

Advanced Trading Bot Using Deep Learning

Vaishnav Kalbhor¹, Siddhant Welinkar², Rashi Agarwal³ and Prof. Neha Chaube⁴

^{1,2,3}MIT-ADT University, Loni Kalbhor

⁴Guide, MIT-ADT University, Loni Kalbhor

Abstract -The Era of Training Data through Machine Learning processes is a popular and effective way to deduce the future outcome of things with variable data plots. Trading with trading algorithms commonly known as trading bots have been in the market for a few years now, Algorithmic Trading accounts for a surprising 35.1% of the stocks that are listed in the USA stock exchange, which is much higher than that of the assets managed by consumer traders. Machine learning with its specifications in deep learning is a remarkable way for data analytics and prediction of future plots, hence leveraging the vast data of historic prices and the predicting capabilities we plan to create a trading bot for the industry.

Trading is a very old concept of commerce; the origin of trading is the barter system, having a fixed price of a commodity depending on the price of a superior commodity well described as cross elasticity in economics. The first stock exchange in this world is the “Amsterdam Stock Exchange” which was established in the year 1602. The volatility in the market prices of stocks/indexes may seem random but it is caused by an event or simple demand-supply changes. There are two main stock exchanges in India: NSE- National Stock Exchange and BSE- Bombay Stock Exchange. The historic trading data of each stock/Index is available on BSE and NSE websites and you can also use a private paid service for live in market data. The dataset consists of the opening and closing prices of the stocks/ indexes from 1994 to date. The use of multiple algorithms on the same data sets to get and find the accuracy for each of the algorithms and then a way to use reinforced learning to improve the accuracy by cross-referencing the prediction with different algorithms.

The method we are proposing involves multiple algorithms and one deep learning model, in order to get the highest accuracy, we are using some indicator functions of the values EMA – Exponential moving average over discreet time periods, Stochastic RSI- Relative Strength Index from the set ranges of. So we will be making a combination of these and other significant indicators to help increase the accuracy of the model.

INTRODUCTION

Stock Market prediction and analysis is the act of trying to determine the future value of company stock or other financial instruments traded on an exchange. The stock market is an important part of the economy of the country and plays a vital role in the growth of the industry and commerce of the country that eventually affects the economy of the country. Both investors and industry are involved within the stock exchange and need to understand whether some stock will rise or go over a particular period of your time . The stock exchange is that the primary source for any company to boost funds for business expansions. Stock market is based on the very concept of demand and supply. If the demand for a company's stock is higher, then the company share price increases and if the demand for the company's stock is low then the company share price decrease.

The NSE is one of the leading stock markets of India, located in Mumbai. The NSE was established in 1992. NSE was the primary exchange within the country to supply a fully automated screen-based electronic trading system, that offered easy trading facilities to the investors spread across the length and breadth of the country. The NIFTY 50 index is the National Stock Exchange of India for the Indian equity market. It represents the weighted average of fifty of the Indian company stocks in 12 sectors and is one among the 2 main stock indices utilized in India, the opposite being the BSE Sensex.II.

Due to the involvement of many industries and companies, it contains very large sets of data from which it is difficult to extract information and analyze their trend of work manually. The application we have developed, not only will help in predicting the future movement of the stock, but also will automate the data retrieval, trend analysis, predictive analysis, and insights generation of stock, just at the click of a button. Stock market analysis and prediction will successfully predict a stock's future price which could yield significant profit. This is

done using large historic market data of 12 months in this project, to represent varying conditions and confirm that the time series patterns have statistically significant predictive power for a high probability of profitable trades and high profitable returns for the competitive business investment.

LSTM

Long Short Term Memory networks – generally called “LSTMs”. They are a special kind of RNN that are capable of learning long-term dependencies. They were introduced by Hochreiter&Schmidhuber (1997), and were refined and popularized by many people in the following work. They work well with a large variety of problems and hence are now widely used.

LSTMs are designed to avoid the long-term dependency problem. Remembering the information for long periods of time is their default behavior.

In this all neural networks that are recurrent have the form of a chain of repeating modules of the neural networks.

For comparison between standard RNN’s repeating structure of the modules. It has a very simple architecture, that is a single tanh layer(Fig 4.1).IV.

UNITS

Use either SI (MKS) or CGS as primary units. (SI units are strongly encouraged.) English units may be used as secondary units (in parentheses). This applies to papers in data storage. For example, write —15 Gb/cm² (100 Gb/in²).[¶] An exception is when English units are used as identifiers in trade, such as —3½ in disk drive.[¶] Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity in an equation.

The SI unit for magnetic field strength H is A/m. However, if you wish to use units of T, either refer to magnetic flux density B or magnetic field strength symbolized as $\mu_0 H$. Use the center dot to separate compound units, e.g., —A·m².[¶]

FEATURES OF THE PROJECT

- The program is trained with a very large dataset over different parameters. It predicts the results

accurately over a large set of time frames, days, weeks, and months

- Using Random Forest Linear Regression, KNN, and other machine learning algorithms accuracy was found and by doing cross-validation the final algorithm was chosen – Deep Learning Model – LSTM.
- The data once downloaded is stored in a folder, similarly the training-testing data and its plotting and predicted output and its graphical representation are stored in a folder.
- The dataset used is from Yahoo Finance and contains many parameters, Data is cleaned for the parameters to be used
- With the current LSTM model, we are able to train, test and predict on any given instrument trading on the market.
- An approach towards the use of LSTM i.e ~ Long Short Term Memory neural networks with a much better accuracy rate in comparison to the other machine learning approaches used.

LITERATURE SURVEY

1] Predictive Analytics a review of Trends and Techniques.- Vaibhav Kumar Department of Computer Science & Engineering, DIT University, Dehradun, India M. L. Garg Department of Computer Science & Engineering, DIT University, Dehradun, India “Time series analysis is a statistical technique that uses time series data which is collected over a time period at a particular interval. It combines traditional data mining techniques and forecasting. The time series analysis is divided into two categories, namely the frequency domain and the time domain. It predicts the future of a variable at future time intervals based on the analysis of values at past time intervals. It is used in stock market prediction and weather forecasting very popularly

2]Kunjir, D. Joshi, R. Chadha, T. Wadiwala and V. Trikha, "A Comparative Study of Predictive Machine Learning Algorithms for COVID-19 Trends and Analysis," 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2020, pp. 3407-3412. In this paper, the authors have discussed the use of CNN, LSTM, and Decision Trees. These algorithms were used to Predict and Analyse data of India and China. For the result and evaluation, the

developed models were tested on the r2 score. There were two tables for both datasets one for China and one for Ontario's dataset.

3]Rong, Shen & Bao-wen, Zhang. The research of regression model in the machine learning field. MATEC Web of Conferences. 2018 This paper introduces the algorithm and model of the field of machine learning. A linear regression model is used to analyze the sale of iced products of the company and the effect of temperature variation on the sale. The data was cleansed and analyzed at the same time. Then the authors chose the forecast temperature as the independent variable and the sale of iced products as the dependent variable to establish a simple linear regression model for analysis.

4]D. Karmiani, R. Kazi, A. Nambisan, A. Shah and V. Kamble, "Comparison of Predictive Algorithms: Backpropagation, SVM, LSTM and Kalman Filter for In this research paper, the authors have given a comparison between Backpropagation, SVM, LSTM, and Kalman Filter for the stock market prediction. A dataset comprising of data of 9 companies is taken and then the different algorithms were implemented, For 10 runs of the SVM algorithm we got approximately 66.9823 mean accuracy with 0.05256 standard deviations. This shows that SVM performance is consistent for 10 runs this is due to the nature of the SVM algorithm. For 10 runs of the LSTM algorithm, we got approximately 68.51635 mean accuracy with a standard deviation of 0.71779. the Backpropagation algorithm, we got approximately 68.649 mean accuracy with a standard deviation of 0.55375. As it can be observed Backpropagation performs similar to SVM but is faster. It has a higher fluctuation in accuracy as compared to LSTM so that might cause an issue when a steady accuracy is required.

5]Ali, Jehad & Khan, Rehanullah& Ahmad, Nasir & Maqsood, Imran. (2012). Random Forests and Decision Trees. International Journal of Computer Science Issues(IJCSI) This paper there is a comparison of the classification results of two models I.e. random forests and the J48 for classifying versatile datasets Classification parameters consist of correctly classified instances, incorrectly classified instances, F-Measure, Precision, Accuracy and

Recall. There is a discussion on the pros and cons of using these models for large and small data sets. The classification results show that Random Forest gives better results for the same number of attributes and large data sets i.e. with a greater number of instances, while J48 is handy with small data sets (less number of instances)

6]Pricope, Tidor-Vlad. (2021). Deep Reinforcement Learning in Quantitative Algorithmic Trading: A Review This paper reviews the progress made so far with deep reinforcement learning in the subdomain of AI in finance, more precisely, automated low-frequency quantitative stock trading. Many of the reviewed studies had only proof-of-concept ideals with experiments conducted in unrealistic settings and no real-time trading applications. We conclude that DRL in stock trading has shown huge applicability potential rivaling professional traders under strong assumptions, but the research is still in the very early stages of development.

7]De la Torre, Oscar &Aguilasocho-Montoya, Dora & Álvarez García, José. (2021). Testing an Algorithm with Asymmetric Markov-Switching GARCH Models in US Stock Trading. Symmetry. Vol. 13. 2346. The use of Markov Switching here was very crudely used for the advantage of Risk forecasting. With Markov Switching GARCH models their risk management accuracy was increased by a considerable percentage. The backtest with a two-core 1.4GZ processor 8GB(3200gz processing) RAM computer could have been better done by 8 core RIZEN 7 5900 RTX 3060 graphics.

8]Bada, Oualid&Kneip, Alois &Liebl, Dominik &Mensing, T. & Gualtieri, J. & Sickles, Robin. (2021). A wavelet method for panel models with jump discontinuities in the parameters. Journal of Econometrics. The basic idea of the estimation approach consists of using a structurally adapted Haar wavelet expansion of to control for its piecewise changing character. The aim is to estimate the regression function g from noisy observations z_t . with considerable time bracket representations from $T = 1, 2, \dots, T_n$. , helps the training of a larger data set by eliminating a considerable amount of overfitting in the model.

9)Henry, Kasereka, Role of Predictive Analytics in Business (April 19, 2021). Available at SSRN: The decision tree helps the author assess a chance that an event could occur. They used it in their scenario to visually and explicitly represent decision and decision making. As a result, the author observes that the customer's chance will be loyal is very high, which leads the tree to tell the authors that the customer will be loyal. Based on this decision, they get the big picture of customer loyalty.

FLOWCHART

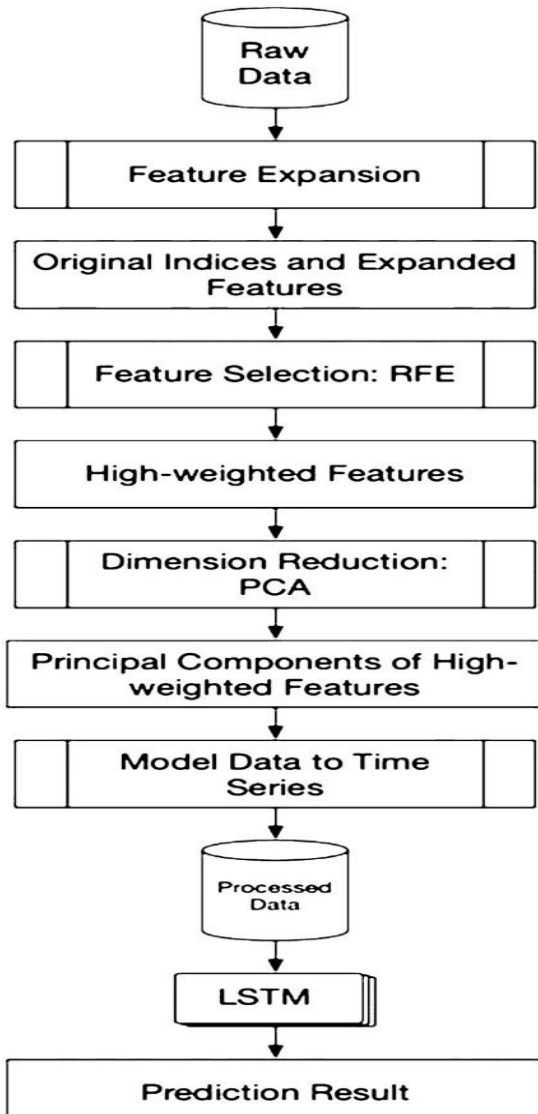


Fig1.1 This flowchart shows the sequential flow of the steps from data preprocessing to training and prediction

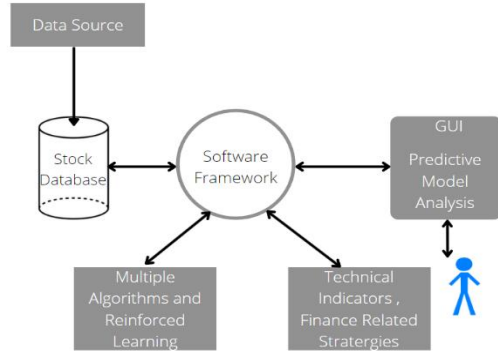


Fig1.2

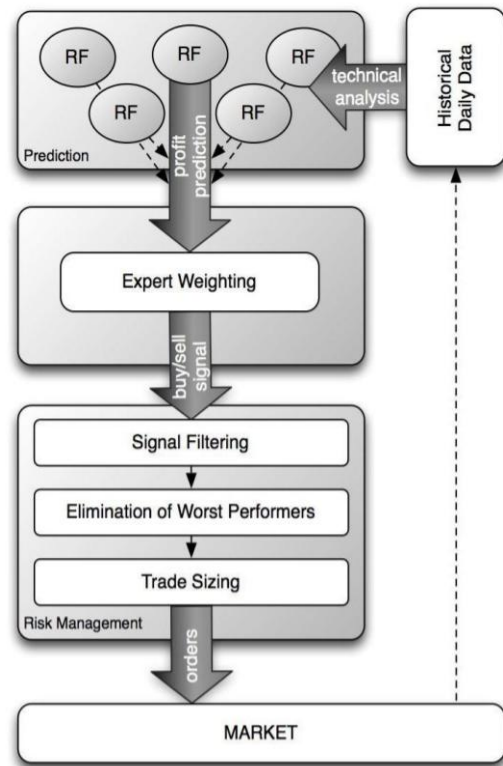


Fig 1.3 – Block Diagram

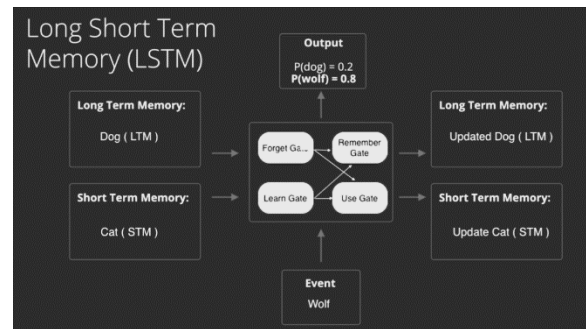


Fig 2.1 Architecture of LSTM

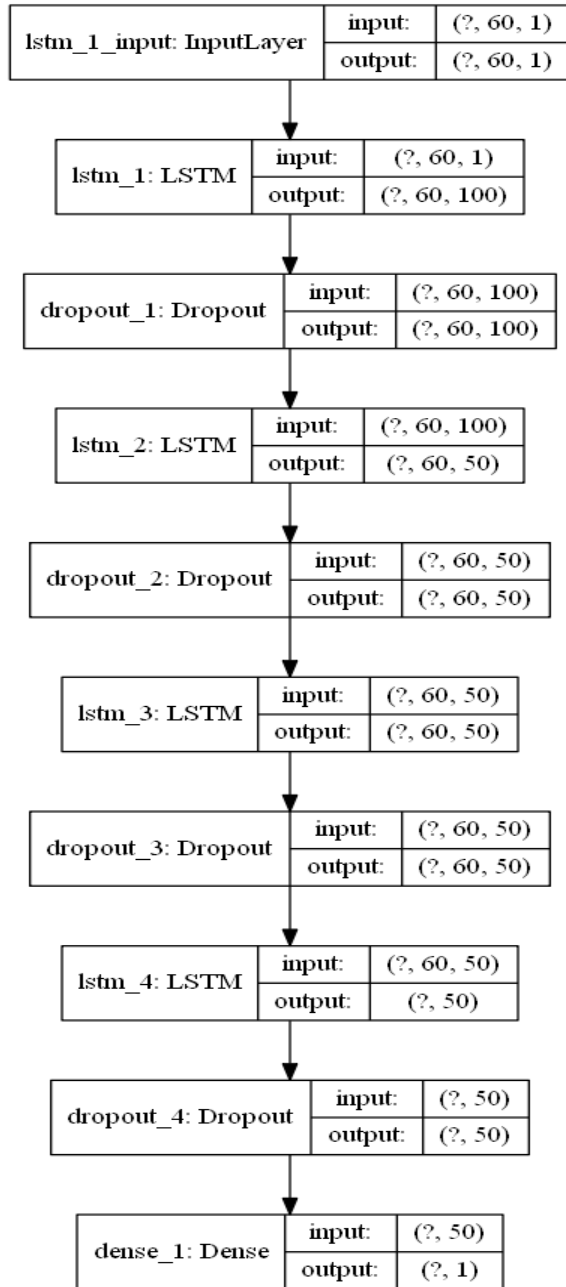


Fig 2.2 LSTM Model representation

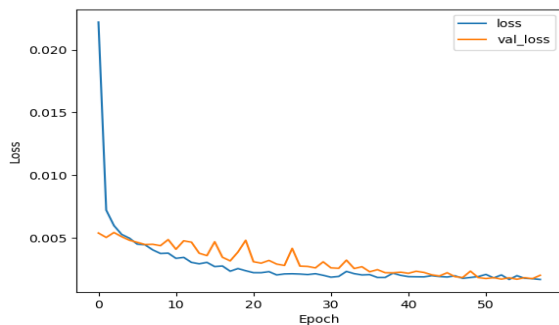


Fig 3.1 Validation loss

REFERENCES

- [1] Ash Booth, Enrico Gerding, and Frank MCGroarty. 2014. Automated trading with performance-weighted random forests and seasonality. *Expert Syst. Appl.* 41, 8 (June 2014), 3651-3661. DOI:
- [2] Younes Chihab, Zineb Bousbaa, Marouane Chihab, Omar Bencharef, Soumia Ziti, and Miin-Shen Yang. 2019. Algo-Trading Strategy for Intra-week Foreign Exchange Speculation Based on Random Forest and Probit Regression. *Appl. Comp. Intell. Soft Comput.* 2019 (2019). DOI:
- [3] SuryodayBasak, Saibal Kar, SnehanShuSaha, LuckysonKhaidem, Sudeepa Roy Dey, Predicting the direction of stock market prices using tree-based classifiers, *The North American Journal of Economics and Finance*, Volume 47, 2019, Pages 552- 567, ISSN 1062-9408,
- [4] Bruno Miranda Henrique, Vinicius Amorim Sobreiro, Herbert Kimura, Stock price prediction using support vector regression on daily and up to the minute prices, *The Journal of Finance and Data Science*, Volume 4, Issue 3, 2018, Pages 183- 201, ISSN 2405-9188,
- [5] Omer BeratSezer, Murat Ozbayoglu, Erdogan Dogdu, A Deep Neural-Network Based Stock Trading System Based on Evolutionary Optimized Technical Analysis Parameters, *Procedia Computer Science*, Volume 114, 2017, Pages 473-480, ISSN 1877-0509,
- [6] Manoj Thakur and Deepak Kumar. 2018. A hybrid financial trading support system using multi-category classifiers and random forest. *Appl. Soft Comput.* 67, C (June 2018), 337349. DOI:
- [7] M. Nabipour, P. Nayyeri, H. Jabani, S. S. and A. Mosavi, "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis," in *IEEE Access*, vol. 8, pp. 150199- 150212, 2020.
- [8] M. H. L. B. Abdullah and V. Ganapathy, "Neural network ensemble for financial trend prediction," 2000 *TENCON Proceedings. Intelligent Systems and Technologies for the New Millennium (Cat. No.00CH37119)*, 2000, pp. 157-161 vol.3, doi: 10.1109/TENCON.2000.892242.

- [9] Wen Long, Zhichen Lu, Lingxiao Cui, Deep learning-based feature engineering for stock price movement prediction, Knowledge-Based Systems, Volume 164, 2019, Pages 163-173, ISSN 0950-7051,
- [10] Cain Evans, Konstantinos Pappas, FatosXhafa, Utilizing artificial neural networks and genetic algorithms to build an algo-trading model for intra-day foreign exchange speculation, Mathematical and Computer Modelling, Volume 58, Issues 56, 2013, Pages 1249-1266, ISSN 0895-7177.
- [11] C. Brownlees, G. Gallo Financial econometric analysis at ultra-high frequency: data handling concerns Comput Stat Data Anal, 51 (4) (2006), pp. 2232-2245
- [12] M.A. Goldstein, P. Kumar, F.C. Graves Computerized and high-frequency trading Financ Rev, 49 (2) (2014), pp. 177-202
- [13] Algorithmic Trading Bot ICACC-2021 Medha Mathur, Satyam Mhadalekar, Sahil Mhatre, Vanita Mane Department of Computer Engineering, RamraoAdikInsitute of Technology, Nerul, Navi Mumbai
- [14] P Mondal, L. Shit, and S. Goswami, "Study of the effectiveness of time series modeling (ARMA) in forecasting stock prices," International Journal of Computer Science, Engineering and Applications, vol 4, pp. 13 - 29, 2014.
- [15] J. Sen and T. Datta Chaudhuri, "An alternative framework for time series decomposition and forecasting and its relevance for portfolio choice - a Comparative study of the Indian consumer durable and small-cap sector," Journal Economics Library, vol 3, no. 2, pp. 303 - 326, 2016.
- [16] Understanding LSTM Networks Posted on August 27, 2015, Colan's blog