NISHITHA, S.N.T., BANO, S., REDDY, G.G., ARJA, P. and NIHARIKA, G.L. 2020. Stock price prognosticator using machine learning techniques. In *Proceedings of the 4th International conference on electronics, communication and aerospace technology (ICECA 2020)*, 5-7 November 2020, Coimbatore, India. Piscataway: IEEE [online], pages 1636-1642. Available from: https://doi.org/10.1109/ICECA49313.2020.9297644

Stock price prognosticator using machine learning techniques.

NISHITHA, S.N.T., BANO, S., REDDY, G.G., ARJA, P. and NIHARIKA, G.L.

2020

Stock Price Prognosticator using Machine Learning Techniques

S. Nithya Tanvi Nishitha
Department Of CSE
Koneru Lakshmaiah Education Foundation,Vaddeswaram,India
nithyatanvi9999@gmail.com

Shahana Bano
Department Of CSE
Koneru Lakshmaiah Education Foundation, Vaddeswaram,India
shahanabano@icloud.com

G. Greeshmanth Reddy
Department Of CSE
Koneru Lakshmaiah Education Foundation, Vaddeswaram,India
greeshmanthreddy61@gmail.com

Pujitha Arja
Department Of CSE
Koneru Lakshmaiah Education Foundation
Vaddeswaram,India
pujithaarja@gmail.com

Gorsa Lakshmi Niharika
Department Of CSE
Koneru Lakshmaiah Education Foundation
Vaddeswaram,India.
niharikagorsa2000@gmail.com

Abstract— **Stock market price prediction is one of the favourite research topics under consideration for professionals from various fields like mathematics, statistics, history, finance, computer science engineering etc., as it requires a set of skills to predict variation of price of shares in a very volatile and challenging share market scenario. Share market trading is mostly dependent on sentiments of investors and other factors like economic policies, political changes, natural disasters etc., Many theories were forwarded, mathematical and statistical applications in conjunction with probability, to simplify the complex process. After the advent of computers, it got further simplified but still challenging due to various external influential factors ruling the volatility of the market prices. Thus, AI and ML algorithms were being developed, but for only for next day using Linear Regression procedures.**

Our project aims to predict the prices of shares more precisely and accurately using special algorithms using RNN by improvising the back propagation, feedback routines to overcome the short-term memory loss involved in RNN thus providing efficiency in LSTM applications.

Our project emphasizes how the LSTM applications perform with datasets of extreme, larger and minimal fluctuating data.

Keywords— Stock Market, Stock Price Prediction, Recurrent Neural Networks, Long-Short Term Memory.

I.INTRODUCTION

The impact of Stock Markets in the economy of a nation is so much that it bears the indication of that country's performance in the business. The businesses, either public or private owned, find rich sources of their capital from these stock exchanges along with the commercial lenders who usually dominate the debt markets. Thus, these businesses can provide large scale employment and also expand their operations. By issuing shares they can raise funds in cash without any burden of re-payment. Investors usually find it very difficult to decide the very likely probability of increase in their invested shares. So, the investor is always in the search of finding a suitable and most reliable Stock Broker who can predict to him whether the price of the shares he invested is going to increase or decrease so that he can decide to buy more stocks or sell his stocks with the sole motive of profit making.

The advent of Artificial Intelligence is helping this highly chaotic market to predict the price of any share by using the advanced Machine Learning algorithms. But still this has its own limitations. Different algorithms have different approach and they are categorically divided into Supervised, Un-supervised and Reinforcement Learning with different levels of accuracy

Stock Market :

The corporation may be a Private Limited Company or a Public Limited Company. The Business is then registered in a Stock Exchange and is entitled to issue a legal entity called share certificate commonly referred to as "STOCK" or more generally "Common Stock". It indicates ownership. The person who owns such share certificates is called a share- holder.

Stock is thus defined as a share of ownership of a company.

Stock Trade :

As companies grow, they require more funds to keep the companies in operations and they look for more investors to invest their money in the form of share certificates or shares. If the investor decides he no longer in need of the shares he then decides to sell them to another investor for a fair price that is different from the face value of the share. It may be traded for more value or lesser value of the share. This is the essence of the Stock Trade.

Stock Price Trends :

If the investor tries to sell or another investor tries to buy the shares for more value than the face value of the share, the sentiment is bullish and the prospects of the company increases. If there is a positive talk on the future projects taken over by the company, the share price further increases and the investors flock to buy the shares of that particular company. This is called BULL Trend. External factors like economic policies and reforms of the government, economic conditions both international and domestic circumstances, price-to-earnings ratio, political circumstances, natural disasters etc., also influence the stock prices.

Stock prices are thus most volatile asset class in the financial markets. The shares may either loss or gain value more erratically.

Stock Price Prediction :

The volatility structure of the prices of the shares in the share market throws challenges in predicting their future price, thus attracting investors in guiding their emotional and investment decisions. This uncertainty makes it nearly impossible to estimate the price of shares with utmost accuracy.

II.RELATED WORK

With the advancements in technology and involvement of computers, it is imperative that modern decision support systems are well-equipped to deal with the prediction of stock prices. These DSS's consider the large amount of unique data generated from various sources, process the data and provide us the necessary information.

Making use of Machine Learning, the machines are trained to analyze the patterns from old (previous) data and make predictions using the newly generated data.

The machine is first trained with historical labeled or unlabeled data based on the problem statement. It is then evaluated on the test data set. It is necessary to define the metrics used for the prediction as the improper use may mislead the prediction process.

Deep Learning techniques like Long Short Term Memory network built on Recursive Neural Network architecture (RNN) is widely used in implementing Machine Learning algorithms like Linear Regression in the implementation of predictions.

Upon Research in the Internet, majority of the numerous Papers submitted to National and International Conferences were able to predict only the next day's price of a company's stock. The challenge lies in the fact that the price of the stock predicted must be close to the actual price. Then that algorithm is the most reliable algorithm. Yuiquing Dai and Yuning Zhang [1] in their research paper concluded that by using all the 16 features in the SVM model predicted highest accuracy of 79% after the next 44 days. Still, they mentioned that they have to test the robustness of their predictor on different categories of stocks. Osman Hegzay, Osman S Soliman , Mustafa Abdul Salam [2] shown that compared to LS-SVM, LS-SVM-PSO produced minimum errors. Rajashree Dash, Dr. Pradipta Kishore Dash [3] concluded that output of CEFLANN has higher accurate results compared to SVM, KNN, DT models. Rajesh, N. Srinivas [4] has opined that to obtain the result having maximum efficiency with minimum classification, complicated stock trend need to be normalized in line of probability distributions. P. Lakshmi Prasanna, D. Rajeswara Rao [5] proved that the accuracy of PRO-RNN algorithm did increase opposing the loss of information due to memory inconsistency. N. Sirisha, K.V.D Kiran [6] concluded that with help of Hadoop derived from BigData can improve the performance of stock exchange analysis by learning through the previous results. Poonam Somani, Shreyas Talele, Suraj Sawant [7] opined that by using SVM, Hidden Markov Model , Neural Networks has given more accuracy than the traditional techniques. Zaho, Lei, Wang, Lin [8] tested the data by using the Outlier Mining algorithm and found that it was consistent with K-means algorithm. Feng Wang, Zhiyong Zaho [9] created a new algorithm i.e., MKL – ELM to achieve high accuracy and prediction speed but failed with respect to complicated data. Mustain Billah, Sajjad Waheed, Abu Hanifa [10] reported that the Levenberg Marquard (LM) algorithm of neural network is more accurate with accuracy of 53% than using the Fuzzy Inference System (ANFIS).

Numerous attempts were made from time to time to predict stock price using Machine Learning. This research focuses on three parameters:

(1) The change in the target price varies from near_term (< 1 minute to a few minutes), short-term (next day to a few days) and long-term (more than 5 days, may be months later).

(2) Limited number of the set of stocks.

(3) The predictors used can range from investors sentiments and economic policies to time series data of prices of stock.

To get better results than those solely operating on time series data, financial technical indicators were applied along with machine learning algorithms. The focus is on short-term price prediction which has the known history of worst-prediction accuracy

Convolution Neural Networks have no memory cells thus cannot memorize the previous inputs and the state of previous inputs are the outputs, but the RNN can. The first output is dependent on the first input but the second output is dependent on a second input as well as the first input and similarly a third output is dependent on a third input as well as the first and second inputs. This is the mechanism of working of the engine. This is a recurrent neural network. In neural networks there are different types of recurrent neural networks like one-to-one, one to many, many to many, many to one.

In the existing system, Support Vector Machine and Back-Propagation algorithm are found to be least effective in handling non-linear data. So, LSTM, a higher accurate algorithm in giving out a future stock price forecasting is proposed.

III.PROPOSED WORK

In the proposed system, the model predicts future stock price accurately employing two different architecture layers or modules viz., engineering as well as prediction layers.

The first layer studies and derives a set of complex and simple rules of trade from the input data set by analyzing time series data. The output consists of discrete values 0, 1 or –1 represents 'hold', 'buy' or 'sell' signals which serve as inputs to the second layer. This layer then implements Random Forest for classification and learning.

Data :

Required dataset was extracted from National Stock Exchange. The data is of Tata Global whose intra-day broking has highly liquid equity market. Its index consists of almost 20% of the market capitalization of the Indian equity markets. The data comprises the date, open – high – low, last, close price levels, volume traded and turnover for each stock. The entire data set covers the period from July 21, 2010 to September 28, 2018. In this study, the data used was from July 21, 2010 to December 31, 2015 for training and then January 1, 2016 to September 28, 2018 for testing.

Implementation:

LSTM will overcome the disadvantage of the RNN i.e., RNN suffers from short memory loss. The LSTM works on gates. Here the gates will help the LSTM decide which data is important and which need to be discarded. LSTM consists of three gates i.e, (1) input gate (2) forget gate (3) output gate(4) cell state.

All the important data which was filtered from gates is carried out to cell state and cell state will carry out the whole data until required output is produced.

(1) **Input gate** : The input gate work is to update the cell state. It will concatenate the previous hidden state and the input vector into the sigmoid activation.

(2) **Forget gate** : The forget gate allows the needed information and discard the necessary data.

(3) **Output gate** : The output gate decides the value for the next hidden state. This information is calculated from previous input state of the cell.

(4) **Cell state** : The cell state is effected by the result of the other gates in the cell.

Five popular technical indicators that are highly successful in predicting the direction of stock prices viz., Moving Average (MA), Weighted Moving Average (WMA), Relative Strength Index (RSI), Moving Average Convergence (MAC) or Divergence (MAD), Bollinger Band Commodity Channel Index (BCI) are used in this model.

These indicators deploy input variables as a combination of continuous values, single technical rules of trade as well as complex rules.

The dataset used is NSE_TATAGLOBAL.csv. In the dataset the attributes are open, close, high, low, last, total trade quantity, turnover. The attribute – open is only used here for the purpose of prediction. RNN is used to fed a sequence of time series data comprising almost over eight years into the network. To predict the next day price, a scale is so chosen in terms of Boolean variables which scales the prices within scale the prices from 0 to 1. By default the scale is set to True that helps the RNN to learn effectively and much faster. To predict the future lookup steps, Lookup step is set to the default value 1.

First the dataset is loaded using the function stock_info.get_data(). If the given argument is True, then using sklearn's MinMaxScaler class it scales all the prices within zero to one inclusive. To scale the data every column will have its own Scaler. It then inserts a column, by the name future column, which shows the target values. From that point onward, it splits and rearranges the data then returns the output. Since a suitable function to stack and set up the dataset is implemented, another suitable function to construct our model is found to be required. Again, this function is so pliable that the optimizer in order to compile the data, the input arguments like dropout-rate, RNN, loss, n-steps etc can be varied. Then the above function builds a RNN with 1 neuron dense layer as the output layer. It also requires other sequence of features i.e., sequence_length of about 50 to 100, and days in the dataset used i.e, consecutive time steps. This results in an output as a single value which is the price of the next_time_step. The results i.e., output can be subjected to experimentation by changing the default parameters mentioned below.

Test_Size: It tests the sample rate. For example, 0.3 means 30% of the total dataset.

Feature_Columns: It uses the feature to predict the next price value.

N_Layers: It is the number of RNN layers to stack.

Cell: The default RNN cell to be used is LSTM. Simple RNN or GRU can also be used.

Dropout: It is the dropout rate after each RNN layer.it is the likelihood of not training the given node in a layer. The value 0.0 means null dropout. This regularization can make the model to not over-fit on the data given for training.

Bidirectional: It is a Boolean. It decides whether to use the functionality of bidirectional RNN's are not.

Optimizer: It selects an optimization algorithm. 'Adam' is the most prominent algorithm.

Loss: Huberloss scores on the mean absolute error function to use in this regression problem.

Batch_Size: It prompts for the number of data samples to be used for every training iteration

Epochs: It is the frequency of the learning algorithm to go through the whole training dataset.

ModelCheckPoint is utilized that spares our model in every epoch during the training process. Additionally TensorBoard was used to picturize the model presentation in the training cycle.

The number of epochs given to prepare the model are 100. After the training terminates or during the training itself, if attempted to run tensorboard, by using this command, the loss is the MeanSquaredError as represented by the create_model(). The orange line depicts the training loss, though the blue line is the required line which represents the validation loss that diminishes over the time. Before beginning to test the model, the raw data i.e., the information without rearranging is reloaded, as it plots the stock curve in the right order.

VI.ALGORITHM

Step 1 : Start.

Step 2: The user downloads the dataset of the required company or organization from the Stock Exchange website.

Step 3: The user then inputs the dataset to the program.

Step 4: The program then reads the data from the dataset.

Step 5: As the date is unique, it is taken as an index for the purpose of Time Series Analytics and Forecasting.

Step 6: The closing Price history is then displayed as a graph.

Step 7: The required libraries are then imported; the data frame is created and then the index is set.

The program is now ready for prediction of the stock price using Deep Learning Algorithm. Long and Short-Term Memory Algorithm (LSTM), which is most popular to predict the Stock Price was used in this project.

Step 8: The program then creates the training and testing datasets and converts them into x and y trains combined respectively using the MinMaxScaler algorithm. It then transforms individual features so that the created datasets scales in the given range i.e. between zero and one.

Step 9: As the program is now trained, LSTM then performs a deep learning process and analyses the data to predict the next day probable price,

Step 10: It displays the result in a graph.

Step 11: The user then conveys the inference to the customer and concludes the session.

Step 12: End.

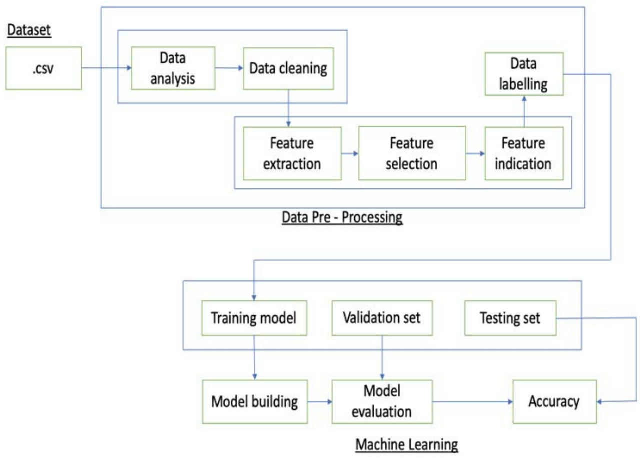

Fig.2 : Block diagram

The data from the dataset in the form of .csv extension is extracted by the algorithm and is pre-processed. It will be first analysed in tandem to the columns provided to be considered. It checks for the data fields for any voids and suitably cleans the data by removing those rows containing empty fields. The algorithm then loads the features of LSTM algorithm and selects only those features with values given to arguments. It then operates on the data by dividing the data into blocks keeping aside 60% for training, 20% for validating and then 20% for testing and then predicting the next day price along with accuracy but first it is processed with 70% training and for 30% testing.

V. Flow Chart

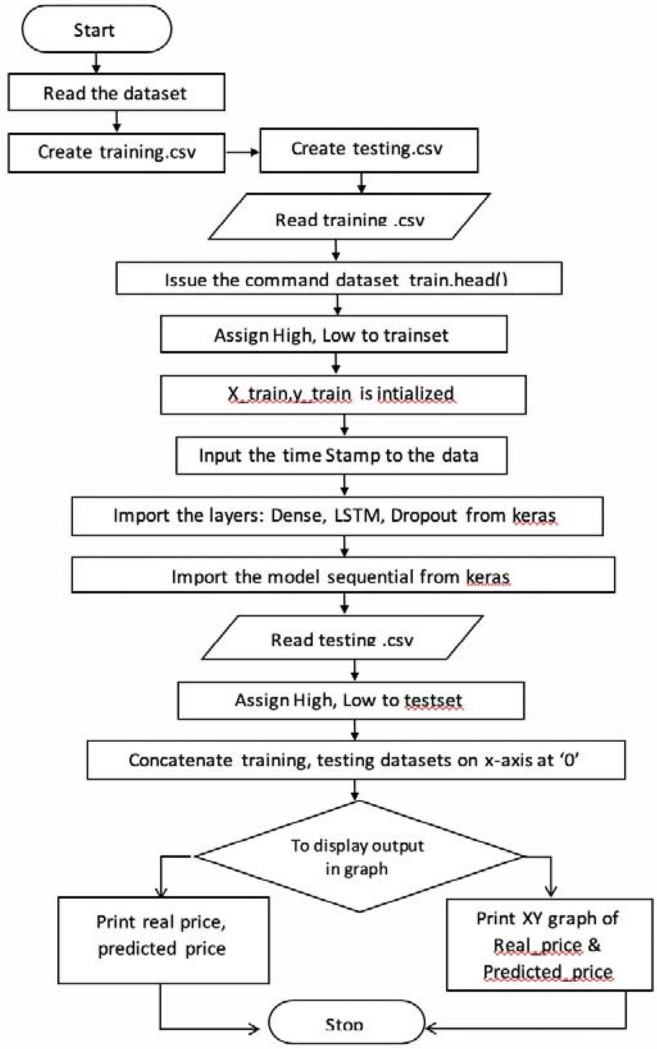

Fig. 1: Overview of the process

VI. RESULTS

The function utilizes the last_sequence variable which was previously stored in the load_data() function. This function has the sequence of prices, in order to predict the following day price from the previously stored data. This function used without validation graphs the last 30 months data of the test set and the predicted price. The blue line is the real test set, and the red line is the prediction of prices. Observation revealed that the stock price as of late is drastically expanding, the model accurately predicted the price as Rs. 231.85 for the following day i.e., September 29. 2018 which is a nearby estimation of Rs. 234.00 that was referenced by NSE in its stock broking for the first dataset. Similarly, two other datasets of Reliance communications RCom and ONGC LTD were considered and the whole above given processes was repeated.

The purpose to use three datasets of varied nature is for testing the accuracy of the proposed model. The first and third datasets are of highly scattered i.e., the price of the share under study have greater fluctuations throughout the period of study and the second dataset has very minimal fluctuations in the price of the share. In the actual practice majority of the other models were found to fail if there are extreme fluctuations in the dataset in the form of large deviations and large value of coefficient of variations in training as well as testing data in their respective datasets. However after normalizing the training and testing sets of datasets, their models predicted nearly 99% accuracy. With respect to long term prices the prediction made by this model reaches about 70±0.5% accuracy for 30 lookup steps), and it reaches about 86±0.6% accuracy for 50 lookup steps with n_steps = 70 i.e., sequence length = 70 for the dataset having extreme fluctuations in the price of the shares throughout the period of study but up to 95±0.5% accuracy for the dataset having minimal fluctuations.

The graphs presented below confirm the results produced. Graphing is done first with 70% of dataset for training and remaining 30% for testing. These are Fig 3.1, 4.1 and 5.1.

Graphs are then derived from that of 60% of dataset for training, 20% for validating and 20% for testing with lookup steps 30 and n_steps = 70 in Fig. 3.2, 4.2 and 5.2.

Graphs of Fig. 3.2, 4.2 and 5.2. are also derived from that of 60% of dataset for training, 20% for validating and 20% for testing with lookup steps 50 and n_steps = 70.

Finally the accuracy is calculated and then displayed.

It was observed that as the parameters passed to arguments were changed, learning mechanism of LSTM improved further and produced better accuracy.

First dataset: **NSE-TATAGLOBAL**

Header Rows

	Date	Open	High	Low	Last	Close
0	2018-09-28	234.05	235.95	230.20	233.50	233.75
1	2018-09-27	234.55	236.80	231.10	233.80	233.25
2	2018-09-26	240.00	240.00	232.50	235.00	234.25
3	2018-09-25	233.30	236.75	232.00	236.25	236.10
4	2018-09-24	233.55	239.20	230.75	234.00	233.30

Tail Rows

2030	2010-07-27	117.6	119.50	112.00	118.80	118.65
2031	2010-07-26	120.1	121.00	117.10	117.10	117.60
2032	2010-07-23	121.8	121.95	120.25	120.35	120.65
2033	2010-07-22	120.3	122.00	120.25	120.75	120.90
2034	2010-07-21	122.1	123.00	121.05	121.10	121.55

Fig. 3.1: Without validation.

Fig.3.2: With lookup steps 30

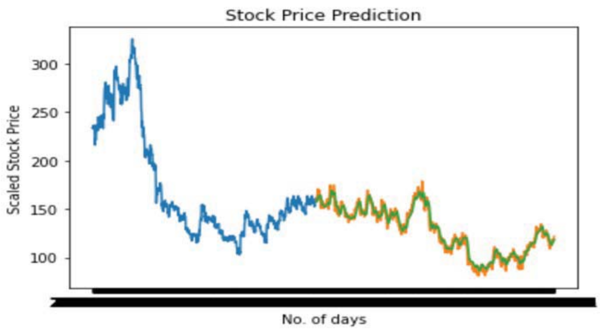

Fig. 3.3: With lookup steps 50

Accuracy is 86.6%.

Second dataset: **Reliance communications**

Header rows

	Open	High	Low	Close	Volume
0	18300	18640	18220	18630	640000
1	18350	18430	18260	18350	440600
2	18250	18460	18160	18460	666200
3	18120	18180	17990	18050	358200
4	18000	18140	17940	18030	572000

Tail rows

	Open	High	Low	Close	Volume
484	47410	47540	47020	47540	537700
485	47395	47620	47235	47300	465200
486	47280	47765	47200	47610	511000
487	46865	47280	46505	47115	666400
488	47180	47200	46410	46770	883600

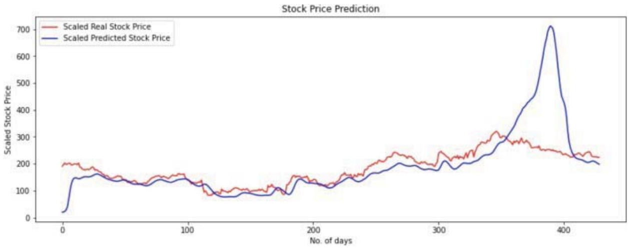

Fig. 4.1: Without validation.

Fig.4.2: With lookup steps 30

Fig. 4.3: With lookup steps 50

Accuracy is 95.5%.

Third dataset: **ONGC India Ltd.**

Header Rows

	Date	Open	High	Low	Close	Volume
0	2016-07-01	17924.240234	18002.380859	17916.910156	17949.369141	82160000
1	2016-06-30	17712.759766	17930.609375	17711.800781	17929.990234	133030000
2	2016-06-29	17456.019531	17704.509766	17456.019531	17694.679688	106380000
3	2016-06-28	17190.509766	17409.720703	17190.509766	17409.720703	112190000
4	2016-06-27	17355.210938	17355.210938	17063.080078	17140.240234	138740000

Tail Rows

	Date	Open	High	Low	Close	Volume
1984	2008-08-14	11532.070312	11718.280273	11450.889648	11615.929688	159790000
1985	2008-08-13	11632.809570	11633.780273	11453.339844	11532.959961	182550000
1986	2008-08-12	11781.700195	11782.349609	11601.519531	11642.469727	173590000
1987	2008-08-11	11729.669922	11867.110352	11675.530273	11782.349609	183190000
1988	2008-08-08	11432.089844	11759.959961	11388.040039	11734.320312	212830000

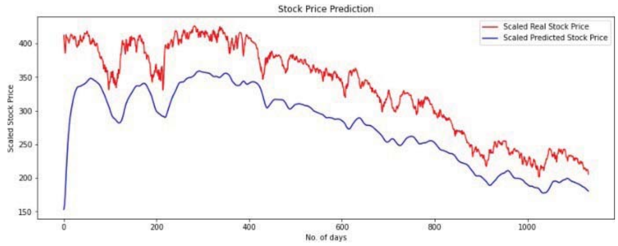

Fig. 5.1: Without validation

Fig.5.2: With lookup steps 30

Fig. 5.3: With lookup steps 50

Accuracy is 86.6%

VII.CONCLUSION

The study attempted to predict accurately the next day price of the listed share. Hence, our active day trading strategy outperforms when compared to other papers in the present domain of study. Using different lookup steps, it has been attempted to build the other models. In tensor board it was observed that the green line is the model to learn the fluctuations in the stock price, which uses the next time-step stock price as the label, whereas for validation an orange curve was used, for normalizing and validating the scaled stock price. The green line is about testing the normalized data. This is a great model which can be used usually in investing in more long term investments. One of the two ways mentioned here can be used when it comes to predict the rise or fall of the price of the share but not its actual price. The first one is to compare the predicted price with the current price and make the decision. The second one is to build an entire model and change the previous output's activation function to sigmoid, as well as the loss and the metrics. This function thus estimates the accuracy score by scaling the predicted price either to 0 or to 1.

VIII.CHALLENGES AHEAD

The potential of LSTM in Machine Learning for the purpose of predictions can be further explored if the algorithm can be modified and improved or developed in such a way that it can learn how the price fluctuations occur with respect to various other factors even beyond the circumstances that are out of control either man-made or because of natural disasters. This is possible, in my opinion, by feeding the datasets of all kinds and types with all extremities viz.., negligible to extreme fluctuations and to train the machines to learn more precisely and with high accuracy.

REFERENCES

[1] "Machine Learning in Stock Price Trend Forecasting" – Dai, Yuqing and Yuning Zhang – 1993.

[2] "A Machine Learning Model for Stock Price Prediction" – Osman, Omar S. Soliman, Hegazy, Mustafa Abdul Salam – 2014.

[3] "A hybrid stock trading framework integrating technical analysis with machine learning techniques". – The Journal of Finance and Data Science – Armstrong, P. and Keevil, S.- 2016

[4] Rajesh, P., Srinivas, N., Vamshikrishna Reddy, K., Vamsipriya, G., Vakula Dwija, M., & Himaja, D. (2019). Stock trend prediction using ensemble learning techniques in python. International Journal of Innovative Technology and Exploring Engineering, 8(5), 150-154.

[5] Lakshmi Prasanna, P., & Rajeswara Rao, D. (2019). Probabilistic recurrent neural network for topic modeling. International Journal of Innovative Technology and Exploring Engineering, 8(4), 165-168.

[6] Sirisha, N., & Kiran, K. V. D. (2018). Stock exchange analysis using hadoop user experience (hue). Paper presented at the Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS 2017, 1141-1144. doi:10.1109/ISS1.2017.8389363.

[7] Stock Market Prediction Using Hidden Markov Model, PoonamSomani, ShreyasTalele and SurajSawant.

[8] Price Trend Prediction of Stock Market Using Outlier Data Mining Algorithm, Lei Zhao ; Lin Wang

[9] Stock Volatility Prediction using Multi-Kernel Learning based Extreme Learning Machine, Zhiyong Zhao, Xiaodong Li, FeiYu and Hao Zhang

[10] Billah, M., Waheed, S., & Hanifa, A. (2016). Stock market prediction using an improved training algorithm of neural network. 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE). https://doi.org/10.1109/icecte.2016.7879611