International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 4, ISSN No. 2455-2143, Pages 245-250
Published Online August 2020 in IJEAST (http://www.ijeast.com)

STOCK PRICE FORECASTING: A MACHINE LEARNING MODEL

Nadar Aishwarya
Department of IT
PCE, New Panvel, Maharashtra, India

Nair Divya
Department of IT
PCE, New Panvel, Maharashtra, India

Karkera Poornima
Department of IT
PCE, New Panvel, Maharashtra, India

***Abstract*— In this paper, the work done is the future Stock prices are predicted. The effective prediction of a future stock's price could yield major profit. The technique mainly used to predict stock price are fundamental and technical analysis. In this paper, Technical analysis is performed on historical data of stock prices are collected from yahoo finance by applying LSTM with RNN (a machine learning algorithm) on close values obtained from yahoo finance and the stock values is analysed. Then the learned models are ready to predict the stock future values.**

Keywords—**Machine Learning, Prediction Model, LSTM, RNN, Yahoo Finance.**

I. INTRODUCTION

Stock Market serves as a mediator for companies to add up to their capital by introducing their company shares to the market as it turns out to be a great platform for investors to earn more profit. The ability to predict the value of stock price efficiently is very important for the market dealers or investors to earn huge gains. This ability to predict future price of stocks forecasting, can eventually result in either profits or loss for low or high volumes of share transactions. Deciding the instant of buying or selling the securities for generating revenue are often conceived as our usual problem. Currently, most of them apply machine learning strategies to solve the profitable problem of accurately indicating the future price of the markets. Machine learning is a domain of artificial intelligence in which systems gain the ability to automatically learn and improve from past experience. Effective prediction models for the stock market help investors, analysts, and traders by providing strong information like the future indication of the stock market. The main goal of the paper is to predict the stock prices of various companies by building a model using suitable machine learning techniques and algorithms which will predict the stock prices by mapping it to its past behaviour under similar conditions. In this paper, we present Long Short-Term Memory[1] (LSTM)[1] with recurrent neural network[1] (RNN)[1] .

The rest of the paper is organized as follows. Background at II. Existing architecture is presented at III. Proposed architecture is presented in section IV. Proposed algorithms are presented in section V. Experiment and results are presented in VI. Evaluation parameters are presented in VII. Concluding remarks are given in section VIII.

II. BACKGROUND

Hiransha. M et al.(2018) used A time series is a set of data measured over time to acquire the status of some activity. The algorithms used are ANN, MLP , RNN, LSTM.
Information is used to decide how they react to the new set of data. This is done with the help of a feedback loop where output at each instant is an input to the next moment. The accuracy results of this paper is 80%. It works for a limited period.
M.Ramaswami et al. studied, a novel approach is proposed by combining both Support Vector Machine (SVM) and Artificial Neural Networks (ANN) in predicting stock trends. SVM technique is introduced to remove irrelevant and redundant variables and subsequently neural network based classification technique is used to forecast stock trends with the reduced feature set. It is 80% accurate. It is used for a limited period of time.
Vivek Kanade et al. used both fundamental and technical analysis. fundamental analysis is done using social media data by applying sentiment analysis processes. Social media data has a higher impact today than ever, it can help in predicting the trend of the stock market and Technical analysis is done using historical data of stock prices by applying machine learning algorithms. The method involves collecting news and also collecting social media data and extracting sentiments expressed by individuals. Then the correlation between the sentiments and stock values is analysed. The learned model can then be used to make future predictions about stock value. It can be shown that this method is able to predict the stock performance and sentiment and social data are also closely correlated with recent news. The algorithms used are CS-SVM and ANN. The Cuckoo Search (CS) is based on Swarm Intelligence optimization technique and is very simple to tune the parameters of SVM. The accuracy of prediction ANN SVM and CS-SVM for Indian stock

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 4, ISSN No. 2455-2143, Pages 245-250
Published Online August 2020 in IJEAST (http://www.ijeast.com)

market was analysed. The accuracy results are 80%. Mediocre live results**.**

Lohith N et al.(2017) comprises analysis and implementation results of stock market prediction. We have chosen 4 strategies for comparative study. They are Neural Network(NN), Random Forest ensemble method, Linear Regression and Technical Indicators as predictors. Algorithms used are Hybrid, SVM, ANN [3]. It is 50% accurate. Only comparative study done for choosing the best algorithm.

Aparna Nayak et al.(2016) an attempt is made for prediction of stock market trends. Two models are built one for daily prediction and the other one is for monthly prediction. Supervised machine learning algorithms are used to build the models. As part of the daily prediction model, historical prices are combined with sentiments. The algorithms used are RBF, ANN and SVM [4] Two different models have been built to predict stock market trends. First model predicts the stock market trend for the next day (Daily prediction model) by considering all the available data on a daily basis as input. Second model predicts the stock market trend for the next month (Monthly prediction model) by considering the available data on a monthly basis. Daily prediction model is built by considering historical price dataset and sentiment dataset. Up to 70% of accuracy is observed using supervised machine learning algorithms on daily prediction model. Monthly prediction model tries to evaluate whether there is any similarity between any two months trend. Evaluation proves that the trend of one month is least correlated with the trend of another month. Two models used can predict upto a month.

III. EXISTING ARCHITECTURE

The existing system of the stock market is explained in below sections:

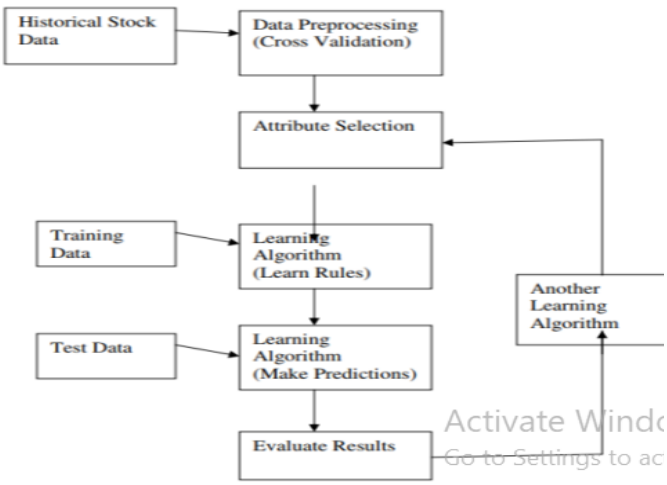

Fig.3.1 Existing architecture

The Weka and the YALE Data Mining tools were used.
In this model Stock Market is predicted using Machine Learning models such as Support Vector Machine (SVM) by RBF kernel.

Steps for Stock Market Prediction:

Step 1: Downloading data from the web. The predictions are made using the share value up to the closing date.
Step 2: Conversion of data into csv format.
Step 3: Select data from the gui from the database.
Step 4: Graph is displayed of Stock before mapping and after mapping.
Step 5: Calculation of the log2c and log2g value for minimizing error.
Step 6: Displays the predicted value graph of select stock which shows the original value and predicted value of the stock.
The models yield 95% of accuracy and can be applicable onto only on certain datasets chosen.

IV. PROPOSED ARCHITECTURE

In the proposed system, the architecture of the system follows a client-server model, where the server and the client are loosely coupled. After relevant stock data is retrieved from the third-party data provider Yahoo Finance, the backend pre-processes the data and builds the models. After that, predictions are made and the prediction results will be stored, which can be retrieved from the web application.

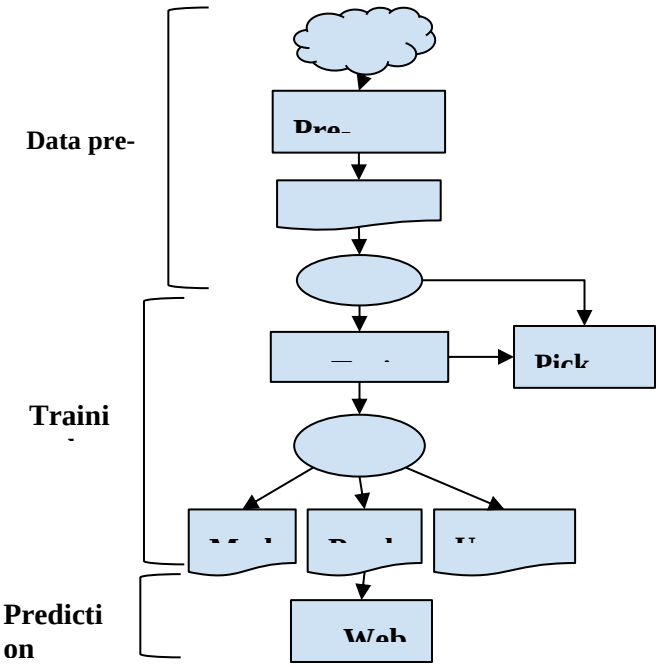

Fig.4.1 System Architecture

V. PROPOSED ALGORITHM

Stock market prediction is the fact of trying to determine the future value of a company's stock or other financial instrument

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 4, ISSN No. 2455-2143, Pages 245-250
Published Online August 2020 in IJEAST (http://www.ijeast.com)

traded on an exchange in the stock market. The successful prediction of a stock's future price could yield significant profit. With the advent of the digital computer, stock market prediction has since moved into the technological realm. Stock price prediction is to determine the future value of a company's stock is trading for on the market. The effective prediction of a future stock's price could yield major profit. The technique mainly used to predict stock price are fundamental and technical analysis. The existing methods have some limitations of having less accuracy and precision. In this paper, inbuilt algorithms are used in order to do the predictions; From those algorithms LSTM is having the highest accuracy rate for long term dependencies. Technical analysis is performed on historical data of stock prices by applying machine learning. Then the learned models are ready to predict the stock future values. The objectives of the proposed technique are: To study various stock value prediction algorithms.

To propose an approach for forecasting future stock prices of the company for n-days using improved LSTM with RNN.

To compare and analyse the results of the proposed approach with the existing on the basis of parameters viz. Accuracy, Root Mean Square Error and Precision.

The methodology for proposed technique is as follows:

1.Raw Data: In this, the historical datasets of stocks of various companies are generated from the Yahoo Finance and this historical dataset is used for predicting future stock prices.

2.Data Pre-processing: The pre-processing step involves various steps:

a) Data discretization: Reduction of numerical data.

b) Data transformation: Normalization.

c) Data cleaning: Data cleaning is the process of preparing data for analysis by removing or modifying data that is incorrect or improperly formatted.

d) Data integration: Integration of data.

Once the dataset is modified into a clean dataset, this dataset is split into training and testing datasets to evaluate. Now, create

a data structure with suitable timesteps and one(1) output.

3.Feature Extraction: In this step, only the features which are to be fed to the neurons into the network are chosen. The feature from Date, open, low, high, close, Adj Close and volume are chosen.

4.Training Neural Network: The data is supplied to the neural network and trained for prediction by assigning random weights

and biases. This LSTM [1]model is made of a sequential input layer, three LSTM [1] layers and a dense layer with activation function and at the end a dense output layer with linear activation function.

5.Optimization: The type of optimizer used greatly affects how fast the algorithm meets the minimum value. From a few great algorithms for optimization, in this paper, Adam optimizer is chosen. The Adaptive Moment Estimation(Adam) optimizer combines the best of two other optimizers:

[a]ADAgrad and

[b]RMSprop.

Terms:

θ:Parameter to be updated.

η:The learning rate.

g :Gradient estimate.

ϵ :Avoid division of zero.

G :Matrix of sum of squares of past gradients.

t :Time-step

α:Step-rate

β:Momentum term

γ :Decay Term

v :First moment of gradients.

m :Second moment of gradients.

The ADAgrad optimizes stochastic gradient descent update for ADAgrad and becomes

$$\theta_{t+1,i} = \theta_{t,i} - \eta g_{t,i}$$

here,

$$g_{t,i} = \nabla_{\theta_t} J(\theta_{t,i})$$

The learning rate (θ)[1] is calculated based on the past gradients.

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{G_t + \epsilon}} \cdot g_t$$

RMSprop fixes the disappearing learning rate(θ) by using a particular number of previous gradients.The result becomes

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} \cdot g_t$$

Where,

$$E[g^2]_t = 0.9E[g^2]_{t-1} + 0.1g_t^2$$

Adam, is the other method that computes each parameter the adaptive learning rates by considering the exponentially decaying average of past squared gradients and past gradients.This is represented as

$$v_t = \beta v_{t-1} + (1-\beta)g_t^2$$
$$m_t = \beta m_{t-1} + (1-\beta)g_t$$

researchers observed an inherent bias towards zero(0) at first and hence they made following estimates:

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t}$$
$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}$$

Hence, gradient update rule becomes

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \cdot \hat{m}_t$$

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 4, ISSN No. 2455-2143, Pages 245-250
Published Online August 2020 in IJEAST (http://www.ijeast.com)

Benefits of optimizer are summarized as follows:

- Every iteration and parameter has a different learning rate.
- With the ADAgrad, the learning does not disappear.
- The gradient update uses the moments of the distribution of weights by allowing more statistically sound descent.

6.Regularization: Another important aspect of training the model is overfitting.
This can be done by including a penalty for large weights. In this paper, Tikhonov regularization is chosen, used as the below minimization problem:

$$\underset{f \in \mathcal{H}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} V(f(x_i), y_i) + \gamma ||f||_K^2$$

The function space in a Reproducing Kernel Hilbert Space (RKHS)[1] makes sure that there exists notion of a norm, allowing us to encode the notion of the norm into the regularizer.

7.Dropouts: Dropouts have been helpful to make the neurons more robust and thereby allowing them to predict the trend by not only focusing on any 1 neuron. It is a new method for preventing overfitting.

8.Output Generation: The target value is compared with the output value generated by the output layer of the RNN[1]. The error (difference between the target value and the output value) is minimized by using the back propagation algorithm by adjusting the biases and the weights of the network.

The above steps are followed for the dataset used for testing.

9. Visualization: All of the above analysis can be implemented with keras and their functional API.

VI. EXPERIMENT RESULT

The web application is made using web html,css,bootstrap,js,chart js as frontend and flask,jupyter notebook as backend development.

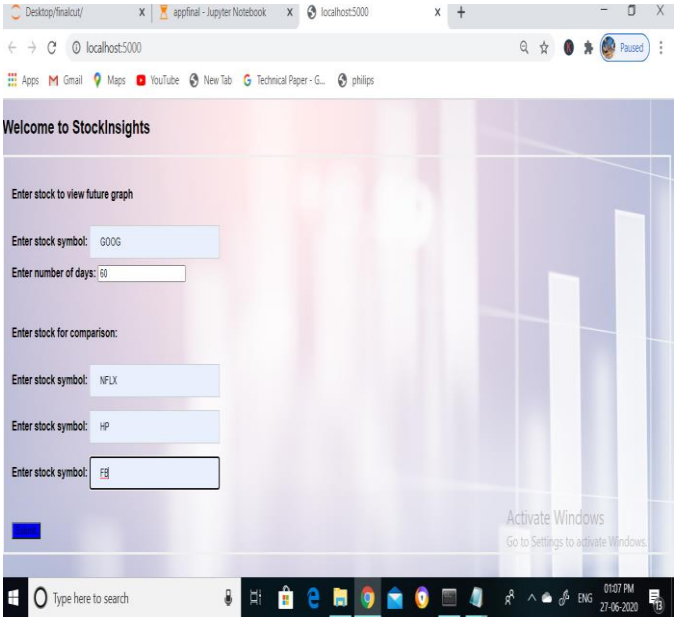

Fig.6.1 Landing page

The landing page collects two types of data, the first one collects data to represent a graphical view of a company stock along with the future number of days as input for which you want to see future prices of the stock, the second data is collected to display a buy or sell graph based on the current situation of the stock prices it takes three inputs which are stock symbol.

	Date	Open	High	Low	Close	Adj Close	Volume
0	2004-08-01	49.813286	56.528118	47.800831	50.993862	50.993862	134241100
1	2004-09-01	51.158245	67.257904	49.285267	64.558022	64.558022	213503200
2	2004-10-01	65.155777	99.601669	64.209328	94.964050	94.964050	516060900
3	2004-11-01	96.413620	100.423584	80.353813	90.650223	90.650223	557267200
4	2004-12-01	90.635277	99.566803	83.920448	96.035034	96.035034	291772100
...	...	...	...	...	...	...	...
186	2020-02-01	1462.000000	1532.105957	1271.000000	1339.329956	1339.329956	37093300
187	2020-03-01	1351.609985	1410.150024	1013.536011	1162.810059	1162.810059	71392100
188	2020-04-01	1122.000000	1359.989990	1079.810059	1348.660034	1348.660034	46333100
189	2020-05-01	1328.500000	1441.000000	1299.000000	1428.920044	1428.920044	31867600
190	2020-05-29	1416.939941	1432.569946	1413.680054	1428.920044	1428.920044	1383501

191 rows × 7 columns

Fig.6.2 Data fetched from yahoo finance

After entering the data and hitting the submit button, the stock historical data are fetched from yahoo finance using the backend techniques. Then the data is cleaned and normalized only the close column is used to predict the future prices of the stock as this column lies contains close values of high and low column from previous day margin value. The data is the fed to the LSTM model and the future predictions are made.

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 4, ISSN No. 2455-2143, Pages 245-250
Published Online August 2020 in IJEAST (http://www.ijeast.com)

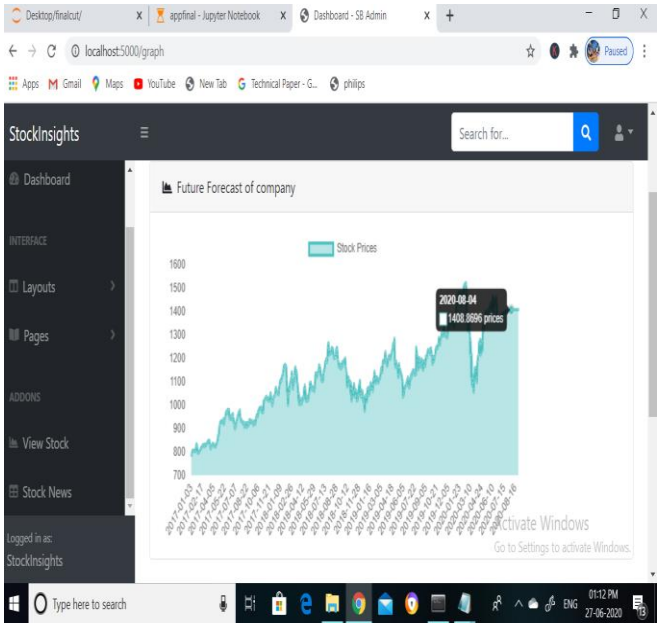

Fig6.3 Result of LSTM algorithm

The above output is generated from the input 1 where the future prices for n days are predicted and displayed as per the inputs taken. The graph has date on x-axis and stock prices on y-axis one can decide to buy or sell stock by viewing the future conditions of the stock prices in the market.

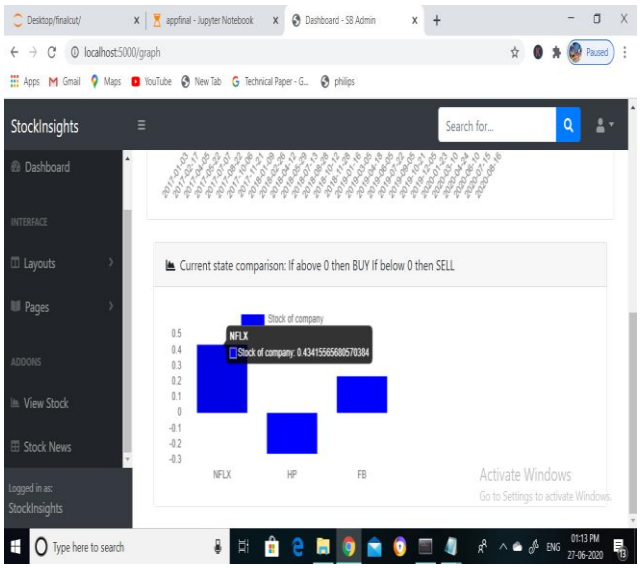

Fig6.4 Buy or sell graph of 3 stocks

The above chart is generated from the 2nd input given the three stock historical data are fetched from yahoo finance and based on the analysis made on the close column of the dataset based on the current situation of the market the return values are mapped and the comparison graph is generated the x-axis contains stock name and y-axis contains the returns of the stock expected if the y-axis value is less than zero then sell the stock if owned or else if the value is greater than zero then buy the stock.

VII. CONCLUSION AND ACKNOWLEDGEMENT

In this paper, an efficient and effective analysis of stock prices prediction model is build. Various studies conducted so far shows that SVM algorithm is the fastest unsupervised approach that can efficiently work on small dataset without overlapping and is resistant to noise and outliers, But for long term dependencies and large dataset LSTM are more preferable and accurate. The results of SVM are initially carried out which shows the accuracy of 95%. The proposed LSTM based model
demonstrated higher performance than existing SVM along with a low error rate and the performance of the LSTM is 98.39%. LSTM helped in choosing the best features from the large dataset. The proposed model gives better results over SVM. In future work, as stock prices are affected by many factors like political and economical news, public opinion, climate, etc. so these many factors are to be considered to predict more effective prices by taking into consideration.

We are thankful to all who provide us with the opportunity to complete this report. We would also like to thank our **Guide Prof. Sheetal Gawande** for our guidance and to encourage us as well as help us to solve our queries.
We also thank our **H.O.D** of Information Technology Department **Dr.Satish Kumar Varma** for this opportunity.
We give special gratitude to our **Principal Dr.Sandeep Joshi** who always encourages and motivates us to do innovative things that will increase our knowledge.

VIII. REFERENCE

[1] Hiransha M.(2018), NSE Stock Market Prediction Using Deep-Learning Models, in Proceedings of International Conference on Computational Intelligence and Data Science.

[2] Rao Sushrut Deepak(2017), Machine Learning approach in stock market, in proceedings of International of pure and applied mathematics.

[3] Lohith N(2017),Comparative study on machine learning technologies for stock prediction, in proceedings of (IJCMS).

[4] Aparna Nayak(2016), Prediction model for Indian stock market, in proceedings of the twelfth International Multi conference on Information processing.

[5] W.D Potter(2015), Sentimental analysis Machine Learning Stock Prediction analysis.

[6] S. Shen, H. Jiang, & T. Zhang(2012), Stock market forecasting using machine learning algorithms.

[7]J. Bollen, H. Mao, & X. Zeng(2011), Twitter mood predicts the stock market. Journal of Computational Science vol. 2 no. 1, (pp. 1-8).

[8] S. Asur, & B. A. Huberman(2010), Predicting the future with social media, in International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT).

[9] S. Qamar, & S. H. Adil(2009), Comparative analysis of data mining techniques for financial data using parallel processing, in Proceedings of the 7th International Conference on Frontiers of Information Technology.

[10]S.H. Adil, & S. Qamar(2009), Implementation of association rule mining using CUDA, in International Conference on Emerging Technologies (ICET).

[11] S. Fatima, & G. Hussain(2008), Statistical models of KSE100 index using hybrid financial systems, in Neurocomputing vol. 71 no. 13 (pp. 2742-2746).

[12]T.Wilson,P.Hoffmann,S.Somasundaran, J. Kessler, J. Wiebe, Y. Choi,& S. Patwardhan(2005), OpinionFinder: A system for subjectivity analysis, in Proceedings of emnlp on interactive demonstrations, (pp. 34-35).