Swami Vivekananda Institute of Science & Technology
Dakshin Gobindapur. P.S.: Sonarpur. Kolkata – 700145.

STOCK MARKET PREDICTOR

A project-report submitted in partial fulfilment for the degree
of
B. Tech in Computer Science and Engineering
of
Maulana Abul Kalam Azad University of Technology (MAKAUT)

(Formerly Known as West Bengal University of Technology)

Under the supervision of
Prof Kishore Ghosh
Submitted by

MRIGANKA DAS
Roll No.: 2400118050
Regn. No.: 182410110144 of 2018-19

PADALA NITISH KUMAR
Roll No.: 2400118034
Regn. No.: 182410110116 of 2018-19

SUCHITA PAUL
Roll No.: 2400118015
Regn. No.: 182410110135 of 2018-19

SOMDUTTA PAUL
Roll No.: 2400118020
Regn. No.: 182410110130 of 2018-19

RAJDIP PAL
Roll No.: 2400118029
Regn. No.: 182410110121 of 2018-19

Department of Computer Science & Engineering Swami Vivekananda Institute of Science & Technology Dakshin Gobindapur. P.S.: Sonarpur
Kolkata – 700 145.

ACKNOWLEDGEMENTS

It is a great pleasure for me/us to express my/our earnest and great appreciation to

_____, my project guide. I/We am/are very much grateful to

him/her for his/her kind guidance, encouragement, valuable suggestions, innovative ideas, and supervision throughout this project work, without which the completion of the project work would have been difficult one.

I/We would like to express my/our thanks to Head of the Department,

_____ for his/her active support.

I/We also express my/our sincere thanks to all the teachers of the department for their precious help, encouragement, kind cooperation and suggestions throughout the development of the project work.

I/We would like to express my/our gratitude to the library staff and laboratory staff for providing me/us with a congenial working environment.

Date:

(Full Signature of the Student(s))
B. Tech in Comp. Sc. & Engg.

Department of Computer Science & Engg.
Swami Vivekananda Institute of Science & Technology

SWAMI VIVEKANANDA INSTITUTE
OF SCIENCE & TECHNOLOGY
Dakshin Gobindapur, P.S. - Sonarpur
Kolkata - 700 145

Approved by AICTE &
Affiliated to MAKAUT
College Code : 241
(033) 2437 9913
Fax : (033) 2437 9913
info @svist.org
www.svist.org

CERTIFICATE

This is to certify that the project work, entitled "_____" submitted by _____ has/have been prepared according to the regulation of the degree B. Tech in Computer Science and Engineering of the West Bengal University of Technology. The candidate(s) has/have partially fulfilled the requirements for the submission of the project work.

(Signature of HOD)
Dept. of Computer Sc. & Engg.

(Signature of the Supervisor)
Dept. of Computer Sc. & Engg.

(Signature of External Examiner)

Contents

***

1.Abstract

In today's world, when we think of investment, stock market seems to be the best option for the same. But, investing without proper planning and future knowledge of any stock, the result can be devastating. So, we need a proper model that can predict the future trend of a stock. In this project we have used LSTM based neural network to build our prediction model. The accuracy of LSTM based model is higher than other models like Recurrent Neural Network model, naive machine learning model etc. We have used four hidden layers in the neural network and trained our model using Backpropagation Algorithm with mean-squared-error as our objective function. Prediction of trends by our model is very close the actual trends and it can also predict the future trend of a particular stock.

Beside good predictions, our model should be easy to use. So, we integrated our model with a python's streamlit API based web application. This web application has an interactive User Interface and besides future prediction, user can also see the historical data, charts and statistics of a particular stock.

***

3.Keywords

1. RNN: Recurrent Neural Network
2. LSTM: Long Short-Term Memory
3. AAPL: Apple
4. SBIN.NS: State Bank of India
5. GOOG: Google

***

4. List of Figures

***

5.List of Tables

***

6.Introduction

Stock Market Prediction is all about prediction of future trends of the various stock tickers based on their past data. Over the years, various machine learning techniques have been used in stock market prediction, but with the increased amount of data and expectation of more accurate prediction, the deep learning models are being used nowadays which have proven their advantage over traditional machine learning methods in terms of accuracy and speed of prediction. In this article, we will discuss the Long-Short-Term Memory (LSTM) Recurrent Neural Network, one of the popular deep learning models, used in stock market prediction. In this task, we will fetch the historical data of stock automatically using python libraries and fit the LSTM model on this data to predict the future prices of the stock.

6.1. Motivation and Objective

In today's world, the stock market can be the best option for investment if we invest in it with proper planning. The stock market is considered to be very volatile and complex in nature, investing without proper planning and future knowledge of any stock, the result can be devastating. An accurate prediction of future prices may lead to a higher yield of profit for investors through stock investments. As per the predictions, investors will be able to pick the stocks that may give a higher return. Investment in the Stock Market is now becoming more relevant in this covid-19 and lockdown era when people are losing their job or working from home.

***

7. Basic Underlying Concept

In this project, we will use the Long-Short-Term Memory (LSTM) Recurrent Neural Network, one of the popular deep learning models, used in stock market prediction and to build beautiful user-friendly web application we have used streamlit -a python framework.

7.1. LSTM Recurrent Neural Network

Long-Short-Term Memory Recurrent Neural Network belongs to the family of deep learning algorithms. It is a recurrent network because of the feedback connections in its architecture. It has an advantage over traditional neural networks due to its capability to process the entire sequence of data. Its architecture comprises the cell, input gate, output gate and forget gate.

The cell remembers values over arbitrary time intervals, and the three gates regulate the flow of information into and out of the cell. The cell of the model is responsible for keeping track of the dependencies between the elements in the input sequence. The input gate controls the extent to which a new value flows into the cell, the forget gate controls the extent to which a value remains in the cell, and the output gate controls the extent to which the value in the cell is used to compute the output activation of the LSTM unit.

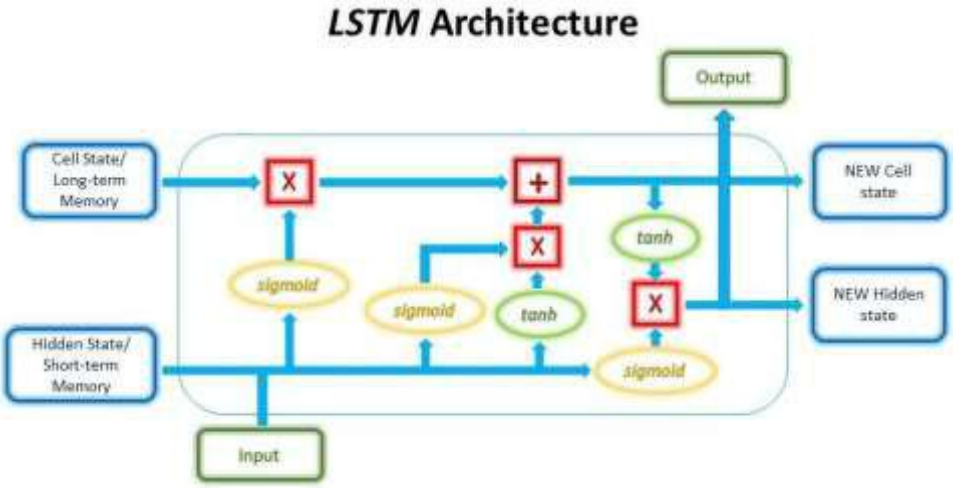


Fig. 7.1: LSTM Architecture

Image credit: https://blog.floydhub.com/long-short-term-memory-from-zero-to-hero-with-pytorch/

LSTM Networks are popularly used on time-series data for classification, processing, and making predictions. The reason for its popularity in time-series application is that there can be several lags of unknown duration between important events in a time series.

7.2. Streamlit

Streamlit is an open-source python framework for building web apps for Machine Learning and Data Science. We can instantly develop web apps and deploy them easily using Streamlit. Streamlit allows you to write an app the same way you write a python code. Streamlit makes it seamless to work on the interactive loop of coding and viewing results in the web app.

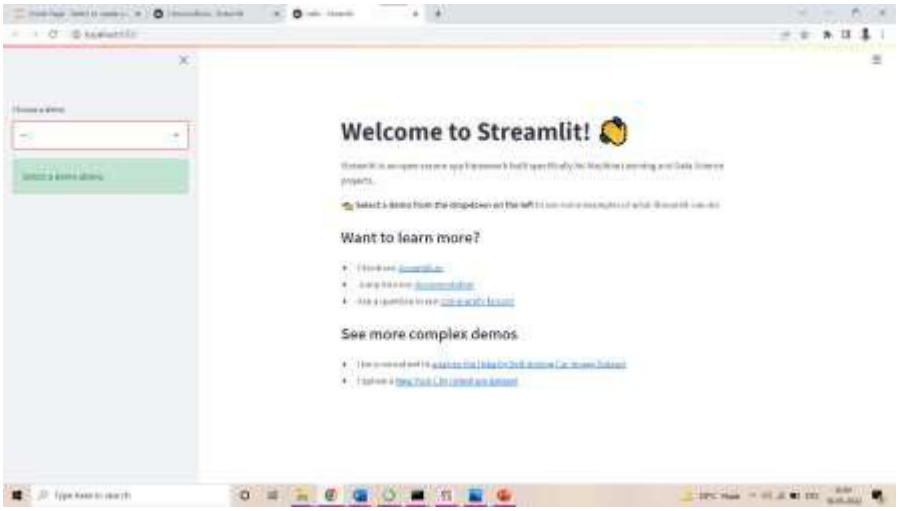


Fig.7.2: Streamlit App Demo

8. Proposed Approach

Over the years, various machine learning techniques have been used in stock market prediction, but with the increased amount of data and expectation of more accurate prediction, the deep learning models are being used nowadays which have proven their advantage over traditional machine learning methods in terms of accuracy and speed of prediction. In this project, we will use the Long-Short-Term Memory (LSTM) Recurrent Neural Network, one of the popular deep learning models, used in stock market prediction. In this task, we will fetch the historical data of stock automatically using python libraries and fit the LSTM

model on this data to predict the future prices of the stock**.**

8.1. Model Training

Steps:

1. Load the data
2. Pre-process the data in order to prepare it for the LSTM model.
3. Define the LSTM Recurrent Neural Network. Here, we can add more LSTM layers and adjust the dropout in order to improve the accuracy of the model.
4. Compile and train the model defined in the above step. Iteratively, we can increase or decrease the epochs and batch size to get more accuracy.
5. Now, our model is trained.

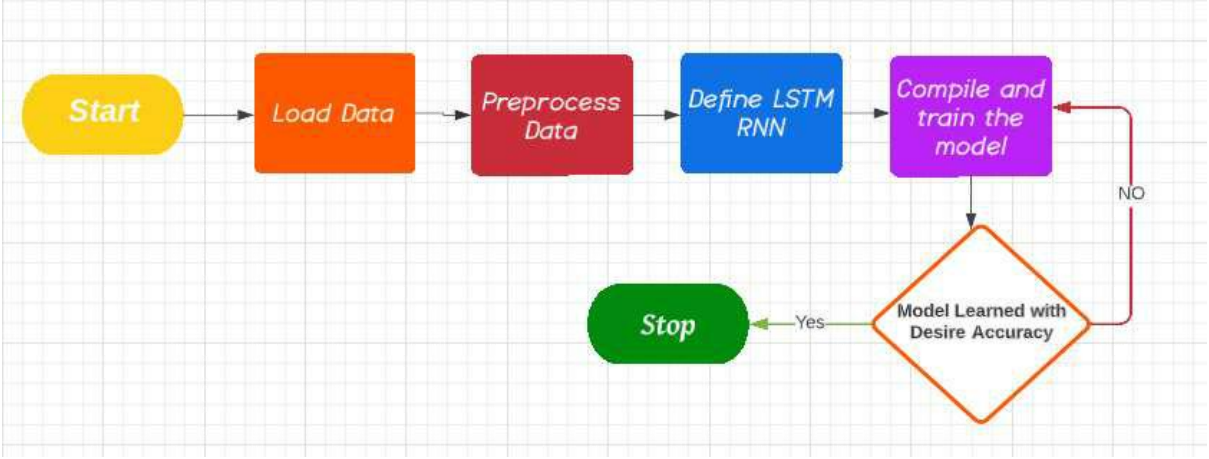


Fig. 8.1: Flow Chart for Model Training

8.2. Prediction

Steps:

1. Load the data
2. Pre-process the data
3. Load the LSTM model
4. Predict the values
5. Plot the predicted values

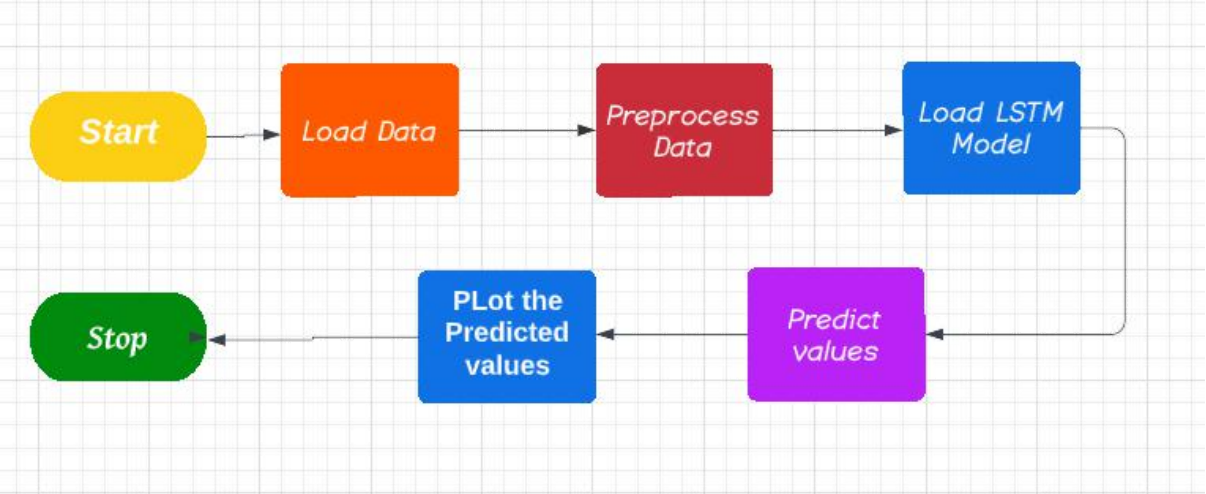


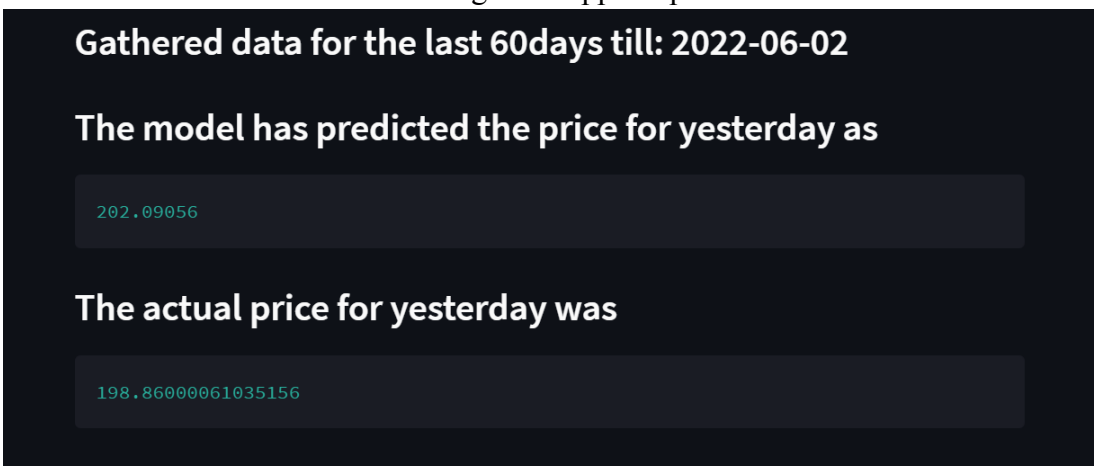
Fig. 8.2: Flow Chart for Prediction

***

9. Outcome

We have built a simple, user-friendly web-application using streamlit which is a python framework which is integrated with our stock market prediction model. In this app stock prediction is just a matter of a few clicks. Besides prediction, users can view the top stock tickers data and chart on the Home Page, can view the historical data, moving averages of a stock ticker and can also compare the data of two stock tickers. The application can be used both in light and dark modes as per the convenience of the user. Some snapshots of application are as follows:

Fig. 9.1: App Snapshot1



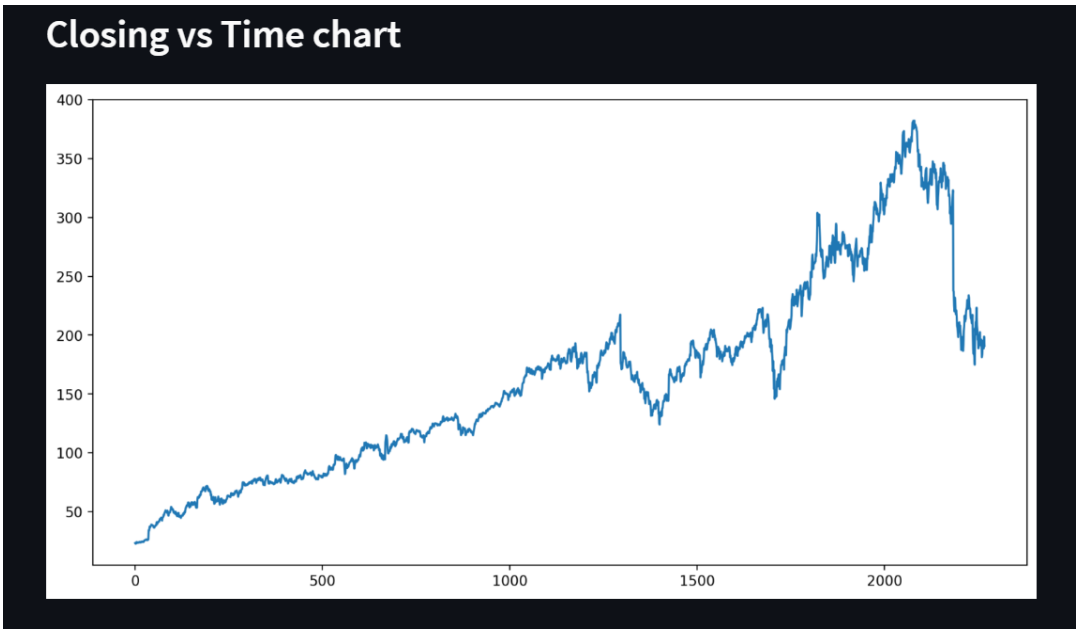


Fig. 9.2: App Snapshot2

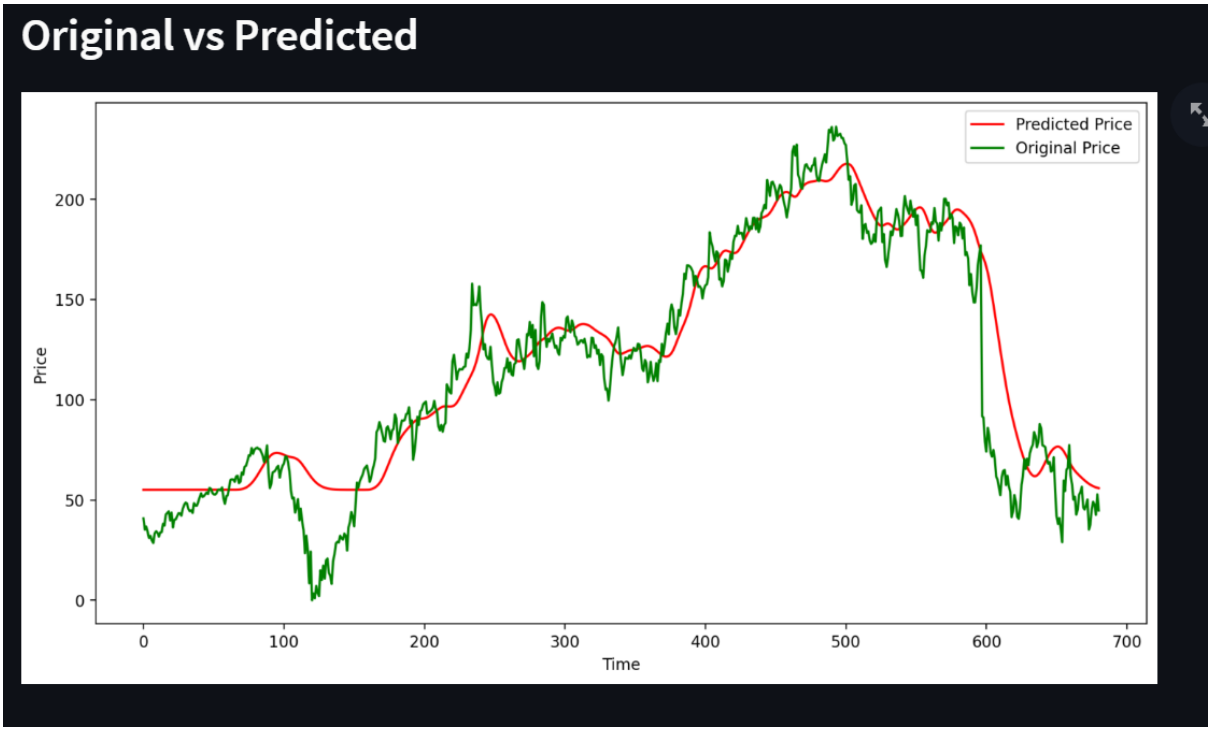


Fig. 9.3: App Snapshot3

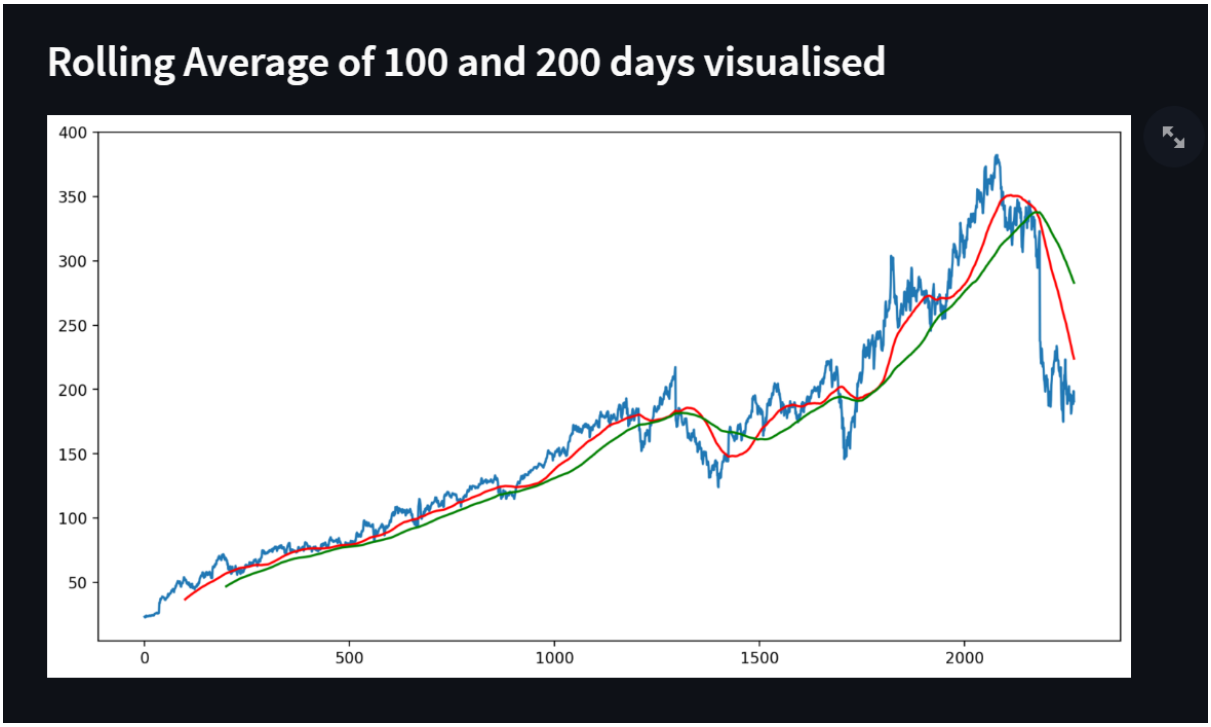


Fig. 9.4: App Snapshot4

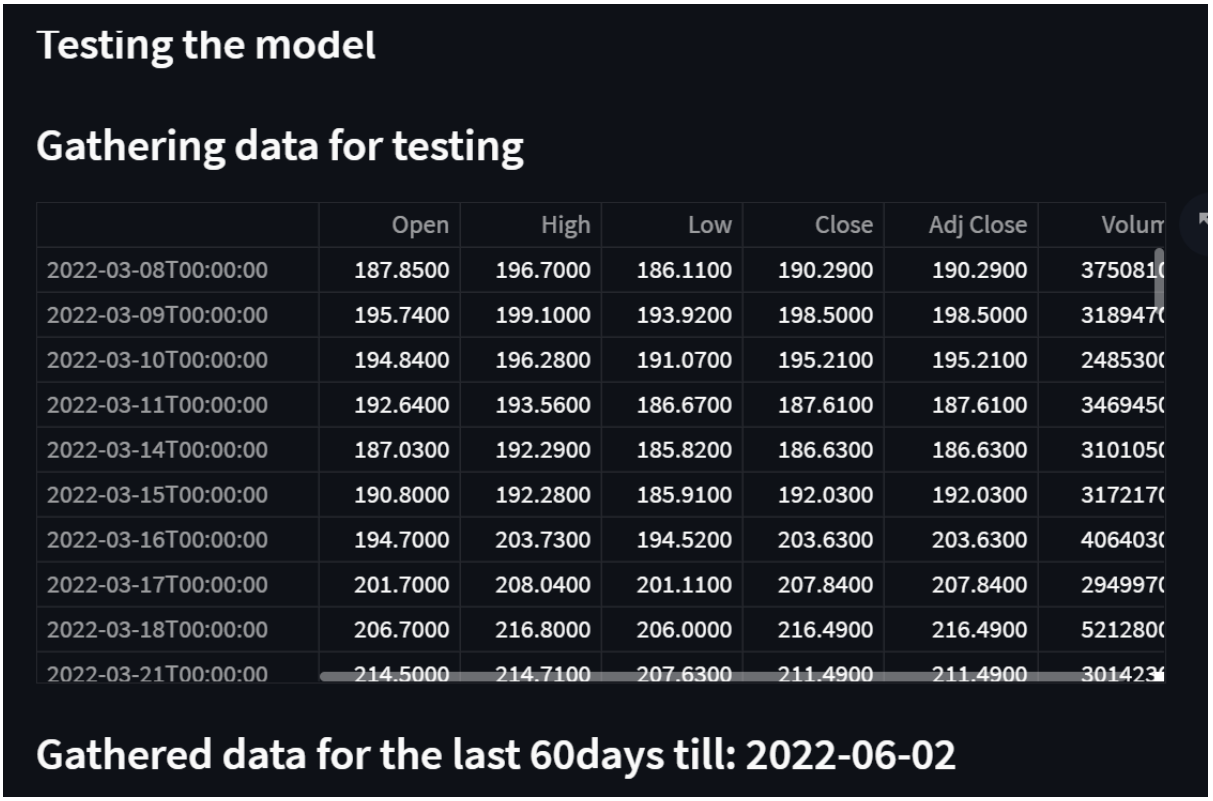


Fig. 9.5: App Snapshot5

10.Illustration

1. The historical prices of AAPL are collected automatically using the pandas_datareader of python. We have used 10 years of historical price data, from 01.01.2010 to 31.12.2018.

Date	High	Low	Open	Close	Volume	Adj Close
2009-12-31	7.619643	7.520000	7.611786	7.526071	352410800.0	6.434926
2010-01-04	7.660714	7.585000	7.622500	7.643214	493729600.0	6.535085
2010-01-05	7.699643	7.616071	7.664286	7.656429	601904800.0	6.546384
2010-01-06	7.686786	7.526786	7.656429	7.534643	552160000.0	6.442256
2010-01-07	7.571429	7.466071	7.562500	7.520714	477131200.0	6.430345
...	...	...	...	...	...	...
2019-12-24	71.222504	70.730003	71.172501	71.067497	48478800.0	69.835587
2019-12-26	72.495003	71.175003	71.205002	72.477501	93121200.0	71.221161
2019-12-27	73.492500	72.029999	72.779999	72.449997	146266000.0	71.194138
2019-12-30	73.172501	71.305000	72.364998	72.879997	144114400.0	71.616669
2019-12-31	73.419998	72.379997	72.482498	73.412498	100805600.0	72.139946

2517 rows × 6 columns

Table 10.1: Stock Data of AAPL from 2010-01-01 to 2019-12-31

2. Before Processing the data, we need to understand and visualise the it to get a sense of what the data entails as well as the identification of features that will be important to us.

	High	Low	Open	Close	Volume	Adj Close
count	2517.000000	2517.000000	2517.000000	2517.000000	2.517000e+03	2517.000000
mean	27.816644	27.322083	27.571889	27.578353	2.996141e+08	25.556786
std	14.147728	13.923162	14.025156	14.046141	2.352356e+08	14.157137
min	7.000000	6.794643	6.870357	6.858929	4.544800e+07	5.864506
25%	16.343929	16.043928	16.235001	16.177500	1.247604e+08	14.028781
50%	25.055000	24.592501	24.770000	24.860001	2.220316e+08	22.638868
75%	37.957500	37.290001	37.567501	37.637501	4.067224e+08	35.803486
max	73.492500	72.379997	72.779999	73.412498	1.880998e+09	72.139946

Table 10.2: Description of Stock Data of AAPL from 2010-01-01 to 2019-12-31

Visualization of Data:

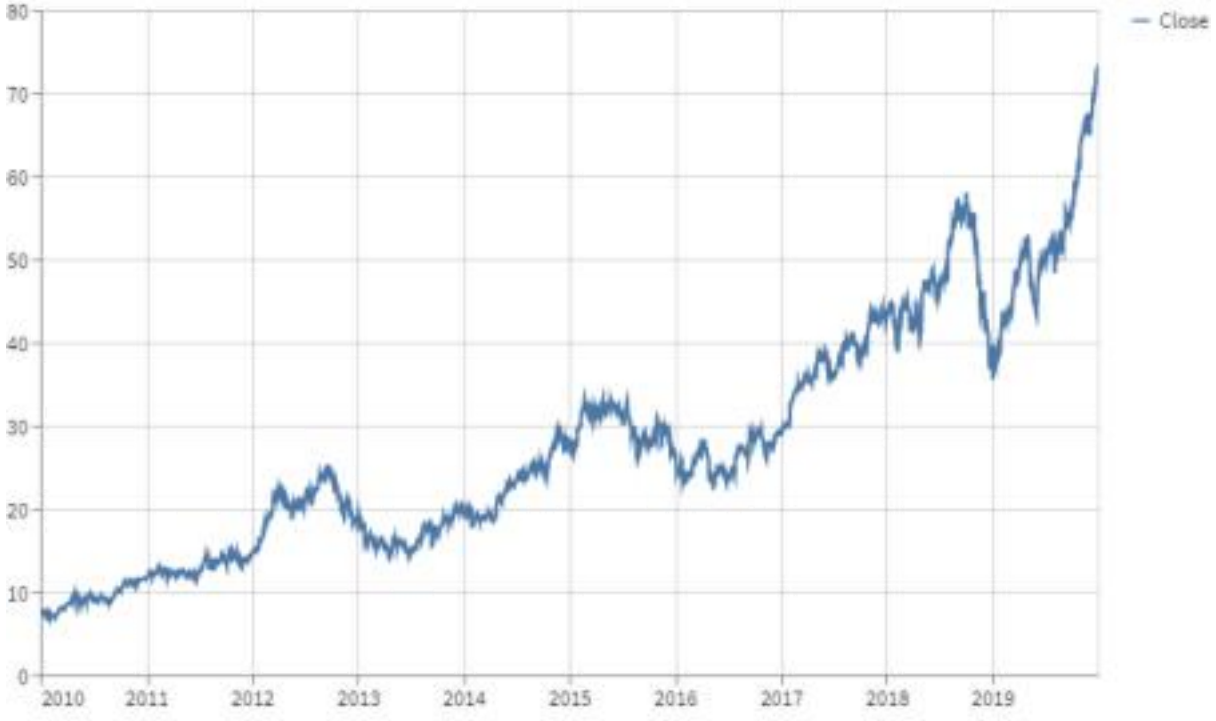


Fig. 10.1: Close vs Time Plot

Fig.10.3: Low vs Time Plot

Fig.10.4: High vs Time Plot

Fig 10.5: (Close, MA100, MA200) vs Time Plot

3. We Pre-process the data in order to prepare it for the LSTM model.

4. Then, we split the dataset into training and testing set. We use 70% data for training and 30% for testing.

5. Define the LSTM Recurrent Neural Network. Here, you can add more LSTM layers and adjust the dropout in order to improve the accuracy of the model.

Fig. 10.6.: Code snippet for Defining LSTM RNN

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 100, 50)           10400

 dropout (Dropout)           (None, 100, 50)           0

 lstm_1 (LSTM)               (None, 100, 60)           26640

 dropout_1 (Dropout)         (None, 100, 60)           0

 lstm_2 (LSTM)               (None, 100, 80)           45120

 dropout_2 (Dropout)         (None, 100, 80)           0

 lstm_3 (LSTM)               (None, 120)               96480

 dropout_3 (Dropout)         (None, 120)               0

 dense (Dense)               (None, 1)                 121

=================================================================
Total params: 178,761
Trainable params: 178,761
Non-trainable params: 0
```

Fig. 10.7: Summary of LSTM model

6. Next, Compile and train the model defined in the above step. Iteratively, we can increase or decrease the epochs and batch size to get more accuracy.

```
model.compile(optimizer = 'adam', loss = 'mean_squared_error')
model.fit(x_train,y_train,epochs = 50)
```

Fig 10.8: Code Snippet for compiling model

7. Now, our model is trained and needs to be tested on the testing data.

8. Lastly, we visualize the predicted stock prices with original stock prices.

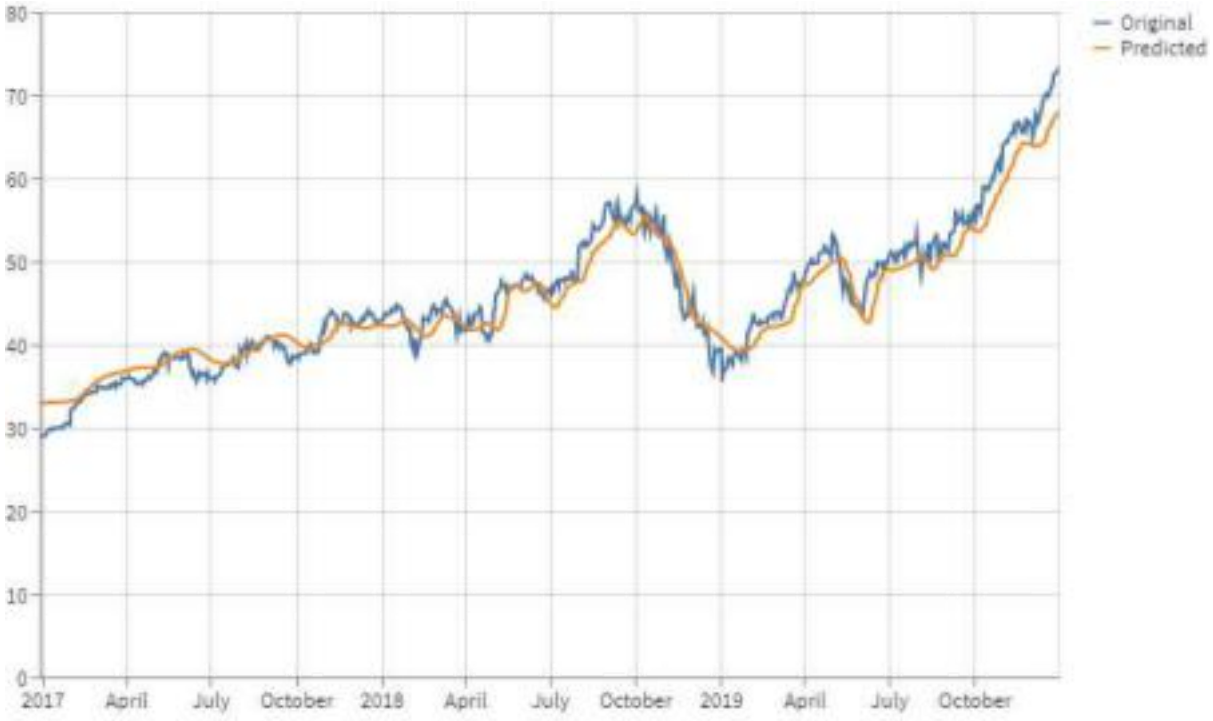


Fig. 10.9: predicted stock prices with original stock prices

9. We can also predict the price of the stock for upcoming days. Here we have predicted AAPL stock price for next 30 days.

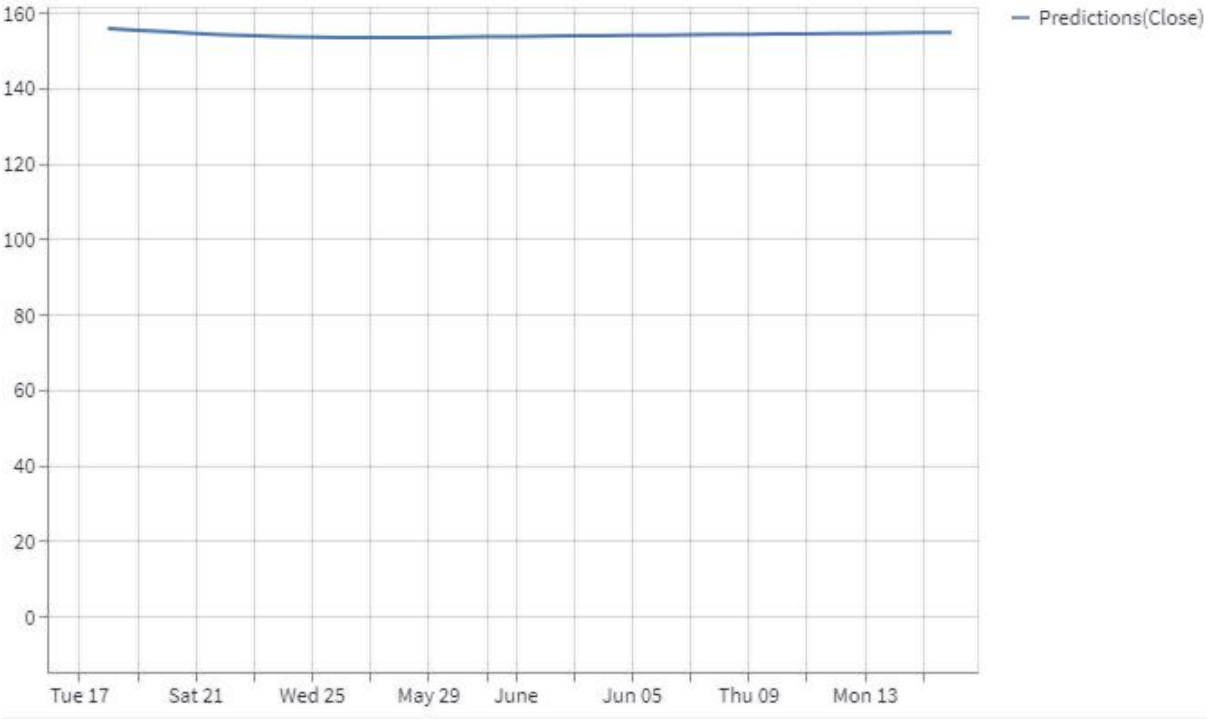


Fig. 10.10: Future Prediction vs Time Plot

	Predictions(Close)
2022-05-18T00:00:00	155.7799835205078
2022-05-19T00:00:00	155.3458709716797
2022-05-20T00:00:00	154.9078369140625
2022-05-21T00:00:00	154.49560546875
2022-05-22T00:00:00	154.13539123535156
2022-05-23T00:00:00	153.84486389160156
2022-05-24T00:00:00	153.630859375
2022-05-25T00:00:00	153.48928833007812
2022-05-26T00:00:00	153.4113311767578
2022-05-27T00:00:00	153.38441467285156

Table 10.3.: Future Prediction Table

***

11. Conclusion

LSTM model for stock prediction has been trained using AAPL stock data from 01.01.2010 to 31.12.2018. We have used 4 hidden layers in LSTM model and ‘relu’ activation function of neurons. The loss function was ‘mean-squared-error’ and we have used ‘adam’ optimizer.

We have tested our model on different data set of various stock-tickers and over different period of time. The performance of our model is quite good. We have integrated this model with a web-application so that the user can easily get access to our model.

***

12. References

[1] Deep Learning Mooc by IIT Madras, NPTEL
https://www.youtube.com/watch?v=aPfkYu_qiF4&list=PLyqSpQzTE6M9gCgajvQbc68Hk_JKGBAYT

[2] Article on Hands-On Guide To LSTM Recurrent Neural Network For Stock Market Prediction
https://analyticsindiamag.com/hands-on-guide-to-lstm-recurrent-neural-network-for-stock-market-prediction/

[3] Analysis of Stock Price Predictions using LSTM models by Yu Hao Lee
https://medium.com/analytics-vidhya/analysis-of-stock-price-predictions-using-lstm-models-f993faa524c4

[4] Goodfellow, I., Bengio,Y., and Courville, A., Deep Learning, MIT Press, 2016.

[5] Bishop, C.M., Pattern Recognition and Machine Learning, Springer, 2006.

[6] Yegnanarayana, B., Artificial Neural Networks PHI Learning Pvt. Ltd, 2009.

[7] Golub, G.H. and Van Loan,C.,F., Matrix Computations, JHU Press,2013.

[8] Satish Kumar, Neural Networks: A Classroom Approach, Tata McGraw-Hill Education, 2004

[9] What is LSTM? By IBM Technology

https://www.youtube.com/watch?v=b61DPVFX03I

[10] Simple Explanation of LSTM by Codebasics
https://www.youtube.com/watch?v=LfnrRPFhkuY&t=5s

[11] Optimizing LSTM for time series prediction in Indian stock market by Anita Yadava, C K Jha, Aditi Sharan

***