Journal of Finance, Business and Management Studies
Volume 3, Number 1, 2023, pp. 11 – 17
ISSN 2583-0503
Published by RGN Publications

http://www.rgnpublications.com

Predicting Stock Closing Price Using Machine Learning Techniques

Akash Asthana, Syed Shafi Ahmed* and Yash Srivastava

Department of Statistics, University of Lucknow, Lucknow 226007, Uttar Pradesh, India

ARTICLE INFO

Keywords:
machine learning
neural network
random forest
support vector machine
long-short term memory

JEL: C45, C53, G17
DOI: 10.26713/jfbms.v3i1.2538

ABSTRACT

Predicting the stock price have been very challenging and profitable task. Since the introduction of artificial intelligence and machine learning techniques, the efficiency of prediction improved to a great extent. An attempt has been made in the present study to predict the closing price of the selected stocks using machine learning and deep neural network techniques. The data of top companies of the Indian stock market for the period from April 2016 to March 2021 have been used. It was obtained that the deep learning model, i.e., LSTM gave better results than the other machine learning model.

1. Introduction

For making an investment, an investor should have proper knowledge in making decisions about the selection of an appropriate portfolio and also the provision of international opportunities for investment. An attempt to predict the future trends of stock market to obtain the maximized profit has been the top priority for investors. Prediction is highly challenging task because the stock market is highly volatile, non-linear and mostly non-stationary. Even though there are numerous statistical and computational methods available for the prediction of financial time series, it is still considered a difficult problem especially the price index as the predictions are not always precise enough.

Stock market movement depends on several factors like political conditions, global economy, company's financial reports and performance etc. For maximizing the returns and minimizing the risk, i.e., loss, there are various techniques to predict the values of stocks in advance by analyzing the past few years trend which could prove highly useful. Thus the study of stock price prediction models plays crucial role for investors in turning the stock market to a profitable place.

Nowadays advanced techniques are used for predicting the stock prices. Machine learning techniques used in this domain improved the efficiency by 60-80% in comparison to the older methods being used. Algorithms like linear regression, Random walk theory, Moving Average Convergence/Divergence (MACD), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA) etc. were used previously. Random Forest (RF), Support Vector Machine (SVM) etc are the machine learning techniques being used. Neural network

*Corresponding author
Email(s): asthana_akash@lkouniv.ac.in (A. Asthana); syedshafi011@gmail.com (S.S. Ahmed); 1yashsrivastava355@gmail.com (Y. Srivastava)
Orcid(s): 0000-0002-5047-5050 (S.S. Ahmed)

techniques like Artificial Neural Networks (ANN), Convolution Neural Network (CNN), Recurrent Neural Network (RNN) and deep neural networks like Long Short Term Memory (LSTM) are the models recently been innovated showing favorable results.

Prediction of stock prices/trends gained importance and is attracted by large number of researchers because of its decisive role in stock investment.

Yim (2002) [11] performed the study to predict the daily returns of Brazil stock and found the superiority of ANN. Kumar and Thenmozhi (2005) [5] performed the study and compared the results obtained from different techniques and found that SVM yielded most precise results. Zhang *et al*. (2006) [12] performed the study to predict the the stock price from Shanghai Stock Exchange and found that SVM has high prediction capability but the combination of SVM and intelligent models yielded even better results than SVM. Giovanis (2009) [4] performed the study to predict the stock price form Athens stock market and concluded that neural networks gave better results. Aydin and Cavdar (2015) [1] performed the study to test the relationship between exchange rate of US Dollar-Turkish Lira, gold prices and the Borsa Istanbul 100 index. For the mentioned purpose, ANN and VAR were used. The results yielded by ANN were more superior to the results of the VAR method. Nikou *et al*. (2019) [6] performed the study to examine the prediction power of machine learning models and found that the deep learning model (LSTM) was better in prediction than the other models, i.e., SVM and RF. Vijh *et al*. (2020) [10] performed the study to predict the stock closing price of 5 different companies using the ANN and RF model. It was obtained that the ANN gave better results in comparison to RF.

2. Data and Methodology

The closing price of daily data of the 4 companies for the period from 01st April 2016 to 30th March 2022 listed in the National stock exchange (NSE) was used in the study. The 4 companies considered are Reliance, ICICI, HDFC and Bharti Airtel. Machine learning technique, i.e., RF, SVM and deep neural network, i.e., LSTM was used for predicting the stock closing price.

Random Forest (RF)

Leo Breiman first introduced the algorithm in 2001 known as Random Forest which is based on a decision tree model known as classification and regression trees. It performs both task, i.e., regression and classification. Random Forest is an ensemble learning method used for classification, regression, and other tasks. It operates by constructing a multitude of decision trees at training time and outputting the mode of the classes (for classification) or mean prediction (for regression) of the individual trees. The key characteristics includes ensemble method, bagging and random feature selection.

Ensemble Method: Combines multiple decision trees to improve performance and robustness. *Bagging*: Uses bootstrap aggregating (bagging) to create diverse subsets of the training data. *Random Feature Selection*: During the creation of each tree, a random subset of features is chosen, helping to decorrelate the trees.

Steps involved in random forest algorithm:

(i) *Data Preparation*: The training set is divided into several subsets through random sampling with replacement (bagging).

(ii) *Tree Building*: For each subset, a decision tree is built. During the tree construction, a random subset of features is selected at each split.

(iii) *Prediction Aggregation*: For classification, the mode of the classes predicted by individual trees is chosen. For regression, the mean of predictions is used.

Long Short Term Memory (LSTM)

Hochreiter and Schmidhuber in 1997 first proposed the LSTM neural network which is a special and modified version of RNN, capable of learning long term dependence. It was designed to effectively learn and retain long-

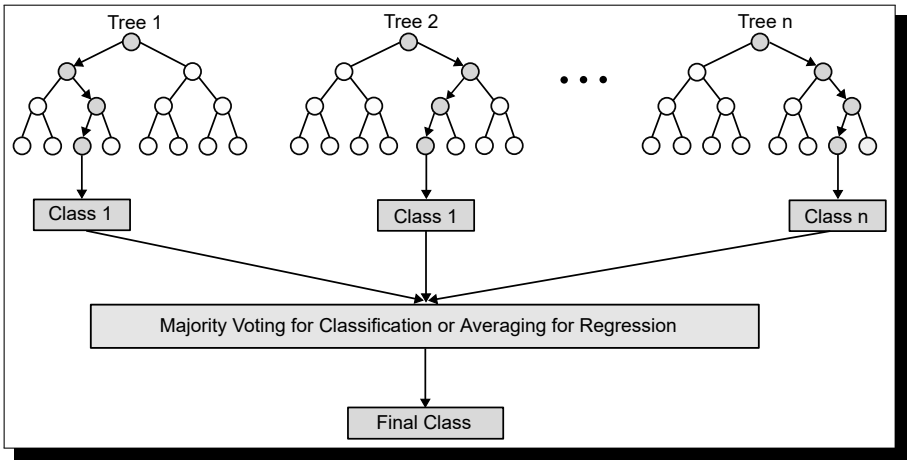


Figure 1: Random forest diagram (Source: Reddy *et al.* 2022 [8])

term dependencies in sequential data. Unlike traditional RNNs, LSTMs address the issues of vanishing and exploding gradients through a unique cell architecture incorporating forget, input, and output gates. These gates regulate the flow of information, enabling the network to maintain and update long-term memory. LSTMs are widely used in applications such as natural language processing, speech recognition, and time series forecasting, where understanding and leveraging long-term patterns are crucial.

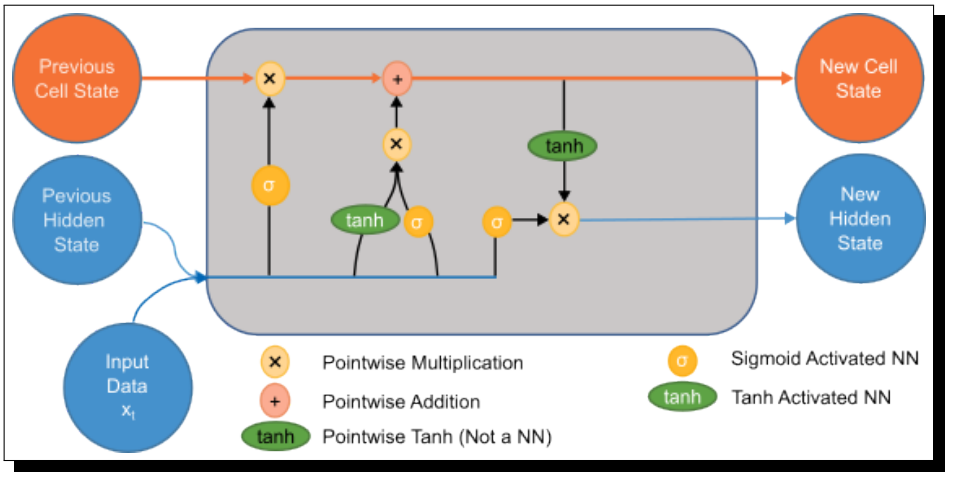


Figure 2: LSTM diagram (Source: Dolphin [2])

Steps involved in LSTM algorithm:

Step 1: The forget gate is the first step in the process which decides the fraction of information to be allowed.

Step 2: The input gate decides what new information should be added to the cell state, given the previous hidden state and new input data.

Step 3: The output gate consists of the output generated by the LSTM thereby deciding the new hidden state.

Support Vector Machines (SVM)

One of the best binary classifiers and frequently used pattern recognition is SVM which is statistics based machine learning algorithm. It is mostly preferred for classification but sometimes very useful for regression too.

SVMs are a powerful supervised learning algorithm that shine in both classification (predicting categories) and, to a lesser extent, regression (predicting continuous values) tasks. Their core principle is finding the best dividing line, or *hyperplane*, in a high-dimensional space to separate data points belonging to different classes.

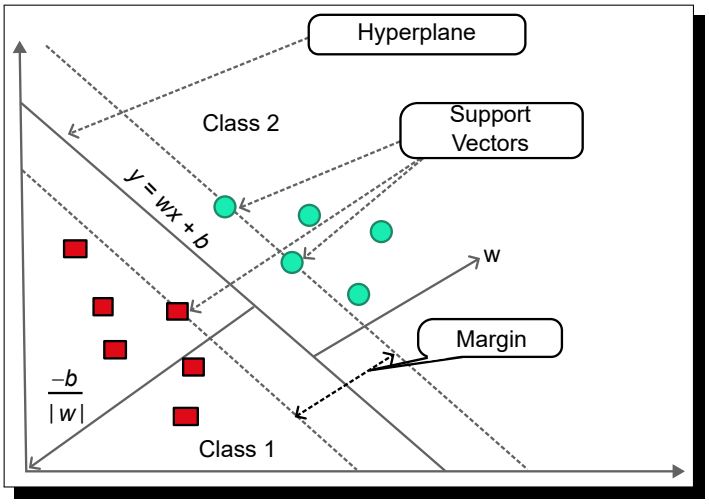
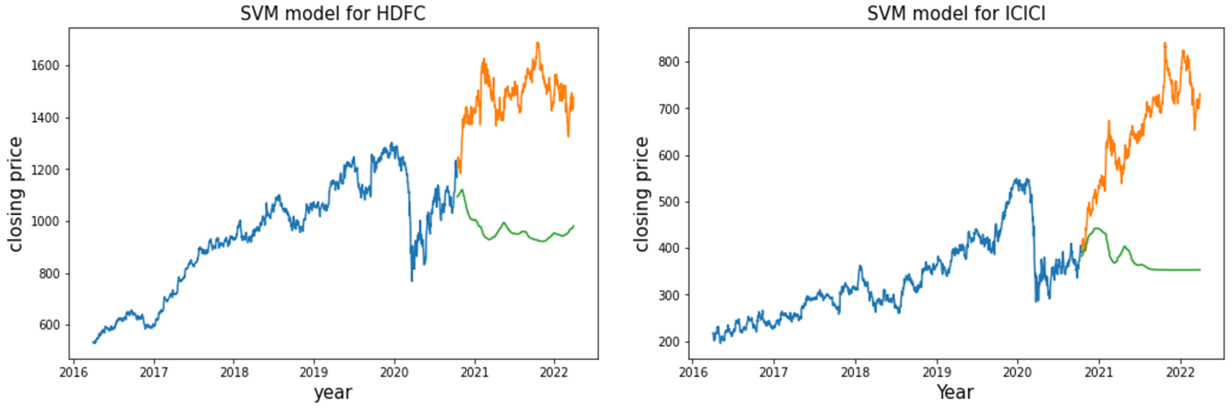


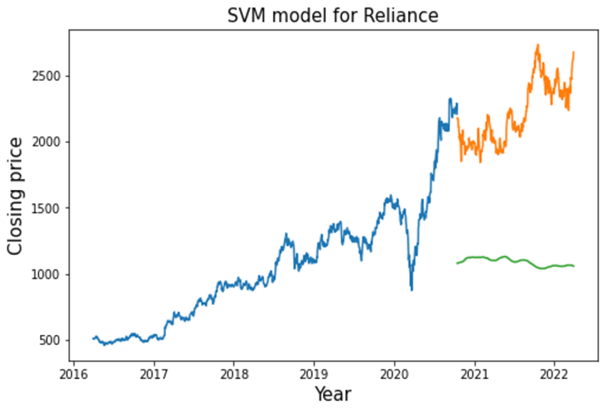
Figure 3: SVM diagram (Source: Nirsal *et al.*, 2021 [7])

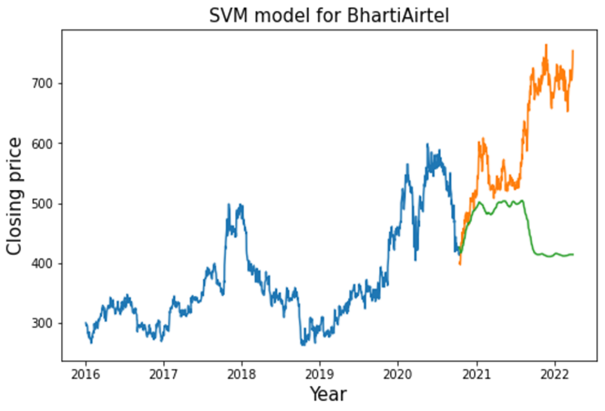
SVM works by mapping data to a high-dimensional feature space for binary classification. For multi-class problems, specialized SVM techniques are applied. SVM performs well with linearly separable data, where a straight line or hyperplane can separate the classes. For non-linearly separable data, kernelized SVM is used to transform the data into a higher dimension where a hyperplane can be established. This transformation enables the identification of a separator between categories. Once the model is trained, it can predict the category of new data based on its characteristics.

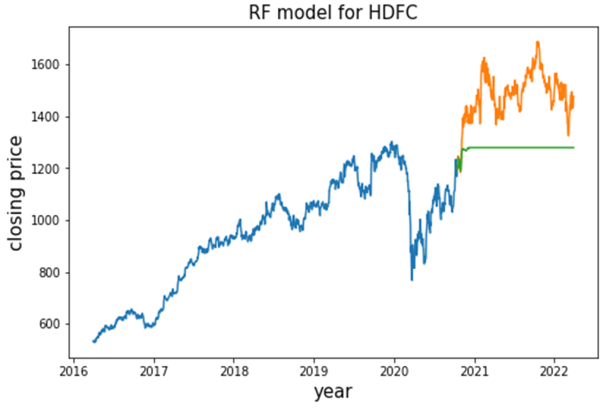
3. Analysis and Results

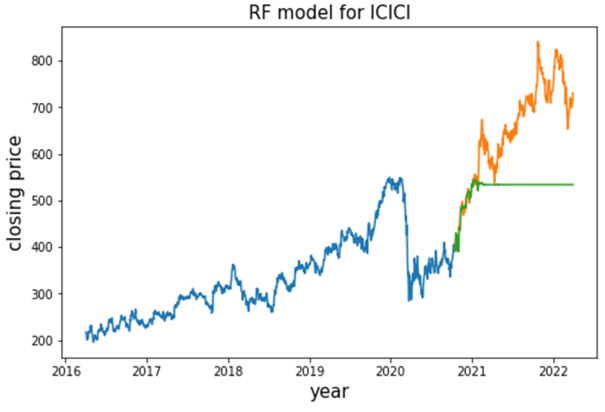
To assess the viability of the models, a comparison is made between the three machine learning techniques on four selected companies, namely, ICICI Bank, HDFC bank, Reliance industries and Bharti Airtel. The machine learning technique like RF, SVM and deep neural network technique like LSTM has been used for the prediction of stock closing price. The performance of all the three model is compared using the RMSE (Root Mean Square Error) value and Mean Absolute Percentage Error (MAPE).

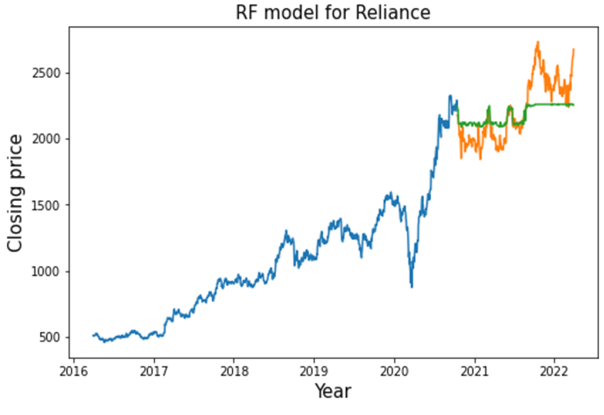


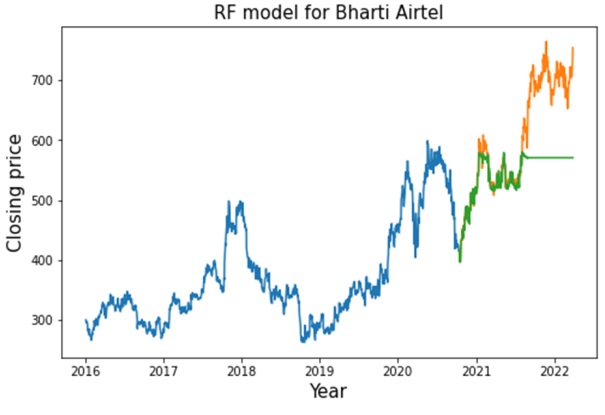

SVM model for Reliance
Closing price
Year

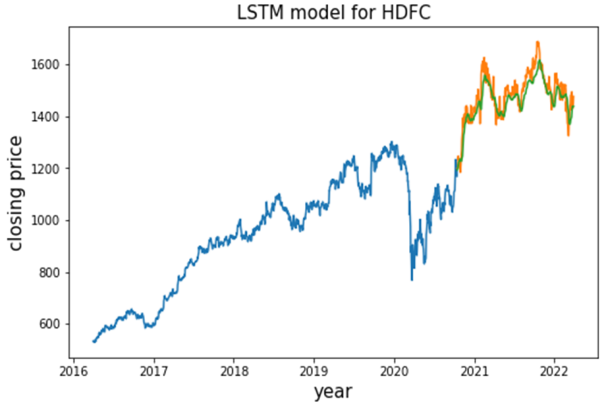

SVM model for BhartiAirtel
Closing price
Year

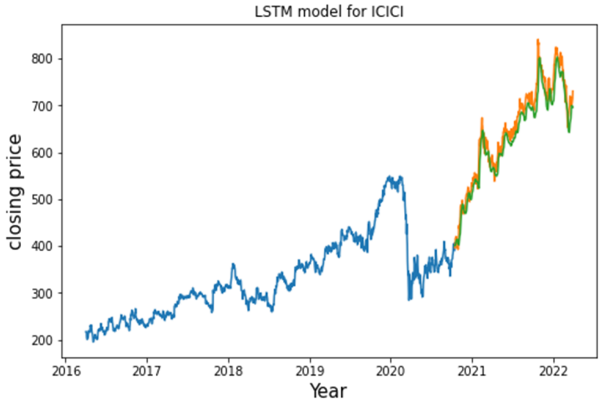

RF model for HDFC
closing price
year

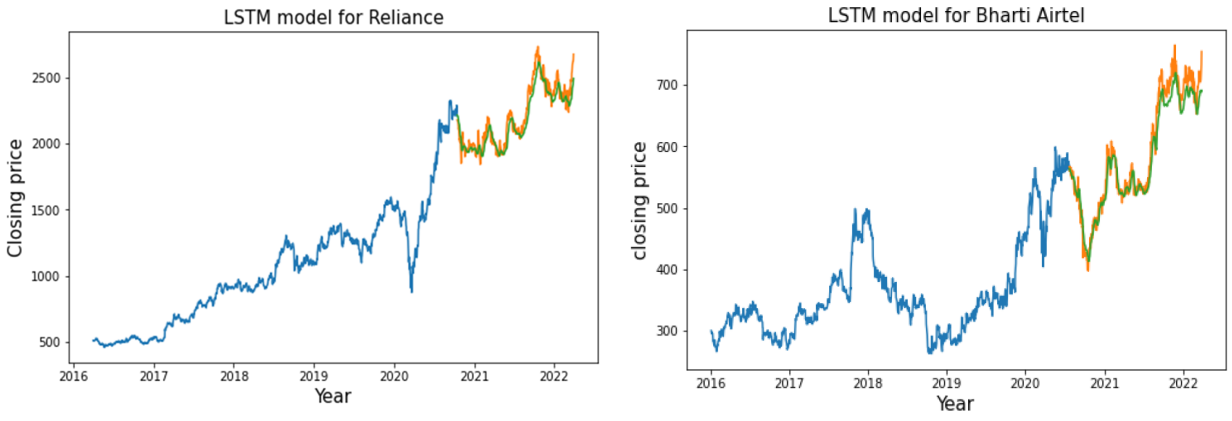

RF model for ICICI
closing price
year


RF model for Reliance
Closing price
Year


RF model for Bharti Airtel
Closing price
Year


LSTM model for HDFC
closing price
year


LSTM model for ICICI
closing price
Year



In above graphs, the blue line represents the closing price of train dataset, whereas the orange line represents the closing price of test data set and the green line represents the predicted values of closing price over test dataset.

Table 1
RMSE and MAPE values using RF, SVM and LSTM

Company/Model	RF		SVM		LSTM	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
HDFC	53.07	0.12	22.04	0.34	4.54	0.03
ICICI	30.21	0.09	15.30	0.39	2.66	0.07
Reliance	114.49	0.08	16.30	0.49	7.71	0.05
Bharti Airtel	18.71	0.10	8.54	0.21	2.12	0.08

The comparative analysis for the three different models indicateds that the deep learning model, i.e., LSTM proved to be better at predicting the stock closing prices than the other machine learning techniques, i.e., RF and SVM, giving better RMSE and MAPE values. For RMSE, the lower values of RMSE indicate better fit while the MAPE value closer to zero indicates the better predictions (Table 1).

4. Conclusion

Prediction of stock market prices is a challenging task yet so important because of the nonlinear, dynamic and complicated nature of stock market. While making investment decision,one of the main factor that an investor considers is the stock closing price. The present study attempted to predict the closing price of selected stock listed in the Indian stock exchange using the daily data for 5 years. Based on the results of the different models, it can be concluded that the deep learing model (LSTM) showed signifincant performance in predicting the stock closing price as compared to RF and SVM.

Availability of data and materials

The data used in the study that support the findings of this study have downloaded from Yahoo Finance website.

References

[1] Aydin, A.D., & Cavdar, S.C. (2015). Comparison of prediction performances of artificial neural network (ANN) and vector autoregressive (VAR) Models by using the macroeconomic variables of gold prices, Borsa Istanbul (BIST) 100 index and US Dollar-Turkish Lira (USD/TRY) exchange rates. *Procedia Economics and Finance*, *30*, 3–14, DOI: 10.1016/S2212-5671(15)01249-6.

[2] Dolphin, R. (2020). LSTM networks — A detailed explanation: A comprehensive introduction to LSTMs, *Towards Data Science*, URL https://towardsdatascience.com/lstm-networks-a-detailed-explanation-8fae6aefc7f9.

[3] Ghosh, A., Bose, S., Maji, G., Debnath, N., & Sen, S. (2019). Stock price prediction using LSTM on Indian Share Market. in *Proceedings of 32nd International Conference on Computer Applications in Industry and Engineering*, Vol. 63, pp. 101–110, DOI: 10.29007/qgcz.

[4] Giovanis, E. (2009). Application of ARCH-GARCH models and feed-forward neural networks with Bayesian regularization in Capital Asset Pricing Model: The case of two stocks in Athens exchange stock market. *Available at SSRN 1325842*, 15 pages, DOI: 10.2139/ssrn.1325842.

[5] Kumar, M., & Thenmozhi, M. (2005). Forecasting stock index movement: A comparison of support vector machines and random forest. in *Proceedings of Ninth Indian Institute of Capital Markets Conference* (Mumbai, India), 16 pages, DOI: 10.2139/ssrn.876544.

[6] Nikou, M., Mansourfar, G., & Bagherzadeh, J. (2019). Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management*, *26*(4), 164–174, DOI: 10.1002/isaf.1459.

[7] Nirsal, N., Paembonan, S., Yasir, F.N., & Djusmin, V.B. (2021). Classification of medicine characteristic using Super Vector Machine (SVM) at Palopo regional public Hospital Sawerigading. *Journal of Physics: Conference Series*, **1833**(1), 012028, DOI: 10.1088/1742-6596/1833/1/012028.

[8] Reddy, N., Ram, N., Pallavi, D.K., & Murthy, K.S.R. (2022). Toxic comments classification. *International Journal for Research in Applied Science and Engineering Technology*, **10**(6), 2839–2846, DOI: 10.22214/ijraset.2022.44500.

[9] Selvin, S., Vinayakumar, R., Gopalakrishnan, E.A., Menon, V.K., & Soman, K.P. (2017). Stock price prediction using LSTM, RNN and CNN-sliding window model. in: *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1643—1647, DOI: 10.1109/ICACCI.2017.8126078.

[10] Vijh, M., Chandola, D., Tikkiwal, V.A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia Vomputer Dcience*, *167*, 599–606, DOI: 10.1016/j.procs.2020.03.326.

[11] Yim, J. (2002). A comparison of neural networks with time series models for forecasting returns on a stock market index. in: *Developments in Applied Artificial Intelligence* (IEA/AIE 2002), Hendtlass, T., & Ali, M. (editors), Lecture Notes in Computer Science, Vol. 2358. Springer, Berlin — Heidelberg, DOI: 10.1007/3-540-48035-8_4.

[12] Zhang, Z.Y., Shi, C., Zhang, S.L., & Shi, Z.Z. (2006). Stock time series forecasting using support vector machines employing analyst recommendations. in: *Advances in Neural Networks - ISNN 2006* (ISNN 2006), Wang, J., Yi, Z., Zurada, J.M., Lu, BL., & Yin, H. (editors), Lecture Notes in Computer Science, Vol. 3973, Springer, Berlin — Heidelberg, DOI: 10.1007/11760191_66.