Deep learning based non-linear regression for Stock Prediction

Subhash Chand Agrawal
Department of Computer Engineering and Application, GLA University
subhash.agrawal@gla.ac.in

Abstract. Stock market prediction is an activity to estimate the future value of a stock. The accurate prediction of particular share’s future price can lead to significant profit margins for an investor. The efficient market hypothesis states that prices of the stock depend on the available information and price changes, do not consider any hidden information. Therefore, prediction of stock plays a significant role to influence the investor’s decisions. It also acts a recommend system for investment related decision in stock market for short term investors and financial suffering system for long term shareholders. In this paper, we propose a stock market prediction system using machine learning algorithms. This paper first explores a few machine learning algorithms for estimating stock value and then proposes a solution that can predict the future stock value with higher accuracy. In this paper, we propose a deep learning based non-linear regression method to predict the stock price. The experiments are performed on two publically available datasets: Tesla Stock Price and New York Stock Exchange which consist of stock data from 2010 to 2020. The analysis of experimentation reveals that the proposed method performed better than existing machine learning based approaches.

1. Introduction
Everyday billions of dollars are invested by the individual to sell and buy stock shares of different companies. Each individual expects that he/she will gain the profit on these shares. Stock market faces the uncertainty and different companies grow up and go down according to the pattern of the market daily. Investor must consider this behavior of the market in his/her investing decision.

Better stock price prediction is helpful in decision making for investors and financial experts [1]. The behavior of the market is judged on the basis of various stocks constituting the market. Accurate movement of stock is an indicator for investor with proving complete information about each and every stock and return gained by a specific portfolio [2]. Therefore, Stock market prediction is an attempt to predict the future value of a company stock more precisely so that an investor can gain profit margin on investment.

Data mining techniques also play a significant role for the new investors that do not have enough knowledge about Stock market. This paper will provide software that will analyze sensex based on company’s stock value. The following factors are considered to decide the stock values of company:

1) Demand and Supply: It is the biggest factor for changing the price of a share. When there is a huge demand of a company’s product and supply is unable to bridge the gap between demand and supply. In this case, price rises and vice versa.

2) Corporate results: This is referred to as growth of the company in terms of profit over a span of time.
3) Popularity: The buyer’s decision of buying stock is also highly influenced by the popularity of the company. Any good news can rise the stock price of a company and any bad news will reduce the stock price.

There are also other factors for deciding the stock value, but we are considering only these main factors.

In today’s world, all the information pertaining to stock market is available. Analysing all this information individually or manually is tremendously difficult. Therefore, automation of the process is required.

The emergence of various machine learning techniques has made stock market price prediction task possible in a efficient way. Various machine learning based methods such as support vector machine (SVM) [3], artificial neural network (ANN) [4], Bayesian networks (BN) [5] and Hidden Markov Model (HMM)[6] are introduced in the past. Amongst them, neural network had drawn a significant attraction of the researcher due to handling of highly non-linear and complex data in case of stock market. It significantly increased the prediction accuracy.

2. Related Work
In the past various methods using artificial intelligence and machine learning came into existence. Some of the important methods are covered in this section.

Wang et al.[7] proposed a method that employ price data to estimate the direction of market index and stock market price. Early method focuses on macroeconomic indicators such as CPI and FDP to train the model. Daily data of the macroeconomic indicators is not available; make these methods unsuitable in the real world. The experimental results are performed on Korean composite stock price prediction and Hang Seng Index (HSI). This method takes stock data as an input and with the help of PCA dimensionality of data is reduced. PCA data is combined with various factors (internal and external). Finally model is trained using SVM classifier.

The existing methods use only historical stock data to predict the stock price. These methods do not consider the stock market fluctuation. Zhang et al.[6] proposed an event driven prediction approach that leverages the multiple data sources in the predictions. It utilized Hidden Markov model that considers the historical trading data as well as new events related to market fluctuations. In the experimental analysis, they showed the importance of these events in the predictions. Experiments are performed on China A share data and shows the better performance with the existing methods.

Huang et al.[8] presented a comparative study of various methods used for stock prediction based on Neural network and fuzzy inference system using fundamental analysis. They identified that neural network performs better than fuzzy inference system in both buy and sell portfolios.

Ye et al.[9] suggested that existing method ignores the impact of other stock information in predicting the stock price. Therefore, they presented a deep learning based method called as graph convolution network and gated recurrent unit to predict movement of stock by considering stock cross effect. The gated convolution network produces the cross-correlation features that are combined with historical trading data. Experiments are performed on two stock indexes in China market.

Li et al.[10] proposed a ensemble deep learning based method that utilized two recurrent neural networks and fully connected neural network. Experiments are performed on S&P 500index. The experimental analysis shows that MSE is reduced by 57.55% from 438.94 to 186.32, recall rate increased by 50%, precision by 40%, F1 score by 44.78% and movement direction accuracy by 33.14% respectively.

Drenska et al.[11] proposed a online machine learning system with the help of expert advice. This method used binary sequence of stock prediction problem, treated as price history and predictor as investors. In this method an investor and market play against each other. The investor combines the advice of two or more expert to reach a final decision to minimize the risk. Expert uses historical data in the analysis.

This paper focuses exclusively on predicting the daily trend (price movement) of individual stocks by considering the various parameters in prediction and acts as a recommended system for investors in
taking their financial decision related to a particular stock. The paper will make no attempt to deciding how much money to allocate to each prediction. Moreover, it will analyze the accuracy of these predictions.

3. The Proposed Methodology
In this section, first we explore few existing machine learning strategies and then proposed a solution for stock market price prediction using deep learning based non-linear regression.

3.1. The machine Learning based approaches
Machine learning strategies are divided into supervised learning and unsupervised learning. In supervised learning, the stock predicting model is trained on past trading data. In the training phase of the model, we clearly define the inputs and their corresponding output. Finally model takes decision on the basis of learning from the training data. Classification and regression problems both are considered in the supervised learning. Regression algorithms as well as classification algorithm can be utilized in stock prediction. Simply, if we are predicting that we should invest in particular stock, it will be considered as classification problem. If we are predicting the future price of a stock which will is a continuous output within a range. Various regression algorithms are considered in the past to address the problem of stock prediction. Some of them are listed below:

3.1.1. Linear Regression. It is one of the simplest and popular machine learning method used for forecasting. It is used in stock market to predict the future price of a product on the basis of the different parameters. It represents the straight line relationship between dependent variable and one or more independent variables [12]. It is given by:

\[ y = a_0 + a_1 x \]  

where \( a_0 \) is the intercept of the line and \( a_1 \) is the regression coefficient. The mean square error (MSE) is the objective cost function used to minimize the error between actual and predictable values. Mathematically given as:

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - (a_0 + a_1 x_i))^2 \]  

where \( n \) is the total number of samples and \( y_i \) is the actual value and \( a_0 + a_1 x_i \) is the predicted value. Furthermore, gradient descent is used to minimize the error between actual and predicted value by calculating the gradient of the cost function.

3.1.2. Logistic Regression. It is another technique in the field of machine learning. It uses more complex cost function as compared to linear regression. The sigmoid function is utilized here whose values are within the range \([0,1]\) [13]. The sigmoid function is defined as:

\[ f(x) = \frac{1}{1 + e^{-x}} \]  

and logistic regression is given by:

\[ h(x) = \sigma(a_0 + a_1 x) \]  

or it can be written as:

\[ h(x) = \frac{1}{1 + e^{-(a_0 + a_1 x)}} \text{ and } 0 \leq h(x) \leq 1 \]
Cost function is also different from the linear regression and given as:

\[
j = -\frac{1}{m} \sum_{i=1}^{m} y_i \log(h(x_i)) + (1 - y_i) \log(1 - h(x_i))
\]  

(6)

In unsupervised learning, there is no training phase (user does not provide the output label), decision is made on the basis of optimization or loss function for unseen data. Cluster analysis and principle component analysis mainly cover in this category. Stock prediction uses supervised learning because model is trained on historical data.

3.1.3. Support Vector Machine

Several researchers also utilized support vector machine that divides the data into two regions separated by a separating boundaries. There are various choices of the kernel such as polynomial, radial basis function, linear, etc. A major drawback with SVM is that its input variables are stored in high dimensional feature space. Therefore, it requires high computational time and a lot of memory for storing features. In case of stock market where market consists of thousands of stocks, it becomes very difficult to process high volume of data. As a solution, high dimensionality of feature space is required to be reduced. Several researchers used principle component analysis (PCA) for reduction of high dimension feature space. However, this dimensionality reduction may lead to reduce the accuracy of stock prediction.

3.2. The Proposed Solution

Figure 1 represents the architecture of deep learning model used in the proposed work. It consists of 3 hidden layers. We used deep learning based non-linear regression to predict the price of stock market. The Linear regression is unable to grasp the changes in the stock market data due to high non-linearity. We used five layers in the model. There are 1 input layer, 3 hidden layers and 1 output layer.

The proposed method uses RELU (Rectified Linear Unit) activating function in the hidden layers that output the input directly, if input is positive, otherwise, output is zero, if input is negative. We have not used sigmoid and hyperbolic tangent activation function due to problem of vanishing gradient problem. RELU removes the problem of vanishing gradient, easier to train the model and also achieve better performance in prediction. Mathematically, it is represented as follows:

\[
R(z) = \begin{cases} 
  z, & \text{if } z > 0 \\
  0, & \text{if } z < 0 
\end{cases}
\]  

(7)

where \( z \) is sum of weighted inputs and bias, is given by-

\[
z = \sum wx + b
\]  

(8)

Initially random weights are assigned and in the subsequent layers, they are adjusted according to the learning. We use the mean absolute error as a loss function. The numbers of neurons in all 5 layers are 128, 256, 256, 256 and 1 respectively. The proposed method used linear activation function for the output layer and output consists of predicted price of the stock as a regression output. The number of epochs is the most important parameter for the model. It should be decided carefully. A large number of epochs may result into over-fitting of training data while a smaller number of epochs may cause the under-fitting problem. Therefore, early stopping method is used which automatically decide the required number of epochs for optimal performance and stops training the model and reduce the error between actual and predicted values.

We divide the stock price data into training and testing. The proposed method used 75% data for training and 25% for testing.
4. The Experimental Results
We have investigated our experiments on two datasets namely New York Stock Exchange [14] and Tesla Stock Price [15]. New York stock Exchange contains the data related to historical and fundamental analysis. It consists of seven columns: date, symbol, open, close, low, high and volume related to different stocks. It contains huge amount of market price data (851K) of different companies.

The second dataset, Tesla Stock Price dataset presents stock price from its initial public offering 2010 to 2020. It consists of seven columns: Date, Open, High, Low, Close, Volume, Adj Close. It consists of historical price of 2416 record related to stock.

We divide the total data into two parts: training 75% and testing 25%.

4.1. Results of Prediction on Tesla Dataset
In this section, we perform the experiments on Tesla’s stock price data and results are recorded against linear regression and the proposed method. Forecasting is done on close column of dataset. The result of prediction for 10 samples from the testing dataset are is shown in figure 2.

![Fig. 2. Results of stock price prediction on Tesla Dataset (a) Linear Regression (b) Non-linear Regression using Deep Learning](image)

The predicted value in figure 2 shows that predicted stock prices are closer to actual values. The predicted values are shown in the graph itself for the comparison. In addition, root mean squared error (RMSE) is shown in table 1.

4.2. Company wise Results of Prediction on New York Stock Exchange Dataset
This dataset contain the company wise stock data. We perform experiment on top two companies Google and Apple. Again, results are compared with linear regression in figure 3 and 4.
Again Figure 3 and 4 prove that in company wise prediction also the proposed method perform better than the linear regression. There is less error in prediction as compared to linear regression. The RMSE values for these two companies are illustrated in Table 1.

### Table 2: Comparisons of RMSE values on two datasets

| Dataset           | New York Stock Exchange | Tesla Stock Price |
|-------------------|-------------------------|-------------------|
|                   | Google | Apple | Google | Apple |              |
| Linear Regression | 5.76   | 0.56  | 2.12   |       |              |
| Khan et al.[16]   | 4.7    | 0.43  |        |       |              |
| The Proposed Method | 3.25   | 0.25  | 1.24   |       |              |

The RMSE values on these two datasets clearly show that improvement of proposed method over the linear regression and Khan et al.[16] method.

### 5. Experimental Results

In this paper, first we have analyzed various machine learning based approaches used in the stock market. Then, we proposed a method using deep learning that overcome the inefficiency of these approaches in handling non-linear stock data. The proposed method is trained on two existing stock datasets. Based on this data, the deep learning based non-linear regression model predicts stock price according to a particular company. The experiments are performed on two publically available datasets: Tesla and New York stock exchange. The experiments results show that the proposed method is better than existing machine learning approaches. The future work will consider the impact of sudden changes in market by considering internal, external factors or sentimental analysis.

### References

1. Kao, L-J., Chiu, C-C., Lu, C-J. and Yang, J-L. ‘Integration of nonlinear independent component analysis and support vector regression for stock price forecasting’, *Neurocomputing*, Vol. 99, No. 1, pp.534–542, 2013.

2. Pathak, B.V. et al. *The Indian Financial System: Markets, Institutions and Services*, Pearson
Education India, 2011.

[3] Bruno Miranda Henrique, Vinicius Amorim Sobreiro, Herbert Kimura, Stock price prediction using support vector regression on daily and up to the minute prices, The Journal of Finance and Data Science, Volume 4, Issue 3, Pages 183-201, ISSN 2405-9188, 2018.

[4] Selvamuthu, D., Kumar, V. & Mishra, A. Indian stock market prediction using artificial neural networks on tick data. Financ Inno 5, 16, 2019.

[5] Luciana S. Malagrino, Norton T. Roman, Ana M. Monteiro, Forecasting stock market index daily direction: A Bayesian Network approach, Expert Systems with Applications, Volume 105, Pages 11-22, ISSN 0957-4174, 2018.

[6] Zhang, X., Li, Y., Wang, S. et al. Enhancing stock market prediction with extended coupled hidden Markov model over multi-sourced data. Knowl Inf Syst 61, 1071–1090 (2019). https://doi.org/10.1007/s10115-018-1315-6

[7] Y. Wang, “Stock price direction prediction by directly using prices data: An empirical study on the KOSPI and HIS,” International Journal of Business Intelligence and Data Mining, vol. 9, no. 2, pp. 145-160, 2014.

[8] Y. Huang, L. F. Capretz and D. Ho, "Neural Network Models for Stock Selection Based on Fundamental Analysis," 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), Edmonton, AB, Canada, 2019, pp. 1-4, doi: 10.1109/CCECE.2019.8861550.

[9] Jiexia Ye, Juanjuan Zhao, Kejiang Ye, Chengzhong Xu, “Multi-Graph Convolutional Network for Relationship-Driven Stock Movement Prediction,” arXiv:2005.04955, 2020.

[10] Y. Li and Y. Pan, “A novel ensemble deep learning model for stock prediction based on stock prices and news” in arXiv:2007.12620, 2020, [online] Available: http://arxiv.org/abs/2007.12620.

[11] Nadejda Drenska, Jeff Calder, "Online Prediction With History-Dependent Experts: The General Case," arXiv:2008.00052, 2020.

[12] Shalabh (2013): "A revisit to the efficient forecasting in linear regression models", Journal of Multivariate Analysis, 114, pp. 161-169.

[13] Alzen, J.L., Langdon, L.S. & Otero, V.K. A logistic regression investigation of the relationship between the Learning Assistant model and failure rates in introductory STEM courses. IJ STEM Ed 5, 56 (2018). https://doi.org/10.1186/s40594-018-0152-1.

[14] New York Stock Exchange Dataset[Online Available at] https://www.kaggle.com/dgawlik/nyse Accessed 15 Oct, 2020.

[15] Tesla Stock Price Data[Online available at] https://www.kaggle.com/timoboz/tesla-stock-data-from-2010-to-2020. Accessed 10 Oct, 2020.

[16] Khan, U.; Aadil, F.; Ghazanfar, M.A.; Khan, S.; Metawa, N.; Muhammad, K.; Mehmood, I.; Nam, Y. A Robust Regression-Based Stock Exchange Forecasting and Determination of Correlation between Stock Markets. Sustainability 2018, 10, 3702.