International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 05 | May 2019 www.irjet.net p-ISSN: 2395-0072

Stock Market Prediction using Deep Learning and Sentiment Analysis

Subin Duresh[1], John Wesley[2], Utkarsh Agarwal[3], G. Paavai Anand[4]

[1,2,3,4]*Department of Computer Science Engineering, SRM Institute of Science and Technology*

Abstract **-** Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on a financial exchange (Wikipedia - Stock Market Prediction). Prediction in the stock market is challenging and complicated for investors. The successful prediction of a stock's future price will maximize investor's gains. In the stock market, social media sentiments have a high impact today more than ever. In this work, various prediction algorithms are analyzed to build a prediction model. The proposed model consists of two phases. Phase I deals with sentiment analysis in combination with historical analysis and phase II deals with Deep Learning. Sentiment Analysis is used to identify and extract sentiments of tweets on social media. The correlation between the sentiments and the stock prices is to be determined. The Deep learning module trains the model on the correlation so as to obtain a better model for prediction. The obtained results showed that the proposed model has better prediction with increase in accuracy and consistent results.

Index Terms **-** *Stock Market; Sentiment Analysis; Deep Learning; Artificial Neural Network (ANN); Stock Price Prediction; Historical Analysis.*

I. INTRODUCTION

STOCK price prediction has been at focus for years since it can yield significant growth by doing various investment. Predicting the stock market is not a simple task, mainly as a consequence of the close to random walk behavior of a stock time series.The **exchange is usually one in every of the foremost standard** investments **thanks to** its high profit. As the level of investing and trading grew, people searched for tools and methods that would increase their gains while minimizing the risk. The technical analysis deals with the sentimental analysis in relation to the tweets and a trend analysis on the stock prices. The model is trained on metrics of both these analysis so as to achieve good accuracy of the model.

From twitter the model will classify the tweets into positive and negative tweets using a feature extraction method built into the tool. Sentiment Classifier will take the respective tweets from twitter and pass it to the model. The model uses the information from feature extractor to polarise the tweets. While classifying the tweets, the model will check for the positive and negative tweets using the learned dictionary mapping of words and their positive or negative impact score and generate the input tr.

In parallel with above operation through tweets and history analysis it checks all the past history data.
That history data consists of the variation of the prices of stocks and help us understand the general trend of a selected stock.

The main role of historical analysis is to check the history of specific stock (the gradual up's and down's) in the past few years which will work as advantage while predicting the stocks. Performing both sentimental analysis and historical analysis are parts of Phase I

Phase II consists of predicting the model using deep learning. For stock prediction the module will be trained using ANN. The training of ANN will be done on the basis of results from phase I. The result obtained from sentimental analysis and historical analysis will be taken as data set for deep learning and the model will be trained to predict the future result. The combination of sentimental analysis and deep learning is used to increase the efficiency of the model.

II. PREDICTION METHOD

Commodities and shares were totally based on feeling . In the modern era of trading, people tried to find methods and tools which can accurately predict the stock prices for the selective shares increasing their profit and minimize their risk of getting loose. Fundamental analysis does not play a vital role in prediction but with the help of technical analysis that include sentimental analysis and deep learning made a great effect on the prediction of stocks.

A. Fundamental Analysis

Fundamental analysis is to get the information about respective organization by looking over their services and product quality, infrastructure, manpower and other factors as well. However most of the market runs on logical analysis rather than psychological factors.

B. Technical Analysis

The technical analysis comprises of two factor like twitter data and archive data and make certain assumptions about

the high and low for a stock. One of the popular approaches used is sentimental analysis. It deals with polarization of tweets by using sentiment analysis tool. This analysis can be used to predict the market price on daily basis but we will not use this approach because error rate for predicting is high. Another approach in technical analysis is use deep learning techniques on the basis of input feeds and has some assumption about the stock prices. The problem with both the techniques alone are not sufficient enough to get a precise prediction. But on combining both the approaches, sentimental analysis and deep learning error rate for the prediction value was reduced. Thus combination of both techniques also improved the consistency of the predicted prices of stock predicted by the model.

C. Sentimental Analysis

Sentiment analysis is a technique of analyzing the sentiment of a given set of input generally in the form of text. According to work by (Rui Ren, Desheng Dash & Tianxiang Liu, 2018), the twitter data is split into sentences instead of a collective document of a given day in order to get a better polarisation and a better data set; according to the work, parsing the tweets individually would yield better determinant value than parsing the tweet document as a whole. However we do not factor in the Monday Effect as it's more of a behavioural perspective than a intellectual one.

We determine the polarity of each tweet in the entire day and use the formula 1.1 to determine the day's polarity.

$$\frac{{}_{\Sigma}p_i - \Sigma n_i}{\Sigma t_i}$$

where pi is the ith positive tweet, ni is the ith negative tweet and ti is the ith tweet.

D. Deep Learning

Deep Learning is a subset of ML which deals with algorithms inspired by the structure and function of the brain called artificial neural networks. It consists of two different phases. One is training phase and another is prediction phase. The training phase involves training the model on the existing data with mean-squared error as the error metric. Error rectification was performed by back propagation of the mean squared error multiplied by a stochastic gradient ,calculated by the optimizer, to modify the weights of each neuron.

III. SYSTEM ARCHITECTURE

The situation at hand is an unstable system.There is a necessity for trend and pattern recognition is needed to know the system higher. Neural Network only needs input and output to learn. It creates weights to determine the steps taken to reach the output.

This system contains two modules separated at initial polarized/normalization phase where data from Twitter and stock archives are polarized/normalized separately. The neural network module analyses the polarized twitter data and the normalized stock data and maps to the given output. It changes the weights in the learning phase to give the most accurate output possible.

A. Sentimental Analysis Model

Sentiment analysis identifies whether a given statement is positive or negative sentiment. Our system uses this to get a set of twitter data and find how positive or negative each of the tweets in the data set is. The data set is refreshed per day and the degree of polarity is pushed into a new data set.

B. Stock Analysis Model

The stock price analysis tool helps visualize the general trend of the stock prices. It determines the regression coefficients of the closing price of stocks and plots a graph of the regression slope. This helps in understanding the predicted prices and serves as a sanity check for the prediction made by the deep learning model.

C. Deep Learning Model

The deep learning model is the backbone of the implemented system. It consists of a mixed input multi-branched deep learning neural network. The outputs of each branch are concatenated and correlated to obtain the output stock prices. It consists of hidden layers with 50 neurons each to find deep relationships between stock data, twitter data and their correlation to each other. With the combined knowledge of both types of the data, the system is empowered to provide precise and consistent predictions on the stock prices.

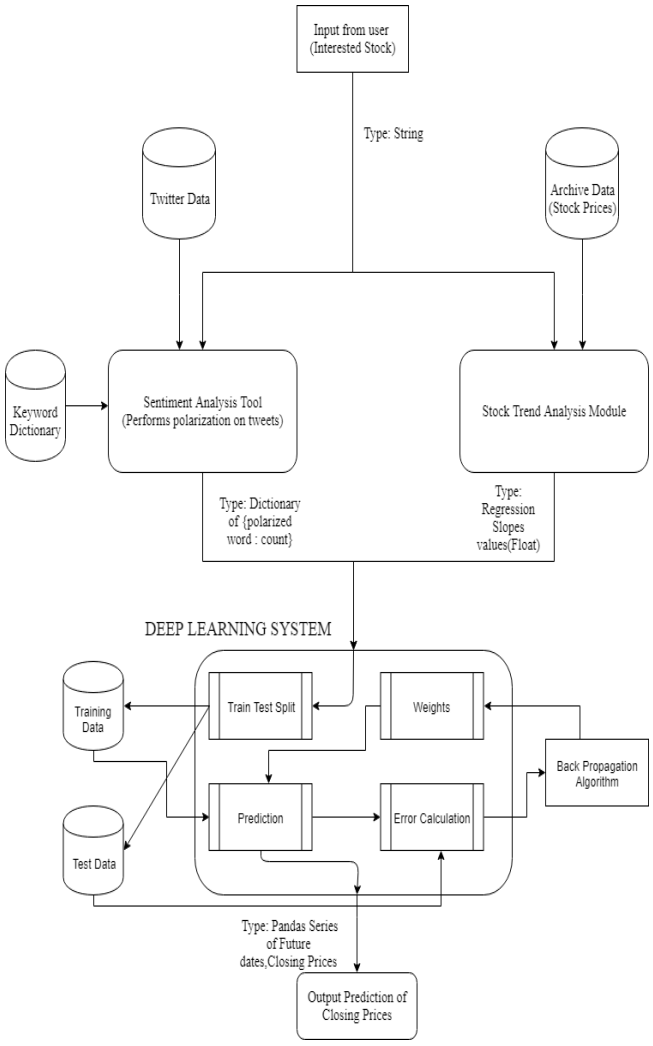

Fig. 1 Stock prediction - Architecture Diagram

IV. MODEL ANALYSIS

We used a feed forward neural network which has an input layer each for the two inputs - Stock data and Twitter sentiments data which results in two independent branches with three hidden layers each. Each layers has 150 neurons. An output concatenation layer with 2 neurons to concatenate the two branches is also present. And finally an output layer with a single neuron is also present in the ANN(fig. 2). The Backpropagation algorithm was used for modifying the network weights based on the mean squared errors. The network optimizer used is a stochastic gradient optimizer - Adam[1].

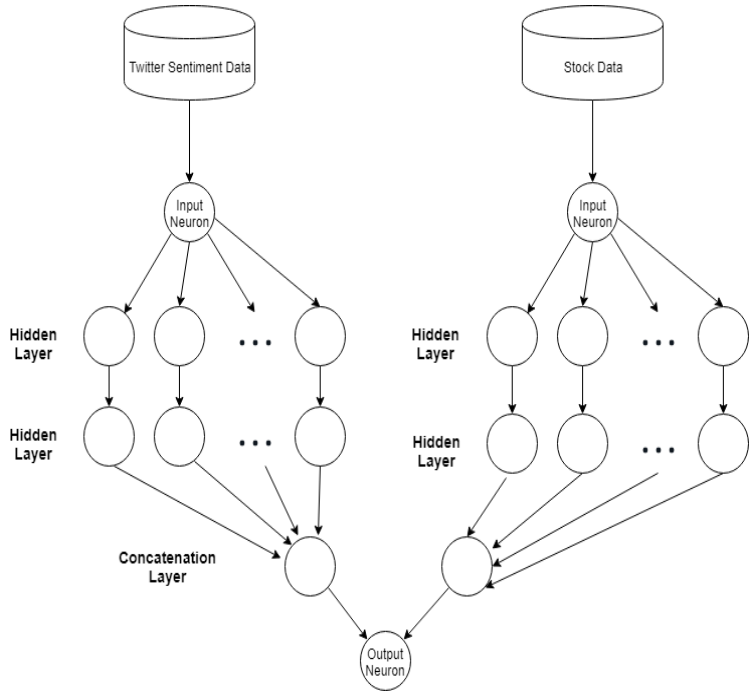

Fig.2. Deep Learning Model Architecture

A. Model Training Phase

During the model training phase, the normalized stock prices data and standardized twitter sentiment data was fed to their respective inputs. The data is a time-series data and therefore it was manually split as top 80% training data and bottom 20% validation data. The model was fit on to the data and compiled. An early stopping system was used to stop the model training at the best possible model performance.

B. Model Prediction Phase

The saved model is loaded and input to the model is the stock price of the current date. The target is the 'future' column that represents the future closing price of the stock. The model can be modified to predict the price for a preordained number days in the future.

C. Model Performance Analysis

The model showed a maximum loss of 2.71 and a minimum loss 0.071 mean square error during the training.

The prediction was tested for both 1 day in the future as well as 5 days in the future. The model was able to predict 1 day into the future with better accuracy but with no significant difference. The other performance metrics for an

average case model are mean absolute error of 0.484,mean squared error of 0.627, mean absolute percentage error of 9.486 amd a cosine proximity of -0.999.

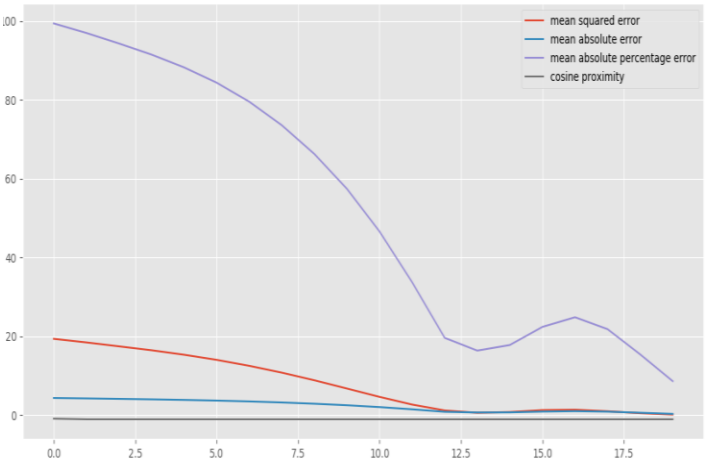

Fig.3 Stock prediction - Performance metrics graph of model

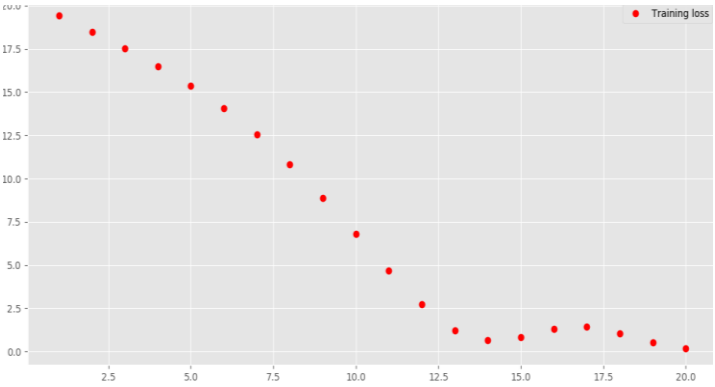

Fig.4 Stock prediction - Training loss graph of model

Figure 3 (performance metrics graph) is an example of the training metrics of the model which resulted in a prediction error score of 0.454 (mean squared error). Figure 4 (training loss graph) is the overall training loss of the model during the training phase. The graph shows a clear downward trend as the model trains well to the data and arrives at a minimum loss of 0.355. The model is saved and has a consistently low error score between 0.3 and 0.5.

The model is constantly undergoing performance improvements and can be used to obtain even more accuracy and consistency with more tweaking and trials.

V. CONCLUSION

The purpose of this study is to increase the accuracy to forecast the prediction of stock using the combination of sentimental analysis, stock trend analysis and deep learning. The ultimate goal was to increase the yield from the investment. The accuracy was found to lean towards precision with increased data . In conclusion we can say that if we train our system with more input data set, modify the deep learning module to perform more trial and errors tests, it has the potential to generate more accurate and significantly more consistent minimal error predictions. .

REFERENCES

[1] Diederik P. Kingma,Jimmy Lei Ba "Adam: A Method for Stochastic Optimization" ICLR Conference,2015.
[2] S. Nnewi - PREDICTING THE NIGERIAN STOCK MARKET USING ARTIFICIAL NEURAL NETWORK
[3] H.B.Kekre, Hitesh Makhija, Pallavi N.Halarnkar. Prediction of Closing Price of Stock Using Artificial Neural Network
[4] Zabir Haider Khan. Price Prediction of Share Market using Artificial Neural Network (ANN)
[5] Abhishek, Kumar, Anshul Khairwa, Tej Pratap, and Surya Prakash.A stock market prediction model using Artificial Neural Network."
[6] Khan, Zabir Haider, Tasnim Sharmin Alin, and Md Akter Hussain."Price prediction of share market using artificial neural network(ANN)."
[7] Stock market prediction, https://en.wikipedia.org/wiki/Stock_market_prediction