Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 4- Issue 2, Paper 10, July 2021

FUTURE-TREND-INDICATORS-FOR-TRADING-STOCKS USING MACHINE LEARNING

Waheeda Dhokley[1], Ninad Sawant[2], Om Chavan[3], Nabeel Wagle[4], Aleem Tajir[5]
[1]Professor, Department of Computer Engineering,
M. H. Saboo Siddik College ofEngineering,
Mumbai, India

Abstract— *Stock price prediction is one among the complex machine learning problems. It depends on a large number of factors which contribute to changes in the supply and demand. This paper presents the technical analysis of the various strategies proposed in the past, for predicting the price of a stock, and evaluation of a novel approach for the same. Stock prices are represented as time series data and neural networks are trained to learn the patterns from trends. Along with the numerical analysis of the stock trend. In this work, we propose an approach of hybrid modeling for stock price prediction building different machine learning and models. For the purpose of our study, we have used 'AAPL' stock values during the period March 15, 2016 till March 12, 2021. Using these regression models, we predicted the close values of AAPLE for 441 datasets. We exploit the power of LSTM regression models in forecasting the future AAPLE close values. Extensive results are presented on various metrics for the all the regression models. The results clearly indicate that the LSTM-based univariate model that uses one-week prior data as input for predicting the next week close value of the AAPLE time series is the most accurate model.*

I. INTRODUCTION

Usually the very first thing that comes to our mind when we are thinking of stocks is a price chart. You can see how the price changes overtime for a specific stock. However, there is a lot of other important charts or statistics that may improve your understanding of stock's behavior e.g. its trend or momentum. It may also help you to try to predict future prices (up to some accuracy). Researchers have proposed models on technical analysis of stock prices wherein the goal is to detect patterns in stock movements that lead to profit for the investors. For this purpose, various economic and stock price-related indicators have been proposed in the literature. Some of these indicators are: Bollinger Band, moving average convergence divergence (MACD), relative strength index (RSI), moving average (MA), momentum stochastics (MS), meta sine wave (MSW). In addition to these indicators, some of the well-known patterns in stock price movements like head and shoulders, triangle, flag, Fibonacci fan, Andrew's pitchfork, etc., are also considered as important indicators for investment in the stock market. These approaches provide effective visualizations to potential investors in making the right investment decisions. The rest of the paper is organized as follows. In Section 2, we explicitly define the problem at hand. Section3 provides a brief review of the related work on stock price movement prediction. In Section 4, we describe our research methodology. Extensive results on the performance of the predictive models are presented in Section 5. This section describes the details of all the predictive models that are built in this work and the results they have produced. Finally, Section 6 concludes the paper.

II. PROBLEM STATEMENT

The goal of our work is to collect the stock price of AAPLE from the Tiingo API over a reasonably long period of five and half years and develop a robust forecasting framework for forecasting the AAPLE stock values. We hypothesize that it is possible for a machine learning model to learn from the

Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 2, Paper 10, July 2021

features of the past movement patterns of daily AAPLE stock values, and these learned features can be effectively exploited in accurately forecasting the future stock values of the AAPLE series. In the current proposition, wehave chosen a forecast horizon of one year forthe machine learning models, demonstrated that the future AAPLE stock values can be predicted using these models with a fairly high level of accuracy. In the present work, we follow four different approaches in building long and short-term memory (LSTM) network-based models in order to augment the predictive power of our forecasting models. It must be noted that in this work, we are not addressing the issues ofshort-term forecasting which are of interest tothe intra-day traders. Instead, the propositionsin this paper are relevant for medium-term investors who might be interested in a weeklyforecast of the AAPLE stock values.

III. RELATED WORK

The currently existing work in the literature on time series forecasting and stock priceprediction can be broadly categorized in three clusters, based on the use of variables and the approach to modeling the problem. The first category of work mainly consists of models that use bivariate or multivariate regression on cross-sectional data. Due to their inherentsimplicity and invalidity of the linearity assumptions that they make, these models failto produce highly accurate results most of thetime. The propositions in the second categoryutilize the concepts of time series and other econometric techniques like autoregressiveintegrated moving average (ARIMA), Granger Causality Test, autoregressive distributed lag (ARDL), vector autoregression (VAR), and quantile regression to forecast stock prices. The third category of work includes learning-based approaches propositions using machinelearning.

Except for the category of work that utilizes learning-based approaches, one of the major shortcomings of the current propositions in literature for stock price prediction is their inability to accurately predict highly dynamic and fast-changing patterns in stock price movement. In this work, we attempt to address the problem by exploiting the power of machine learning models in building a very robust, reliable, and accurate framework for stock index prediction. In particular, we have used a long-and-short- term memory (LSTM) network-based deep learning model and studied its performance inpredicting future stock index values.

IV. METHODOLOGY

In Section 2, we mentioned that the goal of this work is to develop a predictive framework for forecasting the daily price movement of AAPLE STOCK. We collect the historical index values of AAPLE STOCK for the period: March 15, 2016 till March 12, 2021 from the Tiingo website [25].The raw AAPLE STOCK index valuesconsist of the following variables: (i) date, (ii)open value of the index, (iii) high value of theindex, (iv) low value of the index, (v) close value of the index, and (vi) volume of the stock traded on a given date.

We followed the approach of regression in forecasting the AAPLE STOCK index values. For this purpose, we used the variableopen as the response variable and the other variables as the predictors. We carried out some pre-processing of the data before usingit in training and testing the regression models. We design the following derived variables using the six variables in the raw AAPLE STOCK index records. These derived variables will be used for building predictive models.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 2, Paper 10, July 2021

The following five variables are derived and used in our forecasting models:

a)high_norm: it refers to the normalized values of the variable high. We use min- max normalization to normalize the values. Thus, if the maximum and the minimum values of the variable high are H_{max} and H_{min} respectively, then the normalized value high_norm is computed as: high_norm = (high - H_{min})/(H_{max} – H_{min}). After the normalization operation, all values of high_norm lie inside the interval [0, 1].

b) low_norm: this normalized variable is computed from the variable low in a similar way as high_norm is computed: low_norm =(low – L_{min})/(L_{max} - L_{min}). The values of low_norm also lie in the interval [0, 1].

c) close_norm: it is the normalized version of the variable close, and is computed as: close_norm = (close - C_{min}) / (C_{max} – C_{min}). The interval in which the values of this variable lie is [0, 1].

d) volume_norm: this variable is the normalized value of the variable volume. It is computed in a similar way as high_norm, low_norm, and the close_norm, and its values also lie in the interval [0, 1].

e) range_norm: this variable is the normalized counterpart of the variable range. The range for a given index record is computed as the difference between the high and the low values for that index record. Like all other normalized variables e.g., high_norm, low_norm, or close_norm, the variable range_norm also lies in the closed interval [0,1].

After we carry out the pre-processing andtransformation of the variables on the AAPLE STOCK data for the period March 15, 2016 - March 12, 2021, we use the processed data for building and testing the regression models based on machine learningand deep learning.

For training the regression models, we use thedata for the period March 2016 till March 2019 (which was a Friday). The models are then tested on the data for the period March 2019 - till March 2021 The data is collected from the Tiingo website and these are daily AAPLE STOCK values. The training dataset consisted of 817 records. On the other hand, there were 441 records in the test dataset encompassing 88 weeks. For the machine learning-based models, we used the daily data in the training set to construct the models, and then we predicted the open values of the AAPLE STOCK index for every day in the test dataset.

For building the deep learning- based LSTM models, however, we follow a different approach. The approach is called multi-step forecasting with walk-forward validation [27]. Following this approach, we build the models using the records in thetraining dataset and then deploy the model for forecasting the open value of the AAPLE STOCK index on a weekly basis for the records in the test dataset. As soon as the week for which the last round of forecasting was made was over, the actual records for that week were included in the training dataset forthe purpose of forecasting the next week’s open values of the AAPLE STOCK index.

For building the machine learning-based regression models, we considered two cases, which we discuss below.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 2, Paper 10, July 2021

Case I: As already been mentioned earlier, thetraining dataset included historical records ofAAPLE STOCK index values for the period March 2016 till March 2019 The trainingdataset included index values for 817 days. InCase I, the performance of the models was tested in terms of the accuracy with which they could predict the open values forAAPLE STOCK index records of the trainingdataset. In other words, in Case I, we evaluatethe training performance of the machine learning-based regression models. The predictions are made on daily basis.

Case II: In this case, the predictive models aretested on the test dataset and theirperformance is evaluated. The test data consists of historical records of AAPLE STOCK index values for the period March 2019 till March 2021. The performances of the models are evaluated in terms of their prediction accuracy of open values for each of the 441 days included in the test dataset. Hence, in Case II, we have evaluated the testperformance of the machine learning models.

In this work, we designed and evaluated eight machine learning-based regression models. These models are: (i) Relative Strength Index (RSI) (ii) Moving Average Convergence Divergence (MACD), (iii)regression tree, (iv) Stochastic. For thepurpose of evaluation of performances of these model, we use two metrics. The first metric that we use for evaluating a regression model is the value of the product-moment correlation coefficient between the actual andthe predicted values of the open values of theAAPLE STOCK index. The models exhibiting higher values of correlation coefficient are supposed to be more accurate. The second metric that we use for model evaluation is the ratio of the root mean square error (RMSE) values to the mean of the actualopen values in the dataset. The models that yield lower values of this ratio are more accurate.

V. RESULT

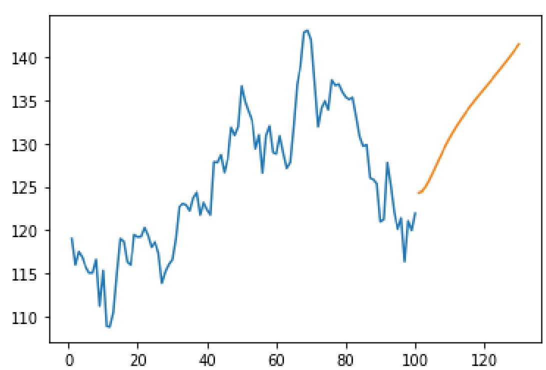

Fig 1. 1

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 2, Paper 10, July 2021

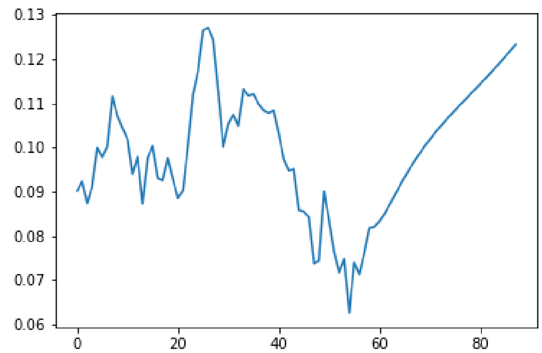

Fig 2. 1

V. CONCLUSION

In this paper, we have presented several approaches to prediction of stock index values its movement patterns on a weekly forecast horizon using machine learning, and LSTM-based deep learning regression models. Using the daily historical data of AAPL stock values during the period March 15, 2016 till March 12, 2021, we constructed, optimized, and then tested the predictive models. Data pre-processing and data wrangling operations were carried on the raw data, and a set of derived variables are created for building the models. Among all the machine learning and deep learning-based regression models, the performances of the LSTM- based deep learning regression models were found to be far too superior to that of the machine-learning-based predictive models. The study has conclusively proved our conjecture that deep learning-based models have much higher capability in extracting and learning the features of a time series data than their corresponding machine learning counterparts. It also reveals the fact that multivariate analysis is not a good idea in LSTM-based regression, as univariate models are more accurate and faster in their execution. As a future scope of work, we will investigate the possibility of using generative adversarial networks (GANs) in time series analysis and forecasting of stock prices.

REFERENCE

[1] Klaus Adam, Albert Marcet, and Juan Pablo Nicolini. 2016. Stock market volatility and learning. The Journal of Finance 71, 1 (2016), 33–82.

[2] Ayodele Ariyo Adebiyi, Aderemi Oluyinka Adewumi, and Charles Korede Ayo. 2014. Comparison of ARIMA and artificial neural networks models for stock price prediction. Journal of Applied Mathematics 2014 (2014).

[3] Ryo Akita, Akira Yoshihara, Takashi Matsubara, and Kuniaki Uehara. 2016. Deep learning for stock prediction using numerical and textual information. In Computer and Information Science (ICIS), 2016IEEE/ACIS 15th International Conference on. IEEE, 1–6.

[4] Kyunghyun Cho, Bart Van Merri¨enboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoderdecoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014). [5] Qiyuan Gao. 2016. Stock market forecasting using recurrent neural network. Ph.D. Dissertation. University of

Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 4- Issue 2, Paper 10, July 2021

Missouri– Columbia.

[6] Felix A Gers and J¨urgen Schmidhuber. 2000. Recurrent nets that time and count. In Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on, Vol. 3.IEEE, 189–194.

(2015), 212–222.

[7]Xavier Glorot and Yoshua Bengio. 2010.Understanding the difficulty of training deepfeedforward neural networks.. In Aistats, Vol. 9. 249–256.

[8] Mustafa G¨o¸cken, Mehmet Oz¸calıcı, Aslı Boru, and Ay¸se Tu˘gba ¨ Dosdo˘gru. 2016. Integrating metaheuristics and Artificial Neural Networks for improved stock price prediction. Expert Systems with Applications 44 (2016), 320–331.

[9] Alex Graves and J¨urgen Schmidhuber. 2005. Framewise phoneme classification with bidirectional LSTM and other neuralnetwork architectures. Neural Networks 18, 5(2005), 602–610.

[10]

[11] Klaus Greff, Rupesh Kumar Srivastava,Jan Koutn´ık, Bas R Steunebrink, and J¨urgen Schmidhuber. 2015. LSTM: A Search SpaceOdyssey. arXiv preprint arXiv:1503.04069(2015). [11] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, and J¨urgen Schmidhuber. 2001. Gradient flow inrecurrent nets: the difficulty of learning long-term dependencies. (2001).

[12] Sepp Hochreiter and J¨urgen Schmidhuber. 1997. Long short-termmemory. Neural computation 9, 8 (1997),1735–1780.

[13] Yong Hu, Bin Feng, Xiangzhou Zhang, EWT Ngai, and Mei Liu. 2015. Stock tradingrule discovery with an evolutionary trend following model. Expert Systems with Applications 42, 1

[14] G Kavitha, A Udhayakumar, and D Nagarajan. 2013. Stock Market Trend Analysis Using Hidden Markov Models. arXiv preprint arXiv:1311.4771 (2013).

[15] Kyoung-Jae Kim and Hyunchul Ahn. 2012. Simultaneous optimization of artificial neural networks for financial forecasting. Applied Intelligence 36, 4 (2012), 887–898.

[16] Leonel A Laboissiere, Ricardo AS Fernandes, and Guilherme G Lage. 2015.Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks. Applied Soft Computing 35 (2015), 66–74.

[17] Lili Li, Shan Leng, Jun Yang, and Mei Yu. 2016. Stock Market Autoregressive Dynamics: A Multinational ComparativeStudy with Quantile Regression. Mathematical Problems in Engineering 2016 (2016).