INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 4 - V7I4-1335)
Available online at: https://www.ijariit.com

Prediction of stocks and options in stock market using Machine Learning

Yathisha T. S.
yathishats.te17@rvce.edu.in
RV College of Engineering, Bengaluru, Karnataka

Nirmal Jain C. M.
nirmaljaincm.te17@rvce.edu.in
RV College of Engineering, Bengaluru, Karnataka

Mithun T. P.
mithuntp@rvce.edu.in
RV College of Engineering, Bengaluru, Karnataka

ABSTRACT

Machine learning is a type of artificial intelligence that generates predictions based on data by using past stock market indexes as training ground values. Machine learning makes use of a variety of models to make prediction more accurate and easier. The use of machine learning in stock market prediction technology is a trend which helps make better decisions while choosing assets to own. The aim of stock market prediction is to analyze trends and be able to predict the future value of an asset value over time. The analysis in this paper concentrates on asset value estimation utilizing LSTM-based machine learning. The results for estimation of prices are conducted on assets like Gold, Bitcoin and Apple Inc. stock. Open, close, low, and high data of assets, obtained from Yahoo Finance are taken into account while training the model. The proposed model can estimate the stock market's movement accurately for the next 30 days.

Keywords—*Yahoo Finance, LSTM, Open, Close, High, Low*

1. INTRODUCTION

Machine Learning is a field of instructing computers to understand and act like human individuals, and to demonstrate their learning curve over time in an intelligent and autonomous manner, by utilizing data and information available or fed to the computers, in the form of observations and real-world engagements.

A stock (sometimes called equity) is a business entity that implies control of a fraction of an organization. This authorizes the investor/stockholder to the fraction of the organization's assets and also profits of the organization proportional to the number of shares possessed by the individual stakeholder. These contracts typically adhere to strict guidelines established to protect investors from unscrupulous activities. They have historically excelled in almost all other asset portfolios over time.

Correct stock predictions can give sellers and buyers great benefits. Prediction is often said to be chaotic and not random, meaning it may be predicted by closely investigating the history of the stock market in question. Machine education is an useful way of expressing these kinds of activities. It predicts an increase in precision of the market value that is close to the tangible value. Machine learning has attracted the interest of many researchers by its efficient and precise measurements in the field of stock prediction.

This paper discusses the development of an LSTM model with features that can save significant bits of data throughout training, which are utilized to estimate prices of different assets for the near future.

2. RELATED WORK

Stock exchanges are large financial organizations where everyday trades involving millions and billions of dollars are made some see it as a quick way to make money while others take a more traditional approach to investing in a company and reaping long-term benefits a thorough investigation of the company and its performance can aid in increasing financial gains machine learning systems have shown to be effective at forecasting stock values a thorough analysis and documentation were used in paper [1] to show the performance of Apple INC's stock price using multiple linear regression and root mean squared error. The results are promising but can be improved by taking into consideration more parameters.

Adaptive boosting, or AdaBoost, a boosting-based machine learning technique, is a technique which allows weak classifiers to assemble into a strong classifier algorithm. Paper [2] proposes using conventional technical indicators to forecast rise or fall in the futures market and compiling these indicators using this algorithm in a unique way in order to discover underlying trends. This study demonstrates that weak classifiers can effectively filter data noise in futures markets, demonstrating that machine learning can produce a better result.

The reliability analysis and the Causality cause-effect test were utilized in paper [3] to assess the causality of the 3-month copper futures on the SHFE and LME market. The results demonstrate that the Regression analysis of copper in SHFE is the future price of copper in LME. It illustrates that the global market influence

has risen significantly attributable to the SHFE price of copper future.

By using tick data from various stock prices, Paper [4] evaluates the efficacy of commonly used technical indicators for intraday recommendations. It is demonstrated that utilizing evolutionary computation, the optimal combination of a few indicators chosen for each stock yields a good prognosis on the future price level a few ticks ahead.

The signal noise difference method is introduced in Paper [5], and it is used to predict commodity futures price prediction. A related transaction strategy is built based on the prediction rules mined from the data of 25 potential prediction indicators of SHFE CU. The proposed transaction method is tested using market data from 2009 to 2013, yielding a 147.85 percent yearly yield. In addition, the study discusses a number of changes that could be made to improve this model.

In study [6], both stock prices and Twitter data were included in the prediction method. Regression and LSTM model is used to predict stock. The purpose of this research is to compare the approaches used in the past for predicting stock market value. The assessment of relevant experience results aids in defining the most efficient algorithm and constructing the system model basis from the selected algorithm.

The stock market is a significant component in determining the economy of a country. The usage of technology in stock trading is increasing, which means that stock price volatility is also rising. The goal of Stock Market Prediction is to forecast a company's financial stock prices in the future. Machine learning, which produces forecasts based on current stock market indices by training on their prior values, has become a popular trend in stock market prediction technologies recently. Machine learning makes use of a variety of models in order to create accurate predictions. The application of logistic regression-based machine learning to predict stock values is the subject of paper [7]. The following factors are taken into account: open, close, low, high, last, total trading quantity, and turnover. Stock market institutions will benefit greatly from good stock prediction since it will bring real-world solutions to the challenges that stock investors encounter.

The paper [8] uses multivariate time series analysis to try to anticipate global agricultural commodities futures prices. The idea behind this paper is that datasets of agricultural commodity futures prices contain a mix of long- and short term information, as well as linear and non-linear structure, for which traditional approaches may fail. The Long- and Short-Term Time-series Network (LSTNet) is used to predict this issue. Based on three performance evaluation measures and two performance difference tests, empirical results reveal that LSTNet outperforms various state-of-the-art baseline approaches.

3. DATA MODELLING

Stock market forecasting is a complicated area to analyze due to various variables that have yet to be taken into consideration. However, with the proper use of machine learning techniques, it is possible to link prior data to current data and train the system to learn from it and make necessary predictions.

1. Data Collection

The data for the analysis in this paper is obtained from Yahoo Finance. The needed stock prices and other important variables are recorded in roughly 9 lakh entries in the dataset. For each day of the year, the data represented stock prices at certain time periods. It is divided into several parts, including date, symbol, open, close, low and high. Assets such as Gold, Apple INC stock and Bitcoin data are considered for the analysis of this paper. The data was in the form of a csv file, which was read from and converted into a data-frame using Python's Pandas module. Following that, the data was normalized using Python's TensorFlow package, and the data was separated into training and testing sets.

2. LSTM Model

Although there are many other machine learning models which can be used to replicate the proposed approach to get similar results, this particular study focuses on the LSTM model.

LSTM networks are a form of recurring neural system which might be dependent on the order in succession. In complicated problem areas, such as machine translation, speech recognition and more, this behavior is necessary. The model's architecture consists of a forget gate ,where LSTM ends up forgetting relevant data after reaching it that isn't beneficial in a learn gate where existing data and STM are coupled as then vital knowledge obtained that has recently been taught from STM can be applied to the input,.The information from LTM which was not lost was put in combination with the event and STM in the remember gate. Use gate also uses LTM, STM, and Event to predict the output of the current event which works as an updated STM as shown in Fig. 1

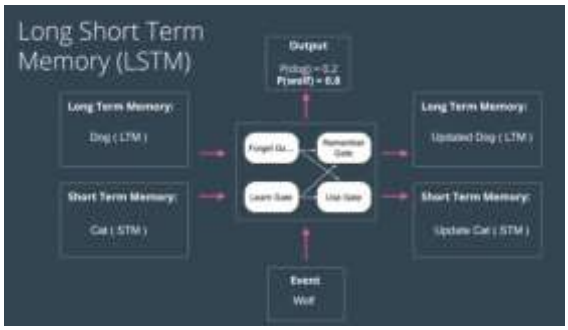

Fig. 1. Architecture of LSTM

4. PROPOSED ARCHITECTURE FOR LSTM

The enhanced version of Recurrent-Neural Networks (RNN) is LSTM, which keeps information from past states. These vary from RNNs in that they have long-term dependencies, whereas RNNs work on a short-term basis. Establishing a link between the recent and the current information. This denotes that the information interval is Compared to LSTM, it's a lot smaller. The major reason for keeping this model on hand is to test it out. The problem with market forecasting is that it is dependent on enormous amounts of data. A large quantity of data and are mostly reliant on the market's long-term history. As a result, LSTM reduces error by assisting RNNs by keeping knowledge from earlier stages, making prediction more accurate. Thus, demonstrating that it is far more dependable than other methods

Since the stock market includes the processing of large amounts of data, the gradients in the weight matrix may become very small, lowering the learning rate. This is the Vanishing Gradient issue. This is avoided by using LSTM.

Remembering the cell, input forget and output gate are all components of the LSTM. For long-term propagation, the cell remembers the value, which the gates control. A sequential

model has been created in this study, as illustrated in Fig. 2., that includes stacking two LSTM layers on top of each other with a 100-output value. Layer[0], layer[1], and layer[2] are the three inputs to the layer. A dropout value of 0.3 has been set, which implies that throughout the training process, 0.3 of the total nodes will be frozen to avoid data overfitting and also speed up the training. Ultimately, the core dense layer is incorporated, which outputs as 1 dense layer is a deeply connected regular neural network layer. The model is built utilizing the 'mean squared error' loss function and the 'adam' optimizer to keep the error low throughout the process, and accuracy is chosen as the prediction metric.

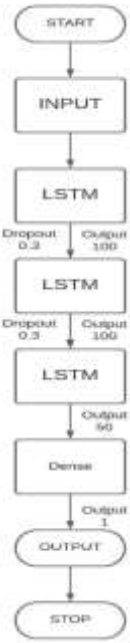

Fig. 2. Architecture of proposed LSTM Model

5. RESULTS AND DISCUSSION

The proposed system is trained and tested on data from Yahoo Finance. It is divided into training and testing sets, and when run through the model, it generates the following results.

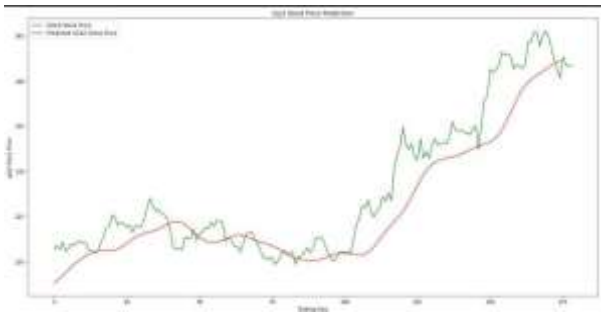

Fig. 3. Plotting the predicted and original asset - GOLD (@15 epochs)

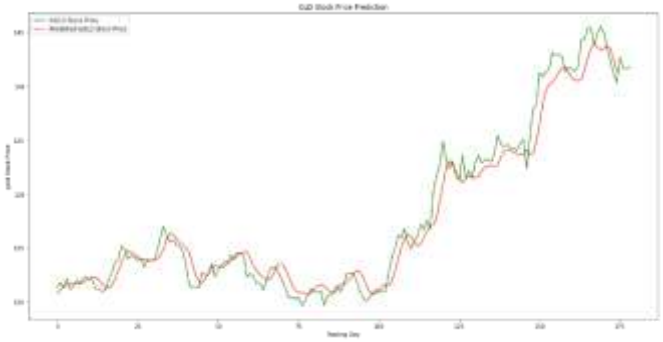

Fig. 4. Plotting the predicted and original asset – GOLD (@100 epochs)

The prediction is plotted in Fig. 3 and Fig. 4. for Gold (using different weight files generated by training the model for 10 and 100 epochs). where the green line represents the actual stock price and the red line represents the predicted price of the gold. Similarly, the graph for bitcoin is shown in Fig. 5 and Fig. 6.and the comparison between the gold and bitcoin about the loss of each individual epoch is shown in Table 1.

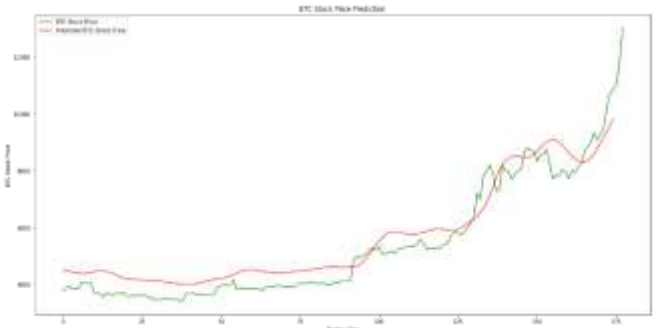

Fig. 5. Plotting the predicted and original asset –Bitcoin (@15 epochs)

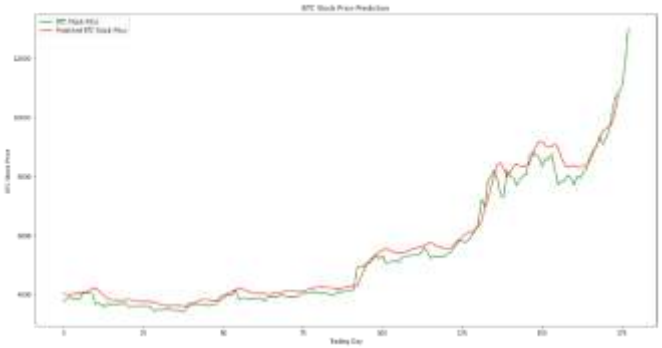

Fig. 6. Plotting the predicted and original asset – Bitcoin (@100 epochs)

Comparison of losses while training the models with Gold and Bitcoin assets is done for 15 and 100 epochs respectively and the results are tabulated in Table 1.

Table 1. Comparison of loss while training

Assets	Loss (@15 epochs)	Loss (@100 epochs)
Gold	0.002088	0.0004897
Bitcoin	0.001139	0.000369

The proposed model is then used to estimate the asset's price for the next 30 days. The chosen asset is Apple Inc. Fig. 7. depicts value of Apple stock for next 30 days, with the blue colour depicting the trained data and the orange colour depicts the future 30 days estimated values of the Apple stock.

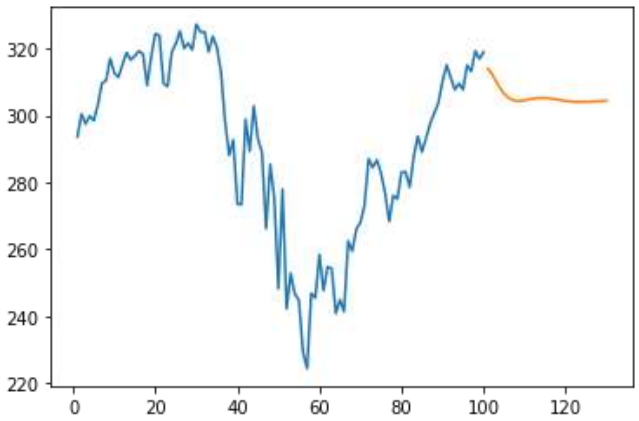

Fig. 7. Prediction of Apple stocks for next 30 days

6. CONCLUSION

The proposed model in this paper is based on LSTM architecture. It is a type of Neural Network, whose working is based on grasping key aspects of the data during the training process, and keeping it in memory for further use. This is especially helpful whenever the data is predictable based on past events (or occurrences) and has a pattern to it. This mechanism helps the model proposed in this paper achieve its accuracy.

By using this LSTM architecture, it is able to predict the assets' prices for a varied number of days. In this paper the analysis is done for 30 days in the future. When it comes to prediction of stock or asset movement for the future 30 days, the proposed model tends to use the data that is stored in the memory and tries to predict the values. The proposed model accurately estimates the price of some assets which are tabulated in the previous section.

7. FUTURE SCOPE

The accuracy of the stock market prediction system can be improved in the future by using a much large - scale dataset than currently used. Furthermore, the combination of technical analysis and various indicators and functions into the training of the model, as well as the use of other emerging machine learning models, could be studied to determine the prediction accuracy.

Sentiment analysis from social media platforms like Twitter, Instagram, and Facebook can be used through machine learning to learn how news affects a company's stock prices.

8. REFERENCES

[1] Shruti Shakhla, Bhavya Shah, Niket Unadkat, Vyom Kanani, and Pratik Kanani, " Stock Price Trend Prediction Using Multiple Linear Regression", *International Journal of Engineering Science Invention (IJESI),* pp. 29-33, 2018.

[2] Y. Wu, J. Mao and W. Li, "Prediction of Futures Market by Using Boosting Algorithm", *International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET),* pp. 1-4, 2018.

[3] Z. Zhi-bo and S. Tong-hua, "An Empirical Analysis on the Casual Relationship of the Futures Copper Price between SHFE and LME", *International Conference on Management Science and Engineering*, pp. 2289-2292, 2006.

[4] Mieko Tanaka-Yamawaki and Seiji Tokuoka,"Adaptive use of technical indicators for the prediction of intra-day stock prices", *Physica A: Statistical Mechanics and its Applications,Volume 383, Issue 1,* 2007

[5] J. Zheng and S. Peng, "Commodity Futures Price Prediction and Trading Strategies -- A Signal Noise Difference Approach", *2013 International Conference on Information Science and Cloud Computing Companion, Guangzhou*, China, pp. 411-417, 2013

[6] Ishita Parmar *et al*., "Stock Market Prediction Using Machine Learning", *First International Conference on Secure Cyber Computing and Communication (ICSCCC),* pp. 574-576, 2018.

[7] M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques", *3rd International Conference on Computer and Information Sciences (ICCOINS),* pp. 322-327, 2016.

[8] Hongbing Ouyang, Xiaolu Wei and Qiufeng Wu , "Agricultural commodity futures prices prediction via long- and short-term time series network", *Journal of Applied Economics, 22:1*, pp. 468-483, 2019.

