Stock Market Prediction on High-Frequency Data Using ANN

Arafat Jahan Nova*, Zahada Qurashi Mim¹, Sanjida Rowshan¹, Md. Riad Ul Islam¹, Md Nurullah¹ and Milon Biswas¹

¹Bangladesh University of Business and Technology, Computer Science and Engineering, Dhaka, Bangladesh.

Authors’ contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJRCOS/2021/v10i330241
Editorial:
(1) Dr. Hasibun Naher, BRAC University, Bangladesh.
(2) Sadeq J. A. Abul, Kuwait.
Reviewers:
(1) Widodo. S. E. M. Si, Unissula University, Indonesia.
Complete Peer review History: http://www.sdiarticle4.com/review-history/70166

Received 25 April 2021
Accepted 30 June 2021
Published 05 July 2021

ABSTRACT

A stock market is a place where company shares are traded to the stockbrokers. Stock price prediction is one of the most challenging problems as a high level of accuracy is the key factor in predicting a stock market. Many methods are used to predict the price in the stock market but none of those methods are proved as a consistently acceptable prediction tool due to its volatile nature. In this paper, we proposed Artificial Neural Network (ANN) technique because ANN can generalize and predict data after learning and analyzing from the initial inputs and their relationships. We used feed forward network and backward propagation algorithm to predict stock prices. In this paper, we introduced a method that can find out the future value of stock prices in a particular day based on some input using ANN back propagation algorithm.

Keywords: Artificial neural networks; stock market; stock price; feed forward artificial neural networks; backward propagation algorithm.
1. INTRODUCTION

Share market is a place where investors buy and sell their shares to benefit financially. But it is very confusing where investors don’t know what is going on with their investments. The movement of share market is unpredictable and uncertain. The ability to predict the stock prices meet the fundamental needs of investors, since a reasonably accurate prediction has the possibility to profit monetarily by contributing their assets on different companies. To estimate the stock prices researchers found many techniques for the uncertain flow of share prices, such as fundamental analysis, technical analysis and statistical analysis with different methods. After analyzing of various models Artificial Neural Network (ANN) has most accurate prediction results to nonparametric, nonlinear, regression models. The share value calculation sometimes done in daily basis, sometimes in monthly basis or sometimes in yearly basis. Every prediction depends on some historical data. By using more data, the more accurate the prediction will be. So high frequency dataset or daily basis datasets are more significant in this field because it can get more perfect result as more historical data is using for stock market prediction. The linear (AR, MA, ARIMA) and non-linear forecasting algorithms (ARCH, GARCH, Neural Networks) both focus on predicting the stock prices for a single company using the daily closing prices [1]. Based on the history data, the neural network model is successfully applied to predict the daily highest, lowest price and closing price of a company stocks in short time, but it is ineffective for predicting the return rate of the stocks [2]. Various features such as stochastic indicator, moving averages, RSI are extracted from the historical stock data to train the ANN model. The dataset is then divided into training and testing sets which are used for the accuracy of the ANN model [3]. Application of ANN in forecasting problems is very successful because ANN’s are good function approximates [4]. Back propagation neural network is a multi-layer perceptron algorithm which can handle non-linear and time series data and the error rate on BPNN can be reduced compared to single layer [5,6]. This paper demonstrates back propagation method for training the multilayer feed forward network to forecast the share prices. The aim of this paper is to use the power of ANN to forecast Bangladesh Stock Exchange market index values with a reasonable degree of accuracy. In last section the Data mining and neural network combine in one system for predict the stock value more accurately. In last experiment our proposed system first collect previous data then analysis these data with ANN method and provide result of the input data in prediction for using data mining and LSTM algorithm.

2. LITERATURE REVIEW

Stock market is a place where prediction of stock prices is unpredictable and uncertain. To make some sort of sense about the stock prices, researchers found some technique to estimate the share prices. Accurate prediction of stock market prices is considered almost impossible [7]. Even small improvement in predictive performance of stock prices is very beneficial [8]. With development of machine learning algorithms and powerful computers, the long-lasting debate on predictability of financial markets is re-invigorated again in the last few years [9]. Different methods and models are applied to predict stock market. Technical, fundamental and statistical measures have been proposed and used in financial forecasting such as simple moving average, linear regression, Support Vector Machine (SVM) and Back Propagation Neural Network (BPNN) [10]. A wide range of time series need to deal with linear and nonlinear forecasting [11]. After researches on various algorithms Artificial Neural Network (ANN) was found to be most efficient for different problem domain [12]. ARIMA models can compete well with existing forecasting techniques in short-term prediction [13,14]. That’s why a new method based on feed forward artificial neural networks has proposed to analyze multivariate high order fuzzy time series forecasting models because it avoids intense computation and saves time [15]. Badrul, Zakir and Amjad demonstrates a hybrid model of artificial neural network and fuzzy inference system and used back propagation method for training the neural network and multilayer feed forward network in order to forecast the share values [16]. Multilayer network and the related backpropagation training algorithm is one of the most popular algorithm in artificial neural network [17]. Abhishek, Anshul, Tej and Surya conducted an experiment to stock prediction using back-propagation with feed forward network with an accuracy of 99 percent [2]. These all of the works inspire us to apply models in this genre to find remarkable technique. In this recent world everything is automated like E-voting [18,19], supply chain management [20], robotics [21], vehicle registration [22], national identity card management [23], online transaction
METHODOLOGY AND DEVELOPMENT

In this methodology phase, our main focus on dataset that we used for our experiments along with machine learning mechanisms and ANN technique. In this paper we provide an effective methodology to predict stock price value.

3.1 Description of Data Section

We collected data for our experiments from yahoo finance and Google trends from two different sources. The attributes of dataset we used in our experiments are: Date, Open, High, Low, Close. Over period of time in the price of a financial instrument, an open-high-low-close is used to illustrate the movement.

Date: This attribute represent the corresponding date of the event. In our experiments we used per day of week to represent the attribute of date.

Open: This attribute represent the price at which the stock opens for the trading day.

High: This attribute represent the highest price of the stock trade during the day.

Low: This attribute represent the lowest price of the stock trade during the day.

Close: This attribute represent the closing price of the stock trade during the day. This is the average price of trading which occurred in the last 15 minutes. It is the reference point which is used by the investors to compare performance of the stock over time period.

Adjacent close: This is used for taking a track or analyzes the historical returns.

3.2 Data Processing

In this process, we first normalized the data and convert it into the range of 0 to 1 or -1 to 1 depending upon which transfer function we want to use. In our project we used 0 to 1 and filter some table and column. We dispute missing value, repeated value, and substitute it with proper values from the dataset. Also, we identify the important attribute from the dataset, the unnecessary value like the values that can't affect our model. Finally, we prepared our desired dataset for our ANN technique.

The data sets, we used for our experiment are open accessible from online recourses.

Fig. 1. Sample of the dataset
3.3 Normalized Data

Normalizing means rescaling the data through ranging and minimizing of the vector. To make all of the values of numeric columns to a common scale in the dataset thus bringing all the elements lie among 0 to 1. When feature of the dataset in the different range we need to normalized ourdataset. We create our dataset in Excel sheet. To normalize the data from 0 to 1 we follow bellow steps-

- At first we calculate the mean value and standard deviation of the raw scores (values of the dataset) for the variables in the dataset.
- After that from each obtained score we subtract the mean value
- In last we divide this result through the standard deviation.

3.4 Missing Value

For real world situation a common scenario is missing value in dataset. Various methods can be following to handle the missing value.

- Remove the data with the missing values depending on their occurrence.
- Deletion with list wise which has more than one missing values by removes all the data for an observation.
- Dropping variables is an easy way to cope with missing values, it is always better then discard to keep the data. If the missing data is more than 60percent observations, we can drop the variables.
- Imputation is always a preferred than dropping variables. Imputation using mean, media works well with numerical small dataset. Here, m= mean value of the dataset, n= total num of values, MV= messing value, x= sum of all values.

\[ m = (\text{values n}) \] and \[ MV = (m – x) \]

Mean, Median and Mode: Computing the general imply, median or mode is a completely primary imputation method. It may be very fast, however have clean disadvantages. One drawback is that imply imputation reduces variance within side the dataset.

Training and Testing split: For assessment purpose, we divide the dataset into two segments: training and testing. Understandably, those segments can’t comprise identical records factors to keep away from bias. Thus, we carry out 7:3 split at the whole dataset wherein 70percent of the data might be used to train the model and 30percent might be stored to test the model.

3.5 Methods

There are several types of forecasting in the stock market; among them three are most popular. Fundamental analysis, Technical analysis and Machine learning approach also called technological methods mostly common and popular in this field.

ANN: Two types of technique are generally used to forecast stock market future value prediction one is artificial neural networks (ANN) and another one is genetic algorithm. In big data analysis or high frequency data analysis, the back propagation of errors algorithm is commonly referred. So, in this project we try to find out the best type of data to predict the stock value based on some input using ANN back propagation algorithm. In this project we used data analysis tools and some mathematical operation to find out our result. Artificial neural network mimics the human brain by learning from a training dataset and applying the learning to generalized patterns for both classification and prediction [22]. In this Fig. 2 the input connected with the weight then output goes to summarizing unit which summarized all of inputs and weights then goes to threshold unit and active an activation function after that it goes to the final phase which is output unit.

Here is the mathematical calculation:

\[ \text{Sum} = x^* w \] (where x is input and w is weights)

\[ = x1*w1 + x2*w2 + x3*w3 + ... + xn * wn \]

\[ = \text{sum of ( xi * wi )} \]

then it goes to a threshold unit, active and activation function. After that we got our desired result of output.

\[ Y(\text{o/p}) = Q \left( \text{sum} \right) \] (here Q is the activation function).

ANN must have contained three-layer technique to find out the original output where first one is input layer, second one is hidden layer and last or final one is output layer. We can see the whole process in this below figure:
3.6 Proposed Methodology

Our work is predicted stock market value from a database on daily basis. The whole proposed model can be define in three phases: data splitting phase, training phase, testing and evaluation phase. Fig. 3 shows our proposed process. In our proposed method first phase is about the dataset and data processing. We split our dataset into train and test, which processing is going into phase2 and in phase3. The training dataset is consisting of 70 percent of the dataset. The train data is fed into multilayer perceptron of ANN prediction model. After that the remaining 30 percent of test data fitted into learn model of ANN to predict the result, which is going on phase3 and also count the error or error evaluate.

Fig. 2. Artificial Neural Network (ANN) model

Fig. 3. Our proposed methodology
4. ANALYSIS SECTION

We have used Knime and Weka in our experiment. Knime and Weka are free software for data mining research. These software addresses pre-processing, classification, regression, clustering, visualization and rule mining association data. We try to evaluate the prediction value of stock market in a particular day in supervised learning using Multilayer perceptron analysis. We used k-fold cross validation technique, where Knime divided 70 percent of the data for training and 30 percent of the others data were for testing. The main objective of using this approach is found out the strong correlation between independent variable and dependent variable that can predict the more reliable outcome. We can see in the Fig. 4 that knime randomly selected row which was make cross validation technique in a particular day predict the future outcomes.

In last for comparison prospective we use LSTM algorithm and bigger dataset than previous section, also calculate root mean squared error.

4.1 Result using Knime

From Fig. 4 ‘Prediction (close)’ column is our predicted output. That means particular day of close value is near of our ‘Prediction (close)’ column. Knime chose a particular day using cross validation technique and using three input as high, open and low. Finally the output from prediction (close) column appear. Fig. 4 shows the list of actual output and predicted output.

Fig. 5 shows the rate of error in our database, which we use for the experiment.

4.2 Comparison between High, Average and Low Frequencies Database

In this section we use Weka software to find out the differences in high, average and low frequencies data.

High frequency data: High frequency data is time-series data which generally used in research of finance and in analysis sector of stock market. It refers daily based data of a liquid market. For training, we set cross validation 70. In Fig. 6 (a) shows the output of the work. In Fig. 7 (b) the output of the work, where correlation coefficient, mean absolute error, relative absolute error values etc shows.

Average frequency data: We used weekly based data in this section. The output is in Fig. 8.

![Fig. 4. The output of prediction stock value of the database.](image-url)
Fig. 5. Error Rate for the dataset

Fig. 6. Output part1 of high frequency data
Fig. 7. Output part2 of high frequency data Prediction on High frequency data

| Attrib Date         | Attrib | Attrib High | Attrib Low |
|---------------------|--------|-------------|------------|
| Thursday, December 31, 2015 | 0.1629285203608273 |
| Friday, January 1, 2016 | 0.05355352757049971 |
| 1.6262959988878183   | 1.7310762080563804 |
| 1.5518902327176494   |         |

Class
Input
Node 0

Time taken to build model: 5.24 seconds

--- Cross-validation ---
--- Summary ---

Correlation coefficient  -0.027
Mean absolute error  0.0052
Root mean squared error  0.0071
Relative absolute error  117.1239 %
Root relative squared error  118.7549 %
Total Number of Instances  366

Fig. 8. Prediction on average frequency data

| Attrib Date         | Attrib | Attrib High | Attrib Low |
|---------------------|--------|-------------|------------|
| Monday, December 23, 2019 | -0.11847276794243106 |
| Monday, December 30, 2019 | 0.2168940816306707 |
| -2.3709587485695347   | -2.8459825550232885 |
| -3.0326062626122975   |         |

Class
Input
Node 0

Time taken to build model: 2.51 seconds

--- Cross-validation ---
--- Summary ---

Correlation coefficient  0.9673
Mean absolute error  0.063
Root mean squared error  0.1011
Relative absolute error  18.3291 %
Root relative squared error  25.2879 %
Total Number of Instances  262
the blue line represents the data that the model closing price or final prices of the stocks. Here whereas the horizontal scale represents the predicted stock prices of the event, whereas the horizontal scale vertical axis represents the corresponding date of the event, whereas the horizontal scale.

Our proposed model generates a satisfactory result in stock market prediction. It almost shows accurate prediction in most of the datasets.

We have taken the stocks market data of Beximco Pharmaceuticals Limited and created a LSTM network for prediction. Here we have two file train and test, having its Beximco share prices for a particular day. Using the past 1000 day’s data in our LSTM model, we will predict the closing prices for the stocks. To build the LSTM, we have imported couple of modules from Keras and use Matplotlib to plot the result of the predicted stock price and the real stock price into a graph. Our proposed model generates a satisfactory result in stock market prediction. It almost shows accurate prediction in most of the datasets.

In Fig. 10 the graph illustrates the actual stock prices of Beximco Pharmaceuticals Limited. The vertical axis represents the corresponding date of the event, whereas the horizontal scale represents the closing price or final prices of the stocks. In Fig. 11 the graph illustrates the predicted stock prices of Beximco Pharmaceuticals Limited. The vertical axis represents the corresponding date of the event, whereas the horizontal scale represents the closing price or final prices of the stocks. Here the blue line represents the data that the model was train on, the orange line is the actual value for the rest of the days that means it represents the actual closing stock prices of Beximco for rest of the days and the yellow line represents the predictions, what our model predicted the values to be.

The plot shows that predicted value is very close to actual value. When the real stock price increased, while our model projected that the stock price will increase as well. This clearly demonstrates the utility of LSTMs in the analysis of time series and sequential data. Our model worked admirably, as can be shown. It can accurately follow the majority of unexpected leaps and drops.

**Root mean squared error (RMSE):** We used RMSE to evaluate the performance of our model. In RMSE, each predicted and actual values differences are squared then averaged and then the square root of the average is taken. As the values are squared so RMSE gives a relatively high value to larger errors. So RMSE is very useful in measuring the accuracy of a model. Our model’s RMSE value is 2.353284796843162, which is pretty good.

**4.4 Comparison**

In Table 1, it is clearly seen that high frequency dataset gives better performance than others. High frequency has minimum correlation coefficient value among them as well as it has less mean absolute error than others. Finally it seems that prediction using high frequency data is perform better than others dataset.
Fig. 12. Getting the RMSE while using LSTM algorithm

Table 1. Comparison among three types of datasets

| Data Type                  | Correlation coefficient | Mean absolute error | Relative absolute error |
|----------------------------|-------------------------|---------------------|-------------------------|
| High frequency data        | -0.027                  | 0.0052              | 117.1239%               |
| Average Frequency Data     | 0.9673                  | 0.063               | 18.3291%                |
| Low frequency data         | 0.9761                  | 0.1095              | 23.581%                 |

Since RMSE = 2.353284796843162 of LSTM algorithm gives a relatively high value to larger errors, so it has shown better performance on predicting stock value than the different type of data sets gives on prediction value.

5. CONCLUSION

In our experiment we used three types of dataset than comparing between them also shows comparison with LSTM algorithm. We train our model for predict the stock market future value using liquid data of stock market, which is useful in analysis sector of stock market value. In last we compare the outcome of our dataset between different dataset, where we find High frequency data perform better between three type dataset, but the LSTM algorithm gives better performance than others. Our work is time consume and useful for an analyst of the stock market.
Future Works and Limitations: Although the stock market prediction using LSTM performs excellently but accurate prediction of stock market prices is considered almost impossible [7]. Even small improvement in predictive performance of stock prices is very beneficial [8]. For our future work we will try to increase the hidden layers in the LSTM or add another layer for better performance and accuracy and get more and best sets of data length by rescaling and transforming the data and changing the number of training epochs that suit our assets to increase the accuracy of our predictions.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Selvin S, Vinayakumar R, Gopalakrishnan EA, Menon VK, Soman KP. Stock price prediction using LSTM, RNN and CNN-sliding window model. In 2017 international conference on advances in computing, communications and informatics (icaccci). 2017;1643-1647. IEEE.
2. Bing Y, Hao JK, Zhang SC. Stock market prediction using artificial neural networks. In Advanced Engineering Forum. Trans Tech Publications Ltd. 2012;6:1055-1060.
3. Gurjar M, Naik P, Mujumdar G, Vaidya T. Stock market prediction using ANN. Int. Res. J. Eng. Technol.(IRJET). 2018;5:2758-2761.
4. Hiransha M, Gopalakrishnan EA, Menon VK, Soman KP. NSE stock market prediction using deep-learning models. Procedia computer science. 2018;132:1351-1362.
5. Utomo D. Stock price prediction using back propagation neural network based on gradient descent with momentum and adaptive learning rate. Journal of Internet Banking and Commerce. 2017;22(3):1-16.
6. Sureshkumar KK, Elango NM. Performance analysis of stock price prediction using artificial neural network. Global journal of computer science and Technology; 2012.
7. Singh R, Srivastava S. Stock prediction using deep learning. Multimedia Tools and Applications. 2017;76(18):18569-18584.
8. de Oliveira FA, Nobre CN, Zarate LE. Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index—Case study of PETR4, Petrobras, Brazil. Expert Systems with Applications. 2013;40(18):7596-7606.
9. Chander SK, Sumathi M, Sivanandam SN. Prediction of stock market price using hybrid of wavelet transform and artificial neural network. Indian journal of Science and Technology. 2016;9(8):1-5.
10. Borovkova S, Tsiamas I. An ensemble of LSTM neural networks for high-frequency stock market classification. Journal of Forecasting. 2019;38(6):600-619.
11. Araujo RDA, Oliveira AL, Meira S. A hybrid model for high-frequency stock market forecasting. Expert Systems with Applications. 2015;42(8):4081-4096.
12. Deepak RS, Uday SI, Malathi D. Machine learning approach in stock market prediction. International Journal of Pure and Applied Mathematics. 2017;115(8):71-77.
13. Egrioglu E, Aladag CH, Yolcu U, Uslu VR, Basaran MA. A new approach based on artificial neural networks for high order multivariate fuzzy time series. Expert Systems with Applications. 2009;36(7):10589-10594.
14. Miah MBA, Hossein MZ, Hossain MA, Islam MM. Price prediction of stock market using hybrid model of artificial intelligence. International Journal of Computer Applications. 2015;111(3).
15. Khashei M, Bijari M, Ardali GAR. Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs). Neurocomputing. 2009;72(4-6):956-967.
16. Nitta T. An extension of the back-propagation algorithm to complex numbers. Neural Networks. 1997;10(8):1391-1415.
17. Abhishek K, Khairwa A, Pratap T, Prakash S. July. A stock market prediction model using Artificial Neural Network. In 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT’12). 2012;1-5. IEEE.
18. Biswas M, Mahi M, Nayeen J, Hossen R, Acharjee UK, Md W. BUVOTS: A blockchain based unmanipulated voting scheme. Rakib and Acharjee, Uzzal Kumar and Md, Whaiduzzaman, BUVOTS: A Blockchain Based Unmanipulated Voting Scheme; 2020.
19. Mukherjee PP, Boshra AA, Ashraf MM, Biswas M. A hyperledger fabric framework as a service for improved quality e-voting system. In 2020 IEEE Region 10 Symposium (TENSYMP). 2020;394-397. IEEE.

20. Al-Amin S, Sharker SR, Kaiser MS, Biswas M. Towards a blockchain-based supply chain management for e-agro business system. In Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Springer, Singapore. 2021;329-339.

21. Akib AA, Ferdous MF, Biswas M, Khondokar HM. Artificial intelligence humanoid BONGO robot in Bangladesh. In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT). 2019;31-6. IEEE.

22. Hossain MP, Khaled M, Saju SA, Roy S, Biswas M. Vehicle registration and information management using blockchain based distributed ledger from bangladesh perspective. In: IEEE Region 10 Symposium (TENSYMP); 2020. IEEE.

23. Datta P, Bhowmik A, Shome A, Biswas M. A Secured Smart national identity card management design using blockchain. In 2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT). 2020;291-296. IEEE.

24. Ahamed S, Siddika M, Islam S, Anika S, Anjum A, Biswas M. BPS: Blockchain based decentralized secure and versatile light payment system. Asian Journal of Research in Computer Science. 2021;12-20.

25. Mahi MJ, Hossain KM, Biswas M, Whaiduzzaman M. SENTRAC: A novel real time sentiment analysis approach through twitter cloud environment. In Advances in Electrical and Computer Technologies. Springer, Singapore. 2020;21-32.

26. Biswas M, Shome A, Islam MA, Nova AJ, Ahmed S. Predicting stock market price: A logical strategy using deep learning. In 2021 IEEE 11th IEEE Symposium on Computer Applications Industrial Electronics (ISCAIE). 2021;218-223. IEEE.

27. Islam M, Biswas M, Mahi M, Nayeen J, Whaiduzzaman M. LBRP: A resilient energy harvesting noise aware routing protocol for under water sensor networks (UWSNS). International Journal in Foundations of Computer Science Technology (IJFCST). 2018;8.

28. Biswas M, Whaiduzzaman MD.. Efficient mobile cloud computing through computation offloading. Int. J. Adv. Technol. 2018;10(2).

29. Rahman MT, Mahi MJN, Biswas M, Kaiser MS, Al Mamun S. May. Performance evaluation of a portable PABX system through developing new bandwidth optimization technique. In 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT). 2015;1-5. IEEE.

30. Ray B, Saha KK, Biswas M, Rahman MM. December. User perspective on usages and privacy of e health systems in Bangladesh: A Dhaka based Survey. In 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE). 2020;1-5. IEEE.

31. Mahi MJ, Rahad KA, Biswas M, Islam R, Chowdhury ZI. An accident detection system for a single VANET at low cost module. IEEE, TechSym. 2016;4(1): 44-45.

32. Mahi M, Nayeen J, Biswas M, Kushum O, Whaiduzzaman M, Al Mamun S. A New unified communication approach to comply bandwidth optimization technique using dynamic channel allocation. International Journal of Computing and Network Technology. 2018;6(01):1-11.