ISBN: 978-981-18-3959-7
2022 the 12th International Workshop on Computer Science and Engineering (WCSE 2022)
doi: 10.18178/wcse.2022.06.025

Estimating Prediction Accuracy of Stock Price Using Artificial Neural Network

H L Gururaj [1+], Hong Lin [2] B R Sunil Kumar [1], Manu M N [3] and Ravi Kumar V [1]

[1] Department of CSE, Vidyavardhaka College of Engineering, Mysuru, Karnataka, India
[2] Department of CS, University of Houston, Downtown, USA
[3] Department of ISE, SJBIT, Bengaluru, Karnataka, India

Abstract. Presently, all over the world, an enormous amount of investment is being made to the Stock Markets. Nationwide economic systems are sturdily associated and closely inclined to the achievement of their Stock Markets. Additionally, nowadays, trading has become too reachable capital expenditure medium, for both planned investors as well as common man also. Artificial Neural Networks (ANN), which is a subset of Artificial Intelligence (AI), is a way that is anticipated to select out samples (styles) and gain an information model. Significant dispositions of ANN are its capability for precise troubles with step by step analyzing and input-output mapping. The neural network is a superior big approach to categorize anonymous, unnoticed samples in input values that is suitable to be looking for the inventory market. Feed forward neural networks with back propagation schooling algorithms were taken via way of the use of way humans to make predictions.

Keywords: price prediction; national stock exchange; artificial neural networks; machine learning;

1. Introduction

A stock market is a public market for the looking for and promoting employer industrial business organization inventory. It is a prepared unit with a regulatory frame, and those who exchange in stocks are registered with the inventory market and regulatory frame Securities and Exchange Board of India (SEBI). Since stock market facts are quite time-version and are generally in an unsystematic pattern, predicting the destiny rate in inventory is pretty tough. Prediction offers knowledgeable data regarding the cutting-edge-day-day popularity of the stock charge motion. Thus, this will be finished in desire making for clients in finalizing whether or not or no longer or no longer or not or not to shop for or sell the right stocks of a given stock. Many kinds of research were completed for predicting stock market charge the usage of several data mining strategies. This paintings goals at using artificial neural network techniques to anticipate the inventory charge of groups indexed under the National Stock Exchange (NSE). The beyond the information of the selected stock is probably used for building and schooling the models. The effects from the model may be used for evaluation with the real statistics to have a study the accuracy of the model.

A large portion of the exercises, for example, looking through a specific stock and purchasing and selling of similar stocks in the Indian the financial exchange happens on its stock trades like NSE and BSE. The Bombay Stock Exchange (BSE) is one of the most seasoned stock trades and set up in 1875. Then again, The National Stock Exchange was set up in 1992 as the main dematerialized electronic trade in the country.

Trading at exchanges like NSE and BSE takes vicinity through e-ledger wherein the number of shares that are being bought is matched with the exact number of share that is being sold on either of the exchanges. Once the record matches, the transaction takes place wherein the share transfer between seller and buyer takes place. This mechanism takes place throughout the day, starting from 9:15AM to 3:30PM IST.

The remainder of this paper is prepared as follows. Section 2 provides an analysis of the mechanism of action. Section 3 discusses the comparison between the existing system and proposed system. Section 4 provides a description of the methodology. Section 5 includes the conclusion.

+ Corresponding author. Tel.: +91- 9686418942; fax: +91- 9686418942
E-mail address: gururaj1711@vvce.ac.in

2. Related Work

In this section, the related work of various authors and methodologies used in the corresponding papers are explained. The use of help vector tool (Support Vector Machine, SVM) approach to enhance the general fundamental typical performance of quadratic, cubic, linear and top-notch Gaussian for forecasting inventory price prediction became advanced on this paper. State of inventory marketplace price a hundred and seventy days modified into divided into 119 facts and 51 statistics and the number one 119 records modified into used for education and second 51 information changed into used for trying out to are looking in advance to the close to inventory rate. The four fashions prediction consequences were in comparison with the actual price of the stock market fee to assume the future inventory expenses. The tool changed into implemented using the manual vector device nd tool getting to know device bins of MATLAB 2015(a). The regular performance of the tool modified into evaluated using Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) and Mean Squared Error (MSE) and in evaluation with the models [1].

The statistics have informed the usage of a model after which they take a look at data is administered through the professional model. We collect a confusion matrix. Confusion matrix represents the values of True compelling, fake terrible, faux remarkable, proper brilliant. True excellent in the form of accurate prediction that a price belongs to the same beauty. Proper awful is the form of correct predictions that a price does belong to the identical splendour. False positivity is the form of incorrect predictions that a charge belongs to a category even as it belongs to 3 fantastic beauty. False lousy is the number of wrong predictions that a rate belongs to three tremendous splendour at the same time as it belongs to the identical beauty. Then we calculate great traditional regular average performance metrics represented through a way of the use of accuracy, bear in thoughts, precision and f-score [2].

Conceptual Big information investigation can be utilized in numerous areas for precise forecast and examination of the enormous measure of information. They encourage the revelation of critical data from enormous information, which is concealed something else. In this paper, we portray a methodology for the examination of the financial exchange to comprehend its unstable nature and anticipate its conduct to make benefits by putting resources into it. We initially give writing review of past deals with this space. At that point, we give a strategy of our methodology, which contains information assortment and AI calculations. [3].

The Stock market facts are substantially time-model and are normally in a nonlinear sample, predicting the destiny charge of an inventory is pretty difficult. Prediction offers informed information concerning the present day-day popularity of the stock price motion. In the literature assessment, exceptional records mining strategies for stock marketplace prediction are reviewed. It is determined that artificial neural network method can be advantageous in predicting inventory indices further to the stock price of unique corporations. Many perfect algorithms had been used with neural networks. Feedforward MLP neural network method is considered to are looking beforehand to the inventory charge of corporations listed below LIX15 index of NSE. From the save you quit end result table it is able to be concluded that the MLP neural network approach offers a remarkable output with Median Normalized Error zero.05995, Median Correct Direction % fifty-one.06, Median Standard Deviation, 6.39825 [4].

Implementing Classification and Regression Tree (CART) with (Adelson-Velsky and Landis) AVL tree the use of C language and Visual Studio for photographs. Machine getting to know is one of the maximum superior thoughts within the gift research state of affairs. Therefore, exploring device reading at the issue of statistics mining and its gaining knowledge of algorithms has hundreds of scope to an artwork. In device analyzing and facts mining, elegance is incredible for generating correct, brief and clear-cut outcomes and consequently among several techniques of device analyzing, splendour has been determined on CART is able to cope with discrete/specific competencies and offer quick, right and clean beauty consequences, and consequently, it is decided on for class of Indian inventory marketplace information [6].Firstly we are capable of describing the CART set of guidelines then go together with the float of CART after that insertion of AVL tree set of policies, ultimately benefits of CART [7].

Implementing a device wherein the entire is uncooked statistics in terms of the inventory costs and different factors associated with the stocks. The charges than are processed with the beneficial useful resource of manner of the tool to offer a prediction about the rate and what choice to exercise, i.e. Buy Sell

or Hold. Association mining tips are also finished. Association rule mining, one of the maximum crucial and nicely-researched strategies of data mining, became brought first. It dreams to extract exciting correlations, commonplace styles, institutions or simple structures amongst devices of gadgets inside the transaction databases or outstanding facts repositories. At the begin of our take, a look at we decided on nine signs and symptoms and symptoms and signs and symptoms due to their exactness in addition to they provide the same form of outputs, or their outcomes may be interpreted into standard output. Our proposed set of policies moreover gives comparable kinds of output that are Buy, Sell or Hold in which purchase way the inventory rate is going to grow and the investor is typically recommended to shop for a few shares, promote manner the stock rate will lower speedily, so the investor is usually advocated to sell his/her shares if it is in investor's portfolio and keep approach investor is suggested to keep their inventory unchanged [8].

Two techniques had been carried out on this paper: Long Short-Term Memory (LSTM) and Regression, on the Yahoo finance dataset. Both the strategies have examined a development inside the accuracy of predictions, thereby yielding top-notch consequences. Use of these days added gadget getting to know strategies within the prediction of shares have yielded promising results and thereby marked the usage of them in profitable change schemes. It has triggered the belief that its miles feasible to are expecting inventory markets with greater accuracy and standard average overall performance using device studying strategies [9].

Data mining may be interpreted as a records discovery method. Data mining strategies are devised to deal with the issues by using the manner of offering a reliable version with information mining competencies. To gather a version that investigates the inventory patterns using the beyond inventory exchange dispositions, we use the Auto-Regressive Integrated Moving Average (ARIMA) model [10].

Several algorithms and techniques such as Support Vector Machine (SVM), CART, Regression, Association rule mining and also historical data of stocks have been considered to predict the future price of a particular stock in order to achieve profitable trading.

3. Proposed Work

In this paper the lowest, the highest and the average value of the stock market in the last d days are used to predict the next day’s and next week’s market value. Recurrent Neural Network (RNN) is a popular method used to incorporate technical analysis for making predictions in financial markets. The stock market data have been extracted from NSE Stock Market dataset. Fig. 1 shows the flow of work or data that has been fetched directly from the NSE and stored as a Comma-Separated Values (CSV) file in order to carry out cleansing of data and perform predictions on the cleansed data and present the result in the graphical visualization.

Different kinds of neural networks may be advanced via the mixture of different factors like community topology, schooling technique and so on. For this test, we had been given were taken into consideration. Artificial neural network and Long Short-Term Memory (LSTM).

Stage1: Raw Data In this stage, the historical stock data is agitated from www.quandl.com/data/NSE and this historical data is utilized for the prediction of future stock prices.

Stage2: Data Pre-processing: The pre-processing step includes

- Data discretization: Part of data reduction but with particular importance especially for numerical data.
- Data transformation: Normalization.
- Data cleaning: Fill in missing null values.
- Data integration: Integration of data files.

Stager3: Feature Extraction: In this layer, only the features which are be fed to the neural network are chosen. We will choose the feature from Date, open, high, low, close, and volume.

Stager4: Training Neural Network: In this stage, the data is fed to the neural network and trained for prediction assigning random biases and weights. Our LSTM model is composed of a sequential input layer followed by 2 LSTM layers and a dense layer with Rectified Linear Unit (ReLU) activation and then finally a dense output layer with a linear activation function.

Stager5: Output Generation: In this layer, the output value is shown by the output layer of the ANN, and it is compared with the target value. After the comparison, the error or the difference between the target and derived output which adjusts the weights and the biases of the network.

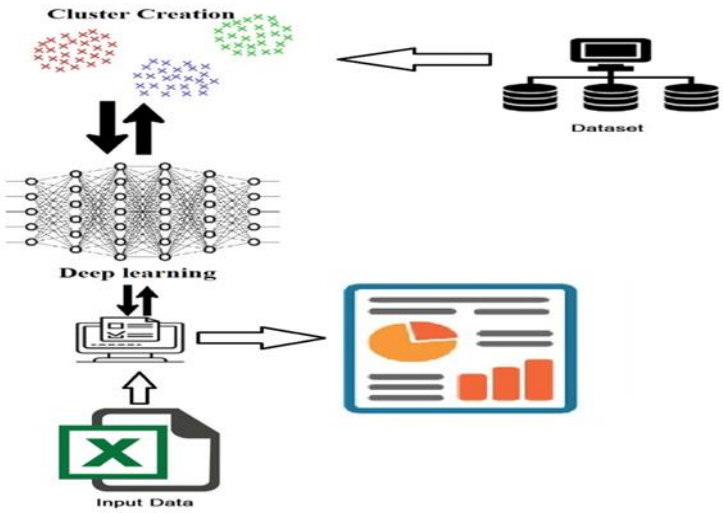

Fig. 1: Proposed system architecture diagram

3.1. Random Forest Algorithm

Random forest is a supervised classification machine learning algorithm which uses ensemble method. A random forest is made up of numerous decision trees and helps to tackle the problem of overfitting in decision trees. These decision trees are randomly constructed by selecting random features from the given dataset.

Random forest arrives at a decision or prediction based on the maximum number of votes received from the decision trees. The outcome which is arrived maximum number of times through the numerous decision trees is considered as the final outcome by the random forest.

Algorithm for Random Forest

Step 1: Start with the selection of random samples from a given stock dataset.
Step 2: Construct a decision tree by specifying the condition for every sample of stock price.
Step 3: Assigning each data point to the closest cluster.
Step 4: Compute cluster centroids.
Step 5: Re-assign each point to the closest cluster centroid.
Step 6: Select the most voted prediction result as the final prediction result.

3.2. Long Short-Term Memory (LSTM) Algorithm

LSTM is a kind of recurrent neural network that has feedback connection in its architecture. It has an advantage over traditional neural network because of its capability to process entire sequence of data. LSTM is very popular in sequence prediction problems because they’re able to store past information.

Algorithm for LSTM

Step 1: Fetch past data of stocks and computation of sigmoid function.
Step 2: Extraction of last n day’s data to detect current state.
Step 3: Define output by considering Data in current cell state.

3.3. Linear Regression Algorithm

Linear regression is the analysis of two separated variables to define a single relationship and it is a useful measure for technical and quantitative analysis in financial markets. One is predictor or independent variable and other is response or dependent variable. It looks for statistical relationship but not deterministic relationship as shown in fig 2.

Algorithm for Linear Regression

Step 1: Analysing the correlation and directionality of the stock data.

Step 2: Estimating the stock data model from the price.

Step 3: Evaluating the validity and usefulness of the stock model.

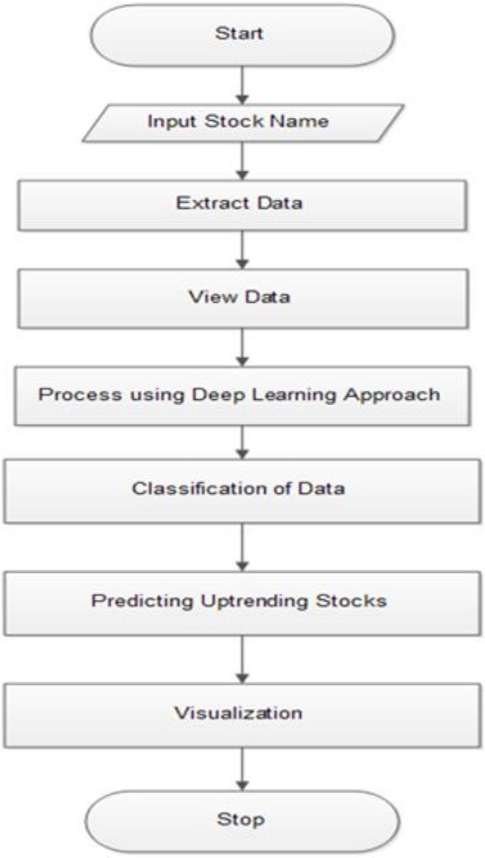

Fig. 2: Flowchart of the proposed work

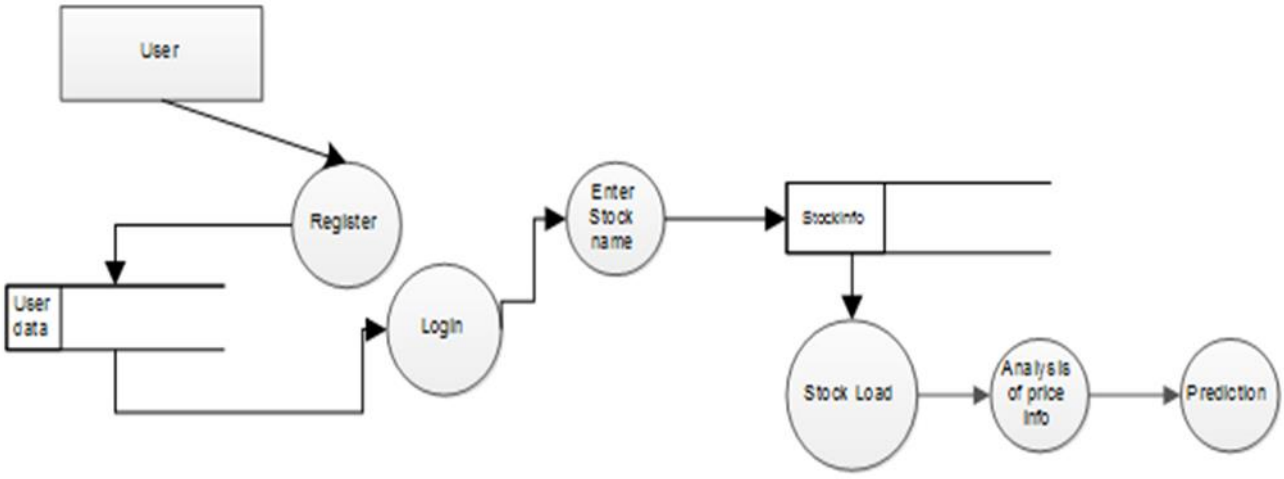

Fig. 3: Data flow diagram

To understand the mathematics behind RNN, have a look at the below image. As discussed inside the first heading, output depends on both current and past inputs. Let I1 be the first input whose dimension is n*1 where n is the length of column. S0 be the hidden state to the first RNN cell having 4 neurons. For each cell, input hidden state should be one previous. For the first cell initialize S0 with zeros or some random

number because no previous state is seen. U be another matrix of dimension d*n where d is the number of neurons in the first RNN cell and n is the input columns of data. W is another matrix of data frame whose dimension is d*d. b is bias whose dimension is d*1 as shown in fig 3.

3.4. Parameters and Gradients

Parameters in the RNN are U, V, b, c, W are shared among all the RNN cells. Parameters are learnable and are responsible for training the model. At each time step, the loss is computing and is back-propagated through the gradient descent algorithm.

3.5. Gradient of Loss with Respect to V

Gradient represents the slope of tangent and points in the direction of the greatest rate of increase of function. From the loss, it means cost function or error. Move is made opposite to the direction of the gradient of the loss with respect to V. Mathematically new value of V is obtained as shown in Eq. (1).

$$Vnew = Vold - \eta d(L)/d(V) , \quad (1)$$

where d(L)/d(V) is the sum of all losses obtained from time steps.

W is multiplied by S. In order to calculate derivative of loss with respect to weight at any time step, the chain rule is applied to take into consideration all the path to reach W from Sn to S0 as shown in fig 4. This means that due to any of the wrong Sn, W is affected. In other words, some wrong information came from some hidden state which leads to loss. Mathematically, weight is updated as shown in Eq. (2) below.

$$Wnew= Wold- \eta d(L)/d(V) . \quad (2)$$

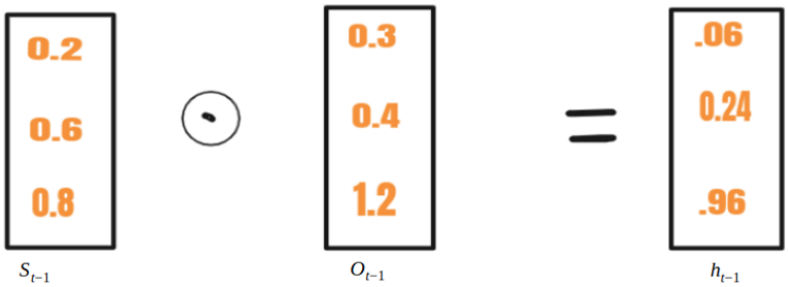

Fig. 4: Example of the selective write, read and forget

In RNN, St-1 is fed along with xt to a cell whereas as, in LSTM St-1 is transformed to ht-1 using another vector Ot-1 as shown in the below Eqs. (3) and (5) respectively. This process is called selective write as shown in Fig. 5. Mathematical equations for selective write are as below.

$$Ot\text{-}1 = \sigma(Uoxt\text{-}1 + W0ht\text{-}2+ b0). \quad (3)$$

$$Ht\text{-}1 = St\text{-}1*Ot\text{-}1. \quad (4)$$

$$Ot = \text{output gate}. \quad (5)$$

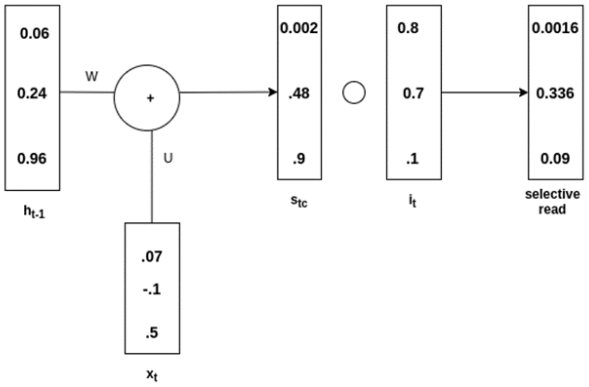

Fig. 5: Example of the selective methodology

As shown in Eq. (4), ht-1 is added with xt to produce st. Then Hadamard product of (written stc in the diagram) and it is made to obtain st.

This is called an input gate. In st only selective information goes and this process is called selective read as shown in Eqs. (6) and (7). Mathematically, equations for selective read are as below.

$$it = \sigma(Ui\ xt + Wi\ ht\text{-}1 + bt). \quad (6)$$

$$\text{Selective read} = st * oit. \quad (7)$$

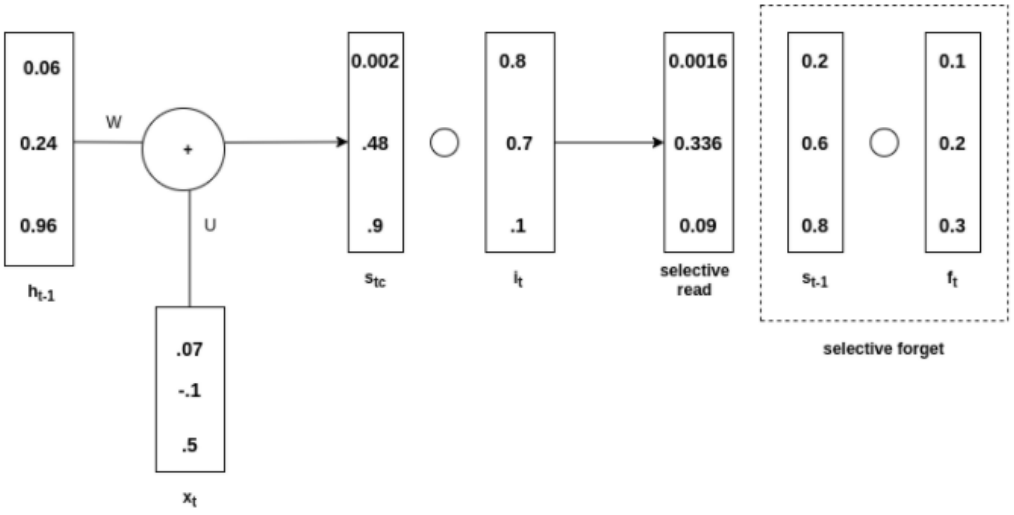

Fig. 6: Illustrative example of selective forget

st-1 is Hadamard product with ft and is called selective forget as shown in Fig. 6. Overall st is obtained from the addition of selective read and selective forget.

4. Simulation Results

Implementation is the process of converting a new or a revised system design into an operational one. The objective is to put the new or revised system that has been tested into operation while holding costs, risks, and personal irritation to the minimum. A critical aspect of the implementation process is to ensure that there will be no disrupting the functioning of the organization. The best method for gaining control while implanting any new system would be to use well planned test for testing all new programs. Before production files are used to test live data, text files must be created on the old system, copied over to the new system, and used for the initial test of each program.

In this application, front end has been developed using HTML, CSS, JavaScript with bootstrap framework. The back end has been developed using C# and MySQL as relational database. The dataset is collected via NEspy API. In addition, numbers of visualization are provided to enhance our system. The models using the features from these external sources along with the traditional stock market data improve the performance for the stock market prediction.

Fig. 7: Homepage

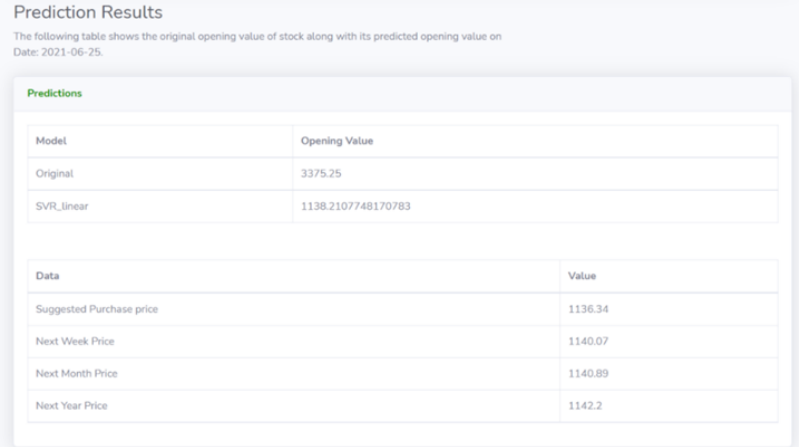

Fig. 8: Random forest prediction result of TCS share

Fig.7 & Fig.8 shows the real time candlestick chart of TCS share price captured from Zerodha Kite Application. The red candle shows the down trend of the price. The green candle shows the uptrend of the price.

Fig. 9: Indicates actual TCS share price

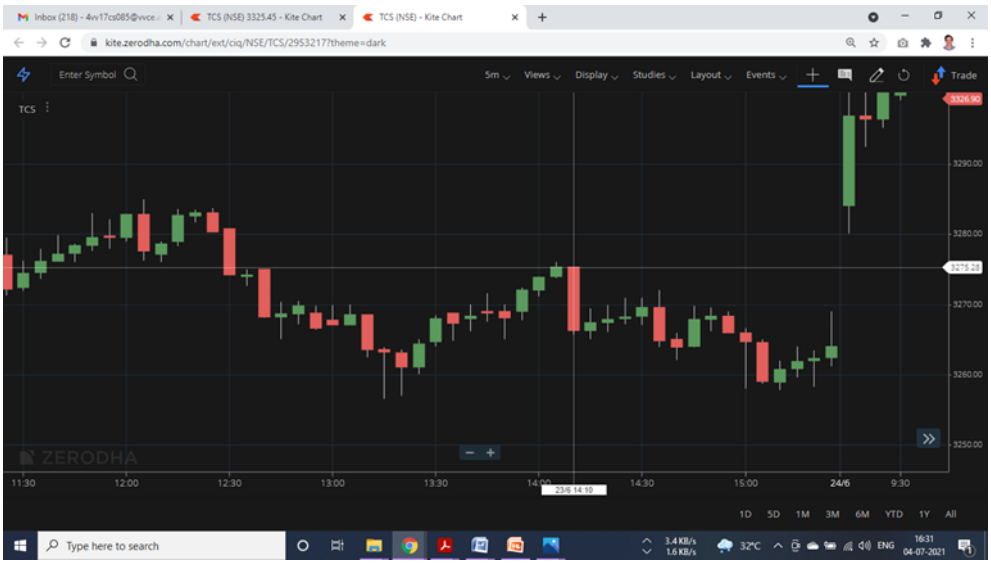

Fig. 10: Indicates actual TCS share prices

Fig. 9 & fig 10. shows the real time candlestick chart of TCS share price captured from Zerodha Kite Application. The red candle shows the down trend of the price. The green candle shows the uptrend of the price.

5. Comparative Analysis

In this project the lowest, the highest and the average value of the stock market in the last days are used to predict the next day's market value. Linear Regression is an approach for predictive modelling to showcase relationship between scalar dependent variable 'Y' and one more independent variable 'X'. RNN & LSTMs are explicitly designed to avoid long-term dependency problem. Remembering information for long period of time as well as easy to learn is practically their default behavior. The random forest algorithm follows an ensemble learning strategy for classification & regression. The random forest takes the average of the various subsamples of the dataset, this increases the predictive accuracy and reduces the over-fitting of the dataset. The stock market data have been extracted from NSE Stock Market dataset. The graphical visual obtained from web application gives the clarity on behavior of market thus helping us understand price action.

6. Conclusion

In this work, it can be concluded that the analysis of the stock market or the prediction of the stock price is a difficult task considering the factors affecting the movement of the market such as sentimental decision made by the trader or the pandemic situation such as Covid-19 which can alter the direction of price action in the market. In order to avoid such anomalies or difficulties, the analysis should be carried out more precisely considering technical, fundamental as well as emotional factors which can directly or indirectly affect various sectors such as financial, agriculture, mining, IT which in turn enables fluctuation in the particular stock price. Hence considering the latest technologies such as ANN can be useful in analysing the movement of the stock market.

7. References

[1] D. Selvamuthu, V. Kumar, A. Mishra, "Indian stock market prediction using artificial neural networks on tick data," *Financial Innovation*, Vol. 5, 16, 2019.

[2] E. Guresen, G. Kayakutlu, T.U. Daim, "Using artificial neural network models in stock market index prediction," *Expert Systems with Applications*, Vol. 38, No. 8, pp. 10389-10397, 2011.

[3] A. S. Chen, M. T. Leung, H. Daouk, "Application of Neural Networks to an Emerging Financial Market: Forecasting and Trading the Taiwan Stock Index," *Computers & Operations Research*, Vol. 30, No. 6, pp. 901-923, 2003.

[4] M. Usmani, S. H. Adil, K. Raza, S. S. A. Ali, "Stock market prediction using machine learning techniques," *2016 3rd International Conference on Computer and Information Sciences* (ICCOINS), Kuala Lumpur, 2016, pp. 322-327.

[5] K. Raza, "Prediction of Stock Market performance by using machine learning techniques," 2017 *International Conference on Innovations in Electrical Engineering and Computational Technologies* (ICIEECT), Karachi, 2017, pp. 1-1.

[6] H. Gunduz, Z. Cataltepe and Y. Yaslan, "Stock market direction prediction using deep neural networks," 2017 25*th Signal Processing and Communications Applications Conference* (SIU), Antalya, 2017, pp. 1-4.

[7] M. Billah, S. Waheed, A. Hanifa, "Stock market prediction using an improved training algorithm of neural network," 2016 2*nd International Conference on Electrical, Computer & Telecommunication Engineering* (ICECTE), Rajshahi, 2016, pp. 1-4.

[8] H. L. Siew, M. J. Nordin, "Regression techniques for the prediction of stock price trend," 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, 2012, pp. 1- 5.

[9] K. V. Sujatha, S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions," INTERACT-2010, Chennai, 2010, pp. 59-63.

[10] Khare, Kaustubh, et al. “Short term stock price prediction using deep learning”, 2017 2*nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology* (RTEICT). IEEE, 2017.

[11] Yaojun Wang, Yaoqing Wang, “Using social media mining technology to assist in price prediction of stock market”, 2016 *IEEE International Conference on Big Data Analysis* (ICBDA)

[12] C. D. R. Aurangzeb, “Factors affecting performance of stock market: Evidence from South Asian countries,” *International Journal of Academic Research in Business and Social Sciences*, Vol. 2, No. 9, 2012.

[13] I. Ali, K. U. Rehman, A. K. Yilmaz, M. A. Khan, H. Afzal, “Causal relationship between macro-economic indicators and stock exchange prices in Pakistan,” *African Journal of Business Management*, Vol. 4. No. 3, pp. 312-319, 2010.

[14] J. Bollen, H. Mao, X. Zeng, “Twitter mood predicts the stock market”, Journal of Computational Science, Vol. 2, No. 1, pp. 1-8, 2011.

[15] Z. Zhang, Y. Shen, G. Zhang, Y. Song, Y. Zhu, “Short-term prediction for opening price of stock market based on self-adapting variant pso-elman neural network”, 2017 8*th IEEE International Conference on Software Engineering and Service Science* (ICSESS), 2017, IEEE.

[16] MustainBillah, Sajjad Waheed, Abu Hanifa, “Stock market prediction using an improved training algorithm of neural network”, 2*nd International Conference on Electrical, Computer & Telecommunication Engineering* (ICECTE), 8-10 December 2016.

[17] A. Izzah, Y. A. Sari, R. Widyastuti, T. A. Cinderatama, “Mobile app for stock prediction using improved multiple linear regression”, 2017 *International Conference on Sustainable Information Engineering and Technology* (SIET), pp. 150-154, 2017.

[18] D. Avramov, “Stock return predictability and model uncertainty,” *Journal of Finance Economics*, Vol 64, No.3, pp. 423458, 2002.

[19] Internet Source (https://www.intelligenteconomist.com)

[20] Internet Source (http://www.acadpubl.eu/)

[21] Internet Source (https://kite.zerodha.com/)