

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

Vinita Aghicha, Rutuja Shevante, Dipali Mane

Department of Computer Engineering, Zeal College of Engineering and Research, Pune, Maharashtra, India

## Article Info

Volume9, Issue 1

Page Number: 404-412

## Publication Issue

January-February-2022

## Article History

Accepted : 15 Feb 2022

Published :25 Feb 2022

## 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

India's financial system consists of capital markets and money markets. Capital markets provide for the conversion of small, scattered household savings into productive investments. Capital markets provide investors with investment security. Well-organized and regulated capital markets provide significant economic development room by making long-term funds available in exchange for financial security. The

Indian stock market is a platform where investors can buy and sell stocks.

A stock exchange is a mutual organization or company that provides both stock brokers and dealers with a "trading" function for trading stocks, other securities, and all other financial instruments. Securities traded on the exchange include stocks, investment trusts, derivatives, pooled investment products and bonds issued by companies. It should be noted that securities can only be traded if that particular stock is listed on a particular stock exchange. Normally, all records are kept in one

central location, but because modern markets are electronic networks, transactions are increasingly less tied to physical locations, gaining speed and transaction cost benefits. .. There are many stock exchanges in India, but only two are important. The old Bombay Stock Exchange (BSE) and the new National Stock Exchange (NSE).

India's stock forecasts are very important because they are used by most ordinary people as well as business people. People are more likely to lose or earn money from their lifetime savings on the stock market. The system is completely chaotic. 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 together to make transactions, but now that IT is on track, almost all operations are done electronically, reducing paper usage. I am. 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.

### 1) Role in the Financial Sector of Information Technology

Information generation performs a chief function in economic offerings for numerous reasons, however most significantly at the digital networks that alternate records associated with the finance sector. Today, economic offerings are all versions primarily based totally on growing a faster, greater green provider for customers. Since cloud-primarily based totally and new generation records are so regularly used, the significance of records generation is enormously critical.

In this twenty-first century, the commercial enterprise global is marked through drastic modifications. These modifications manifest through non-stop improvements in computer & telecommunication technologies. The preference of the perfect IT is a critical selection as it's far certain to have a protracted term & lasting effect at the destiny of the enterprise. Up-gradation of generation enables in growing productivity, decreasing cost & enhancing general quality. It is being helpful & has a first rate effect on commercial enterprise.

- IT can assist to discover the essential regions for aggressive gain enterprise of Business enterprise.
- Competitive blessings can be accomplished through numerous strategies in commercial enterprise with the assistance of IT.
- Helps in coping with the strategic alignment of essential commercial enterprise processes.
- Decision-making and operational manipulation through managers progressed through IT. · I can assist in preserving the conversion dating with customers, suppliers, trials, ability new entrants, and many others. IT in commercial enterprise outcomes in progressed verbal exchange, reduced costs, decreased selection making time, tracking the competition and higher manipulation of transactions.
- IT may be used as innovation withinside the functioning of the entire commercial enterprise gadget in the course of strategic commercial enterprise planning.
- IT enables a boom rate of change, decreasing paperwork & emergence of the worldwide economic gadget.

### 2) Information Technology (It) Shaping Indian Stock Market

Traditionally withinside the Indian inventory marketplace inventory buying and selling is carried out thru inventory agents, individually or thru telephones. As the wide variety of human beings buying and selling withinside the Indian inventory

marketplace has accelerated hastily withinside the previous couple of decades, a few problems like busy telecel smartphone lines, vicinity constraints, neglected verbal exchange and many others begin developing in Indian inventory dealer workplaces or organizations. Information generation (inventory marketplace software) enables inventory agents in fixing those troubles with on-line inventory buying and selling. It is an internet-primarily based totally inventory buying and selling facility. Investors or shareholders can change stocks through an internet site with no guide intervention from inventory agents. In this case, those on-line inventory buying and selling organizations are inventory agents for the investor.

## II. LITERATURE SURVEY

### Deep Learning

Artificial intelligence (AI) and Deep getting to know is a kind of system getting to know that follows the manner human beings benefit from information of sure types. Deep getting to know is a completely critical detail of information science, which includes predictive modeling and statistics.

Advantages	Disadvantages
1. The same neural network-based approach can be applied to many different applications and data types.	1. It is very costly to train due to very complex data models.
2. Deep learning is flexible to be adapted to new problems in the future.	2. No particular theory to guide you in selecting the right deep learning tools as it requires knowledge of topology, training method and other parameters.

### Robotic process automation

Robotic Process Automation (RPA) is a process automation technology that performs repetitive tasks based on the characteristics of software robots (bots) or artificial intelligence (AI) / digital workers.

Advantages	Disadvantages
1. No Coding Anyone doesn't require any programming or coding knowledge as RPA tools are used to automate applications in any department where the day to day office work is performed across an enterprise.	1. Potential Job Losses If a robot can work faster with a more consistent rate than human, then it is assumed that there will be no need for human input. It is the main concern for the employees, and this results in a major threat to the labor market.
2. User-Friendly Robotic Process Automation (RPA) doesn't require a special kind of knowledge, such as coding, or deep IT skills. RPA software is easy to understand and user-friendly, and very easy to use.	2. Initial Investment Cost RPA is still in the stage of innovation and so it can present challenges that may result in unwanted outcomes. So, it isn't easy for organizations to decide whether they should invest in rpa or wait until its expansion.

<b>3. Security</b> It provides options to assign role-based security capabilities to ensure action specific permissions. Further, the entire automated data, audits, and instructions which can be accessed by bots, are encrypted to avoid any malicious tampering.	<b>3. Hiring Skilled Staff</b> Many organizations believe that to work with RPA, the staff must have essential technical knowledge of automation as robots may require some programming skills and an awareness of operating RPA technology.
<b>4. Efficiency gain.</b> RPA can quickly complete tasks than humans, and it's able to do so at a much cheaper cost.	<b>4. Employee Resistance</b> Any change in the organization may cause stress to the employees because people are habitual.

### III. PROPOSED WORK

The project we are working on is stock market forecasting. We have developed a bot that automates this entire process. H. 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.

1. Data collection.
2. Data preprocessing.
3. Creating a stacked LSTM model.
4. Predicting the future and plotting the output.

#### 1. Data collection.

First, you need the data you want to work on to proceed with the task. Therefore, in order 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. Since LSTMs are

sensitive to data scales, we apply a minmax scalar in the range 0f (01) to the data.

#### 2. 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.

#### 3. 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.

#### 4. Predicting the future 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 our graph in which:

Blue represents the complete data set.

Green represents our prediction for our tested data

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

#### Stock Screener

Equity Screener is a set of tools that allow investors or traders to quickly screen a large number of available stocks and increase exchange-traded funds according to investor criteria.

There are two types of cleaners  
Technical Screener and Basic Screener

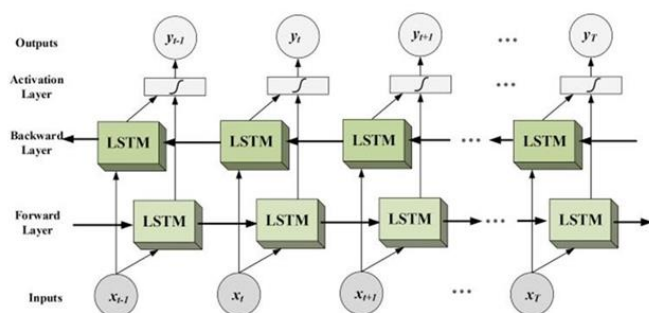
First, use a technical screener to access chartink.com. We selected a 15-minute breakout (live market) for the top stocks and later extracted the stock symbol from the RPA bot into the Jupyter notebook. I used Yahoo Finance's ticker to collect data about specific stock data for two years. We have acquired data for 2 years due to the higher accuracy of certain stocks compared to 5 years.

### Long-term short-term model

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

### Bidirectional LSTM

BIDirectional Long Short Term Memory is a type of LSTM in which input flows in both directions, and is BIDirectional LSTM. Unlike regular LSTMs, where the input flows in only one direction, BIDirectional allows the input to flow in both directions, retaining past and future information.



Description

### Evaluation Metrics

Evaluation of any machine learning model requires to satisfy technical as well as business point of view. If any model does not evaluate and directly launch that model it will affect the business if that product doesn't match to the required parameters. There are lots of evaluation metrics to be used to estimate the

accuracy of a prediction model. But, for Time Series, the following metrics are suitable for our model:

### Root Mean Squared Error (RMSE)

RMSE is a frequently used measure of the differences between the values predicted or estimated by the model. It is a way to measure the error of a model in predicting quantitative data.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$$

Here,

N stands for number of non missing data points

Y<sub>i</sub> stands for actual observation time series

^

Y<sub>i</sub> stands for estimated time series i stands for variable i

### Mean Absolute Percentage Error (MAPE)

MAPE is the ratio of the mean absolute difference between the predicted value and the true value divided by the true value. The true value is A<sub>t</sub> and the expected value is F<sub>t</sub>. The total number of values in the test set refers to the number (n). Using the denominator works better with non-zero and non-extreme data.

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

### R-Squared

This is a statistical measure that represents the percentage of variance of the dependent variable as described by the independent variable or the variable in the regression model. Here, SS<sub>tot</sub> shows the sum of the squares of the deviations from the sample mean of the dependent variable, and SS<sub>res</sub> shows the sum of the squares of the residuals from the expected value.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

#### IV. RESULT ANALYSIS

Since the types of models we have used are depend 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 choose 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 affects the outcome of data is depended upon the no of epochs and the batch size used for developing the model. In table 1 we have taken number of epochs as 100 and tested Evaluation metrics for both the models for stacked lstm and bidirectional 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.

**Table-1: Comparison of RMSE, MAPE and R-squared metrics for different batch sizes for 100 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	0.795	(0, 3)

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	0.795	(0, 3)

**Total = (6,9)**

**Table-2 : Comparison of RMSE, MAPE and R-squared metrics for different batch sizes for 100 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)
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)

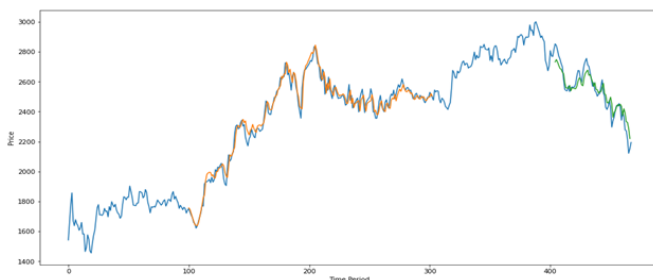


50	1	4	43.	3.5	0.	6	70.	3.	0.	(3,
	6	6	56	01	54	4	98	84	45	0)
			4		8		8	9	6	
								0		
50	8	8	24.	2.9	0.	1	41.	3.	0.	(3,
		2	41	30	69	0	02	13	61	0)
			6		0	7	7	9	7	
50	4	1	14.	2.0	0.	1	15.	2.	0.	(3,
		5	54	59	81	9	93	74	71	0)
		3	7		9	2	5	3	2	

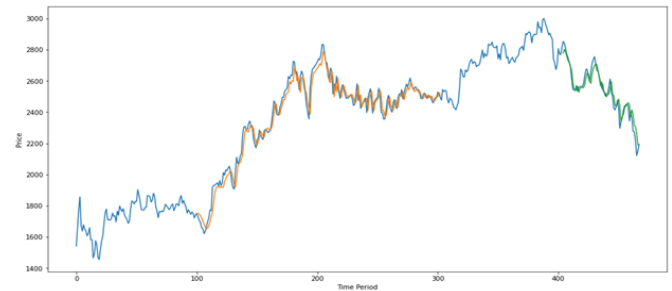
**Total** =(12,3)

Since we are using 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 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.

#### Train and Test Results :



**BI-LSTM : 100 Epochs 4 Batch-size**



**LSTM : 50 Epochs 4 Batch-size**

#### Predicted Plots:



**BI-LSTM : 100 Epochs 4 Batch-size**



**LSTM : 50 Epochs 4 Batch-size**

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.

**Table-3 : Comparison of RMSE,MAPE and R-squared metrics for different sector stocks with selected parameters for stacked LSTM and Bi-directional LSTM model.**

**LSTMBI-LSTM**

Sectors	RMS E	MA PE	R- suar ed	RM SE	MA PE	R- suar ed
TCS	103.2 65	3.13 2	- 0.048	49.2 11	5.81 6	- 0.964
HUL	10.81 0	1.85 5	0.746	4.87 6	2.16 4	0.638
HDFC	14.54 7	2.05 9	0.819	12.3 46	2.23 9	0.790 5
MARU TI SUZU KI	101.3 46	2.83 5	0.717	68.2 4	2.80 8	0.727
SUN PHAR MA	6.345	1.87 0	0.718	39.1 52	4.62 8	- 0.305

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.

**Predicted Plots:**



**BI-LSTM\_HDFC : 100 Epochs 4 Batch-size**



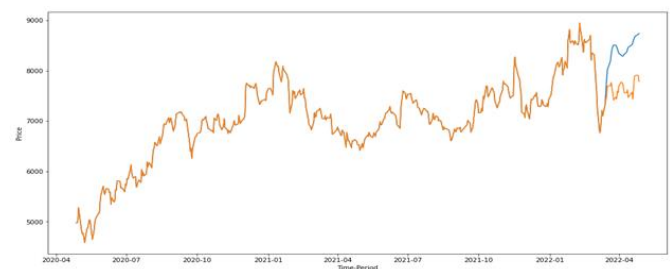
**LSTM\_HDFC : 50 Epochs 4 Batch-size**



**LSTM\_HUL : 50 Epochs 4 Batch-size**



**BI-LSTM\_HUL: 100 Epochs 4 Batch-size**



**LSTM\_MARUTI : 50 Epochs 4 Batch-size**



**LSTM\_SUNPHARMA : 50 Epochs 4 Batch-size**



Upon observing all the predicted plots and data we can 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 wont show you the exact same pattern but you can rely on the direction of the pattern that is been forming through our model.

### Inferences

1. Considering a new breakout happening towards the up or down side 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 next day or two and if the pattern follows the same path you might make some good profits.
2. In some scenarios you could also get early into a breakout trade following our models pattern but you have to maintain a proper risk management strategy in case the market conditions change and trade goes wrong.
3. 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.

### 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 there is a control room which consists features of scheduling the time of execution of particular bot, so by these means we can schedule our bot to run after every 15 minutes and generate our system analysis automatically.Even you can schedule multiple bots which can run for different stock screeners simultaneously and produce results.

## V. 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 a given intervals for multiple times as well.This system can help users to generates analysis reports within minimal time usage without any manual process requirement which can save a lot of time and help users to take decisions more effectively. In future work we can analyze model for more stocks of various categories considering different stock screeners which are best suitable for the type of for particular model along with more parameters such as volume of sock, volatility, etc., for better results.