Volume 11, Special Issue I, May 2020
ISSN No. 0976-5697
International Journal of Advanced Research in Computer Science
CONFERENCE PAPER
Available Online at www.ijarcs.info

STOCK PREDICTOR ANALYSIS USING ML AND DATA SCIENCE

Sooryah P.
School of C&IT,REVA University
Bengaluru, India
sooryahprasath@gmail.com

Anindya Mahapatra
School of C&IT,REVA University
Bengaluru, India
iiscolder@gmail.com

Dhanshree R.
School of C&IT, REVA University
Bengaluru, India
Dhanprit.shree@gmail.com

Surjeet Kumar V.
School of C&IT, REVA University
Bengaluru, India

Ajil A.
School of C&IT,REVA University
Bengaluru, India
Ajil.a@reva.edu.in

Abstract: Stock Market is a dynamic market where the prices vary and it becomes difficult for an investor for predicting the prices considering external factors like factors like political situations, public image on the company according to efficient market hypothesis. Stock market is extremely volatile and ever evolving with constant developments and research in machine learning and deep learning. This paper deals with a comparative analysis between the most commonly used prediction methods such as Linear Regression, LSTM, CNN , both statistical and recursive learning models using tensor flow and machine learning to find the best fit for individual companies based on historical data.

Keywords*:* Linear Regression, LSTM, CNN, prediction, stock market.

INTRODUCTION

Stock Market is a public market for trading of company stock and derivatives at an agreed price. It is generally a dynamic market where the prices vary and it becomes difficult for an investor for predicting the prices considering external factors like factors like political situations, public image on the company according to efficient market hypothesis. Recently, a lot of interesting work has been done in the area of applying Machine Learning Algorithms for analysing price patterns and predicting stock prices and changes. Most stock traders nowadays depend on Intelligent Trading Systems which help them in predicting prices based on various situations and conditions, thereby helping them in making instantaneous investment decisions. Stock Prices are considered to be very dynamic and susceptible to quick
Changes because of the underlying nature of the financial domain and in part because of the mix of known parameters.
This model aims to create a comparative analytical predictive model with acceptable accuracy levels to be used for pragmatic stock market insight and pattern isolations from various trends historically. To attain this, using supervised learning methods including Recursive Neural Networks (feed forward networks) and Statistical Learning Technical from historical data.
Stock market analysis is an ever evolving, fluid methodology and requires multiple analysis methods in tandem for an accurate investment policy for buying or selling stocks in the international market. Mostly the investor sentiments are influenced by two analysis methods: Fundamental Analysis and Technical Analysis. Fundamental Analysis deals with the intangibles which cannot be statistically measured. For example, it can include something as major as economic scenario to something as minor as political scenario. It is left to the investor's acumen and not included in this statistical analysis. Technical analysis deals with tangibles i.e data and data history over time. It is a tedious study made based on past and future stock data put into charts through data visualisation and various algorithms.
Linear regression which is a simple statistical learning method, LSTM which is a recursive neural network model and stands for Long Short Term Memory, and Convolutional Neural Networks (CNN). This method is a part of neural network but it has delayed activation allowing for same signal to flow back into the network. Renewed and repeated flowing of data into the network allows for a more robust and accurate prediction in comparison to linear regression in theory. CNN is used in

this prediction model comparison because CNN allows for a better prediction by allowing for data inputs in layers. That means the sample data can be input periodically, intermittently and over a large duration of time. The cumulative data is then consolidated in a different layer before getting processed.

The main motivation of this model is to find a comprehensive analysis between the methods and find the various levels of accuracy for them individually as well as in tandem with each other by means of graphical overlay which may influence the evolution of stock market prediction models.

A custom field is also provided for individual company predictions over a user chosen time period. This can then be a potential investor by comparing the predictions through the 3 different methods in isolation and even together through graphs. Looking into the existing models and prediction algorithms for stock prediction and why there is room for improvement in this ever-evolving market.

MODELS COMPARED WITH RESPECT TO STOCKS

Machine Learning is a subset of Artificial Intelligence that deals with learning from data provided and making predictions on it's own through explicit programming.

LINEAR REGRESSION

Linear regression is a method used to model a relationship between a dependent variable (y), and an independent variable (x). With simple linear regression, there will only be one independent variable x. There can be many independent variables which would fall under the category of multiple linear regression. In this circumstance, we only have one independent variable which is the date. The date will be represented by an integer starting at 1 for the first date going up to the length of the vector of dates which can vary depending on the time series data. Our dependent variable, of course, will be the price of a stock. [1]

EQUATION : y= a + bx

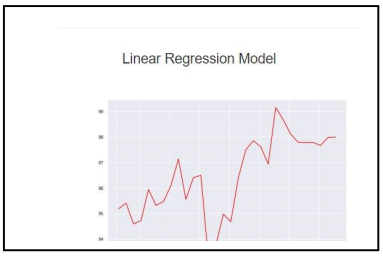
Where a is the y intercept, and y is the final predicted value.

Algorithm :

Input: x train : opening value of a company
Y_train : closing value
X_test : prediction value

1. Define (x_train, y_train,x_test)
2. For i=0 to n
3. X_sum=0,y_sum
4. X_mean=X_sum/n
5. Y_mean =Y_sum/n
6. xy_sum = 00
7. xy_sum + = x[i] * y[i]
8. x2_sum + = x[i] * x[i]
9. end
10. x_sum_sq = x_sum * x_sum
11. l= n * xy_sum - (x_sum * y_sum)
12. m = n * x2_sum - x_sum_sq
13. b = l / m 21: a = y_mean - (b * x_mean)
14. predict_value = a + b * x_test
15. return predict_value.

Fig 1 : Linear regression model prediction for MSFT



.LSTM

An LSTM memory cell, has the following three components, or gates:

1. Forget gate: the forget gate decides when specific portions of the cell state are to be replaced with more recent information. It outputs values close to 1 for parts of the cell state that should be retained, and zero for values that should be neglected.

2. Input gate : based on the input (i.e., previous output o(t-1), input x(t), and previous cell state c(t-1)), this section of the network learns the conditions under which any information should be stored (or updated) in the cell state

3. Output gate: depending on the input and cell state, this portion decides what information is propagated forward (i.e., output o(t) and cell state c(t)) to the next node in the network. [2]

Algorithm :

Input : Historical data from Quandl.

Output : Prediction for a period of time depending on the user requirement of individual company.

1. Start
2. Store the no of training sequences, length of sequence, and feature or dimension of sequence.
3. The no of input layers a is definer, and no of neurons on next layer b is defined and so on.
4. The network is trained with intermittent data.
5. The output of prediction is used as input for next step.
6. Repeat this until accuracy is achieved.
7. Provide test data from quandl.
8. Evaluate the accuracy in comparison to other algorithm.
9. Stop.

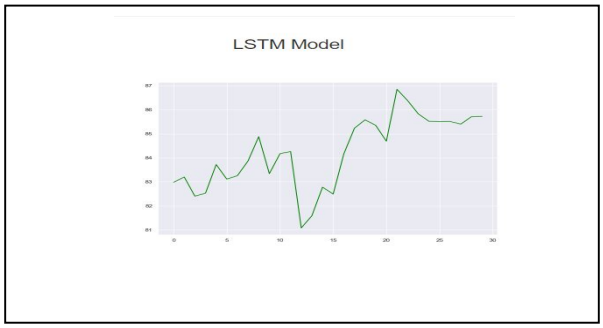


Figure 2 : LSTM model for MSFT

CNN

Convolutional Neural Network is a feed-forward neural network. Convolution model uses pooling and gathers input from the previous convolution model output. The difference is CNN has hidden layers that increases the efficiency of the model.

The more the hidden layers are, the higher feature it can extract and recognize from the input. People always use convolutional neural network in computer vision, such as face recognition, image classification but in this case 1D data from quandl is used for the predictions.[3]

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.[4]

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals)

RESULTS

As per the inputs the project output has similar prediction outcomes for all the methods to ensure accuracy. But depending on the data available and the type of prediction needed a model can be chosen. Below is a comparative overlay of the outputs for a single company for long term prediction. [5]

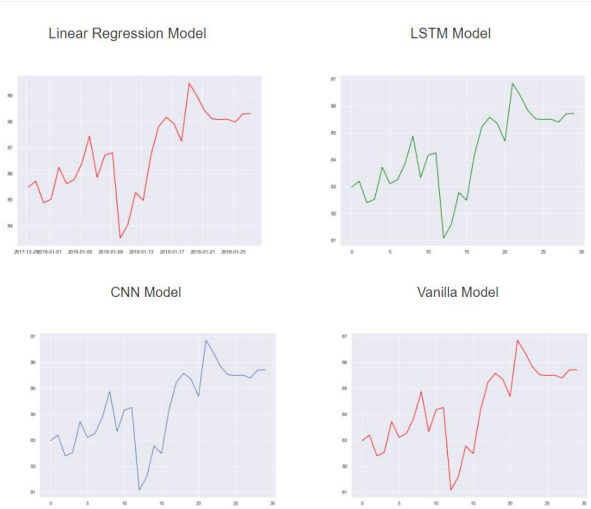


FIGURE 3 : COMPARATIVE GRAPH BETWEEN THE MODELS

CONCLUSION

We have taken historical data and compared the efficiency of prediction using multiple models. The values considered are the high, low, and opening and closing value of stocks. The scope for this includes an in depth study of the existing models and using investors and researchers make important stock market decisions in this ever changing and dynamic market.

For future scope more models and algorithms can be compared and used in conjunction along with multiple data sources for varied results and higher accuracy level.

.

REFERENCES

[1] Rajashree Dash Pradipta Kishore Dash , A hybrid stock trading framework integrating technical analysis with machine learning techniques, The Journal of Finance and Data Science, Vol. 2, issue 1, 2016, pp. 42-57.

[2] Vatsal H Shah, Machine Learning Techniques for Stock Prediction.

[3] Textual Analysis of Stock Market Prediction Using Breaking Financial News.

[4] L. J. Cao and F. E. H. Tay, “Support vector machine with adaptive parameters in financial time

[5]. Aspray, T. (1989). Individual stocks and macd. Technical Analysis of Stocks & Commodities