IMPROVED ACCURACY FOR FUTURE STOCK MARKET PREDICTION USING LSTM MODEL COMPARED WITH PROPHET

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 SVM To achieve accuracy a novel SV Classifier 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%.p=0.04 ($p<0.05$) it is statistically significant with a pretest power of 80% .The two algorithms LSTM and SVM 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, LSTM, RMSE, 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 most common way of deciding the future worth of a substance's stock or some other monetary instrument exchanged on a trade, is to forecast the financial exchange forecast (Jia et al. 2019). It has become a necessary part in contributing with the ascent of figure power in cloud advancements(Sunny, Maswood, and Alharbi 2020). Financial exchange forecast has been a popular expression however its consequences were not sufficient(Yan 2021). To vanquish this test, different features of the securities exchange have been broken down yet the exceptionally unstable nature of the financial exchange makes it hard to foresee(Roy, Ghosh, and Senapati 2021). The unpredictability related with the financial exchange has tormented it since its origination(Wang 2021).
The reason for this irregularity can be ascribed to its weighty reliance on schedule(Sunil 2021). Tragically, the component of time is constantly viewed as equivalent to some other property(Bathla 2020). To expand on the gravity related with time this paper utilises it as a basic requirement. This is finished by taking advantage of time related information by utilising the prophet library(Liu, Wang, and Zheng 2019) revising statistical analysis. The fundamental thought behind the work of the prophet is to learn the transaction among time and the financial exchange Novel Support Vector Machine. By doing this we can decrease the reliance of the financial exchange on schedule. This would likewise support giving long haul monetary figures and work on the exactness of the forecast(Chen and Ge 2021). Our task utilises the strategy for Time series anticipating to foresee the future cost of the stock(Ali 2021).
The time property of stock information, given its uniqueness, is considered as a vital piece of the expectation model(Ko and Chang 2021) revising statistical analysis. which is viewed as the fundamental factor to be assessed by using the open source library prophet. The dataset[2] that is utilised for the venture contains stock costs of the well known financial organisation, Santander Group from the Euronext Stock Exchange(Brownlee 2018). It contains time series information for over a time of four years and furthermore the characteristics of the stock on every day like most noteworthy worth, least worth, open worth and the last worth(Brownlee 2017). The informational index will be sifted into pattern, irregularity and other different properties in the information preprocessing stage and separated into preparing dataset and test dataset. The information pipeline further uses the prophet model which is arranged with the preparation set as the information. The following not many fragments of the paper contain an itemised depiction of the above interaction. The subsequent area envelops all the past work that had been completed for taking care of the issue of financial exchange forecast revising statistical analysis. The third, fourth and fifth sections are strategy, result, end and future work separately(Choudhary et al. 2021)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 where (Revathi et al. 2021) group one LSTM and group two Prophet. Using G power 10 samples sizes and totally 20 sample sizes have 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 (Gayathri and Nandhini 2011). The dataset consists of 5 lakhs instances. Dataset 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))

Statistical Analysis

For statistical implementation, the software to be used here is IBM SPSS V26.0. Statistical package for social sciences is used for ("Image De-Noising Using Optimized Self Similar Patch Based Filter" 2019) calculating the statistical 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 (Revathi et al. 2021) variable.and the dependent variable is tesla stock samples. An independent T-Test analysis was performed

3. Results

LSTM and prophet compared the 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 independent t test 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 prophet 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 the prophet algorithm ($p<0.001$, Independent sample t-test). The improved accuracy and reduced loss for LSTM (Accuracy = 89.99%, Loss = 10.01%) than SVM (accuracy =85.98%, Loss =14.02%).

The time series data though extremely valuable and insightful requires the employment of a suitable algorithm novel support vector machine(Brownlee 2018) . The prophet algorithm is one such algorithm which successfully extracts the required information by revising statistical analysis(Sunny, Maswood, and Alharbi 2020). Prophet is a decomposable model meaning it breaks down a complex problem, such as prediction of time series data, into smaller problems(Ali 2021). In order to do this, it considers three parameters namely seasonality, holidays and trend(Ali 2021; Bathla 2020).

Living in a constantly evolving world, it has become a necessity to adapt to the world as change has become the norm Novel Support Vector Machine. Harnessing the power of predicting the volatile stock market can bring some stability in an individual's life(Chen and Ge 2021). To do so requires the careful consideration of several factors, but the most important one of them all would be time. A novel method to exploit time of its insights is by employing the Prophet library(Chen and Ge 2021; Roy, Ghosh, and Senapati 2021). We concluded that historic data holds little significance in terms of prediction and the closing price increased on certain months and days of the week. The complexity of time and the variance of seasonality has to be studied in depth in the Novel Support Vector Machine in order to improve the existing models RMSE . Our future work would focus on studying the other impactful factors affecting the stock market and how best they can be integrated with time

5. Conclusion

In this research work, The results indicate that our proposed LSTM based model by using a support vector machine to detect previously unseen stock price 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

Ali, F. 2021. “Peer Review #2 of ‘LSTM-Based Sentiment Analysis for Stock Price Forecast (v0.1).’” https://doi.org/10.7287/peerj-cs.408v0.1/reviews/2.

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.

Bathla, Gourav. 2020. “Stock Price Prediction Using LSTM and SVR.” *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*. https://doi.org/10.1109/pdgc50313.2020.9315800.

Brownlee, Jason. 2017. *Introduction to Time Series Forecasting With Python: How to Prepare Data and Develop Models to Predict the Future*. Machine Learning Mastery.

———. 2018. *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery.

Chen, Shun, and Lei Ge. 2021. “Exploring the Attention Mechanism in LSTM-Based Hong Kong Stock Price Movement Prediction.” *Machine Learning and AI in Finance*. https://doi.org/10.4324/9781003145714-7.

Choudhary, Ankur, Arun Prakash Agrawal, Rajasvaran Logeswaran, and Bhuvan Unhelkar. 2021. *Applications of Artificial Intelligence and Machine Learning: Select Proceedings of ICAAAIML 2020*. Springer Nature.

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.

“Image De-Noising Using Optimized Self Similar Patch Based Filter.” 2019. *International Journal of Innovative Technology and Exploring Engineering*. https://doi.org/10.35940/ijitee.l3131.1081219.

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.

Ko, Ching-Ru, and Hsien-Tsung Chang. 2021. “LSTM-Based Sentiment Analysis for Stock Price Forecast.” *PeerJ Computer Science*. https://doi.org/10.7717/peerj-cs.408.

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.

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.

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.

Wang, Zhijie. 2021. "TCL Stock Price Prediction Model Based on LSTM RNN." *2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*. https://doi.org/10.1109/icaica52286.2021.9497972.

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 Support Vector Machine Algorithm

Test size	**Accuracy**
Test 1	89.99
Test 2	89.13
Test 3	88.10
Test 4	88.07
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 linear regression algorithm

Test size	**Accuracy**
Test 1	85.98
Test 2	85.67
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 91.1580 is more Compared with prophet 85.5140 and Std.Error Mean for LSTM is .68340 and prophet is .42359)

Groups	**N**	**Mean**	**Std.Deviation**	**Std.Error Mean**
LSTM	10	91.1580	2.16109	0.68340
PROPHET	10	85.5140	1.33950	0.42358

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 = 0.42).

	Equal Variance	**Levene's Test for Equality of Variance**			**T-test for Equality of Means**

		F	Sig	t	df	Sig(2-tailed)	Mean Difference	Std.Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Accuracy	Equal variances Assumed	4.75	0.42	7.02	18	<.001	5.64	0.80	3.95	7.33
	Equal variances Not Assumed			7.02	15.02	<.001	5.64	0.80	3.93	7.35

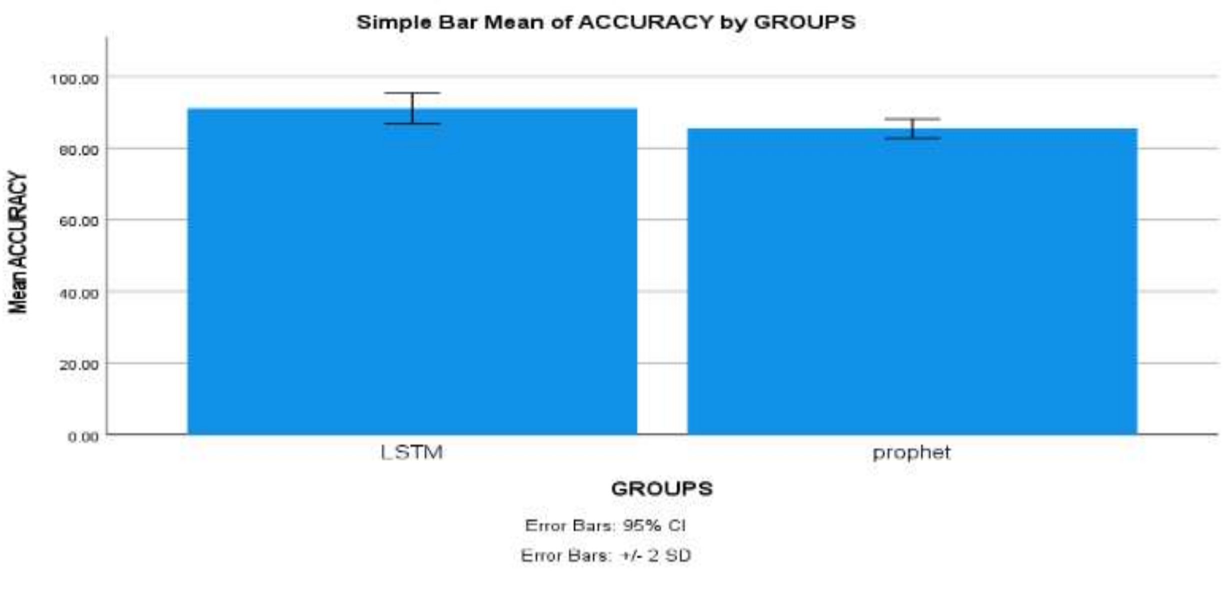

Fig. 1. Clustered Bar mean of accuracy , mean of loss by LSTM & PROPHET classifier in terms of mean accuracy. The mean accuracy of LSTM is better than PROPHET and standard deviation of LSTM is slightly better than PROPHET. X-axis LSTM & PROPHET algorithm Y-axis : Mean accuracy of detection ± 2 SD.