035.

[20] R. Nandakumar, K. R. Uttamraj, R. Vishal and Y. U. Lokeswari, Stock price prediction using long short-term memory, *International Research Journal of Engineering and Technology* **5**(3) (2018), 3342 – 3348.

[21] E. O. Oyeyemi, L.-A. McKinnell and A. W. V. Poole, Neural network-based prediction techniques for global modeling of M(3000)F2 ionospheric parameter, *Advances in Space Research* **39**(5) (2007), 643 – 650, DOI: 10.1016/j.asr.2006.09.038.

[22] B. Panwar, G. Dhuriya, P. Johri, S. S. Yadav and N. Gaur, Stock market prediction using linear regression and SVM, in: *Proceedings of the 2021 International Conference on Advance Computing and Innovative Technologies in Engineering* (ICACITE), Greater Noida, India, 4-5 March 2021, pp. 629 – 631, (2021).

[23] G. G. Rajput and B. H. Kaular, A study of effect of training period duration and share split event on predicting stock price using neural networks, in: *Proceedings of WRFER-IEEEFORUM International Conference*, 21 May 2017, Pune, India, pp. 38 – 43 (2017), URL: https://www.digitalxplore.org/up_proc/pdf/305-149768102438-43.pdf.

[24] J. Rasheed, A. Jamil, A. A. Hameed, M. Ilyas, A. Özyavaş and N. Ajlouni, Improving stock prediction accuracy using CNN and LSTM, in: *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy* (ICDABI), Sakheer, Bahrain (2020), pp. 1 – 5, DOI: 10.1109/ICDABI51230.2020.9325597.

[25] G. A. F. Seber and A. J. Lee, *Linear Regression Analysis*, John Wiley & Sons, xvi + 565 pages (2003).

[26] S. Siami-Namini, N. Tavakoli and A. S. Namin, A comparison of ARIMA and LSTM in forecasting time series, in: *2018 17th IEEE International Conference on Machine Learning and Applications* (ICMLA), Orlando, FL, USA, 2018, pp. 1394 – 1401 (2018), DOI: 10.1109/ICMLA.2018.00227.

[27] F. Sukesti, I. Ghozali, F. Fuad, A. K. Almasyhari and N. Nurcahyono, Factors affecting the stock price: The role of firm performance, *The Journal of Asian Finance, Economics and Business* **8**(2) (2021), 165 – 173, URL: https://koreascience.kr/article/JAKO202104142178581.pdf.

[28] R. Taneja and Vaibhav, Stock market prediction using regression, *International Research Journal of Engineering and Technology* **5**(5) (2018), 813 – 815.

[29] M. Vijh, D. Chandola, V. A. Tikkiwal and A. Kumar, Stock closing price prediction using machine learning techniques, *Procedia Computer Science* **167** (2020), 599 – 606, DOI: 10.1016/j.procs.2020.03.326.

[30] G. W. R. I. Wijesinghe and R. M. K. T. Rathnayaka, ARIMA and ANN approach for forecasting daily stock price fluctuations of industries in Colombo Stock Exchange, Sri Lanka, in: *2020 5th International Conference on Information Technology Research* (ICITR 2020), pp. 1 – 7 (2020), DOI: 10.1109/ICITR51448.2020.9310826.

[31] G. P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing* **50** (2003), 159 – 175, DOI: 10.1016/S0925-2312(01)00702-0.

[32] G. Zhang, B. E. Patuwo and M. Y. Hu, Forecasting with artificial neural networks: The state of the art, *International Journal of Forecasting* **14**(1) (1998), 35 – 62, DOI: 10.1016/S0169-2070(97)00044-7.