International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 09 | September - 2022 Impact Factor: 7.185 ISSN: 2582-3930

PREDICTION OF STOCK MARKET USING MACHINE LEARNING AND DASH

Sakshi Kesharwani

M.Tech

Shikha Singh

Assistant Professor

Saurabh Mitra

Head Of Department

Dept Of Digital Communications, Dr. C.V.Raman University Kota, Bilaspur Chhattisgarh

ABSTRACT:

Stock market is a place where you can buy and sell shares for publicly listed companies. Stocks which are also known as equities represent the ownership in the company. Stock exchange is the mediator that allow buying and selling of shares. Stock market prediction is the act of trying to determine future value of a company stock. Stock market prediction is made possible with the help of machine learning. It allows you to analyze and predict the future values of company stocks. Most of the stock prediction techniques make use of linear regression model of machine learning. It is simple and easy to handle, but the main limitation of linear regression is that, linear regression is the assumption of linearity between the dependent and independent variables which can be incorrect many times. Also it is very sensitive to the outliers and cause over-fitting of the data. This project is about the prediction of the stock market price by using the LSTM model which will overcome the limitation of the method that is implemented using the linear regression. Along with plotting the graphical representation using LSTM model, I also built a dashboard in this project to analyse the stock. By using dash it is possible to view the graphical analysis of the stock market which will be very effective

Keywords: Stock market, LSTM model, Dash, Machine learning, Python

I. INTRODUCTION

The stock market is defined to be the collection of markets and exchanges where regular activities of buying and selling of shares of publicly-held companies take place. It is a place where shares of pubic listed companies are traded. The primary market is the place where companies float shares to the general public in an initial public offering (IPO). This is done to raise capital. People mainly buy the stocks in the expectation that their price may rise in the future. But there is always an uncertainty in the stock market due to which people is not willing to invest their money in the stock markets. Thus we need a technique which can predict the stock market prices, so that people can invest their money in the best stocks. The project is about the prediction of the stock market price using LSTM model and also uses dash to visualize stock market analysis which include the actual price and predicted price as a web application. LSTM is the Long Short Term Memory. LSTM are the type of recurrent neural network for learning long term dependencies. It is mainly used for processing and predicting on the basis of time series data. LSTM mainly have a chain like structure. Instead of having a single neural network layer, there are four interacting layers communicating with each other in a very special way. Working of LSTM mainly consists of four steps:

1) First step in the LSTM is to identify the information that are not required and will be thrown away from the cell state. This decision is made by a sigmoid layer. This layer is also called forget gate layer.

2) The next step is to decide what new information we are going to store in cell state. Key to the LSTM is the cell state. This mainly consists of two steps: a) The sigmoid layer called input gate layer decides which values will be updated. b) tanh layer creates a vector of new candidate values that could be added to that state. tanh function is a squashing function which means it converts the value into between the range -1 and +1.

3) Now we will update old cell state Ct-1 into the new cell state Ct. first we multiply old state Ct-1 by ft, forgetting the things we decided to forget earlier. Then we add it*Ct . This is the new candidate value and it is scaled by how much we decided to update each state value.

4) We will run a sigmoid layer which decides what part of cell state we are going to output. Then we put the cell state through tanh and multiply it by the output of sigmoid state, so we can output only the parts we decided to output. Dash is a wonderful library framework allowing python to build interactive web application dashboards. Layout of dash consists of all HTML contents. To implement dash we need to install dash components.

II. LITERATURE REVIEW

In [1], Kim and Han developed a model for forecasting stock price index that combined artificial neural networks (ANN) and genetic algorithms (GAs) with feature discretization. The technical indicators as well as the direction of change in the daily Korea stock price index were employed in their research (KOSPI). Tey used data from 2928 trade days from January 1989 to December 1998 to determine their preferred features and formulae. Tey also used feature discretization optimization, which is an approach comparable to dimensionality reduction. The fact that they used GA to optimize the ANN is one of their work's merits. First, the buried layer has 12 input characteristics and processing components, not adjustable. Another disadvantage is that the authors only concentrated on two elements in the optimization phase throughout the ANN learning process. They still feel GA has a lot of promise when it comes to feature discretization optimization. Our initialized feature pool refers to the features that have been chosen.

In [2], Qiu and Song provided a method based on an optimized artificial neural network model to forecast the direction of the Japanese stock market. The authors use genetic algorithms in conjunction with artificial neural network-based models to create a hybrid GA-ANN model.

In [3], Piramuthu evaluated several feature selection strategies for data mining applications in depth. He evaluated how different feature selection strategies maximized decision tree performance using datasets such as credit approval data, loan default data, online traffic data, tam, and kiang data. The Bhattacharyya measure, the Matusita measure, the divergence measure, the Mahalanobis distance measure, and the Patrick-Fisher measure were among the probabilistic distance measures he compared. The Minkowski distance measure, city block distance measure, Euclidean distance measure, Chebychev distance measure, and nonlinear (Parzen and hyper-spherical kernel) distance measure are all inter-class distance measures. The author assessed both probabilistic distance-based and multiple inter-class feature selection approaches, which is a strength of this study. Furthermore, the author conducted the study using a variety of datasets, which added to the paper's strength. The evaluation algorithm, on the other hand, was only a decision tree. We can't say if the feature selection approaches would hold up in a bigger dataset or with a more complicated model.

In [4], Hassan and Nath used the Hidden Markov Model (HMM) to estimate stock market values for four major airlines. They divide the model's states into four categories: opening price, closing price, maximum price, and lowest price. The technique used in this study is unique in that it does not require expert knowledge to develop a prediction model. While this research is confined to the airline sector and assessed on a short dataset, it may not result in a generalizable prediction model. To accomplish the comparison job, one of the methodologies used in stock market prediction related activities might be used. The date range of the training and testing datasets was specified at a maximum of two years by the authors which provided us a date range reference for our evaluation part.

Wavelet Neural Network (WNN) was used by **Lei in [5]** to anticipate stock price patterns. As an optimization, the author used Rough Set (RS) for attribute reduction. The stock price trend feature dimensions were reduced using Rough Set. It was also used to identify the Wavelet Neural Network's structure. This study's dataset includes five well-known stock market indices: (1) SSE Composite Index (China), (2) CSI 300 Index (China), (3) All Ordinaries Index (Australia), (4) Nikkei 225 Index (Japan), and (5) Dow Jones Index (USA). The model was evaluated using several stock market indexes, and the results were convincing in terms of generality. The computational complexity is reduced by employing Rough Set to optimize the feature dimension before processing. However, in the discussion section, the author mainly emphasized parameter tweaking and did not mention the model's flaws. Meanwhile, we discovered that while the assessments were done on indices, the same model would not perform as well if applied to an individual stock.

III. PROPOSED SYSTEM

As per prediction system, developed in the stock price prediction to help investors in making financial decisions. In most researches it focuses on "lowest price buy", "highest selling price". On the "lowest buy" and "highest selling" strategy of stocks occurs when stocks are at the lowest price and sell shares when prices are highest. For stock price predictions ANN technique is used with back propagation the dataset was pre-processed and tuned up for real analysis. Hence, our admin can upload stock price history i.e. open price, highest price, lowest price and close price of the day. Paper will also focus on data preprocessing. Secondly, after preprocessing the data, System reads stock price history and gives input to the Back-propagation algorithm. In addition, the proposed paper examines the use of the prediction system in real-world settings and issues associated with the accuracy of the overall values given. The back propagation gives output as final predicted rate comes. The proposed system can get the output of prediction list of stock price and graph of prediction table like that user can view the final predicted result. The successful prediction of the stock will be a great asset for the stock market institutions and will provide real-life solutions to the problems that stock investors face. To test BPNN method, mean square error is used to predict result and data reality. The backward propagation of errors, of back propagation, is a common method of training artificial neural networks and used in conjunction with an optimization method such as gradient descent. The method repeats in two phase cycle, propagation and weight update. During back-propagation phase the output after forward pass is compared with the expected output which is then used to adjust link weights. The output of the proposed system is to predict the list of stock price. At the end of system the user can view the final result of predicted value of stock market.

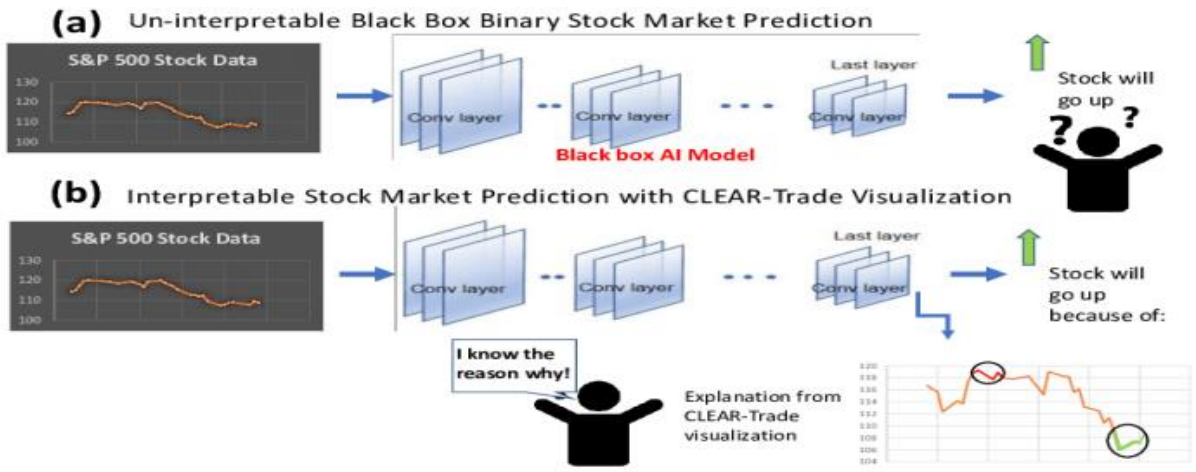

Figure 1: Proposed Algorithm

The stock price prediction using LSTM and dash consist of following stages:

1) Raw Data: As a first step we need to collect the required data sets in order to train our model. Historical stock data is collected. Historical stock data is the data that obtained in the previous year. This is used for the comparison purpose. And this data is used for the prediction of stock prices in the future. As an initialization we use the libraries like numpy, matplotlib, pandas, keras etc. Numpy helps to apply mathematical functions and operations to the array. Matplotlib is used for the visualization purpose. Pandas library is used as data analysis and manipulation tool. It takes the data as CSV and then creates a data frame.

2) Data Preprocessing: The preprocessing stage involves data discretization, data transformation, data cleaning and data integration. After the data set is transformed into a clean dataset, dataset is divided into training and testing sets for the evaluation purpose.

3) Feature Extraction: In feature extraction layer the features that fed in to the neural network are chosen. Here dropout is used which is a regularization technique for reducing over-fitting in neural networks.

4) Train Neural Network: In this stage, data is fed into the neural network and trained for prediction and for assigning random biases and weights.

5) Optimizer: Optimizer is used for the compiling purpose. The type of optimizer used can greatly affect how fast the algorithm converges to minimum value. Here we have chosen to use Adam optimizer. The Adam optimizer combines the perks of two optimizers: ADAgrad and RMSprop. In the ADAgrad learning rate is calculated based on past gradient that that have been computed for each parameter. RMSprop considers fixing the diminishing learning rate by only using a certain number of certain gradients. Adam is the adaptive movement estimation. It is another method that computes the adaptive learning rates for each parameter based on its past gradients.

6) Output Generation: In this layer, output value generated by the output later of the RNN is compared with the target value. The error or the difference between the target and the obtained output is minimized by using back propagation algorithm.

7) Visualization: A rolling analysis of a time series model is often used to assess the stability of the model over time. When analyzing financial statistic data employing a statistical model, a key assumption is that the parameters of the model are constant over time.

IV. RESULT:

When we run code in pycharm using flask following figure are the outcome of our project.

Figure 2: Stock price of google when search using dash panel.

Figure 3: Exponential and predicted Stock price

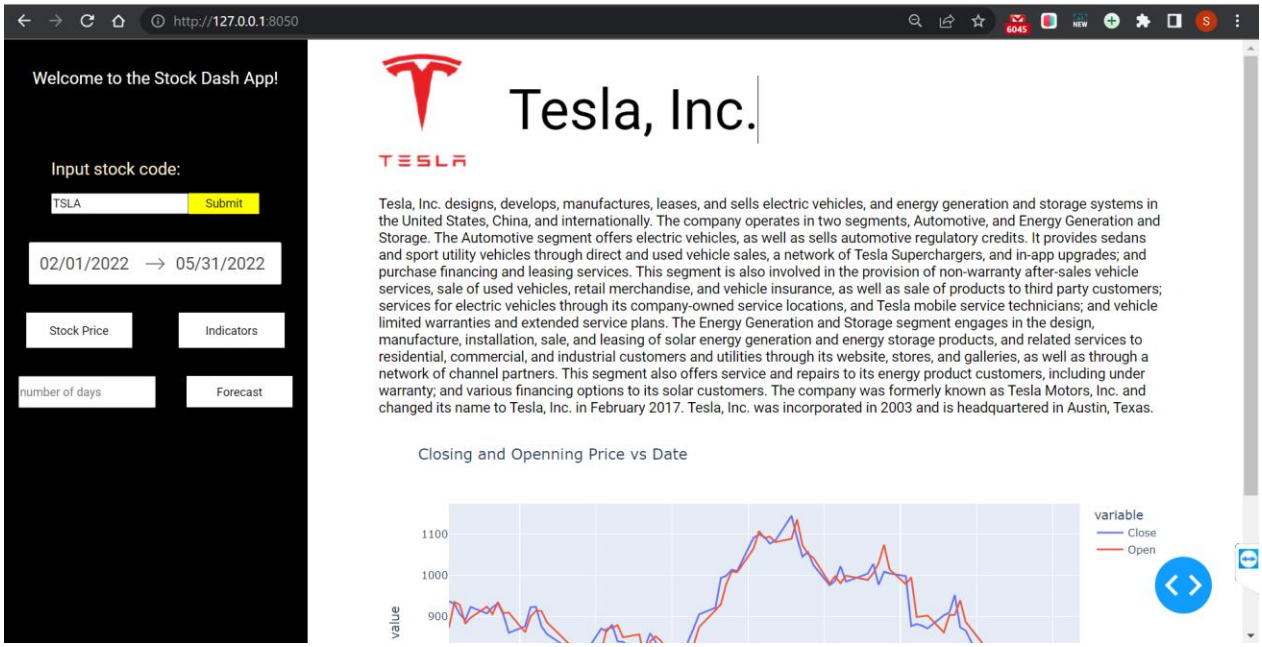

Figure 4: Stock price of Tesla when search using dash panel.

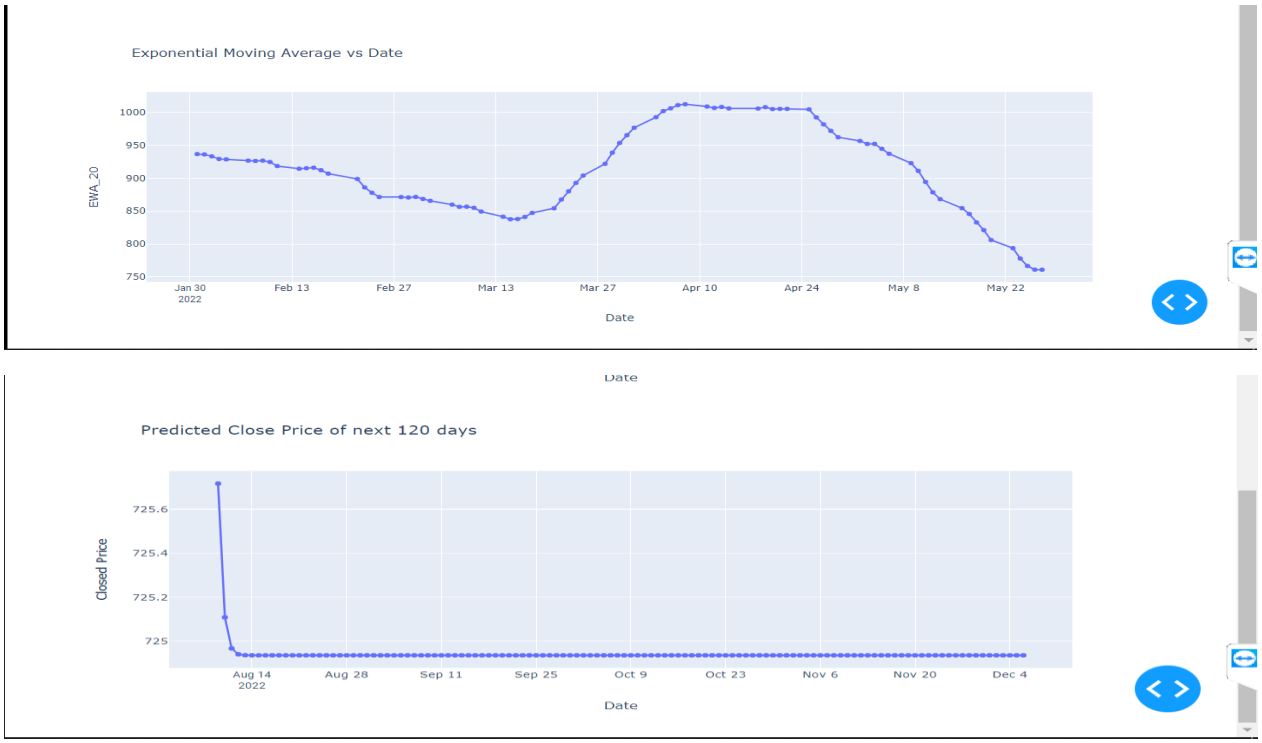

Figure 5: Exponential and predicted Stock price

V. CONCLUSION

In the above paper, we inspected advancements in economic market predictions. By looking at different predictive models, we discovered that Logistic-regression is offering the capacity to predict and analyzing market movement direction more precisely than the other existing methods. Different models, for example, Random Forest and ARIMA have additionally turned out to be well known in stock market prediction. Random Forest demonstrated its fruitful application in classification work, ARIMA on time series prediction and financial related applications. K-NN model is also applied in the experiment which also show some good results in predicting stock market directions.

REFERENCE:

[1]. Kim K, Han I. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. Expert Syst Appl. 2000;19:125–32. https://doi.org/10.1016/S0957-4174(00)00027-0

[2]. Qiu M, Song Y. Predicting the direction of stock market index movement using an optimized artificial neural net-work model. PLoS ONE. 2016;11(5):e0155133.

[3]. Piramuthu S. Evaluating feature selection methods for learning in data mining applications. Eur J Oper Res. 2004;156(2):483–94. https://doi.org/10.1016/S0377-2217(02)00911-6.

[4]. Hassan MR, Nath B. Stock market forecasting using Hidden Markov Model: a new approach. In: Proceedings—5th international conference on intelligent systems design and applications 2005, ISDA'05. 2005. pp. 192–6. https://doi. org/10.1109/ISDA.2005.85.

[5]. Lei L. Wavelet neural network prediction method of stock price trend based on rough set attribute reduction. Appl Soft Comput J. 2018;62:923–32. https://doi.org/10.1016/j.asoc.2017.09.029.

[6]. Lee MC. Using support vector machine with a hybrid feature selection method to the stock trend prediction. Expert Syst Appl. 2009;36(8):10896–904. https://doi.org/10.1016/j.eswa.2009.02.038.

[7]. Sirignano J, Cont R. Universal features of price formation in financial markets: perspectives from deep learning. Ssrn.2018. https://doi.org/10.2139/ssrn.3141294.

[8]. Ni LP, Ni ZW, Gao YZ. Stock trend prediction based on fractal feature selection and support vector machine. Expert Syst Appl. 2011;38(5):5569–76. https://doi.org/10.1016/j.eswa.2010.10.079.