A Novel Long Short Term Memory Algorithm Compared to Random Forest Algorithm for Improved Accuracy Values in Real-Time Stock Market Prediction.

Kodam .Naveen[1], S.S Arumugam[2*]

[1]Research Scholar, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, pincode: 602105.

[2*]Project Guide, Corresponding Author, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, pincode: 602105.

ABSTRACT

Aim: This work is a comparative study of novel long short term memory algorithm and random forest algorithm for optimizing stock market prediction to improve the accuracy of real time stock exchange. **Materials and Methods:** Novel Long time short term memory algorithm (N= 10) and random forest algorithm (N=10) methods are simulated by varying the LSTM parameter and random forest parameter to optimize the pH. Sample size is calculated using Gpower 80% for two groups and there are 20 samples used in this work. **Results and Discussion:** Based on obtained results LSTM has significantly better accuracy (66.80%) compared to Random forest accuracy (62.97%). Statistical significance difference between long short term memory and random forest and it was found to be 0.049($p<0.05$). **Conclusion:** Long short term memory algorithm produces better results in predicting stock market price to improve accuracy percentage than random forest algorithm.

Keywords: Machine Learning, Stock price prediction, Novel Long Short Term Memory Algorithm, Random Forest Algorithm, Supervised learning.

INTRODUCTION

In this research work, the Machine learning model had a great impact in stock price prediction, which makes predictions based on the values of current stock markets by training on their previous values(Sable, Goel, and Chatterjee 2019). A stock price prediction is a trading platform where different investors sell and buy the shares consistently with stock availability(Vui et al. 2013). Prediction of stock exchange trends is taken into account as a crucial task and is of great attention as predicting stock prices with supervised learning in success might result in engaging profits by making correct choices(Verma and Mohapatra, n.d.). Stock market predictions are frequently used in business applications(Kothari 2010). It has continually been a warm spot for buyers and investing groups to comprehend the alternate regularity of the inventory market price and expect its trend(Ding and Qin 2020).In the last 5 years, more than 60 papers have been published on IEEE xplore and google scholar on Stock market predictions which can be greatly invested in stocks for big companies.

A comparison analysis that predicts the investment can result in success (Nabipour et al. 2020). This study (Bosco and Khan

2018) uses an experimental methodology to analyze the high-performance efficiency of the Random forest and LSTM algorithms. The accurate control of novel last short term memory (LSTM) utilizing traditional controllers, such as stock prediction controllers and Stock Price Controller, is compared in this article (Ghosh et al., n.d). (SPC). In order to increase efficiency, a novel LSTM method utilizing the random forest algorithm and stock rate management has been presented in this study (Premkumar and Manikandan 2013). (Nabipour et al. 2020).(Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020)

Previously our team has a rich experience in working on various research projects across multiple disciplines(Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020).The accuracy of the stock market prediction using the support vector machine algorithm was not fully taken into account in a prior study. An innovative long short term memory approach is used to improve the log loss rate of stock market forecasts in order to get around this problem.

MATERIALS AND METHODS

The research work is carried out in the Data Analytics laboratory lab at Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai. The sample size has been calculated using the GPower software by comparing both of the controllers in Supervised learning. Two numbers of groups are selected for comparing the process and their result. In each group, 10 sets of samples and 20 samples in total are selected for this work. The pre-test power value is calculated using GPower 3.1 software (g power setting parameters: statistical test difference between two independent means, α=0.05, power=0.80, Two algorithms (LSTM and random forest algorithm) are implemented using Technical Analysis software. In this work, no human and animal samples were used so no ethical approval is required(Ghosh et al., n.d.).

Long Short Term Memory Algorithm

Equation 2 describes the long short term memory algorithm (LSTM), an artificial neural network (RNN) architecture used in deep learning. Since there may be unknown lags between significant occurrences in a time series as shown in equation 1, LSTM networks are well suited to categorizing, processing, and making predictions based on time series data. A cell, an input gate, an output gate, and a forget gate make up a typical LSTM unit. Three gates control the flow of information into and out of the cell, and the cell remembers the values across arbitrary time intervals.

$$Xnorm = X-min(X) / max(X) - min(X) \quad (1)$$

$$\sum (y_i - y_i)^2 \quad (2)$$

i=1

Where

y_i, is the predicted value

y_i,is the actual value

i , is the time instance

N, is the number of data samples

Random Forest Algorithm

Random forest is a supervised machine learning algorithm. Random forest is used for both regression and classification which is given in equation 3. It is mainly used for classification problems. Random forest calculation makes choice trees on information tests and afterward gets expectation from every one of them and chooses the best arrangement through casting a ballot. It is an outfit technique which is superior to a solitary choice tree since it decreases the over-fitting of the outcome. Random Forest algorithm pseudo code.

$$MSE = 1/N \Sigma(fi - yi)^2 \quad (3)$$

MSE is the Mean squared error N is the number of data points
Fi is the value return by the model Yi is the actual value for data point i

Accuracy for Long Short Term Memory and Random Forest algorithms was calculated based on the equation:

$$Accuracy=TP + TN / TP +TN + FP + FN \quad (4)$$

Where TP, is the model's classification of the number of true positives, TN, is the number of true negatives, FP, is the number of false positives, FN, and is the model's classification of the number of false negatives. An 8GB of RAM and an Intel Core i5 processor made up the hardware setup. The system configuration used a 256GB SSD, a 64bit OS, and an X64-based processor. Windows 10 was used as the operating system, and Google Colab with Python was used as the implementation tool.

Statistical analysis

For statistical study of approaches based on the LSTM and Random Forest algorithms, SPSS software is employed. Efficiency is the dependent variable, whereas LSTM accuracy is the independent variable. For both methods, the accuracy of the LSTM is calculated using independent T test analyses.

RESULTS

Table 1 shows the simulation result of proposed algorithm long short term memory and the existing system random forest were run at different times in the google colab with a sample size of 10. From table 1, it was observed that the mean accuracy of the LSTM algorithm was 66.80% and the Random forest algorithm was 62.97%.

Table 2 represents the T-test comparison of both LSTM algorithm and random forest algorithm. The Mean, Standard Deviation and Standard Error Mean were calculated by taking an independent variable T test among the study groups. The LSTM algorithm produces a significant difference than the random forest algorithm with a value of 0.712 and effect size=1.414.

Table 3 represents the Mean of LSTM algorithm which is better compared with random forest algorithm with a standard deviation of 0.38028 and 0.46643 respectively. From the results,

LSTM algorithm (66.80%) gives better accuracy than the random forest algorithm (62.97%). Figure 1 gives the comparison chart of LSTM of random forest algorithms in terms of mean and accuracy. The mean accuracy of the LSTM algorithm is better than Random forest. Figure 2 shows the error difference of LSTM algorithm (0.12) and random forest algorithm (0.14).

DISCUSSIONS

LSTM and random forest algorithms are implemented and compared for stock market prediction to improve the accuracy by stock pricing. From obtained results it is concluded that the random forest algorithm provides better accuracy results compared to the LSTM algorithm.

In the recent survey, the proposed(Kalbandi et al. 2021)showed that the LSTM algorithm is a promising option for Stock market prediction with root mean square value 0.04.(Ghosh et al., n.d.) proposed a LSTM based model for different companies belonging to the banking sector based on historical data and observed that the error level comes down drastically with the data for longer periods.(Saravagi, Saravagi, and Agrawal 2021) proposed LSTM algorithm for predicting stock prices of selected companies by comparing the daily stock price movement in various sectors.(Raza 2017) implemented six machine learning techniques i.e., ANN, MLP, RBF, SVM, Decision Tree and Naive Bayes and by comparing them concluded that MLP works better with accuracy 77%. Major research contribution supports Implementation and comparative analysis of random forest algorithms to optimize stock gain of LSTM drive with reduced efficiency improvement. Even though few articles listed the disadvantages of proposed random forest algorithms(Maiti and Pushparaj 2020). Further, the random forest algorithm is not suitable for improving accuracy of stock market prediction(Gupta et al., n.d.).

From the above discussion, only a few articles ensure that they provide better performance than the proposed LSTM and random forest algorithm for improving accuracy of stock market prediction. Also, the present price prediction requires no additional cost and therefore received intense attention in recent years. So, we can infer that the proposed LSTM and random forest algorithm can be used to improve the accuracy of price prediction by regulating the stock gain.

Stock market prediction has limited price prediction ability based on future price significant profit which makes better price prediction in future. Deep Learning algorithm can address future stock prediction.

CONCLUSION

The work invovels long short term memory algorithm to find the stock market prediction to be proved with better accuracy of 66.80% when compared to Random Forest accuracy is 62.97% for predicting Stock price.

DECLARATIONS

Conflict of Interests

No conflict of Interest in this manuscript.

Authors Contributions

Author KN was involved in data collection, data analysis, manuscript writing. Author SSA was involved in conceptualization, data validation, and critical review of manuscript**.**

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Soft square pvt.Ltd, chennai
2.Saveetha University
2. Saveetha Institute of Medical And Technical Sciences
3. Saveetha School of Engineering

REFERENCES

1. Benin, S. R., S. Kannan, Renjin J. Bright, and A. Jacob Moses. 2020. "A Review on Mechanical Characterization of Polymer Matrix Composites & Its Effects Reinforced with Various Natural Fibres." *Materials Today: Proceedings* 33 (January): 798–805.
2. Ding, Guangyu, and Liangxi Qin. 2020. "Study on the Prediction of Stock Price Based on the Associated Network Model of LSTM." *International Journal of Machine Learning and Cybernetics*. https://doi.org/10.1007/s13042-019-01041-1.
3. Ghosh, Achyut, Soumik Bose, Giridhar Maji, Narayan Debnath, and Soumya Sen. n.d. "Stock Price Prediction Using LSTM on Indian Share Market." https://doi.org/10.29007/qgcz.
4. Gudipaneni, Ravi Kumar, Mohammad Khursheed Alam, Santosh R. Patil, and Mohmed Isaqali Karobari. 2020. "Measurement of the Maximum Occlusal Bite Force and Its Relation to the Caries Spectrum of First Permanent Molars in Early Permanent Dentition." *The Journal of Clinical Pediatric Dentistry* 44 (6): 423–28.
5. Gupta, Archana, Pranay Bhatia, Kashyap Dave, and Pritesh Jain. n.d. "Stock Market Prediction Using Data Mining Techniques." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3370789.
6. Kalbandi, Ishwarappa, Ashutosh Jare, Om Kale, Himanshu Borole, and Swapnil Navsare. 2021. "Stock Market Prediction Using LSTM." *International Journal of Advanced Research in Science, Communication and Technology*. https://doi.org/10.48175/ijarsct-877.
7. Kothari, Rajesh. 2010. *Financial Services in India: Concept and Application*. SAGE Publications India.
8. Maiti, Ayan, and Shetty D. Pushparaj. 2020. "Indian Stock Market Prediction Using Deep Learning." *2020 IEEE REGION 10 CONFERENCE (TENCON)*. https://doi.org/10.1109/tencon50793.2020.9293712.
9. Nalini, Devarajan, Jayaraman Selvaraj, and Ganesan Senthil Kumar. 2020. "Herbal Nutraceuticals: Safe and Potent Therapeutics to Battle Tumor Hypoxia." *Journal of Cancer Research and Clinical Oncology* 146 (1): 1–18.
10. Raza, Kamran. 2017. "Prediction of Stock Market Performance by Using Machine Learning Techniques." *2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT)*. https://doi.org/10.1109/icieect.2017.7916583.
11. Reddy, Poornima, Jogikalmat Krithikadatta, Valarmathi Srinivasan, Sandhya Raghu, and Natanasabapathy Velumurugan. 2020. "Dental Caries Profile and Associated Risk Factors Among Adolescent School Children in an Urban South-Indian City." *Oral Health & Preventive Dentistry* 18 (1): 379–86.
12. Sable, Rachna, Shivani Goel, and Pradeep Chatterjee. 2019. "Empirical Study on Stock Market Prediction

Using Machine Learning." *2019 International Conference on Advances in Computing, Communication and Control (ICAC3)*. https://doi.org/10.1109/icac347590.2019.9036786.

13. Saravagi, Manisha, Deepika Saravagi, and Shweta Agrawal. 2021. "Indian Stock Market Analysis and Prediction Using LSTM Model during COVID-19." *International Journal of Engineering Systems Modelling and Simulation*. https://doi.org/10.1504/ijesms.2021.10036743.
14. Sathish, T., and S. Karthick. 2020. "Gravity Die Casting Based Analysis of Aluminum Alloy with AC4B Nano-Composite." *Materials Today: Proceedings* 33 (January): 2555–58.
15. Sathish, T., D. Bala Subramanian, R. Saravanan, and V. Dhinakaran. 2020. "Experimental Investigation of Temperature Variation on Flat Plate Collector by Using Silicon Carbide as a Nanofluid." In *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS ENGINEERING: ICRTMME 2019*. AIP Publishing. https://doi.org/10.1063/5.0024965.
16. Sivasamy, Ramesh, Potu Venugopal, and Rodrigo Espinoza-González. 2020. "Structure, Electronic Structure, Optical and Magnetic Studies of Double Perovskite Gd2MnFeO6 Nanoparticles: First Principle and Experimental Studies." *Materials Today Communications* 25 (December): 101603.
17. Venu, Harish, and Prabhu Appavu. 2021. "Experimental Studies on the Influence of Zirconium Nanoparticle on Biodiesel–diesel Fuel Blend in CI Engine." *International Journal of Ambient Energy* 42 (14): 1588–94.
18. Verma, Nitish, and Baibaswata Mohapatra. n.d. "Stock Market Predication Using Machine Learning." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3645875.
19. Vui, Chang Sim, Gan Kim Soon, Chin Kim On, Rayner Alfred, and Patricia Anthony. 2013. "A Review of Stock Market Prediction with Artificial Neural Network (ANN)." In *2013 IEEE International Conference on Control System, Computing and Engineering*. IEEE. https://doi.org/10.1109/iccsce.2013.6720012.

TABLES AND FIGURES

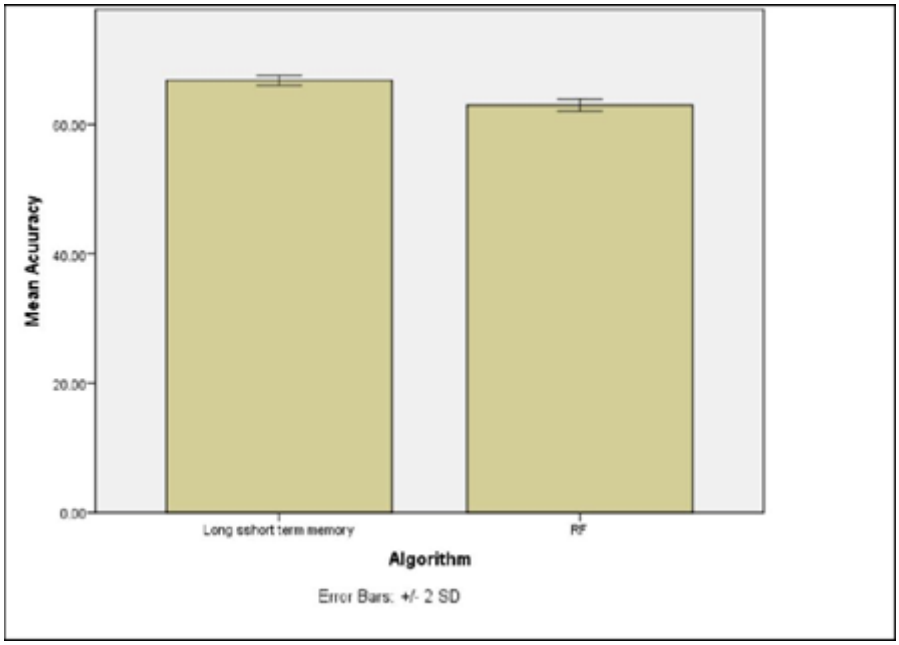

Fig. 1. Comparison of the mean and accuracy of the Long Short Term Memory Algorithm versus the Support Vector Machine Algorithm. The Long Short Term Memory Algorithm has a higher mean accuracy than the Support Vector Machine. The support vector machine algorithm is shown on the X-axis, while the long short term memory algorithm is shown on the Y-axis.

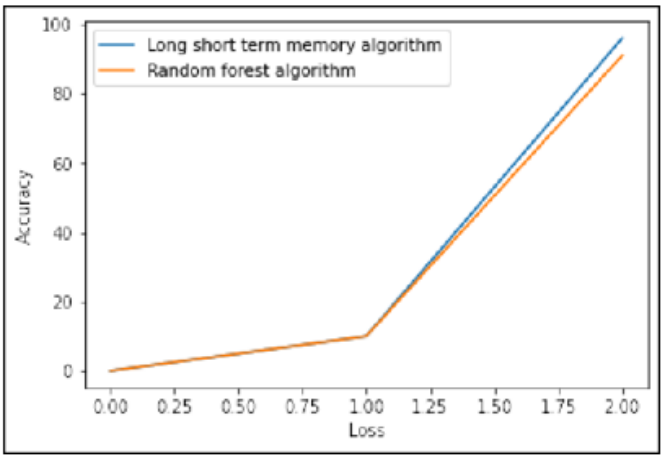

Fig. 2. Comparison of Long short term memory algorithm accuracy of 68.11% and support vector machine with accuracy of 52.46%

Table 1. Predicted Accuracy of stock market prediction

SI.No	Sample_size	Long short term memory algorithm Accuracy in percentage	Random forest algorithm Accuracy in percentage
1	100	66.74	62.98
2	200	66.89	62.45
3	300	66.15	63.01
4	400	67.01	63.16
5	500	67.23	61.96
6	600	67.41	63.45
7	700	66.49	63.46
8	800	66.43	62.82
9	900	66.98	63.26
10	1000	66.76	63.20

Table 2. Group statistics (mean of Long short term memory algorithm is 68.11% more than Support vector machine algorithm 52.46% and Error mean for Long short term memory algorithm is 0.15 and for Support vector machine algorithm is 0.32)

	Algorithm	N	Mean	Std. deviation	Std. Error Mean
Accuracy	Long short term memory algorithm	10	66.8090	0.38028	0.12025
	SVM	10	62.9750	0.46643	0.14750

Table 3. Independent Sample T-test: (Long short term memory algorithm is significantly better than Support vector machine algorithm with P=0.023)

		Levene's Test for Equality of Variances		T-test for Equality of means						
									95% Confidence Interval of the difference.	
		F	Sig.	t	df	Sig.(2-tailed)	Mean Difference	Std. Error Differences	Lower	upper

Accuracy	Equal Variances assumed.	0.141	0.049	20.146	18	0.000	3.83400	0.19031	3.43418	4.23382
	Equal Variances not assumed.			20.146	17.298	0.000	3.83400	0.19031	3.43301	4.23382