

Stock market prediction using Deep Learning

*Patel Aatish, #Dave Parth, \$Patel Aditya

*,#,\$UG Student, A.D. Patel Institute of Technology, Karamsad, Gujarat -India,

*aatishpatel8456@gmail.com, #daveparthn17@gmail.com, \$adityamgopi@gmail.com

Abstract- Stock Markets are very volatile and extremely difficult to predict, the reason for their unpredictable behavior is that they rely on a huge number of variables so that they create a butterfly effect meaning just a small scale change can have a drastic impact on the values of the stocks. Conventionally manual mathematics and feature extraction is used predict the trends in the stock market but that is a quite limiting approach with good but not reliable results. In this paper we use "Deep Learning" technique to predict the overall behavior of the stock market, here none of the features are extracted manually. Meaning all of the "feature Extraction" is done by the "Deep Learning" model. By taking "Deep Learning" approach we are able to predict the stock values quit successfully.

Keywords- Neural network, Deep learning, Back propagation, Stock market, TensorFlow, Google colab , LSTM

I. INTRODUCTION

Stock market prediction is largely outlined as attempting to see the stock price and supply a strong plan for the folks to grasp and predict the market and also the stock worth's. On earlier days investors and analyser do stock price prediction manually by using mathematical calculation and also some other aspect as well. But now due to emerging technology of Artificial intelligent, which allow us to make prediction of company's stock price not only accurately but also rapidly. we generally tend to thinking about closer to the observe of method studying with numerous datasets integration to expect the marketplace and the stock trends. the matter with estimating the stock price can stay a haul if an improved stock market prediction algorithmic rule isn't proposed. However, predicting stock market behaviour is not easy. The movement within the stock market is typically determined by the feelings of thousands of investors. exchange prediction, it can even be a global event like sharp movements in currencies and goods and so forth of these events have an effect on the company earnings, that successively affects the sentiment of investors. of these factors create stock worth prediction terribly difficult. Once the correct knowledge is collected, it then is often used to train a machine and to get a prognosticative result.

Over the years, numerous researchers use machine learning method in order to find optimal technique for stock market prediction. But now due to vast amount of data deep learning models are used for the prediction, which has its own advantage of accuracy and rapidness over the machine learning model. Most popular among all deep learning model Long-Term-Short-term memory (LSTM) is very popular. Long-short-Term Memory (LSTM) is part of deep learning cluster.

II. LITERATURE REVIEW

Paper 1: Stock market prediction technique

Summary-

Generally, data mining in semantic analysis are used to predict D tendencies of the stock market. We all know that stock market is going by the emotion of the peoples so opinion mining is a big area of interest and research in today's day in age. Is more and more opinion-based data increases the accuracy of the prediction and semantic analysis also increases with it. Therefore, to handle such enormous amount of data various semantic analysis techniques are being developed. Heuristic statical and probabilistic approaches are used in this area. Even though enormous amounts of efforts are put into this research areas these techniques have a huge amount of limitations. Here is conclusion after analysing the research paper is that predicting the stock prices is extremely difficult task and requires a lot more knowledge not only about the mathematics but also human behaviour and statistics. And how All this thing can interact with each other to have drastic effects on stock market

Paper 2: Stock Closing Price Prediction using Machine Learning Techniques

Summary-

Accurate prediction of securities market returns could be a terribly difficult task thanks to volatile and non-linear nature of the monetary stock markets. With the introduction of computer science and redoubled procedure capabilities, programmed ways of pre-diction have established to be additional economical in predicting stock costs. Predicting company's stock value may be a difficult task because of systematically dynamic stock values that area unit smitten by multiple parameters. The historical dataset obtainable on company's web site consists of solely few options like high, low, open, close, adjacent shut price of stock costs, volume of shares listed etc., that don't seem to be decent enough. To

get higher accuracy within the foreseen worth price new variables are created by the prevailing variables. So new variable plays significant role in stock market prediction such as volume, which is essential variable for price prediction and apart from variables nowadays most researchers use LSTM model for prediction because LSTM have superior and unique feature of supervised learning. but on the other side LSTM also have some drawback. so instead of using LSTM we can use hybrid model of LSTM and SVM.

III. PROBLEM STATEMENT

Interior Financial analysts investing in stock market usually are not aware of the stock exchange behaviour. They are facing the problem of trading as they do not properly understand which stocks to buy or which stocks to sell for the purpose of get more profits. In today's world, all the information pertaining to stock retail is not available and data is key element to predict stock prices and more data gives more accurate result. So, for the prediction of stock price it's the mandatory that we should have at least 8 year of company's historical database. Data mining is necessary in order to get previous stock database. Once database found and received then still data is not always in acceptable form. it contains missing values, noise, So data pre-processing is required for data cleaning process, which is very time consuming process. After the data cleaning process data is given to deep learning model, where different deep learning model with different number of neurons are used for prediction and it is difficult to tell which model is suitable and gives optimal stock's price prediction. Analysing all this database individually or manually is tremendously difficult. Investigating trends in stock market and factors affecting the stock prices. Stock market is highly volatile. At the most fundamental level, it's far stated that supply and demand in the market determines stock price. But it does not follow any fixed pattern and is also affected by a huge number of highly varying factors. Graphical user interface is one of the vital elements of any stock market prediction application. Because stock market is kind of convoluted system. So, interface should be user friendly and easy to understand, which is challenging.

IV. PROPOSED SYSTEM

The main aim of this project is to develop a stock market prediction application which gives benefits to stock broker, investor, analyser and researchers. our main aim is building a software which gives accurate result of future stock. Stock market application will not only give accurate result but also its provide user friendly interface, Because along with accuracy user experience is highly important and nowadays all kind of people interested to investing into stock market, so it is important that all user gets eject and tractable evaluated information.

User can access application from anywhere, from laptop to mobile and from web interface to software interface. At the

first user have to create their account by using their Gmail id or phone number. By doing this, user can maintain their record of past evaluation, their notes and selling and buying records of stock. once they logged in then one animated video will appear on the screen, which will guide to user that how to use this application in better way and also offer assistance. In the main screen user will have one search bar, in that they can start prediction on according to their preferable company and date.

Transparency is main motto, so user can see carry, error and variation on each prediction after the actual stock price revile, as a result user starting to believe on the software. Also, the new feature of scale zooming will be added, shows more detailed version of prediction result by using scale zooming, so by this feature viewer can privileged particular period of prediction in more detail.

V. METHODS/ TECHNOLOGY

Methodology

Various varieties of neural networks are often developed by the mix of completely different factors like network topology, coaching technique etc.

This section we will discuss the methodology of our system. Our system consists of several stages which are as follows:

Step – 1: Raw Data:

In this stage, the historical stock data is collected from <https://www.Yahoo.Finance.com/data/NSE> and this historical data is very essential for predicting stock price

Step – 2: Data Pre-processing:

The pre-processing stage involves

- information discretization: half of data reduction but with explicit importance, particularly for numerical data
- information transformation: Normalization
- information cleaning: Remove noise and fill missing value.
- Data integration: Integration of information files

when the dataset become clean dataset then, the dataset is divided into training and testing set. Where more recent values are taken as the training values. Testing data take as 5 to 10% of the total dataset

Step – 3: Feature Extraction:

In this layer, different type of feature extracted from the database and given to neural network, in our model we have volume, close-high, etc.

Step – 4: Training Neural Network:

In this stage, the data is fed to the neural network and trained for prediction assign random biases and weight. Our LSTM model is composed of a sequential input layer followed by

different combination of LSTM layers and dense with RELU activation and then finally a dense output with linear activation function.

Technology stack

1. Python

(programming language used to code the whole deep learning model, python has the biggest support of public libraries which are really useful to create deep learning models and tuned hyperparameters for the model architecture)

2. Google Collab

(it's a Jupiter notebook in the cloud which is hosted on Google's server that has all the required libraries pre-installed, you just need to import the libraries that you want to use such as NumPy or pandas and you don't need to install it because it is already in the system of the cloud)

3. Jupiter Notebook

(there interactive python notebooks used for testing, meaning you don't need to compile your code every single time and the environment variables and your variables are already in the system so whenever you want to run your code it is all done dynamically so that it feels much for interactive)

4. TensorFlow

(a high level library to create detailed deep learning model architecture, using TensorFlow you don't need to understand the low level mathematics which is used for the planets arches backpropagation or matrix multiplication, TensorFlow takes care of this so that you just need to defined how many neurones you want in a single layer and how many layers what and by doing so TensorFlow will handle all the rest)

5. Keras

(A wrapper around TensorFlow to increase the abstraction, meaning it is even more simpler than TensorFlow because you just to define what kind of model you want and keras will handle all the rest of it all the best practices and best architecture that is used for a particular model keras will use that)

6. NumPy

(mathematics library for matrix multiplication, NumPy uses vectorisation for multiplication which is much faster than normal multiplication done with a for loop. which results in to faster runtime for your code and no time conception and low time complexity)

7. Matplot lib, Seaborn & Plotly

And these three libraries for data visualisation (it is really hard for humans to understand large number of data sets Surat is better locked seaborn and pottery libraries are used

to make sense of the data by visualising them in different manners such as pie chart bar chart chloroplasts graph Etc, we use Plotly and Matplotlib to see loss over time when we train a neural network model and finally, we plot them in a line chart to see over time how are model was becoming more accurate in our prediction)

VI SYSTEM ARCHITECTURE

System and LSTM model architecture

shows basic system architecture, At the first collect a raw database of any company that we want to perform prediction on it. After that perform feature extraction for finding more variables, divide this data in two parts testing and training, so based on the trained database it will predict stock price of respected company.

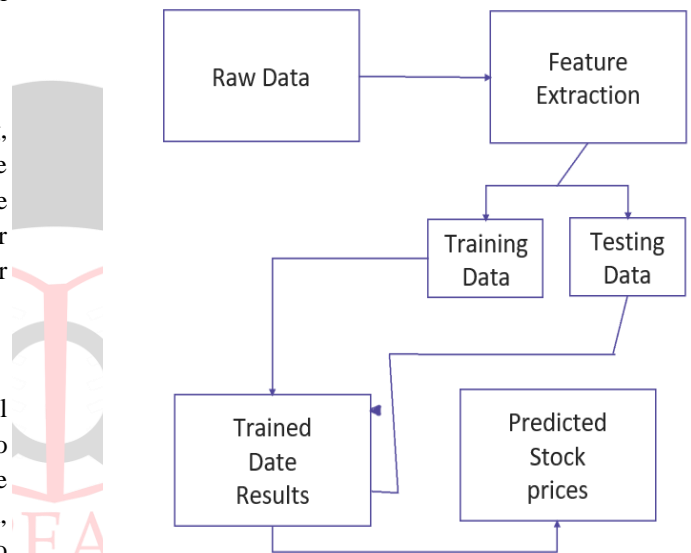


Figure 1 system architecture

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 60, 20)	1760
lstm_2 (LSTM)	(None, 20)	3280
dense_2 (Dense)	(None, 5)	105
dense_3 (Dense)	(None, 1)	6
Total params: 5,151		
Trainable params: 5,151		
Non-trainable params: 0		

Figure 2 LSTM model architecture

Algorithm structure:

1. How to collate input data:

As mention in flowchart that read stock data so for that model needs a vast particular company's data. So now question is from where we get the database?

Input data is collect from Yahoo Finance the use of following steps:

1. For our project, we are considering multiple company's data which are belongs from different in market
2. Use unique symbol character of company to get data from Yahoo Finance.
3. System will take last 8 year's stock data of the company.
4. Further we divide the data into parts, training data and testing data, where 80% of the data will used for training and 20% of the data will used for testing

2. Plotting close price over the years

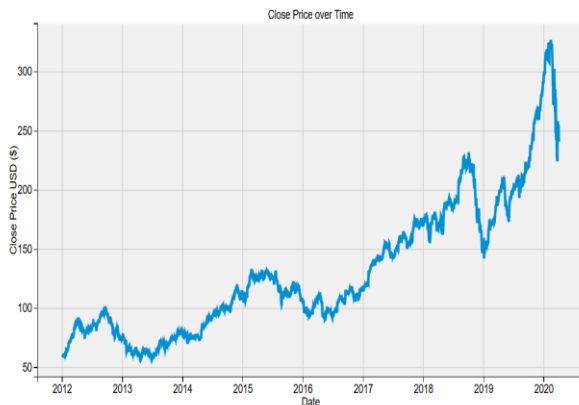


Figure 3

Figure 3 show how plotting of closing price because we take 'close' Colum for prediction

3. Data scaling

This way that you are remodelling your statistics in order that it suits inside a particular scale, like 0-a hundred or 0-1 Deep learning algorithm is very sensitive, if scaling is not done then Deep learning algorithm tend to take higher value, consider as best value, where it's ignore lower value.

4. Testing & Training dataset

In order to train the model, which can perform various computation, training dataset is used.

The training data is use for ensure the system acknowledges styles with inside the information, the cross-validation information is required to make sure higher accuracy and performance of the set of rules used to educate the system, and the test data into look how nicely the system can expect new solutions primarily based totally on its training.

5. Model building & Training

A machine learning prototype is mathematical delineation of a real-world process. The machine learning method will identify hide information pattern within the dataset correspond to target during training process. After the training process, develop machine learning model could use for stock market prediction.

Machine learning consists of algorithms that can automate analytical model building. Using algorithms that iteratively analyse from data, device getting to knows trends and facilitate computer systems to discover hidden insights from Big Data.

Training a model simply means learning (determining) good values for all the weights and the bias from labelled examples. In supervised learning, a machine learning use set of rules to builds a version through inspecting many examples and looking for a version that minimizes loss; this process is called empirical risk minimization.

6. LONG-SHORT TERM MEMORY

For the stock prediction we required a RNN model which is well suited for our requirement and condition and LSTM model is best option for us among all other model because LSTM are very effective in series prediction issues due to the fact, they're capable of save previous information. This is essential in our case due to the fact the preceding rate of a company's stock is important for predicting future price of stock

7. BUILDING LSTM

For the purpose of building LSTM, we have to import multiple modules from KERAS:

1. In order to initializing neural network, add sequence layer
2. Dense layer for inserting a densely connected neural network layer
3. LSTM for inserting the Long Short-Term Memory layer

Model selection

In this project we use LSTM model for the prediction but later on we found that changing layers and neurone in the model gives us some constructive advantage, so what's makes model more beneficial by changing layers and neurone? So which model will gives best prediction is evaluated on three factors RMSE, delta density, worst case and best case.

Following are different LSTM model by modifying their hyper parameters

- 1 Model 1 (Mono-LSTM 10, Hyper parameter)
One LSTM layer each contain 10 neurons
- 2 Model 2 (Mono-LSTM 20, Hyper parameter)
One LSTM layer each contain 20 neurons
- 3 Model 3 (Dual-LSTM 20, Hyper parameters)
Two LSTM layer each contain 20 neurons
- 4 Model 4 (Tri-LSTM 10, Hyper parameters)
Three LSTM layer each contain 10 neurons
- 5 Model 5 (Tri-LSTM 20, Hyper parameters)
Three LSTM layer each contain 20 neurons
- 6 Model 6 (Tri-LSTM 30, Hyper parameters)
Three LSTM layer each contain 30 neurons
- 7 Model 7 (Quad-LSTM 10, Hyper parameter)
Four LSTM layer each contain 10 neurons
- 8 Model 8 (Quad-LSTM 20, Hyper parameters)
Four LSTM layer each contain 20 neurons

We perform prediction by using above various LSTM model on single company in order to get best model, after that we evaluate the result, analysis the output and go

through a particular procedure for accomplishing best and less time-consuming model by root mean square error, Histogram, Density graph. Below mention table 4.1 give more information for model selection.

Model	Number of LSTM layer	Neurons in LSTM	Total Neurons	Neurons in dense layer	RMSE
Model 1	1	10	10	6	18.7072
Model 2	1	20	20	6	11.4665
Model 3	2	20	40	6	11.0582
Model 4	3	10	30	16	28.4152
Model 5	3	20	60	16	12.9471
Model 6	3	30	90	16	13.3540
Model 7	4	10	40	36	20.3652
Model 8	4	20	80	36	38.2821

Table 4

Above Table address us to model's LSTM layers, Neurons in LSTM, Total neuros and RMSE.

RSME indicate difference between actual value and predicated value, it is shows result in term of error. Root mean square error is very vital to check software, model working perfectly or not. Lower RMSE value means model is performing well, model can successfully predict stock price and higher value shows high level of fault; So according to above information it is indicate that model no 3 show less RMSE value compare to others and model no 8 has highest values.

Delta & Delta density

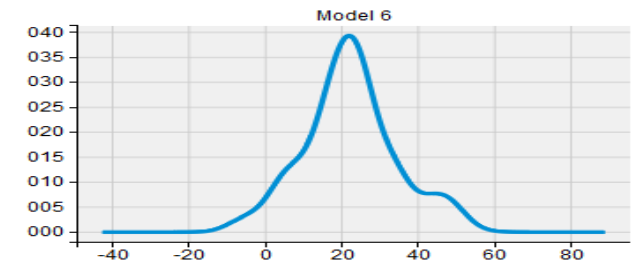
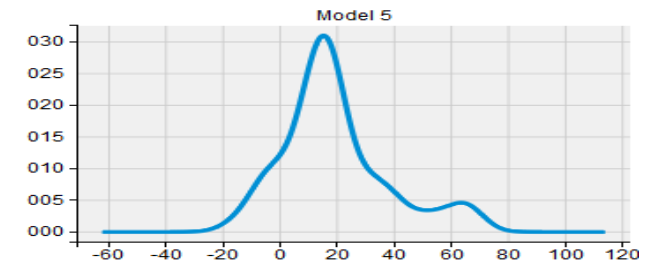
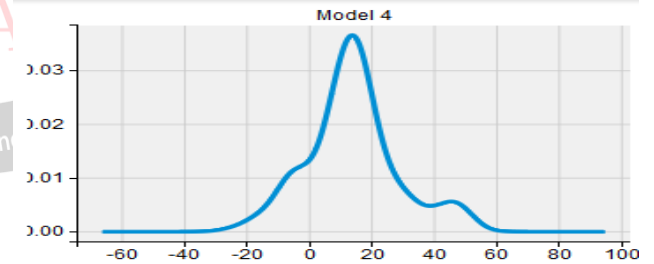
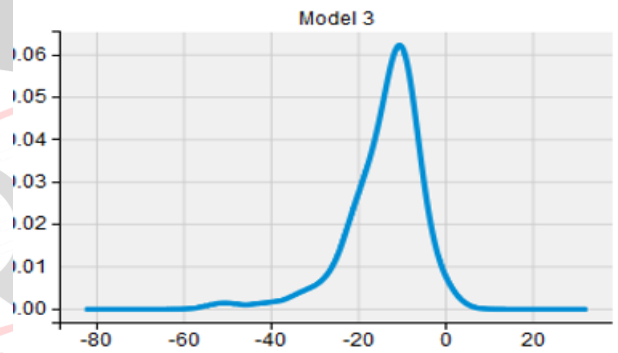
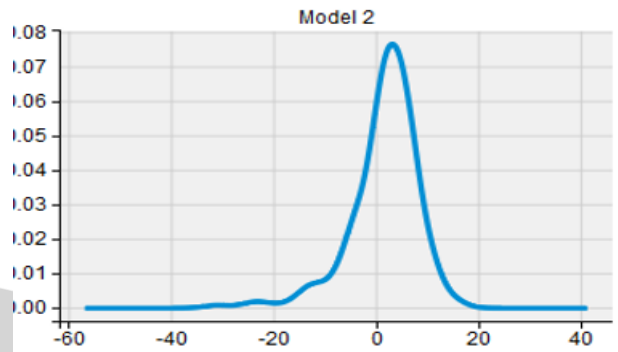
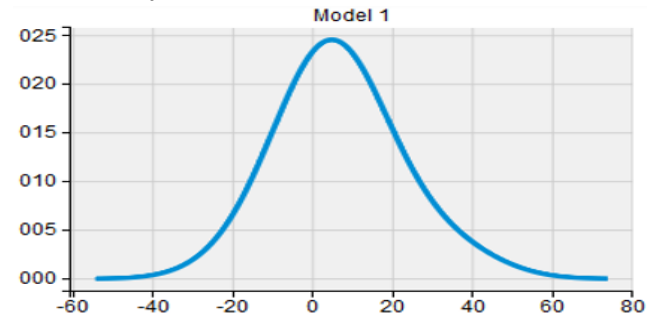
In our case delta indicates a different between actual value and predicted value. So as per delta analysis we perceive that model can work invariably in each condition or it can fluctuate over the time. On the next page figure show delta value of result.

Date	
2018-08-08	3.776047
2018-08-09	4.495178
2018-08-10	2.064163
2018-08-13	2.933975
2018-08-14	3.150452
...	...
2020-03-26	16.112549
2020-03-27	1.683136
2020-03-30	8.399490
2020-03-31	6.089355
2020-04-01	-8.604233

415 rows x 1 columns

Above Figure shows delta of predicted value like on date 2018-08-09 different between actual value and predicted value is +3.776047

Delta density of different model



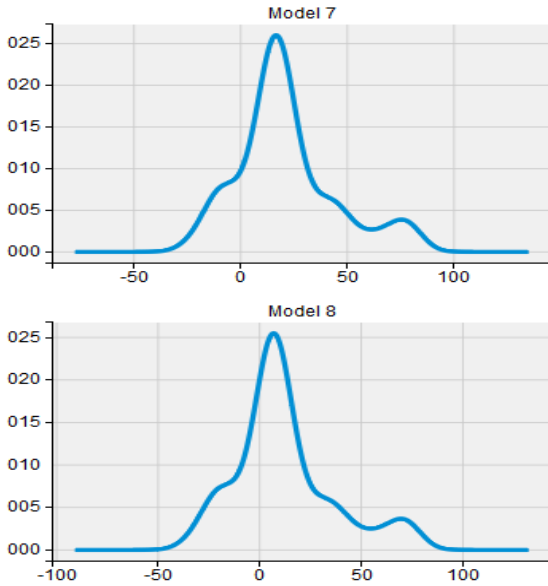


Figure 5 Delta density

So according to delta density graph, model no 3 is better from other we can see 90% of the Delta lies within the range of [-30, 03], so even when market is volatile our model is quit handsomely able to predict the up-coming trends.

Worst case & Best case

We also found worst-case and best-case in order to know performance of each model and find out best prediction model.

Best case shows minimum or null different between predicated and actual value and worst case shows maximum different between predicated and actual value.

Model 1	Delta	Model 5	Delta
Date		Date	
2020-01-13	41.700562	2020-02-12	69.850861
	Delta		Delta
Date		Date	
2020-03-23	-22.276459	2018-12-24	-18.120895

Model 2	Delta	Model 6	Delta
Date		Date	
2020-01-13	16.685272	2020-01-13	56.017456
	Delta		Delta
Date		Date	
2020-03-12	-32.328415	2020-03-23	-9.691467

Model 3	Delta	Model 7	Delta
Date		Date	
2020-01-13	3.670288	2020-01-29	81.879745
	Delta		Delta
Date		Date	
2020-03-12	-53.831039	2018-12-24	-24.301239

Model 4	Delta	Model 8	Delta
Date		Date	
2020-01-13	54.430878	2020-02-12	76.648071
	Delta		Delta
Date		Date	
2020-03-23	-25.99057	2018-12-24	-34.868332

Figure 6 Best & worst case of each model

VI. RESULTS

Hence above all evidence prove that model number 3(2 LSTM layers and 20 neurons each) is best model for prediction among all model so we finalized model 3 for out stock market project



Figure 7 prediction

In above graph shows prediction of stock market, it is Apple stock price to time graph, In this graph red line represent actual stock price and blue line represent predicted Apple stock price. We can see that prediction is so accurate.

Date	Close	Predictions
2018-08-08	207.250000	197.557892
2018-08-09	208.880005	198.998550
2018-08-10	207.529999	200.309174
2018-08-13	208.869995	201.323029
2018-08-14	209.750000	202.214310
...
2020-03-26	258.440002	242.161270
2020-03-27	247.740005	242.282135
2020-03-30	254.809998	242.311935
2020-03-31	254.289993	242.948456
2020-04-01	240.910004	243.730621

Figure 8

Figure 8 indicate difference between actual close price and predicated value

For the purpose of checking that this model is working well in every situation, We selected stocks of 15 companies to train and test the system. Eight years of data is downloaded from Yahoo Finance, of which 80% is used to train the system and the remaining 20% is used for testing

Following are a Stock prediction of different company by using model no 3.

Tata motors;

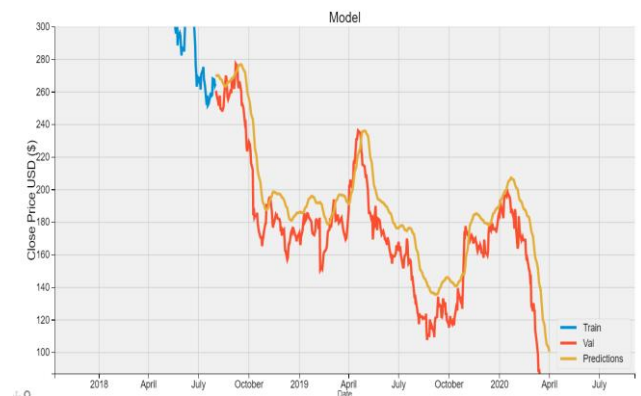


Figure 9 Result 1

SBI bank:



Figure 10 Result 2

ITC company:

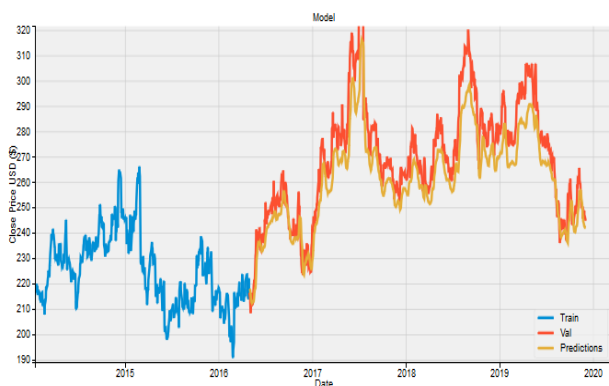


Figure 11 Result 3

Airtel:



Figure 12 Result 4

VII. ADVANTAGES

1. Biases:

Transparent, honestly is biggest issue in stock market, many traders want you to buy stocks because they want to influences their targeted stocks prices, however in the case of stock market prediction software investor directly dealing with stocks
2. No third-party advisor:

According to survey, 93% local stock market investor hires advisor to take advise, whether they buy or sell their stocks.

3. Tractable:

User interface of prediction software is easy to understand. So, it's does not matter how qualified person use prediction software, they all can understand and use effectively.
4. Scheduling:

Software program allows you conduct new task by relief from few responsibilities that may be effortlessly scheduled without fail. With those schedules you're capable of provoke responsibilities with none delays, consequently undertaking your evaluation on time

VIII. LIMITATIONS AND FUTURE WORK

The limitations in this model that we considered is that it requires a lot of data to train and to predict stock prices for a given stock. We would like to add that this is the biggest limitation, take a requirement of data for a model of this small-scale is really large in order to a improve the accuracy of the model we would like to have "One-Shot Learning" or release mall Data sets to extract information and extract the feature of even stock market.

So, in future a we would like to crate the model which can learn really fast from a limited or a quit small set of data. Right now, this model requires 2 to 3 years of past data not to predict "Stock Prices" accurately but in future would like to build a model architecture such that it can predict the accurate stock price with limited amount of data let's say data of 30 to 60 days.

IX. CONCLUSION

The popularity of stock market prediction is growing speedily, that is encouraging researchers to seek out new ways for the prediction by using new techniques. The prediction technique isn't solely serving to the researchers, it also additionally helps investors and someone addressing the stock market. so as stock price is highly volatile, a prediction model with sensible accuracy is needed. during this work, we've got used one among the foremost precise prediction technology of recurrent Neural Network and Long short-term memory that helps investors, analysts or someone fascinated by finance within the share market by providing them a decent data of the long run scenario of the stock market.

This project surveyed the applying of neural networks to financial systems. However, it has many limitations like artificial intelligent technique not always works during unforeseen events. And on the other hand, we cannot always find enough company's historical data, so in that case prediction cannot be possible. Nevertheless, it has certain loophole, which have to solve, but it also offers estimate of future stock price, which is very essential information for investor.

performance is not always satisfactory. Back LSTM algorithm is not only algorithm to be used in Feed forward

neural network, but also can predict stock market price with the use of RNN, ANN, SVM algorithm, LSTM by making hybrid model, and we also planning to make our model functional even in unforeseen situation.

X. REFERENCES

- [1] T. Finley and T. Joachims, "Training structural SVMs when exact inference is intractable," in 25th international conference on Machine learning, Helsinki, Finland, 2018
- [2] Y. S. X. Z. Y. Ding, "Forecasting financial condition of chinese listed companies based on support vector machine," *Expert Syst. Appl.*, vol. 34, no. 4, pp. 3081-3089, 20018.
- [3] Milosevic, Nikola. "Equity forecast: Predicting long term stock price movement using machine learning."arXivpreprint arXiv:1603.00751 (2016)
- [4] Shen, Shunrong, Haomiao Jiang, and Tongda Zhang."Stock market forecasting using machine learning algorithms. "Department of Electrical Engineering, Stanford University, Stanford, CA (2012):
- [5] Leung, Carson Kai-Sang, Richard Kyle MacKinnon, and Yang Wang."A machine learning approach for stock priceprediction."Proceedings of the 18th International Database Engineering & Applications Symposium. ACM, 2017
- [6] Chen, Kai, Yi Zhou, and Fangyan Dai. "A LSTM-based method for stock returns prediction: A case study of Chinastock market."Big Data (Big Data), 2015 IEEE International Conference on. IEEE, 2015
- [7] Lin X, Yang Z, Song Y. Expert systems with applications short-term stock price prediction based on echo state networks. *Expert Syst Appl.* 2009;36(3):7313-7. <https://doi.org/10.1016/j.eswa.2008.09.049>.
- [8] Idrees SM, Alam MA, Agarwal P. A prediction approach for stock market volatility based on time series data. *IEEE Access.*2019;7:17287-98. <https://doi.org/10.1109/ACCESS.2019.2895252>.
- [9] Kim K, Han I. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Syst Appl.* 2000;19:125-32.
- [10]Shen J. Thesis, "Short-term stock market price trend prediction using a customized deep learning system", supervised by M. Omair Shafiq, Carleton University. 2019.