IMPROVED ACCURACY FOR STOCK PREDICTION USING LSTM MODEL COMPARED WITH ARIMA

R.Rohith Kumar[1], Gayathri[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: The objective of the work is to predict the Stock Price Prediction Using LSTM Model Compared with ARIMA. To achieve accuracy a novel SVClassifier is used.
Method and Materials: Accuracy and loss are performed with a DATA dataset from the keras library. The total sample size is 20. The two groups Convolutional linear regression (N=10) and Support Vector Machine algorithms (N=10).
Result: The result proved that Support Vector Machine (SVM) with better accuracy of 97% than linear regression accuracy of 96% and p=0.14 (p<0.05). It is statistically insignificant with a pretest power of 80%. The two algorithms LSTM and ARIMA are statistically satisfied with the independent sample T-Test value (p<0.001) with confidence level of 95%,
Conclusion: prediction of stock price significantly seems to be better in LSTM.

Keywords: Stock market, Machine learning, ARIMA, LSTM, Novel Support Vector Machine, Revising statistical analysis.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, 602105
[2*]Project guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu,602105

1. Introduction

The securities exchange is an unpredictable spot loaded up with vulnerability (Ma 2020). It is additionally where amazing measures of cash change hands each day, in the expectations that the exchanges caused will to create benefits for financial backers (Ho, Darman, and Musa 2021). In case it was feasible to explore this unpredictability and precisely gauge the developments of the market it would set out a freedom to get extraordinary measures of abundance for individuals who can make these projections (Hua 2020).

Albeit investment opportunities – which in some sense structure the premise of quantitative money – have been available since the seventeenth century, it was not until the twentieth century that the field truly took a goliath jump forward. In spite of the fact that there had been some work done by mathematicians in the last part of the 1800s on the properties of monetary business sectors, they had not acquired any importance until the centre of the following century when increasingly more exploration on the point was performed (Jia et al. 2019). The huge insurgency, be that as it may, happened in 1973 when Black and Scholes distributed their paper on the estimating of alternatives, which thus caused an expanding influence on the premium for subordinates exchanging and made the market in its present structure (Sunny, Maswood, and Alharbi 2020). Although this proposal isn't on the subject of alternatives exchanging, it gives some setting on the development of monetary designing and quantitative examination (Yan 2021).

The reason for this postulation is to check out whether the Long Short-Term Memory (LSTM) neural organisation can all the more precisely figure the developments of the securities exchange than a more old style technique (Ho, Darman, and Musa 2021), the Autoregressive incorporated moving normal (ARIMA), which have generally been utilised to attempt to conjecture developments in time series information (Roy, Ghosh, and Senapati 2021). Upon starting the audit it appears to be that a LSTM probably would beat an exemplary model given the measure of boundaries that are thought about (Wei 2019). This be that as it may, isn't authoritative and should be tried and applied on genuine information, explicitly on the Swedish securities exchange which is fundamentally less contemplated and with unexpected attributes in comparison to the US markets (Liu, Wang, and Zheng 2019).Our team has extensive knowledge and research experience that has translated into high quality publications(Pandiyan et al. 2022; Yaashikaa, Devi, and Kumar 2022; Venu et al. 2022; Kumar et al. 2022; Nagaraju et al. 2022; Karpagam et al. 2022; Baraneedharan et al. 2022; Whangchai et al. 2022; Nagarajan et al. 2022; Deena et al. 2022)

2. Materials and Methods

The study setting of the proposed work is done in Saveetha School of Engineering. Two groups were identified for the study setting (Gayathri and Nandhini 2011; B, Dheeraj, and Gayathri 2022) group one LSTM and group two ARIMA. Using G power 10 sample sizes and totally 20 sample sizes have (Revathi et al. 2021) been carried out for our study, 95% confidence and pretest power 80%.

The dataset named 'DATA' is downloaded from the public domain keras library. In our experiments here we used the data.csv dataset. Detailed descriptions of the features/attributes in the dataset can be found below in the form of a Table 1. The dataset consists of 5 lakhs instances. Dataset (Revathi et al. 2021) has two columns: url and label. The dataset was splitted into two parts namely the training part and testing part. 70% of the data was used for training and the remaining 30% was used for testing. The algorithm was implemented by evaluating the train and test. Input dataset collected from the link (tesla.csv).

Lstm Algorithm

Rolling LSTM Inputs: Time series Outputs: RMSE of the forecasted data
Split data into:
70\% training and 30\% testing data
1. size ← length(series) * 0.70
2. train ← series[0...size]
3. test ← series[size...length(size)] # Set the random seed to a fixed value
4. set random.seed(7)
Fit an LSTM model to training data Procedure fit_lstm(train, epoch, neurons) 5.
X ← train
6. y ← train - X
7. model = Sequential()
8. model.add(LSTM(neurons), stateful=True))
9. model.compile(loss='mean_squared_error', optimizer='adam')
10.for each i in range(epoch) do
11. model.fit(X, y, epochs=1, shuffle=False)
12. model.reset_states()
13.end for return model
Make a one-step forecast Procedure forecast_lstm(model, X)
14. yhat ← model.predict(X) return that
15. epoch ← 1
16. neurons ← 4
17. predictions ← empty # Fit the lstm model
18. lstm_model = fit_lstm(train,epoch,neurons) # Forecast the training dataset
19. lstm_model.predict(train) # Walk-forward validation on the test data
20. for each i in range(length(test)) do

21. # make one-step forecast
22. X ← test[i]
23. yhat ← forecast_lstm(lstm_model, X)
24. # record forecast
25. predictions.append(yhat)
26. expected ← test[i]
27. end for
28. MSE ← mean_squared_error(expected, predictions)
29. Return (RMSE ← sqrt(MSE))

Arima Algorithm

Rolling ARIMA Inputs: series Outputs: RMSE of the forecasted data
Split data into:
70% training and 30% testing data
1. size ← length(series) * 0.70
2. train ← series[0...size]
3. test ← series[size...length(size)]
Data structure preparation
4. history ← train
5. predictions ← empty
Forecast 6. for each t in range(length(test)) do
7. model ← ARIMA(history, order=(5, 1, 0))
8. model fit ← model.fit()
9. hat ← model_fit.forecast()
10. predictions.append(hat)
11. observed ← test[t]
12. history.append(observed)
13. end for
14. MSE = mean_squared_error(test, predictions)
15. RMSE = sqrt(MSE) 16. Return RMSE

Statistical Analysis

For statistical implementation, the software to be used here is IBM SPSS V26.0. Statistical package for social sciences is used for calculating the statistical (Gayathri and Nandhini 2011) calculations such as mean, standard deviation, and also to plot the graphs etc.,. The independent variables are Url, Label and the dependent variable is ‘accuracy’. In SPSS, the dataset is prepared using 10 as sample size for each group and accuracy is given as the testing variable.and the dependent variable is tesla stock samples. An independent T-Test analysis was performed

3. Results

LSTM and ARIMA compared both algorithms with their accuracy rate. For both proposed and existing algorithms 10 iterations were taken for each iteration the predicted accuracy was noted for analysing accuracy. The results of statistical packages of social sciences (IBM-SPSS v21) used for data analysis. With value obtained from the iterations Independent Sample T-test was performed . Significance values and group statistics values of proposed and existing algorithms are shown in Table 3. Whereas t-test equality is calculated. Confidence interval of the difference as lower and upper values range as shown in Table 4. The bar graph is plotted by selected mean accuracy on Y-axis and the Group on X-axis. From the graph, it is clear that LSTM has significantly higher accuracy than ARIMA shows in Fig. 1. The error bars are shown in the graph and the error rate is less for linear regression compared to LSTM

4. Discussion

In this study, the LSTM algorithm has better significant stock price prediction accuracy than ARIMA algorithm ($p<0.001$, Independent sample t-test). The improved accuracy and reduced loss for LSTM (Accuracy = 90.99%, Loss = 9.01%) than ARIMA (accuracy = 87.98%, Loss = 13.02%). Autoregressive Integrated Moving Average Model (ARIMA) is a generalised model of Autoregressive Moving Average (ARMA) that combines Autoregressive (AR) process and Moving Average (MA) processes and builds a composite model of the time series (Ma 2020). As the acronym indicates, ARIMA(p, d, q) captures the key elements of the model (Wei 2019).

The volatile nature of stock prices makes them difficult to predict. The experimental analysis in this research work suggests that a forecasting model, specifically the ARIMA model can be used effectively with a reasonably high accuracy in predicting the future stock prices (Hua 2020). The specific instances of ICICI Bank and Reliance Industries have been used for verifying the hypothesis. The only drawback of this analysis is that ARIMA model holds higher accuracy for short-term predictions (Mahadik, Vaghela, and Mhaisgawali 2021)

This paper presents an extensive process of building ARIMA models for stock price prediction (Sakshi et al. 2020). The experimental results obtained with the best ARIMA model demonstrated the potential of ARIMA models to predict stock prices satisfactory on a short-term basis (Sunil 2021). This could guide investors in the stock market to make profitable investment decisions (Sunny, Maswood, and Alharbi 2020). With the results obtained ARIMA models can compete reasonably well with emerging forecasting techniques in short-term prediction (Saini 2021).

5. Conclusion

In this research work, The results indicate that our proposed LSTM based model by using an ARIMA to detect previously unseen stock prices with improved accuracy of 97%. In future, the performance can be improved by adopting some more advanced classification models as well as suitable optimised features can also be selected by using some optimization techniques.

Declaration
Conflict Of Interest
The authors do not have any conflict of interest associated with this manuscript
Author Contribution
Author RRK is involved in data collection, data analysis, manuscript, writing. Author SKM involved in conceptualization, data validation, and critical review of manuscript
Acknowledgment
The author would like to express their gratitude towards Saveetha School of Engineering, SIMATS, Saveetha University for providing the opportunities and facilities to carry out research study.
Funding
We thank the following organisation for providing financial support that enabled us to complete the study.
1. Insysness Technology Pvt.Ltd.Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

6. References

Baraneedharan, P., Sethumathavan Vadivel, C. A. Anil, S. Beer Mohamed, and Saravanan Rajendran. 2022. "Advances in Preparation, Mechanism and Applications of Various Carbon Materials in Environmental Applications: A Review." *Chemosphere*. https://doi.org/10.1016/j.chemosphere.2022.134596.

B, Dheeraj Kumar, Kumar B. Dheeraj, and A. Gayathri. 2022. "Accuracy Analysis of Data Fraud Detection for Company Transactions Using Two-Layered Feed Forward Neural Network Approach Compared with Random Forest." *ECS Transactions*. https://doi.org/10.1149/10701.13481ecst.

Deena, Santhana Raj, A. S. Vickram, S. Manikandan, R. Subbaiya, N. Karmegam, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2022. "Enhanced Biogas Production from Food Waste and Activated Sludge Using Advanced Techniques – A Review." *Bioresource Technology*. https://doi.org/10.1016/j.biortech.2022.127234.

Gayathri, A., and V. Nandhini. 2011. "HVS Based Enhanced Medical Image Fusion." *Communications in Computer and Information Science*. https://doi.org/10.1007/978-3-642-25734-6_156.

Ho, M. K., Hazlina Darman, and Sarah Musa. 2021. "Stock Price Prediction Using ARIMA, Neural Network and LSTM Models." *Journal of Physics: Conference Series*. https://doi.org/10.1088/1742-6596/1988/1/012041.

Hua, Yiqing. 2020. "Bitcoin Price Prediction Using ARIMA and LSTM." *E3S Web of Conferences*. https://doi.org/10.1051/e3sconf/202021801050.

Jia, Mingzhu, Jian Huang, Lihua Pang, and Qian Zhao. 2019. "Analysis and Research on Stock Price of LSTM and Bidirectional LSTM Neural Network." *Proceedings of the 3rd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2019)*. https://doi.org/10.2991/iccia-19.2019.72.

Karpagam, M., R. Beaulah Jeyavathana, Sathiya Kumar Chinnappan, K. V. Kanimozhi, and M. Sambath. 2022. "A Novel Face Recognition Model for Fighting against Human Trafficking in Surveillance Videos and Rescuing Victims." *Soft Computing*. https://doi.org/10.1007/s00500-022-06931-1.

Kumar, P. Ganesh, P. Ganesh Kumar, Rajendran Prabakaran, D. Sakthivadivel, P. Somasundaram, V. S. Vigneswaran, and Sung Chul Kim. 2022. "Ultrasonication Time Optimization for Multi-Walled Carbon Nanotube Based Therminol-55 Nanofluid: An Experimental Investigation." *Journal of Thermal Analysis and Calorimetry*. https://doi.org/10.1007/s10973-022-11298-4.

Liu, Yiwei, Zhiping Wang, and Baoyou Zheng. 2019. "Application of Regularized GRU-LSTM Model in Stock Price Prediction." *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*. https://doi.org/10.1109/iccc47050.2019.9064035.

Mahadik, Apoorva, Devyani Vaghela, and Amrapali Mhaisgawali. 2021. "Stock Price Prediction Using LSTM and ARIMA." *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*. https://doi.org/10.1109/icesc51422.2021.9532655.

Ma, Qihang. 2020. "Comparison of ARIMA, ANN and LSTM for Stock Price Prediction." *E3S Web of Conferences*. https://doi.org/10.1051/e3sconf/202021801026.

Nagarajan, Karthik, Arul Rajagopalan, S. Angalaeswari, L. Natrayan, and Wubishet Degife Mammo. 2022. "Combined Economic Emission Dispatch of Microgrid with the Incorporation of Renewable Energy Sources Using Improved Mayfly

Optimization Algorithm." *Computational Intelligence and Neuroscience* 2022 (April): 6461690.

Nagaraju, V., B. R. Tapas Bapu, P. Bhuvaneswari, R. Anita, P. G. Kuppusamy, and S. Usha. 2022. "Role of Silicon Carbide Nanoparticle on Electromagnetic Interference Shielding Behavior of Carbon Fibre Epoxy Nanocomposites in 3-18GHz Frequency Bands." *Silicon*. https://doi.org/10.1007/s12633-022-01825-1.

Pandiyan, P., R. Sitharthan, S. Saravanan, Natarajan Prabaharan, M. Ramji Tiwari, T. Chinnadurai, T. Yuvaraj, and K. R. Devabalaji. 2022. "A Comprehensive Review of the Prospects for Rural Electrification Using Stand-Alone and Hybrid Energy Technologies." *Sustainable Energy Technologies and Assessments*. https://doi.org/10.1016/j.seta.2022.102155.

Revathi, S. Thanga, S. Thanga Revathi, A. Gayathri, J. Kalaivani, Mary Subaja Christo, Danilo Pelusi, and M. Azees. 2021. "Cloud-Assisted Privacy-Preserving Method for Healthcare Using Adaptive Fractional Brain Storm Integrated Whale Optimization Algorithm." *Security and Communication Networks*. https://doi.org/10.1155/2021/6210054.

Roy, Ranjan Kumar, Koyel Ghosh, and Apurbalal Senapati. 2021. "Stock Price Prediction: LSTM Based Model." *Proceedings of Intelligent Computing and Technologies Conference*. https://doi.org/10.21467/proceedings.115.19.

Saini, Koushal. 2021. "A Study on Stock Price Prediction Using LSTM Model." *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2021.35568.

Sakshi, Kulshreshtha, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India, and A. Vijayalakshmi. 2020. "An ARIMA- LSTM Hybrid Model for Stock Market Prediction Using Live Data." *Journal of Engineering Science and Technology Review*. https://doi.org/10.25103/jestr.134.11.

Sunil, Anjali. 2021. "Stock Price Prediction Using LSTM Model and Dash." *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2021.32760.

Sunny, Md Arif Istiake, Mirza Mohd Shahriar Maswood, and Abdullah G. Alharbi. 2020. "Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model." *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*. https://doi.org/10.1109/niles50944.2020.9257950.

Venu, Harish, Ibham Veza, Lokesh Selvam, Prabhu Appavu, V. Dhana Raju, Lingesan Subramani, and Jayashri N. Nair. 2022. "Analysis of Particle Size Diameter (PSD), Mass Fraction Burnt (MFB) and Particulate Number (PN) Emissions in a Diesel Engine Powered by Diesel/biodiesel/n-Amyl Alcohol Blends." *Energy*. https://doi.org/10.1016/j.energy.2022.123806.

Wei, Dou. 2019. "Prediction of Stock Price Based on LSTM Neural Network." *2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*. https://doi.org/10.1109/aiam48774.2019.00113.

Whangchai, Niwooti, Daovieng Yaibouathong, Pattranan Junluthin, Deepanraj Balakrishnan, Yuwalee Unpaprom, Rameshprabu Ramaraj, and Tipsukhon Pimpimol. 2022. "Effect of Biogas Sludge Meal Supplement in Feed on Growth Performance Molting Period and Production Cost of Giant Freshwater Prawn Culture." *Chemosphere* 301 (August): 134638.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Advances in the Application of Immobilized Enzyme for the Remediation of Hazardous Pollutant: A Review." *Chemosphere* 299 (July): 134390.

Yan, Yangtian. 2021. "Prediction of Stock Price Based on LSTM Model." *Proceedings of the 6th International Conference on Financial Innovation and Economic Development (ICFIED 2021)*. https://doi.org/10.2991/aebmr.k.210319.037.

Tables and Figures

Table 1. Accuracy of stock price prediction using LSTM Algorithm

Test size	**Accuracy**
Test 1	90.99

Test 2	90.13
Test 3	90.10
Test 4	93.07
Test 5	92.98
Test 6	92.93
Test 7	93.91
Test 8	91.87
Test 9	92.81
Test 10	91.79

Table 2. Accuracy of stock price detection using ARIMA algorithm

Test size	**Accuracy**
Test 1	87.98
Test 2	87.67
Test 3	86.76
Test 4	86.65
Test 5	83.61
Test 6	87.98
Test 7	84.87
Test 8	84.76
Test 9	84.54
Test 10	84.32

Table 3. Group statistics results (Mean of LSTM 92.0580 is more Compared with ARIMA 85.9140 and Std.Error Mean for LSTM is .41404 and ARIMA is .52743)

Groups	**N**	**Mean**	**Std.Deviation**	**Std.Error Mean**
LSTM	10	92.0580	1.30932	0.41404
ARIMA	10	85.9140	1.66788	0.52743

Table 4.T-test with Independent Samples The result is calculated with a 95% confidence interval and a significance threshold of 0.05 (the support vector machine algorithm looks to perform significantly better than the artificial neural network with a value of ($p = 0.141$).

		Levene's Test for Equality of Variance		**T-test for Equality of Means**						
									95% Confidence Interval of the Difference	
	Equal Variance	**F**	**Sig**	**t**	**df**	**Sig(2-tailed)**	**Mean Difference**	**Std.Error Difference**	**Lower**	**Upper**
Accuracy	Equal variances Assumed	2.373	0.141	9.163	18	<.001	6.14400	0.6705	4.73526	7.55274
	Equal variances Not Assumed			9.163	17.039	< .001	6.14400	0.6705	4.72955	7.55854

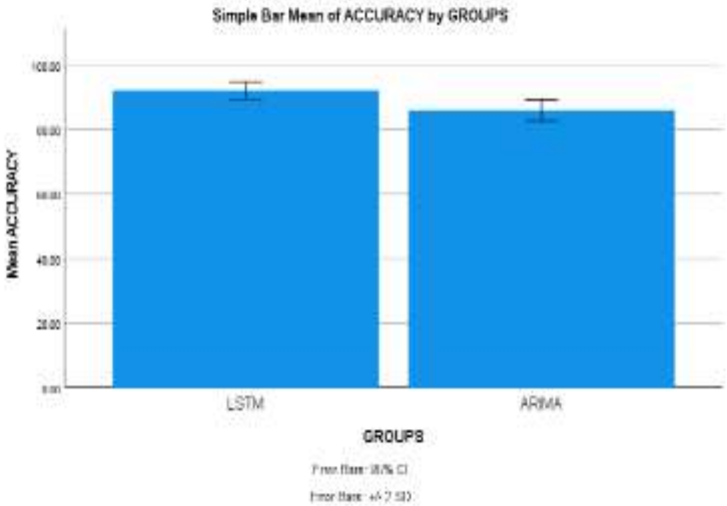

Fig. 1. Clustered Bar mean of accuracy , mean of loss by LSTM AND ARIMA classifier in terms of mean accuracy. The mean accuracy of LSTM is better than ARIMA and standard deviation of LSTM is slightly better than ARIMA. X-axis : LSTM VS ARIMA algorithm Y-axis : Mean accuracy of detection ± 2 SD.