Stock Market Prediction

Arpit Dhamija
2K18/EC/044
DTU
Delhi, India
arpitdhamija_2k18ec044@dtu.ac.in

***Abstract*—Stock market prediction model is made using recurrent neural networks(RNN),Long short term memory (LSTM) and GRU which is an advance form of LSTM. That LSTM cell is made in such a way that it makes patterns in technical trends, remove and update various important and non-important information in the neurons. We have trained 17 to 20 years of data and predicted the next month stock prices, we are getting very good results. In this paper, Stock prices of MARUTI are used which are taken from Nifty 50 dataset from Kaggle.**

Keywords – RNN(Recurrent Neural Networks), LSTM(Long Short Term Memory), GRU, RMSE(Root Mean Square Error)

I. INTRODUCTION

Stock market trading is one of the most important financial activity for investment and capitalization of the economy. It allows enterprises to raise money from the investing public to fund their growth and operations. There is also a large amount of risk is involved due to its unconditional trends which are affected by various scenarios, its fluctuations are highly violent, we will discuss that all later in this paper.

However this trend of stock market is getting increased and it is creating a lot of wealth in out society, made many people millionaires. There have come a lot of organizations who are predicting the stock prices, analyzing stock prices by expert. Stock market prediction is always been a curious field for all And the advancement of today's technology improved computers to predict stock prices through machine learning and deep learning techniques.

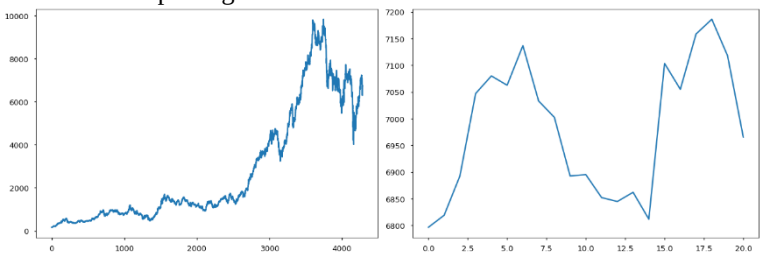
As we all know that stock market is a sequential data as the order of stock prices matters in the future prices.
For this time of time series sequential data, we use neural networks for solving this financial problem. An advance form of neural networks, i.e. Recurrent Neural Networks (RNN) is used for predicting sequential data.
Recurrent Neural Networks suffer from short-term memory. If a sequence is long enough, they'll have a hard time carrying information from earlier time steps to later ones. During back propagation, recurrent neural networks suffer from the vanishing gradient problem.

LSTMs and GRUs (newer generation of RNN) is a solution for this problem - LSTM 's and GRU's were created as the solution to short-term memory. They have internal mechanisms called gates that can regulate the flow of information. These gates can learn which data in a sequence is important to keep or throw away. By doing that, it can pass relevant information down the long chain of sequences to make predictions.

Here in this paper, we have predicted our stock market prices by using LSTM and GRU recurrent networks layers.
Data is taken from Kaggle Nifty 50 dataset, it contains the top 50 Indian companies which are listed in nifty 50. And among those 50, I have mainly focused on MARUTI stock prices. Few datasets were incorrect, so just first find a suitable correct dataset first.
In MARUTI dataset, stock prices are given from July 2003 to October 2020, I have trained last 17 years of data, i.e. from July 2003 to September 2020, and predicted the stock prices of October 2020, predicted prices are showing good results and upwards and downwards trends are also coming out to be very good.
As most of the students and even working professionals do not have proper knowledge of stock market and they don't know the technical and fundamental analysis like when and which stock to buy, So this model can provide them clear cut decision that when are the prices of stocks can rise and fall so that they can invest their money and earn profits.

II. ROADMAP

1. **Data Preprocessing**
   - Getting data from Kaggle
     https://www.kaggle.com/rohanrao/nifty50-stock-market-data
   - Loading libraries and visualizing the data.
   - Splitting the data.



Training set (till sep,20) test set (oct,20)

- These are the graphs of closing price

   - Handling missing data
     - Identify missing data in data-frames.
     - Treat (delete or impute) missing values if found

     No missing values were found, so no need to handle anything and not to fill anything.
     Only "Trades" whose all values are not in our dataset but we actually don't need Trade values as of now as our prediction is not dependent on Trade values. So, Just ignore handling this missing data

- Feature Scaling
  - We scale all the values of 'close', new scaled training set whose values are between 0 and 1.
  - We have done normalization

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

- Creating a data structure with 80 timesteps and 1 output. Model learns from past 80 days (appx 4 months) and predict the 81th day stock price by seeing the trends, model will basically create a pattern like what going to happen after certain movements in stock market. It is like doing the technical analysis of the stock.

- Reshaping to make the dimensions desirable for passing it into LSTM and GRU layers

2. Building and Training RNN

- Loading all important libraries

- Initializing the neural network.

- Adding LSTMs layers with number of neurons and a dropout ratio.
  Inside LSTM cell :
  - Activation = 'Tanh'
  - Recurrent Activation = 'Sigmoid'

- In this research, I have used 4-5 LSTM layers

- Compiling

- Fitting, using 1000 epochs and batch size=32. These parameters are user dependent. Choose as per your choice.

3. Predicting and Visualizing the result

- Predicting the stocks by using my logic and trained neural networks.

- Visualizing

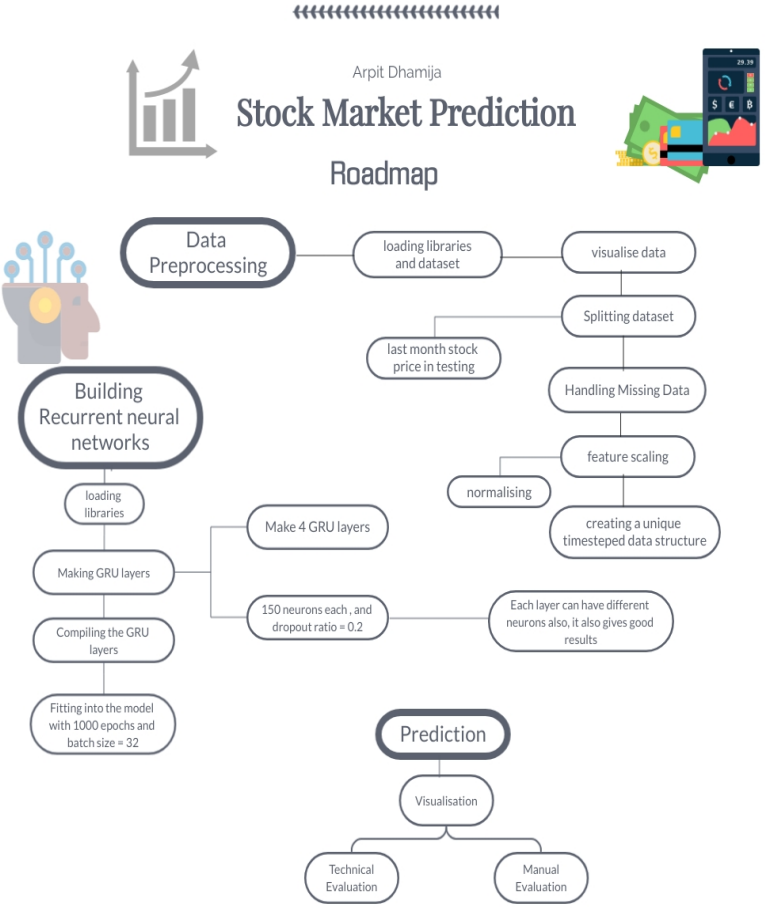
- Technical evaluation
- Manual evaluation

4. Evaluation and Analysis

For analyzing the efficiency of the system we are used the Root Mean Square Error(RMSE). The error or the difference between the target and the obtained output value is minimized by using RMSE value. RMSE is the square root of the mean/average of the square of all of the error.
The use of RMSE is highly common and it makes an excellent general purpose error metric for numerical predictions.
Compared to the similar Mean Absolute Error, RMSE amplifies and severely punishes large errors.

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

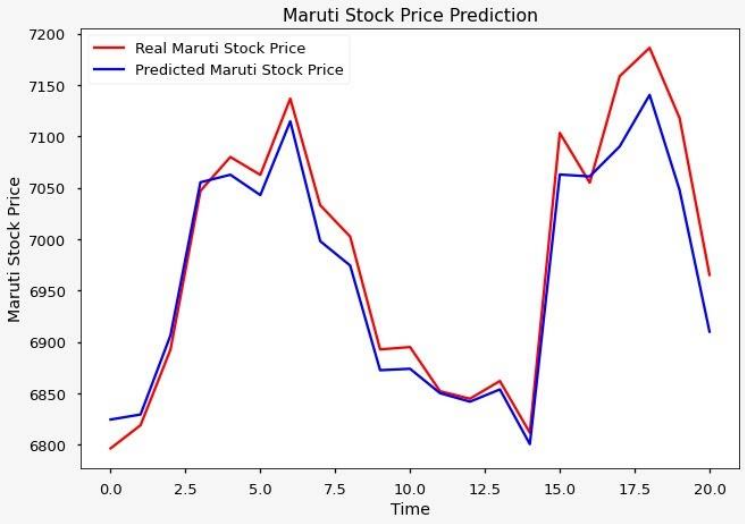
$\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$ are predicted values
$y_1, y_2, \ldots, y_n$ are observed values
n is the number of observations



Do same procedure for GRU based model also.

III. EXPERIMENTATIONS OF OUR DATASET

1. Using GRU

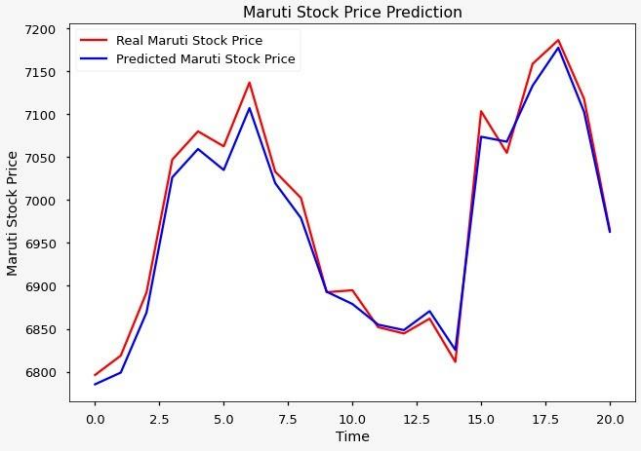


- 500 epochs
- 4 GRU layers with 150 neurons each
- dropout = 0.2
- history data = appx 4 months(80 days)
- data trained = 17 years data, from July 2013 to September 2020.
- It's the prediction of October 2020

RMSE = 32.21746434471314

---

2. Using LSTM

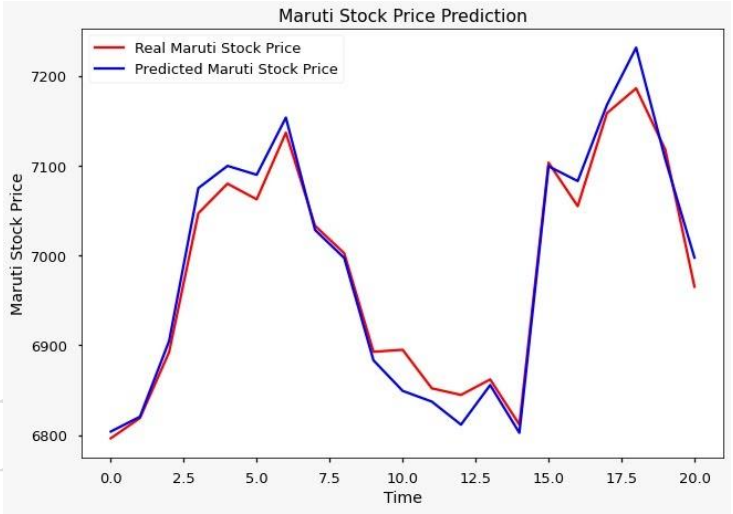


- 500 epochs
- 4 LSTM layers with 140 neurons each
- dropout = 0.2
- history data = appx 4 months(80 days)
- data trained = 17 years data, from July 2013 to September 2020.
- It's the prediction of October 2020

RMSE = 18.021098224779312

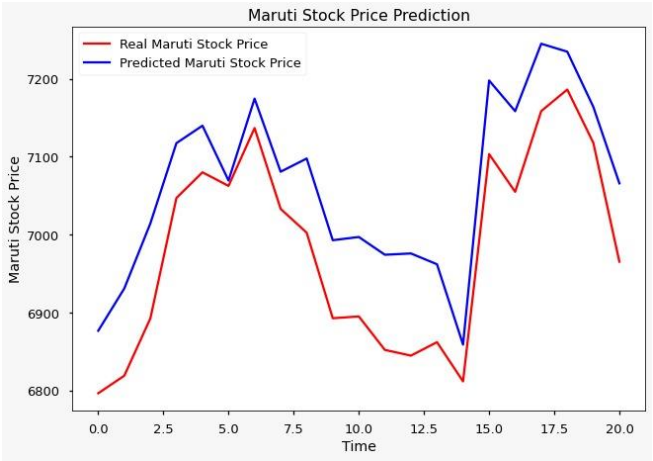
We are getting less root mean square error in GRU system.
In LSTM, RMSE = 18.021098224779312
In GRU, RMSE = 32.21746434471314

LSTM is performing better, let us see do more experiments.



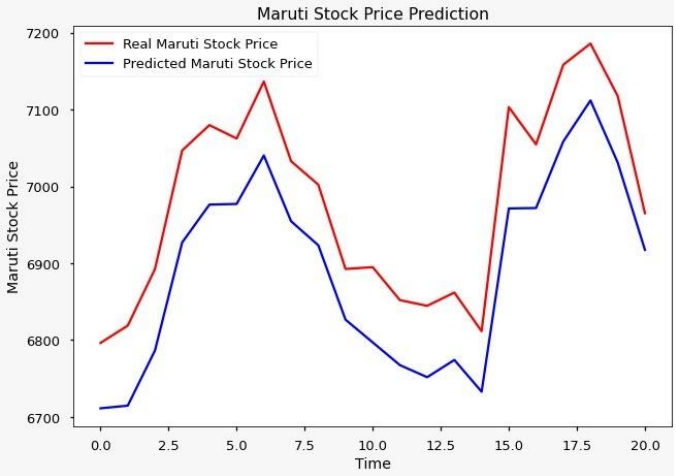
- 1000 epochs
- 4 LSTM layers with 150 neurons each
- dropout = 0.2
- history data = appx 4 months(80 days)
- data trained = 17 years data, from July 2013 to September 2020.
- It's the prediction of October 2020

RMSE = 22.024644990413297



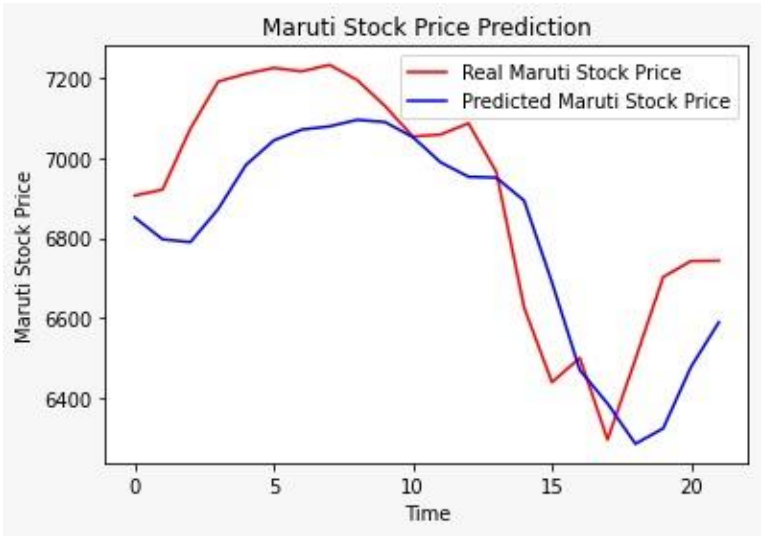
- 4 GRU layers

- 150 neurons each
- 1000 epochs
  RMSE = 87.73010089605076

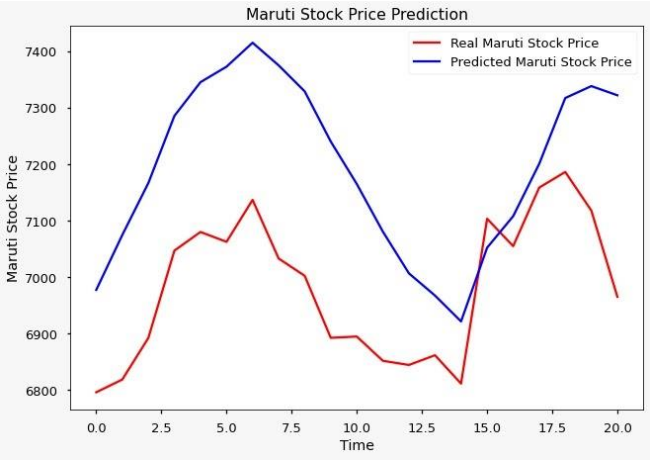


- 250 epochs
- 4 LSTM layers with 150,130,140,125 neurons in each layer respectively.
- dropout = 0.2
- history data = appx 4 months(80 days)
- data trained = 17 years data, from July 2013 to September 2020.
- It's the prediction of October 2020

  RMSE = 91.63977189371019

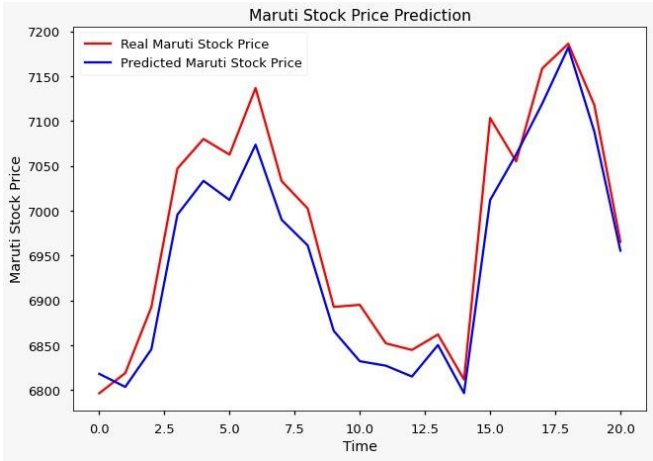


- 120 epochs
- 4 LSTM layers 100,75,95,80 neurons in each layer respectively
- dropout = 0.2
- history data = appx 6 months(120 days)
- data trained = 17 years data, from July 2013 to August 2020.
- It's the prediction of September 2020



- 4 GRU layers
- 150,130,140,125 neurons in each layer respectively
- 250 epochs
- dropout = 0.2
- history data = appx 4 months (80 days), data trained = 17
  years data, from july 2003 to september 2020.
- It's the prediction of october 2020

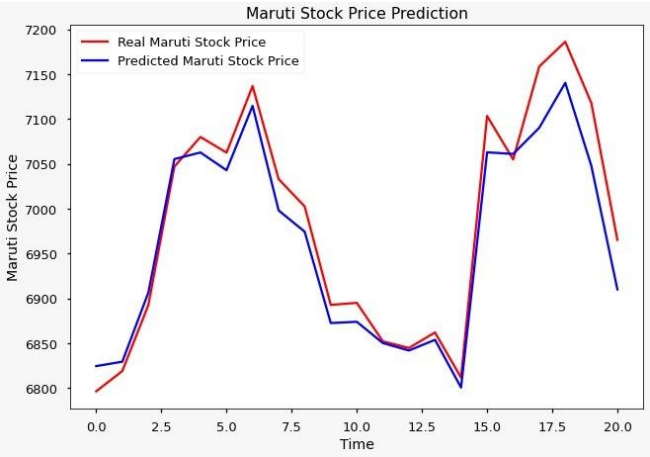
rmse = 238.2181760541056



- 5 LSTM layers
- 150 neurons each
- 1000 epochs
- dropout = 0.2
- It's the prediction of October 2020

RMSE = 41.024846905704585

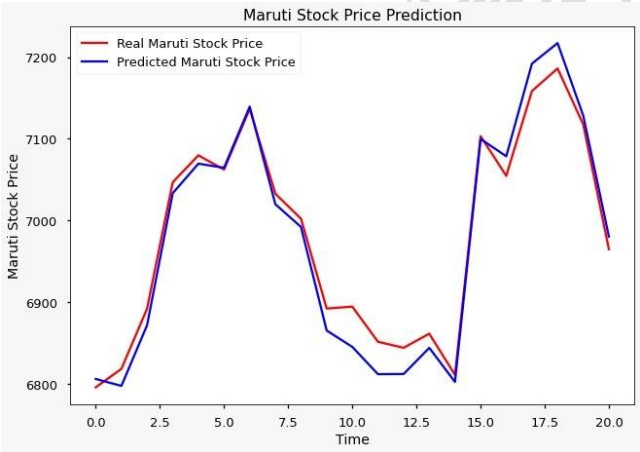
❖ Combining LSTM and GRU together and making a new model



4 Layers in order:

LSTM → GRU → LSTM → GRU with 150 neurons each. 500 epochs.

RMSE = 19.45398867372914
Dropout = 0.2
Prediction of October,20



5 Layers in order:

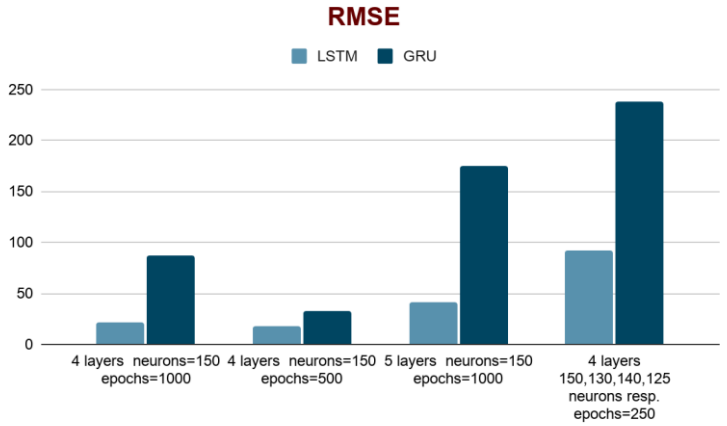
LSTM → GRU → LSTM → GRU → LSTM

- 150 neurons in every layer,
- 500 epochs
- Dropout = 0.2
- Prediction of October,20
- RMSE = 22.50089383603376

Model	Layers	Neurons	Epochs	RMSE
LSTM → GRU → LSTM → GRU	4	150	500	19.45398867372914
LSTM → GRU → LSTM → GRU → LSTM	4	150	500	22.50089383603376

Analytics of single LSTM/GRU models

Model	Layers	No. of Neurons	Epochs	RMSE
LSTM	4	150	1000	22.024644990413297
GRU	4	150	1000	87.73010089605076
LSTM	4	150	500	18.021098224779312
GRU	4	150	500	32.21746434471314
LSTM	5	150	1000	41.024846905704585
GRU	5	150	1000	174.46829375374043
LSTM	4	150,130,140,125	250	91.63977189371019
GRU	4	150,130,140,125	250	238.2181760541056



IV. Conclusion

So, as per our experiments and analysis, we can conclude that this time series data i.e. stock market data can be predicted using recurrent neural networks techniques as these models performed phenomenal as compared to other traditional methods.
Here LSTM and GRU is used where GRU is a modern form of RNN which has been recently discovered.
In each experiment, LSTM has outperformed GRU, but that doesn't mean that GRU is bad predictor, the average price of stocks of MARUTI in month of October was Rs. 6981.495 and error of 0.618% in LSTM and error of 1.9% is observed in GRU. And Combining LSTM and GRU is also giving best results, RMSE averages out to be appx Rs. 20, which is 0.28% error and 99.72% accurate results.
So, according to my calculations, we can conclude that its one of the best and reliable model to invest in stock market.

V. FUTURE WORK AND INNOVATIONS

Integrating this stock market prediction model with the
Fundamental analysis of companies,
Social media movements,
Twitter sentiment analysis,
Political influence

as it highly effects the stock prices of certain companies.
Example -

1. When pandemic due to covid had raised, then there was a massive fall in the stock market.

2. During lockdowns, online classes/meetings were in boom, so stock prices of these software providing companies were exponentially increased and made many chairperson billionaire overnight.

3. When a successful vaccine for covid-19 was announced, then stock prices of oil and gas increased as people will come out of their home and stocks of transportation industry was increased.

4. When Prime Minister of Japan left his position, then Japanese stock Market fall suffered a huge loss of $100 billion.

5. Tesla wealth was dropped down by few billion dollars when Elon Musk(Tesla's CEO) tweeted he is going to sell of all of his assets. Then shares of Tesla dropped down.

6. Elon Musk fortune jumps $15 billion and Tesla stocks

VI. REFERENCES

[1]https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] http://colah.github.io/posts/2015-08-Understanding-LSTMs/K. Elissa, "Title of paper if known," unpublished.

[3] Murtaza Roondiwala , Harshal Patel , Shraddha Varma, "Predicting Stock Prices Using LSTM International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

[4] https://www.analyticsvidhya.com/blog/2018/10/predicting-stock-price-machine-learningnd-deep-learning-techniques-python/

[5] https://www.sciencedirect.com/science/article/pii/S1877050920307924