IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 9, Issue.2, February 2021 E-ISSN: 2347-2693

Stock Market Trend Prediction using Technical Indicators and Deep Learning Methods

Shreshtha Sarkar[1*], Inteshab Nehal[2]

[1,2]Department of Computer Science, St. Xavier's College, Kolkata, India

[*]*Corresponding Author: shreshtha.sarkar5@gmail.com, Tel.: +91-8337-041155*

DOI: https://doi.org/10.26438/ijcse/v9i2.14 | Available online at: www.ijcseonline.org

Received: 09/Feb/2021, Accepted: 11/Feb/2021, Published: 28/Feb/2021

***Abstract*—** The stock market is volatile and is subject to fluctuations. There are many factors like news, fundamental indicators, and heuristic technical indicators et cetera which contribute to such fluctuations. The randomness and volatility have drawn the attention of many researchers and perplexed them. Algorithm trading has been gaining popularity, as machines are able to process tons of data. The ability of an algorithm to predict the price movement gives an opportunity to gain a fortune from the stock market. In this paper, we study the historical prices, calculate the technical indicators based on them, apply feature selection to remove multicollinearity and find the most important features affecting the prices before processing it into the LSTM network to predict the price movement. The prediction of market value can help maximize the profit while keeping the risk comparatively low.

***Keywords*—** Stock prediction, Technical Indicators, Feature Selection, XGBoost, LSTM, Neural Network

I. INTRODUCTION

Stock market is very unpredictable and fascinating. Various stock prediction models exist using macroeconomic factors [1]. One needs to have an understanding of various theories and methods to analyse information regarding the stock market. There are broadly two types of indicators: fundamental and technical. The advanced development in the field of machine learning and artificial intelligence has drawn attention to find its use in the field of finance. Researchers have used Support Vector Machines, Artificial Neural Network, Deep Learning models to predict stock prices in the past [2]. Machine learning algorithms can easily recognize patterns and predict future price trends which are used to evaluate the relationship between a dependent variable and one or more independent variables.

In this paper, we study the historical prices of the stock and their computed technical indicators. We perform feature selection on the technical indicators using an ensemble tree-based method of XGBoost and feed the data into LSTM based neural network to predict the pattern of price movement.

II. RELATED WORK

R.Yamini Nivetha, C. Dhaya (2017) used different prediction algorithms are studied in this research to construct a prediction model. A monthly forecast and a regular estimate for the market value of the next day are the basis of the prediction model. This model forecasts an open price on the market for the next day. Analysis of sentiment involves recognizing and removing emotions on social media from each person. It is important to decide the relationship between sentiment and stock prices. These three algorithms are a comparative study: Multiple Linear Regression, Help Vector Machine and Artificial Neural Network. For the best algorithm, the stock price is predicted.

Vaishali Ingle, Sachin Deshmukh (2017) scraped the form of breaking news from different finance websites, data was gathered for the stock market. For the creation of the HMM model, along with log probability values, the TF-IDF features derived from online news data are used. So, the stock price of the next day is expected to be either higher or lower than the stock price of the present day. Results obtained from their proposed model are compared with results from predictive techniques such as random forest, KNN, multiple regression, bagging and boosting from other machine learning. About 70 percent of the precise forecast is produced by their proposed model.

III. METHODOLOGY

A. Dataset

The raw data are the stock prices which had been fetched from Yahoo Finances, compromising of Close, Low, High price and Volume of everyday, since 2006 to 2019. There exist many different kinds of technical indicators, but we work with a few established ones which more or less fundamentally exhibit the important properties [3] [4]. These values of the technical indicators serve as input to the feature selection model. The dataset is normalized before feeding into the neural network, as they are sensitive to data which is not scaled. The neural network model primarily works on minimizing the RMSE error, so

we segment our data set into Gaussian or non-Gaussian distribution for a certain time interval.

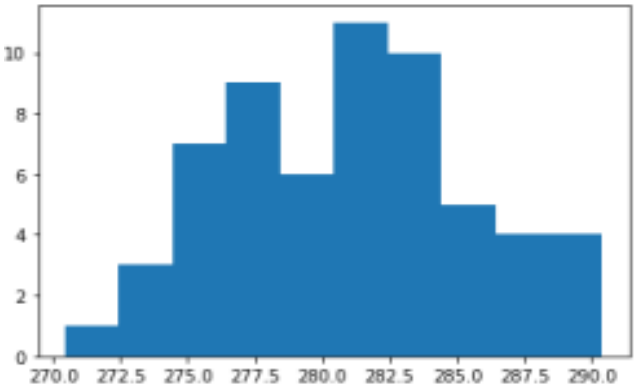

Figure 1. Bar graph of stock prices for 60-day time period which follow Gaussian distribution

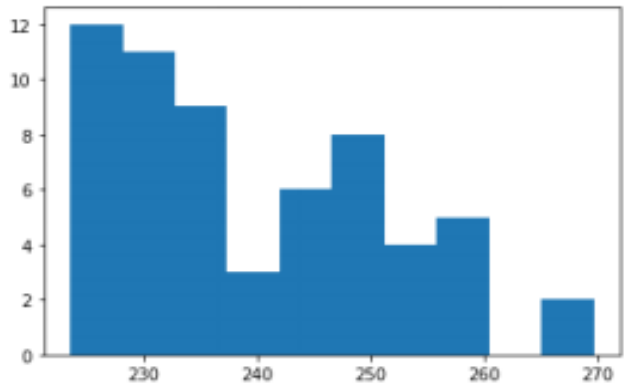

Figure 2. Bar graphs of stock prices for 60-day time period which do not follow Gaussian distribution

B. Feature Selection

Feature selection model is needed because the processing of high dimensional data can be very challenging. We need an algorithm to remove non-informative and redundant features [5] [6]. Neural networks have a tendency to over fit, feature selection helps to overcome this issue, along with increasing accuracy and reducing the time of computation. There are many types of feature selection algorithms like Filter, Wrapper, Embedded based methods.

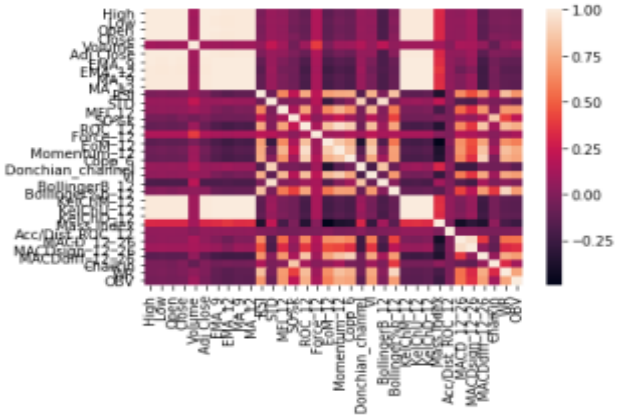

Figure 3. Heat map of the entire feature set

C. XGBoost

XGBoost is gradient boosting method of ensemble learning, which produces a superior output using many models simultaneously. The essential idea behind boosting mechanism involves changing the weak learns into strong learners in a tree-based data structure, to reduce the bias and improve accuracy. The model sequentially adds up iterations as long as needed. Major benefits of using this algorithm are its highly scalable and typically outperform other algorithms. XGBoost forms an importance matrix as a result of model output, which contains the names of all the features and their importance [7]. There are three metrics like

- Gain - relative contribution of the corresponding feature to the model calculated by taking each feature's contribution for each tree in the model.
- Coverage - the relative number of observations related to this feature.
- Frequency - percentage representing the relative number of times a particular feature occurs in the trees of the model.

The Gain is the most considered the most relevant attribute to interpret the relative importance of each feature. In figure 3, we saw 'Close' and 'Adj Close' were highly correlated and in figure 4, we find that the only the Close feature was considered thereby reinstating that it successfully removes redundant, features and reduces multicollinearity from the dataset.

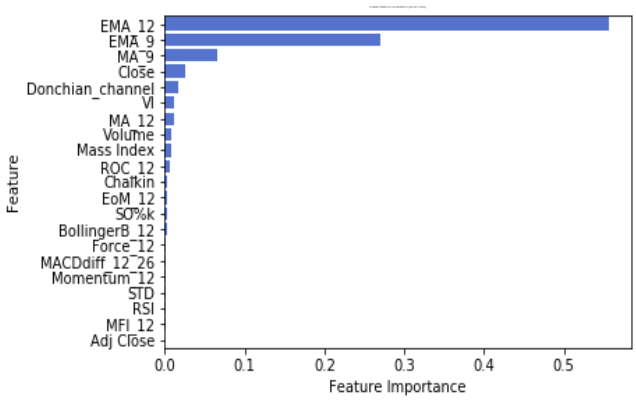

Figure 4. Feature importance plot of technical indicators given by XGBoost

D. LSTM

In recent times neural network has been gaining popularity. It has been proved several times that recurrent neural network is extremely effective to process sequential data. Specifically, for time series problem statements, LSTM is recommended. The uniqueness of LSTM lies in the concept of memory cell. It is a unit of computation which replaces the traditional neurons in the hidden layer of the neural network. Memory cells are successful in associating past inputs with new inputs, which eases it to grasp the pattern of the data dynamically over time ensuring a high forecasting capacity [8]. The training data was transformed into a three-dimensional matrix and converted into 60 days

timestep before feeding to the network. The neural network used in this paper constitutes of two LSTM layers of 50 neurons each with a dropout layer between them to avoid over fitting of the data. Finally, a dense layer with one output was added before compiling the model. Activation function decides if a neuron should fire by calculating weighted sum and further adding bias. The purpose of the activation function is to introduce non-linearity into the output of a neuron. LSTM uses Tanh activation function to overcome the vanishing gradient problem. The speed in which an algorithm converges to the minimum loss depends on the type of optimizer used. The Adam optimizer which is an extension of the stochastic gradient descent was used to fit the model for 6 epochs with a batch size of 64. It is memory efficient requires less hyper-parameter tuning as they have an intuitive interpretation and appropriate for sparse gradients [9].

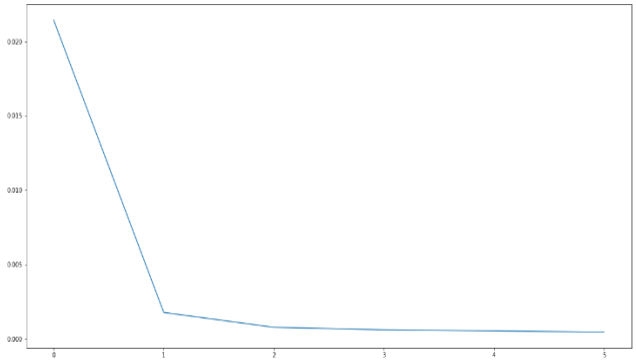

Figure 5. Loss vs. Number of Epochs Plot

Figure 6. Plot of actual and predicted price (LSTM network with batch size = 8 and no dropout).

Figure 7. Plot of actual and predicted price (LSTM network with batch size = 64 and dropout).

IV. RESULTS AND DISCUSSION

The prediction of stock prices is highly volatile and fluctuating, as it tends to mostly lie in the outlier's window [10]. Generally, stock prices are log-normally distributed. We divide the dataset into time steps and perform Shapiro wilk's normality test. It is found that the RMSE is reducing to 1.47 as compared to 9.62 when the predicted price lies within a time period when the prices are distributed normally [11]. The training time also increases when the batch size is small. The small batch size and the stochastic nature of the algorithm indicate that every time it executes the same model, it would generate slightly different mapping of the inputs to outputs. As a result, the outcome would differ when testing the model. In figure 7 we notice that our model is capable of predicting the movement trend of price if not able to accurately pin point the price. The blue line, orange line, green line indicates historical prices, actual prices and predicted prices respectively in the figures above.

V. CONCLUSION AND FUTURE SCOPE

The primary objective of the paper revolves around studying the effect of technical indicators on stock price movement and be able to use neural network model to anticipate the price changes. We have achieved this goal by using the combined knowledge of machine learning techniques and financial sciences. Fifteen years of stock data was taken from yahoo finance. XGBoost was used to find the important technical indicators affecting the stocks primarily. We got an accuracy of 88.2% after adopting Gaussian distribution, feature selection model, dataset normalization and hyper-parameter tuning as compared to an accuracy of 76.1% for non-Gaussian distribution.

To determine the price trend or price of stocks with high precision, we need to incorporate as much data as possible in a model from a variety of sources. Training and hyper parameter tuning of advanced models like fuzzy logic, GAN generally result in improved models for time series problems like price prediction of stocks, but they are difficult to train [12]. Stocks are also drastically affected by the current news. NLP transfer learning models like BERT may be used to act as an indicator of the price movement [13] [14] [15].

REFERENCES

[1] P. Aithal, A. Dinesh and M. Geetha, *"Identifying Significant Macroeconomic Indicators for Indian Stock Markets"*, IEEE Access, vol. **7**, pp. **143829-143840**, **2019**.

[2] M. Firdaus, S. Pratiwi, D. Kowanda and A. Kowanda, *"Literature review on Artificial Neural Networks Techniques Application for Stock Market Prediction and as Decision Support Tools"*, Third International Conference on Informatics and Computing (ICIC), **2018**

[3] Y. Snezhko, *"The use of technical analysis indicators in the Russian stock market"*, Russian Journal of Entrepreneurship, vol. **16**, no. **16**, p. **2681, 2015**.

[4] Blume, L.,Easley, D., O'Hara, M, "*Market Statistics and Technical Analysis: The Role of Volume. The Journal of Finance*", Wiley online Library, 1994.

[5] Iain M. Johnstone[1] and D. Michael Titterington, Padhy S, "*Statistical challenges of high-dimensional data*", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. **367,** Issue.**1906**, pp. **175-195**, **2009**.

[6] Mela, C. and K. Kopalle, "*The impact of collinearity on regression analysis: the asymmetric eVect of negative and positive correlations*", *Applied Economics*, Vol. **34,** pp. **667-677**, **2002**.

[7] Yuqinq He, Kamaladdin Fataliyev, and Lipo Wang, "*Feature Selection for Stock Market Analysis*", International Conference on Neural Information Processing, Vol. **8227,** pp. **737-744**, **2013**.

[8] M. H., E.A., G., Menon, V. and K.P., S., "*NSE Stock Market Prediction Using Deep-Learning Models*", *Procedia Computer Science*, Vol. **132,** Issue.**4**, pp. **1351-1362**, **2018**.

[9] A. Kumar, J. S. Saini, *"Stock Data Analysis and Prediction in Machine Learning"*, IJCSE, Vol.**8**, Issue.**9**, pp**.70-78, 2020.**

[10] P. Kanade, *"Machine Learning Model for Stock Market Prediction",* International Journal for Research in Applied Science and Engineering Technology, vol. **8**, no. **6**, pp. **209-216, 2020.**

[11] J. Patel, S. Shah, P. Thakkar and K. Kotecha, *"Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques", Expert Systems with Applications*, vol. **42**, no. 1, pp. **259-268, 2015**.

[12] Acheme, David & Vincent, Olufunke & Folorunso, Olusegun & Isaac, Olusola, "*A Predictive Stock Market Technical Analysis Using Fuzzy Logic*", International Conference on Neural Information Processing, Vol. **7,** Issue.**3,2014**.

[13] Lee H, Surdeanu M, MacCartney B, Jurafsky D, "*On the Importance of Text Analysis for Stock Price Prediction*", Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pp. 1170–1175, **2014**

[14] A.Sahoo, J.K.Mantri, *"Stock Market Prediction Using Text Mining Approaches: A Survey",* IJCSE, Vol. **7,** Issue.**2**, pp. **443-450**, **2019**.

[15] V. Ingle, S. Deshmukh, *"Predictive mining for stock market based on live news TF-IDF features",* International Journal of Autonomic Computing, Vol.**2** No.**4**, pp. **341-345,2017.**

AUTHORS PROFILE

Ms. Shreshtha Sarkar pursed Bachelor of Science (in Computer Science) from St. Xavier's College, Kolkata in year 2019. She is currently working as Business Analyst in Data Sutram, Location Intelligence start-up. Her main research work focuses on Machine Learning, Deep Learning, Natural Language Processing, IoT and Business Intelligence. She has 3 years of Corporate Experience and 4 years of Research Experience.

Mr. Inteshab Nehal pursed Bachelor of Science (in Computer Science) from St. Xavier's College, Kolkata in year 2019. He is currently working as Data Science Engineer in Spring and River, AI&ML based start-up. His main research work focuses on Industrial Automation, Research and Development, ML, AI, NLP. He has worked with clients in various industries like Manufacturing, Composite Material, Steel, IT, Oil and Petrochemical. He has 2 years of Corporate Experience and 3 years of Research Experience.

