International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5569
***Article Received**: 05 May 2022 **Revised**: 30 May 2022 **Accepted**: 20 June 2022 **Publication**: 31 July 2022*

Robotic Process Automation for Stock Selection Process and Price Prediction Model using Machine Learning Techniques

Vinayak Jadkar
vinayak.21920003@viit.ac.in

Mayur Khandate
mayur.21920020@viit.ac.in

Vedant Gampawar
vedant.21810949@viit.ac.in

Pritesh Bhutada
pritesh.21810832@viit.ac.in

Prof. Leena Deshpande
leena.deshpande@viit.ac.in
Department of Computer Engineering,
Vishwakarma Institute of Information Technology,
Pune – 411048, Maharashtra (India)

Abstract— Among these last few years, we have seen a tremendous increase in the participation in financial markets as well as there are more robotic process automation jobs emerging in recent years. We can clearly see the scope and increased requirement in both these domains. In the stock market, predicting the stock prices/direction and making profits is the main goal whereas in rpa, tasks which are done on a regular basis are converted into automated or semi-automated form. In this paper we have tried to apply both things into the picture such as developing a price prediction model using machine learning techniques and automating the stock selecting process through technical screeners depending on user requirements. Stacked LSTM and Bi-directional LSTM ML techniques are used and for automation part powerful rpa tool Automation Anywhere has been used. Factors such as evaluation metrics and graph plots are compared for models and advantages, and disadvantages are discussed for using systems with RPA and without RPA practices. Price prediction plots have been analyzed for stocks of different sectors with highest market capitalization and results/analysis and inferences have been stated.

Keywords- Stacked LSTM, Bi-directional LSTM, Stock Market Prediction, Robotic Process Automation.

I. INTRODUCTION

Stock market

The Indian stock market is a platform where investors can buy and sell stocks. There are many stock exchanges in India [1], but only two are important. The old Bombay Stock Exchange (BSE) and the new National Stock Exchange (NSE) [2]. India's stock forecasts are very important because they are used by most ordinary people as well as businesspeople. People are more likely to lose or earn money from their lifetime savings on the stock market. The system is completely chaotic [3]. Price fluctuations depend on multiple factors such as government bonds, news, fundamentals, social media data, company production, historical prices, and the country's economy, making it difficult to create an accurate model. It is not advisable to consider only one factor in a predictive model. In that case, the result will not be accurate. Therefore, including multiple factors such as news, social media data, and historical prices can improve the accuracy of your model. In the olden days, buyers and sellers of all kinds gathered to make transactions, but now that IT is on track, almost all operations are done electronically, reducing paper usage. Investors no longer have to gather on the stock exchange and are free to trade from their homes or offices over the phone or the Internet [4].

II. LITERATURE SURVEY

In this project we have made RPA bots to overcome repetitive tasks which are hectic and tedious, and it is the main task if this fails due to human error it can lose our daily earning due to the wrong prediction. In this project we have learned the use of machine learning techniques, LSTM method and RPA bots for automating the selection of stock market in prediction.

Deep Learning

Deep learning is a subfield of machine learning and artificial intelligence (AI) that follows the way humans gain certain types of knowledge [5].

Deep learning getting to know is a completely critical detail of information science, which includes predictive modeling and statistics [6][7].

Robotic process automation

Robotic Process Automation (RPA), also known as robotics, uses automation technology to mimic back-office tasks performed by humans, such as retrieving data, filling out forms, and moving files. Integrate and perform repetitive tasks between enterprise and productivity applications by combining APIs and interactions with user interfaces (UIs) [8]. RPA is process-centric, and it solely replicates human-directed tasks. It is driven by the execution of a series of workflow tasks. RPA can be integrated with existing applications using several methods like Back-end RPA Integrations, which may include business use cases like Banking operation, finance domain, and many more. Front-end RPA Integrations, includes read, write data, and capture events directly from the user interface of a human-like host application Salesforce CRM, SAP etc. RPA is a code free, user friendly language with rich Analytics features [9].

III. PROPOSED WORK

The project we are working on is stock market forecasting. We have developed a bot that automates this entire process. Stock market forecast. Here, the bot automatically retrieves the required inventory data from the Yahoo Finance website. Use the stack LSTM model to predict future stock prices.

Therefore, to predict the stock price, use the following four-step model.

Data collection → Data preprocessing → Creating a stacked LSTM model → Plotting future trend and output

a. Data collection.

First, you need the data you want to work on to proceed with the task. Therefore, to collect the data, we will collect the data in Yahoo Finance through the Panda library. After collecting the data, save it in csv format. file format.

b. Data preprocessing.

After collecting and saving the data. The data needs to be trained (or) preprocessed. Data preprocessing is done in two steps

i. Train First, train your data. Use cross-validation with numbers generated by random seed values. To train the data in chronological format, the next data must depend on the previous data.
ii. Because the test data was trained in chronological format. Test your data using time steps. A time step is a number that a particular number is said to depend on the previous number.

c. Create a stacked LSTM model.

To create a stack of models, you need to convert the data to 3D. Then start building the stack using the sequential model.

d. Predicting the future trend and plotting the output.

To do the prediction, firstly we will check performance metrics. We will also be using inverse transformation to our data and later RMS performance metrics. Finally, we will be plotting a

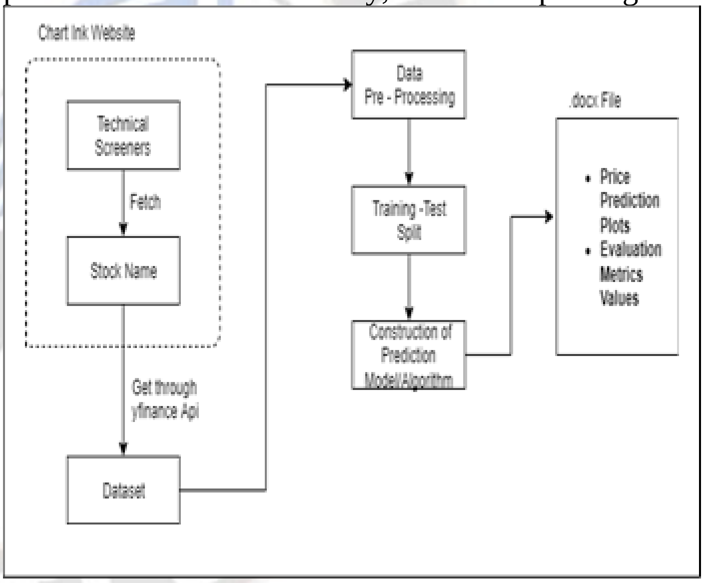

Figure 1. Proposed System Architecture

graph in which Blue represents the complete data set. Green represents our prediction for our tested data and Orange represents trained data. Robot process automation part We have developed a bot that automates this entire stock market forecasting process. Here, the bot automatically retrieves the required inventory data from the Yahoo Finance website and stores and processes the data as needed. Later, the code will run automatically and a chart showing the future of the stock will be drawn. This

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5569
***Article Received**: 05 May 2022 **Revised**: 30 May 2022 **Accepted**: 20 June 2022 **Publication**: 31 July 2022*

allows you to automate the entire process using a single bot.

Advantage through RPA:

Considering a scenario where uses technical screener (15 minutes Breakouts) which shows all the stocks which have given breakout in the last 15 min, now if the user want to run this model for the stock appearing at top of the list after every 15 minutes he will have to do the process manually by visiting the screener site every 15 minutes and then adding the name manually to the model and then the execution will start and generate report, but through RPA all of this manually work can be neglected and user could get his reports every 15 minutes without any human interaction manually. In Automation Anywhere [10], there is a control room which consists of scheduling the time of execution of a particular bot, so by these means we can schedule our bot to run after every 15 minutes and generate our system analysis automatically. You can schedule multiple bots which can run for different stock screeners simultaneously and produce results.

Long-term short-term model

LSTMs are very powerful against sequence prediction problems because they can store historical information [11]. In this case, this is important because the previous price of the stock is important in predicting future prices [12].

Bidirectional LSTM

Bi-Directional Long Short-Term Memory is a type of LSTM in which input flows in both directions and is Bi-Directional LSTM [13]. Unlike regular LSTMs, where the input flows in only one direction, Bi-directional allows the input to flow in both directions, retaining past and future information [14].

Evaluation Metrics

Evaluation of any machine learning model requires to satisfy technical as well as business point of view. For Time Series, the following metrics are suitable for our model:

Root Mean Square Error, Mean Absolute Percentage Error, R-Squared [15][16].

IV. RESULT ANALYSIS

Since the types of models we have used are dependent upon time series data which produces prediction results depending upon the previous patterns from sequence of variable length using recurrent neural networks. So, when we checked data for more than 2 years of period, we always saw that the predicted plot always used to be downwards now this scenario resulted because of unexpected conditions like occurrence of covid which resulted in major downfall trend.so we have chosen dataset for last 2 years so that we can neglect the unfamiliar situations and our model could predict data more accurately. After the development of model and training testing phase we have chosen RMSE, MAPE and R-squared metrics which works best for evaluation of model performance for time series-based data. Now the two factors that affect the outcome of data depend upon the number of epochs and the batch size used for developing the model. In table 1 we have taken the number of epochs as 100 and tested Evaluation metrics for both the models for stacked LSTM and Bi-Directional LSTM Using different batch sizes. Similarly, for table 2 we have used 50 epochs, observing both the tables 1 and 2 we can see that as we start reducing the batch size training time increases but all the metrics give better results.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5569
***Article Received**: 05 May 2022 **Revised**: 30 May 2022 **Accepted**: 20 June 2022 **Publication**: 31 July 2022*

Table I. Comparison of RMSE, MAPE and R-squared metrics for different batch sizes for 100 and 50 epochs for both stacked LSTM and Bi-directional LSTM model.

		LSTM				Bi-LSTM				
No. of Epochs	**Batch-size**	**Time (sec)**	**RMSE**	**MAPE**	**R-squared**	**Time (sec)**	**RMSE**	**MAPE**	**R-squared**	**Points (LSTM,BI-LSTM)**
100	64	49	48.008	4.012	0.459	84	23.743	4.173	0.641	(1,2)
100	32	57	11.005	3.401	0.585	89	85.582	4.099	0.419	(3,0)
100	16	94	57.421	3.703	0.479	121	29.140	2.996	0.650	(0,3)
100	8	166	38.538	2.663	0.733	204	6.919	3.030	0.635	(2,1)
100	4	301	30.707	2.476	0.748	392	12.346	2.239	0.7905	(0,3)
50	64	25	55.696	4.295	0.388	48	13.117	3.285	0.601	(0.3)
50	32	30	11.829	3.648	0.482	49	75.819	4.024	0.427	(3,0)
50	16	46	43.564	3.501	0.548	64	70.988	3.8490	0.456	(3,0)
50	8	82	24.416	2.930	0.690	107	41.027	3.139	0.617	(3,0)
50	4	153	14.547	2.059	0.819	192	15.935	2.743	0.712	(3,0)

As we are using the RPA approach to deduct the human intervention of selecting stocks manually which will reduce the human interaction and save a lot of time when we want to run the model multiple times for different stocks and generate reports. So time also plays a major role for our system, so observing both the tables 1 and 2 we see that there is no much of difference between 100 and 50 epochs rather we can see that for 50 epochs and 4 batch size the results are better compared to 100 epochs of 4 batch size in stacked LSTM and also the main advantage here is less time consumption for model which results in faster execution. No in all the cases Bi-directional LSTM works better than stacked LSTM, considering all the scenarios like evaluation metrics, time constraint and training testing plots we finalized that the best results are formed for Stacked LSTM of 50 epochs for 4 batch size as well as Bi-directional LSTM 100 epochs of 4 batch size.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5569
***Article Received**: 05 May 2022 **Revised**: 30 May 2022 **Accepted**: 20 June 2022 **Publication**: 31 July 2022*

V. TRAIN & TEST RESULTS

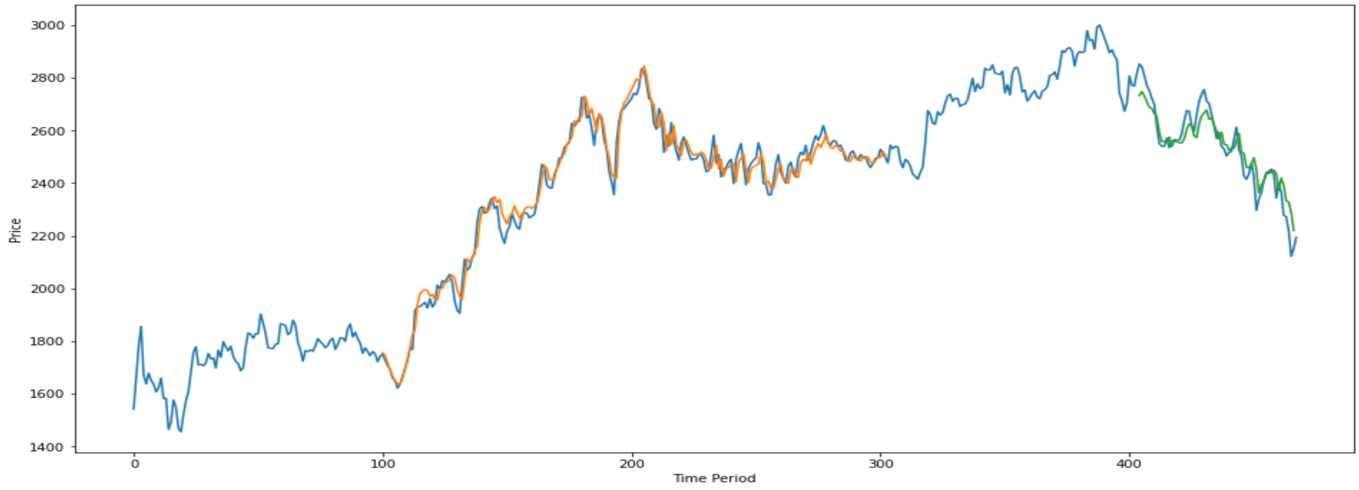

Figure 2. BI-LSTM: 100 Epochs 4 Batch-size

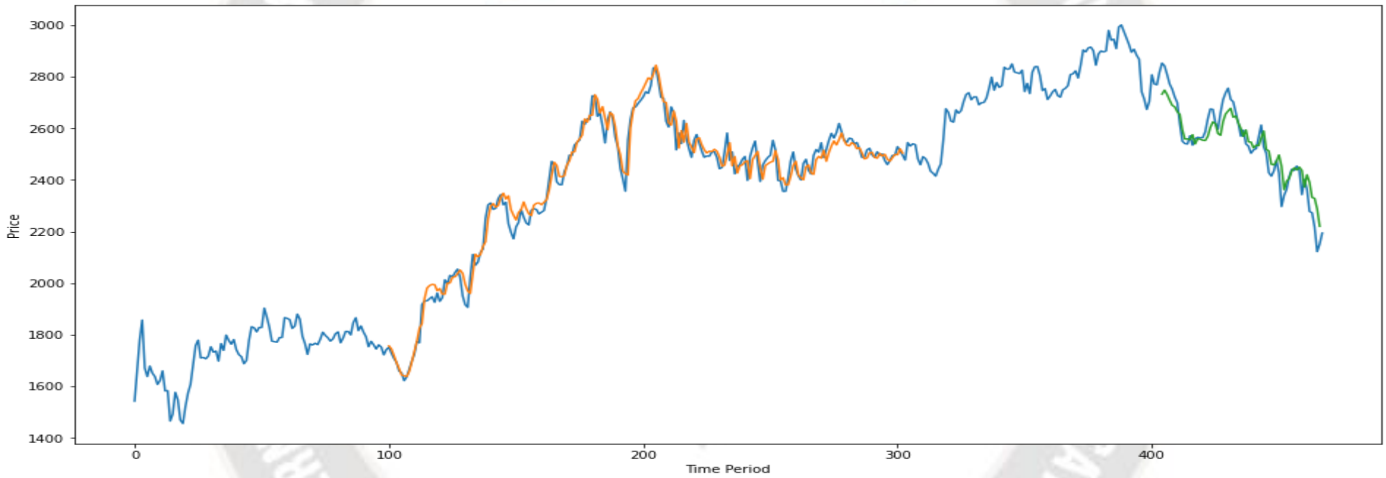

Figure 3. LSTM: 50 Epochs 4 Batch-size

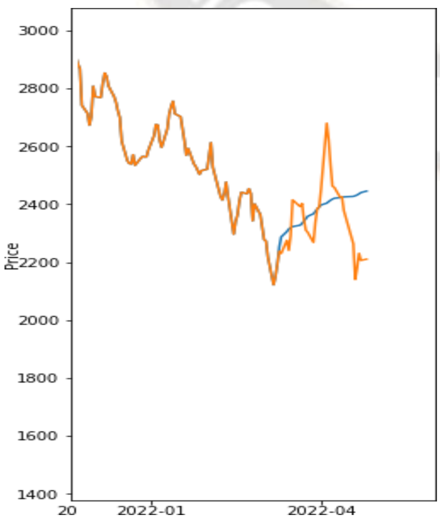

Figure 4. BI-LSTM: 100 Epochs 4 Batch-size

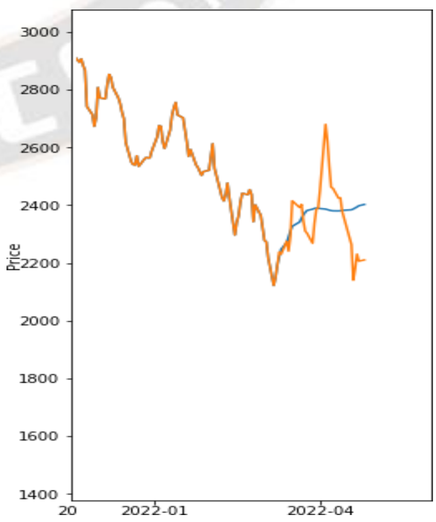

Figure 5. LSTM: 50 Epochs 4 Batch-size

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5569
***Article Received**: 05 May 2022 **Revised**: 30 May 2022 **Accepted**: 20 June 2022 **Publication**: 31 July 2022*

Table II. Comparison of RMSE, MAPE and R-squared metrics for different sector stocks with selected parameters for stacked LSTM and Bi-directional LSTM model.

	LSTM			BI-LSTM		
Sectors	**RMSE**	**MAPE**	**R-squared**	**RMSE**	**MAPE**	**R-squared**
TCS	103.265	3.132	-0.048	49.211	5.816	-0.964
HUL	10.810	1.855	0.746	4.876	2.164	0.638
HDFC	14.547	2.059	0.819	12.346	2.239	0.7905
MARUTI SUZUKI	101.346	2.835	0.717	68.24	2.808	0.727
SUN PHARMA	6.345	1.870	0.718	39.152	4.628	-0.305

Looking at the predicted graph plots and training testing plots for the values we have selected we can see that there is not much of a difference while comparing plots for both the models. So, further we did some more analysis for these models considering different stocks from different sectors which have the highest market cap in their particular domain and a comparison has been shown in table 3 of evaluation metrics for both the models.

Since all the valuation metrics play a major role in analyzing the performance of our model/system in some stock scenarios after training and testing we might not get good scores for our metrics, so it is advised to neglect such cases whenever you don't see good scores forming

VI. PREDICTED PLOTS

LSTM: 50 Epochs 4 Batch-size

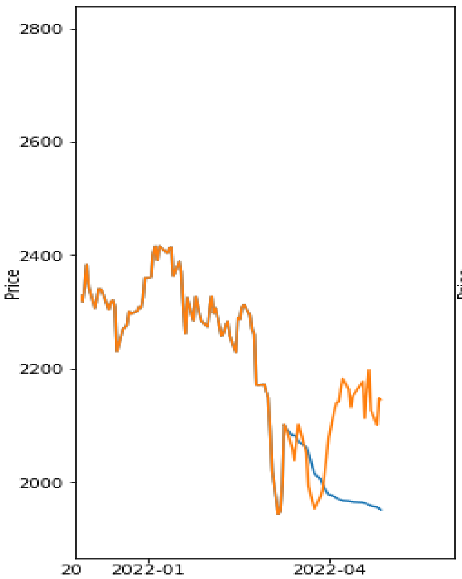

Figure 6. HUL

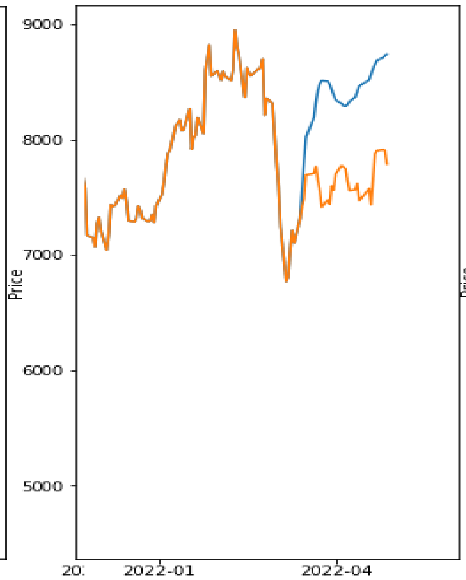

Figure 7. MARUTI

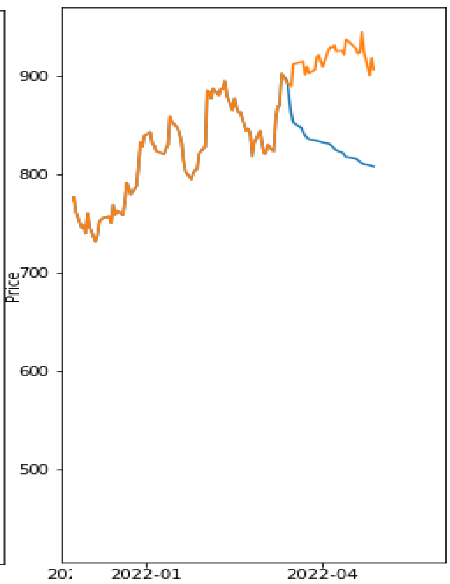

Figure 8. SUN-PHARMA

Upon observing all the predicted plots and evaluation metrics data we found that there is not much of a difference between both the models results for parameters we have chosen i.e. (50 epochs 4 batch size for stacked LSTM and 100 epochs 4 batch size for BI-LSTM) but still in majority cases 50 epochs LSTM have found out to be slightly more better than the 100 epochs, as well as 50 epochs will always take less time for execution process compared to 100 epochs which becomes advantage for our system. Predicted market plots are compared with actual market data that has been formed for the last 30 days. Since the data is dependent on time series it won't show you the exact same pattern, but you can rely on the direction of the pattern that has been forming through our model.

VII. INFERENCES

Considering a new breakout happening towards the up or downside it is advised not to believe that right away as sometimes it may be a false breakout. But you can reconsider the stock again for the next day or two and if the pattern follows the same path, you might make some good profits. In some scenarios you could also get early into a breakout trade following our model's pattern, but you have to maintain a proper risk management strategy in case the market conditions change and trade goes wrong. It is advised that before applying the trades, the user should have some basic knowledge regarding the stock market and technical analysis (support, resistance, all time high, etc.) Which will minimize the risk of choosing the right trades and taking action accordingly.

VIII. CONCLUSION AND FUTURE WORK

Predicting stock prices is a difficult task as there are too many factors involved which comes into play but still one can make good predictions through historical data and stacked LSTM and Bi-directional LSTM are best model for time series forecasting based data, we used RMSE, MAPE and R-squared as evaluation metrics for our model even with different sector of stocks and found that with 50 epochs and 4 batch size LSTM model was producing good results in most of the cases and was the fastest one taking minimal execution time out of all other parameter configurations. Robotic process automation has been very useful for our system when a user wants to run multiple stock screeners simultaneously or at given intervals for multiple times as well. This system can help users to generate analysis reports within minimal time usage without any manual process requirement which can save a lot of time and help users to make decisions more effectively. In future work we can analyze models for more stocks of various categories considering different stock screeners which are best suitable for the type of model along with more parameters such as volume of sock, volatility, etc., for better results.

IX. REFERENCES

[1]. Kumar Abhishek and Anshul Khairwa, Tej Pratap and Surya Prakash, "A stock market prediction model using artificial neural network", ICCCNT12, 26 th 28th July 2012, Coimbatore, India.

[2]. I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," Artificial Intelligence Review, 2019.

[3]. Deepak Mathur, N. K. V. . (2022). Analysis & Prediction of Road Accident Data for NH-19/44. International Journal on Recent Technologies in Mechanical and Electrical Engineering, 9(2), 13–33. https://doi.org/10.17762/ijrmee.v9i2.366 https://www.nseindia.com/ and https://www.bseindia.com/

[4]. N.N. Zhozhuashvili, " The role of informational technologies in modern business and financial sector", 12-14 Oct. 2011, DOI: 10.1109/ICAICT.2011.6110882

[5]. M. S. Kiran and P. Yunusova, "Tree-Seed Programming for Modelling of Turkey Electricity Energy Demand", Int J Intell Syst Appl Eng, vol. 10, no. 1, pp. 142–152, Mar. 2022.

[6]. G. A. H. Moghaddam, M. H. Moghaddam, & M. Esfandyari, "Stock market index prediction using artificial neural network," Journal of Economic

[7]. Shruti Goswami; Sonal Yadav, " Stock Market Prediction Using Deep Learning LSTM Model", 29-30 Oct. 2021, DOI: 10.1109/SMARTGENCON51891.2021.9645837

[8]. https://medium.com/@TalPerry/deep-learning-the-stock-market-df853d139e02

[9]. S. Aguirre and A. Rodriguez, ''Automation of a business process using robotic process automation (RPA): A case study,'' in Proc. Appl. Comput. Sci. Eng., 4th Workshop Eng. Appl. (WEA), 2017, pp. 65.

[10]. C. Osman, "Robotic Process Automation: Lessons Learned from Case Studies". Informatica Economica, 2019, vol. 23, pp. 66. doi: 10.12948/issn14531305/23.4.2019.06.

[11]. https://docs.automationanywhere.com/

[12]. J.Aruna Jasmine; S. Srinivasan; M. Godson; T.P. Rani; S.Susila Sakthy,"Share Market Prediction Using Long Short

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 7
DOI: https://doi.org/10.17762/ijritcc.v10i7.5569
***Article Received**: 05 May 2022 **Revised**: 30 May 2022 **Accepted**: 20 June 2022 **Publication**: 31 July 2022*

Term Memory and Artificial Neural Network",16-17 Dec. 2021, DOI: 10.1109/ICCCT53315.2021.9711849

[13]. https://towardsdatascience.com/predicting-stock-prices-using-a-keras-lstm-model-4225457f0233

[14]. Jaiwin Shah; Rishabh Jain; Vedant Jolly; Anand Godbole,"Stock Market Prediction using Bi-Directional LSTM", 25-27 June 2021, DOI: 10.1109/ICCICT50803.2021.9510147

[15]. https://towardsdatascience.com/lstm-and-bidirectional-lstm-for-regression-4fddf910c655

[16]. Khaled A. Althelaya; El-Sayed M. El-Alfy; Salahadin Mohammed,"Evaluation of bidirectional LSTM for short- and long-term stock market prediction", 3-5 April 2018, DOI: 10.1109/IACS.2018.8355458

[17]. https://medium.com/towards-data-science/recurrent-neural-network-to-predict-multivariate-commodity-prices-8a8202afd853