Algorithmic trading for a buy-sell platform: study and comparison

Aishwarya Sakhare[1]*, Nihar* Mhaskar[2]*, Vicky* Mishra[3]*, and Madhuri* Chavan[4.]
[1]Department of Information Technology, RAIT, Nerul Email: aishwarya.sakhare56@gmail.com
[2]Department of Information Technology, RAIT, Nerul Email: niharmhaskar266@gmail.com
[3]Department of Information Technology, RAIT, Nerul Email: vickymishra1967@gmail.com
[4]Department of Information Technology, RAIT, Nerul Email: madhuri.chavan@rait.ac.in

Abstract. With the high-paced change in the world, the market trends are changing fast and so do the technological innovations. The market is constantly moving and being affected by a number of external factors, making it difficult for investors to make decision. To solve this problem, a technique called Algorithmic Trading is being implemented widely. It refers to use of computer algorithms to make trading decisions without human intervention. In this paper, the algorithm used to predict the stock prices is the LSTM model which is a methodology of neural networks.The existing models are less accurate as they do not take technical factors into consideration. LSTM model, having error of only 0.00036%, overcomes this limitation and also takes temporal factor into consideration.

Keywords: Algorithmic Trading, LSTM, Neural Network, Ridge Regression

1 Introduction

Algorithmic Trading, also known as automated trading, refers to use of computer algorithms to make trading decisions without human intervention. This computer algorithm uses a defined set of instructions to carry out a trade. These instructions are based on some mathematical model involving price of stock, timing, volume, etc. Not only does this make trading instant and accurate but also reduces the chances of mistake by ruling out human intervention. Algorithmic Trading is categorized in two types - High Frequency Trading (HFT) and Low Frequency Trading (LFT). High Frequency Trading is a type of algorithmic trading that is dependent on ultra-fast connections to monitor the market quotes and adjust orders accordingly. It is stated that HFT is done by firms located near data centres when large orders are exchanged. Low Frequency Trading is a slower form yet isfaster than trading floor specialists and other non- algorithmic techniques. LFT focuses on design theory to implement a strategy. The three factors to design an LFT are sentiment analysis, technical analysis and machine learning techniques. [1, 2]

Traditionally, traders calculate gain and loss based on market trends. Also, there are methods which depend on statistical considerations. The Spread or Pair trading method uses two financial measures that are co-related. Another method called Multi-Spread Trading is based on quantitative algorithms and does not depend on market trends. [3] One more methodology widely used is the Machine Learning model. It refers to a set of algorithms and rules that apply mathematical computations to learn from big sets of data and improve its efficiency. [4] This is based on various trend indicators such as Exponential Moving Average (EMA), Average Directional Movement Index (ADX), Relative Strength Index (RSI), Average True Range (ATR), Moving Average Convergence Divergence (MACD), Moving Average Convergence Divergence (MACD Signal). [5]

To predict the future stock prices with maximum accuracy, Long Short Term Memory (LSTM) model is implemented. This is a technique that uses Recurrent Neural Network (RNN).Artificial Neural Networks are information processing models that are inspired by the

human brain. The neuron collects signals from input channels named dendrites, processes information in its nucleus, and then generates an output in a long thin branch called axon.[6] A RNN is a type of ANN that is designed to take temporal dimension into consideration by having a memory (internal state) (feedback loop). RNN contains a temporal loop in which the hidden layer not only gives an output but it feeds itself as well. An extra dimension is added which is time. [7]

2 Methodology

The work presented has taken into consideration the technical indicators that affect the Stock Market. Examples of such technical indicators include Exponential Moving Average(EMA), Moving Average Convergence Divergence (MACD), and Slow Moving Average (SMA), etc.[8]. The work demonstrates the use of the LSTM model considering the technical indicator, SMA. As shown in Fig.1, the results obtained by this prediction model are used by the trading algorithm to decide whether to sell the stock or buys it, based on a threshold value.

Technical indicators are mathematical calculations carried out on history of stock price. They are generally used as visual aidsto help identify the direction the market is going to change in. These technical indicators are also fed to the model through secondary input branches. Since, this model allows to provide more than one input, the data of technical indicators can be provided as well. Hence, the model works better in terms of consideration of sudden changes in the market.

Recurrent neural networks are good at working with time series data. This also applies to LSTM due to their ability to remember the past. This is done by using recurrent nodes to store the old states. RNNs and LSTMs can work on current information as well as information the network had seen in previous steps. Also, implementation of a LSTM cell is handy as they're a default layer in TensorFlow's Keras.

This work has used a basic technical indicator called Slow Moving Average (SMA). To calculate the simple moving average of a stock, the average of the closing price of the stock of previous n days is taken. The processed data containing the calculated SMA goes through the data transformation as required for the model of LSTM. The data is scaled to fit within the values 0 to 1.

SMA is not a time-series piece of data, hence, it is not passed through the LSTM. Instead it is fed to the model at the last step,in before the final prediction is made. It is given as input it in to the penultimate 64-node dense layer.So, we will need a model with two inputs, one concatenation layer and an output. As we had divided the data into training and test dataset, similarly the technical

indicator's data is split. The inputs in this model are the original training data set and the training dataset of technical indicator's data. The output of these two models is then given as concatenated input to the third model which gives the final output.
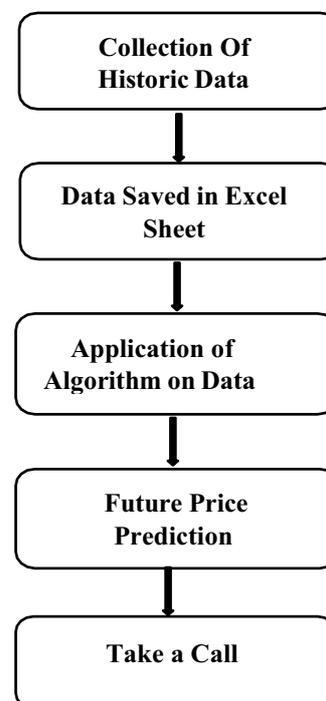

Fig. 1. Formulation and Working

These predictions are then used to perform asset allocation. This refers to process of dividing a given capital in a set of assets such that risk is reduced and profit is increased. For this, the CAPM formula is used. Finally, a trading algorithm named Bollinger bands is used to depict when the system will buy or sell the stock.

Bollinger bands are an indicator which uses slow moving average as the middle band and two other bands named Upper band and Lower band. Lower band is calculated as difference of mean and two times the standard deviation. Similarly, Upper band is calculated as mean summed with two times the standard deviation. [9]

This strategy is used to identify buying and selling position based on the position of stock price as compared to the bands. When the position of stock price is less than the lower band the algorithm will long the stock, while it will short the stock when the position of stock price is greater than the upper band. [10]

3 Results

The methodology used in this work helps to predict the close price of different stocks on n^{th} day (where n is pre-decided value).

On evaluating the model, the evaluation results show an error of0.00036655570147559047 which will magnify on upscaling the predicted data.

The model is evaluated for various stocks. The graph shown inFig.2 depicts the result for close price of Apple.
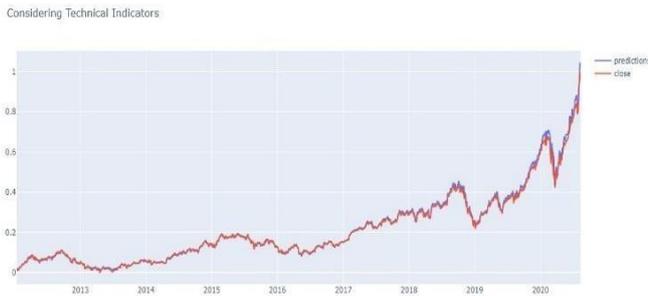

Fig. 2. Prediction using LSTM including technical indicators.

The prediction of close price of Apple stock and comparison with original price. This prediction model uses technical indicators and works well with training data and test data. It also manages to avoid the shift problem that occurs in existing prediction methods.
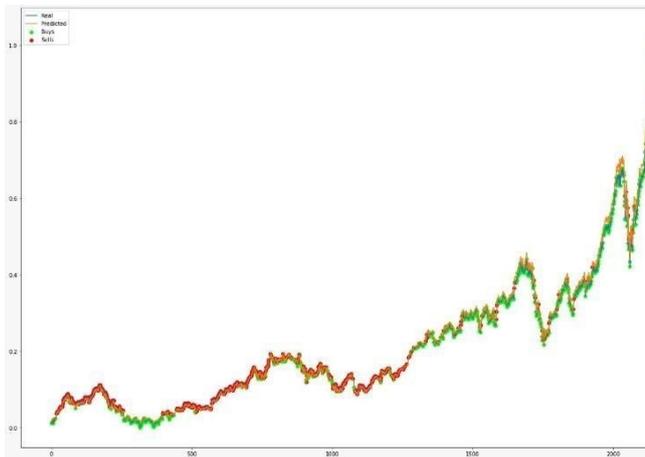

Fig. 3. Buying and Selling points

The green points, as shown in Fig.3 are plotted when the application buys the particular stock and red when it sells the stock. This is done using an algorithm for trading based on a threshold value.
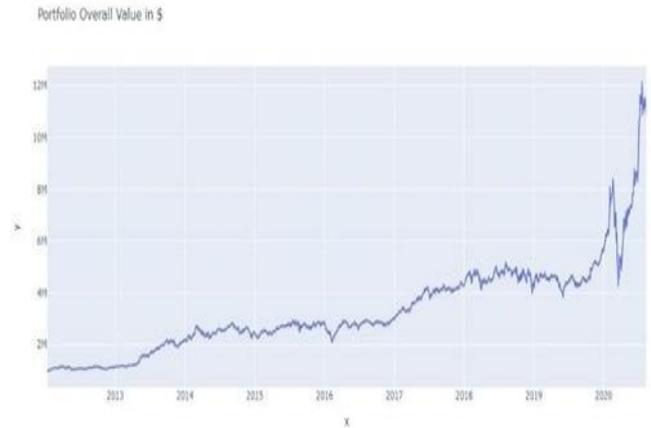

Fig. 4. Portfolio Overall value

The graph shown in Fig.4 depicts how asset allocation formula used performs in 8 years. It shows the net value of portfolio at the end of time span as compared to the beginning. This basically helps users to understand how much return is earned on the given portfolio.

Table 1. Comparison of Algorithms

1.	Algorithm	Ridge Regression	LSTM Time Series Model	LSTM using Technical Indicators
2.	Performance Parameters	'Close' value of stock	'Close' value of stock	Slow Moving Average'Close' value of stock
3.	Score	(Accuracy) 0.7950028030821767	(Error) 0.000197624322026968	(Error) 0.00036655570147559047
4.	Advantages	This model reduces the problem of overfitting.	The prediction results obtained using the LSTM model are accurate to a large extent and are better than the earlier model.	This model proves better than previous model as it avoids the right shift as seen earlier.
5.	Disadvantages	The model becomes less sensitive to the variations of the independent variable.	The model faces the problem of being shifted to right by a constant amount.	The drawback of this model is it is not too much efficient in capturing the sudden changes. [11]

3.1 Comparison with Ridge Regression

In this work, the first model used for predicting was the Ridge Regression Model which reduced the problem of overfitting. In ridge regression, 'Least sum of squares' is applied to obtain the best fit line. Ridge regression works by applying a penalizing term (reducing the weights and biases) to overcome overfitting. It works by attempting at increasing the bias to improve variance (generalization capability). This works by changing the slope of the line, the model performance might be little poor on the training set but it will perform consistently well on both the training and testing datasets. Alpha is part of the penalty term in the equation of ridge regression. As Alpha increases, the slope of the regression line is reduced and becomes more horizontal. The model becomes less sensitive to the variations of the independent variable. [12]

On implementing this model, the score defined by the model is as follows-

Ridge Regression Score: 0.7950028030821767

3.2 Comparison with LSTM

This model uses cells, called as long short term memory cells. These are small models that are structured to store memory in large neural networks. Inside the LSTM cell, a recurrent node is
present that allows this function. The recurrent node has an edge with a weight equal to one. The edge is looping backon itself to ensure that at every feedforward iteration the cell retains the information from all the previous steps.

Feedforward Neural Networks (vanilla networks) map a fixed size input (such as image) to a constant size output (classes or probabilities). A drawback in Feedforward networks is that they do not have any time dependency ormemory effect. RNN can recall what happened in the previous time stamp so it works great with sequence of text. LSTM networks work much better compared to vanilla RNN since they overcome the vanishing gradient problem. [13, 14]

On evaluating the model, the result shows an error value of 0.000197624322026968 on normalized values. On upscaling the values, the error is magnified as well.

4 Conclusion

Share market is a mode of investment for buyers and sellers across the globe. A software which predicts the rise and fall of stock prices can help investors plan their investment wisely. Hence, to fulfil the above requirement, it is important to predict the future trend of a stock using the large pre-existing database. To do this successfully, various factors that determines the state of market need to be understood. To automate this process, algorithmic trading is implemented, which is basically an execution process based on a written algorithm.

As shown in Table 1, this work depicts a comparison of three methods, starting with the traditional Machine Learning Model that uses Ridge Regression. Further to increase the accuracy, time series model called Long Short Term Memory is used. This performs better as the technique is a part of Recurrent Neural Network that imitates the working of brain. Even though the error given by this method is very less, the model fails to capture real-life events. Hence, to make it efficient in terms of predicting future trends, the LSTM Model is fed more data using the technical indicators.

References

[1] F. Obeid Algorithmic Trading: High Frequency & Low Frequency Trading, FERM 6303: Financial Assets and Markets, (2018).

[2] E. Boehmer, K. Fong, J. Wu, International evidence on Algo. Trading, Journal of financial and quantitative analysis,(2015).

[3] M. Bogliardi, F. S. Canepa, G. Frisina , Multiple Spread Trading 60, An investment methodology neutral to financial market trends English Version, (2019).

[4] P. Rys, R. Slepaczuk, ML in algo. trading strategy opt. implementation and efficiency, Faculty of Economic Sciences, University of Warsaw, (2018).

[5] E. Sorhun, How Is a Machine Learning Algorithm Now-Casting Stock Returns? A Test for ASELSAN, Springer Science and Business Media LLC, (2019).

[6] N. Budhani, Dr. C. K. Jha, S. K. Budhani—Prediction of Stock Market Using Artificial Neural Network, International Conference of Soft Computing Techniques for Engineering and Technology (ICSCTET), (2014).

[7] M. Adya and F. Collopy, How effective are neural networks at fore-casting and prediction, A review and evaluation, J. Forecasting, vol. 17, (1998).

[8] M. R. Vargas, C. E. M. dos Anjos, G. L. G. Bichara and A. G. Evsukoff, "Deep Leaming for Stock Market Prediction Using Technical Indicators and Financial News Articles," International Joint Conference on Neural Networks (IJCNN), (2018).

[9] Lo, Andrew W., H. Mamaysky and J. Wang. "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, And Empirical Implementation," Journal of Finance, (2000).

[10] F. Allen, R. Karjalainen, using genetic algorithms to find technical trading rules, Journal of financial economics, vol. 51, issue 2, (1999).

[11] K. Chen, Y. Zhou and F. Dai —A LSTM-based method for stock returns prediction: A case study of China stock market, IEEE International Conference on Big Data (Big Data), (2015).

[12] H. Duzan, NSBM. Shariff, Ridge regression for solving the multicollinearity problem: review of methods and models, Journal of Applied Sciences, 15, (2015).

[13] T. Gao, Y. Chai and Y. Liu, "Applying long short- term memory neural networks for predicting stock closing price," 8th IEEE International conference on software engineering and service science(ICSESS), (2017).

[14] D. Wei, "Prediction of Stock Price Based on LSTM Neural Network," International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), (2019).