A Mini Project Synopsis on

ML based Stock Prediction

Submitted By

VATSAL SINGH	**19104006**
SHUBHANGI TRIPATHI	**19104067**
PUSHKAR TELAVANE	**19104011**

Under The Guidance Of

Prof. Apeksha Mohite

Prof. Sneha Dalvi

DEPARTMENT OF INFORMATION TECHNOLOGY

A.P.SHAH   INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W), Mumbai-400615

UNIVERSITY OF MUMBAI

Academic year : 2021-22

CERTIFICATE

This to certify that the Mini Project report on ML based Stock Prediction has been submitted by Vatsal Singh(19104006), Shubhangi Tripathi(19104067), Pushkar Telavane(19104011) who are a bonafide students of A.P. Shah Institute of Technology, Thane, Mumbai, as a partial fulfilment of the requirement for the degree in Information Technology, during the academic year 2021-2022 in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Prof. Apeksha Mohite
Guide

Prof. Sneha Dalvi
 Co-Guide

Prof. Kiran Deshpande   Dr. Uttam D.Kolekar
Head Department of Information Technology   Principal

External Examiner(s)
1.
2.

Place:A.P.Shah Institute of Technology, Thane
Date:

ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our **guide Prof. Apeksha Mohite and Prof. Sneha Dalvi.** We express gratitude towards our **HOD, Prof. Kiran Deshpande**, and the Department of Information Technology for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our teachers who gave us their valuable suggestions and ideas when we needed them. We would also like to thank our peers for their helpful suggestions.

TABLE OF CONTENTS

Chapter No:1

INTRODUCTION

The share market is a place where the shares of a public company are traded. As discussed in the volatile nature of the stock market makes it an area which needs an abundance of analysis with the old data predicated. The previous stock trend prediction algorithms use the historic time series stock data. the typical scientific stock price forecasting procedures are focused on the statistical analysis of stock data. In the paper will develop a stock data predictor program that uses previous stock prices and data will be treated as training sets for the program to predict the stock prices of a particular share this program develops a procedure.

This model considers the historical equity share price of a company price and applies RNN (Recurrent) technique called Long Short Term Memory (LSTM). The proposed approach considers available historic data of a share and it provides prediction on a particular feature. The features of shares are Opening price, day High, day Low, previous day price, Close price, Date of trading, Total Trade Quantity and Turnover. The proposed model uses the time series analysis in order to predict a share price for a required time span. The proposed will be considering Indian stock exchange Company named as The National Stock Exchange of India Limited (NSE). The National Stock Exchange (NSE) is the Indian stock exchange entity, the NSE was the first exchange in India to provide a modern, provides latest facility to the investors spread across the length and breadth of the country. It has thoroughly modern with all latest facilities, which provides investors with the facility to trade from anywhere in India. This has a decisive role in reforming the Indian equity market to add increased transparency, convergence and efficiency to the capital market. NSE's Common Index, The CNX NIFTY, is used prodigiously by the investor across India as well as globally. It provides accommodation for the exchange, settlement and clearing in equity and debt market and additionally in derivatives. This is one of India's most astronomically enormous mazuma, currency and index options trading exchanges worldwide. There are numerous domestic and ecumenical companies which have an interest in the exchange. Several regional companies include TATA, WIPRO, HDFC and YES BANK ltd.

OBJECTIVES

- To predict the stock price such that he/she can sell it before its value decline or buy the stock before the price rises.

- To determine the future movement of the stock value of a financial exchange.

- To predict the price of the share in the stock using recent/previous timeline data.

- Through this application users can identify the factors affecting the price of the share market. This application generates the pattern from large set of data of stock market for prediction of BSE & NSE.

SCOPE

- The choice of machine learning model has been based in previous research where the LSTM architecture has proven effective in prediction with regards to financial time series data.

- The main aim is to build an application in such a way that it will provide a platform where a stock price prediction of all the companies under BSE&NSE will be displayed.

- This website comparatively analyzes the effectiveness of prediction algorithms on stock price prediction and get general insight on this data through visualization to predict future stock behavior and value at risk for each stock.

- This website will use LSTM method to predict future stock returns based on past.

Chapter No: 2

PROBLEM DEFINITION

This project aims to explore the possibility of predicting stock market movements through one-day ahead financial time series. The study is an analysis of long short-term memory recurrent neural networks. The objective of this thesis is to evaluate the performance of LSTM on financial time series data, and how the number of time steps for LSTM impacts the predictive power of the model. Time Series forecasting & modelling plays an important role in data analysis. Time series analysis is a specialized branch of statistics used extensively in fields such as Econometrics & Operation Research. Time Series is being widely used in analytics & data science. Stock prices are volatile in nature and price depends on various factors. The main aim of this project is to predict stock prices using Long short-term memory (LSTM).

Chapter No: 3

PROPOSED SYSTEM

3.1 Features & Functionality:

- Latest Updates on Prices of Stocks:

  Our website basically provides the latest Updates on prices of BSE and NSE which help the users to stay up to date.

- History of Stocks records:

  There are many sites to get historical stock prices but accessing them may be a bit more difficult. So, we have decided to provide a website which keep all the record perfectly managed.

- Multiple Stocks to Study:

  The purpose of our framework is to analyze which is the best time span to predict the future share price of a company from a particular sector.

- Graphical Stats :

  The graphical charts makes easier to predict the future prices. Rather than from the heap of numerical stock prices.

3.2 LITERATURE SURVEY

[1]The research work done in 2008 in order to predict market movement, investors used to analyze the stock prices and stock indicators using Neuro Fuzzy Logic in addition to the news related to these stocks. Hence, the importance of news on the stock price movement.   Firstly, it helps us understand the impact of increasing / decreasing the dataset vertically or horizontally on computational time. Secondly, it helps us understand the situations or cases where the model fits best. NFL fail to capture correlation between stock prices in the form of long-term temporal dependencies.

[3]In the Year 2018, the research work done traditionally and in order to predict market movement, investors used to analyze the stock prices and stock indicators in addition to the news related to these stocks. Hence, the importance of news on the stock price movement. The use LSTM Algorithm to predict stock prizes after observation he provided some important parts i.e., Very High accuracy in image recognition problems. Automatically detects the important features without any human supervision. LSTM does not encode the position and orientation of objects. Lack of ability to be spatially invariant to the input data and more time is required to train the data.

[2]In the year 2019 has done some research for predicting the price correlation of two assets for future time periods are important in portfolio optimization. The Applied recurrent neural networks (RNN) in predicting the stock price correlation coefficient of two individual stocks. RNN's are competent in understanding temporal dependencies. There is also some important points regarding RNN Algorithm Recurrent Neural Networks (RNNs) are a class of Artificial Neural Networks which are often used with sequential data. Recurrent neural network is even used with convolutional layers to extend the effective pixel neighborhood.Training an RNN is a very Difficult Task.

[6]In 2020, the observation is done that ARIMA (Auto Regressive Integrated Moving Algorithm) is used which robust and efficient and the main disadvantage for this model is that it is only good for short term.

[4]So, the main reason for using LSTM is that it uses three gates are placed in an LSTM unit, LSTM is used instead of RNN to avoid exploding and vanishing gradients. LSTM networks as training methodologies to analyze their effectiveness in forecasting out of-sample directional movements of constituent stocks. He also provided information that the principal advantage of LSTM over ARIMA, Neuro Fuzzy. The main advantage of LSTM is that LSTM can model a collection of records (i.e., time collection) so that each pattern can be assumed to be dependent on previous ones. LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of me is practically their default.

YEAR	NAME	ALGORITHM	ADVANTAGES	DISADVANTAGES
2008	Neuro Fuzzy SP Forecast	Neuro Fuzzy logic	Forecast easy on one trend(seasonal) and small data	Hard to predict for any different train recurring(long) dataset
2019	RNN SP	Recurrent Neural Network	Previous time points input layer contains input	It is possible to feed those words in through much small set of inputs nodes
2020	SP using ARIMA	Auto Regressive Integrated Moving Average	Robust and efficient	Only for short term prediction
2018	SPP ML in Python (TATA Global Dataset, stocks)	LSTM	Used for Higher Accuracy	More memory usage, Longer time period to train

Table1: Summary of Literature Survey

Chapter No: 4

PROJECT OUTCOMES

• The movement in the stock market is usually determined by the sentiments of thousands of investors. Stock market prediction, calls for an ability to predict the effect of recent events on the investors.

• The Opening Value of the stock, the Highest and Lowest values of that stock on the same days, as well as the Closing Value at the end of the day, are all indicated for each date through scrapping.

• Predicting the stock market was a time-consuming and laborious procedure a few years or even a decade ago.

• However, with the application of LSTM for stock price prediction forecasts, the procedure has become much simpler.

• It is beyond the scope of almost all investors to correctly and consistently predict these hyper parameters. All these factors make stock price prediction very difficult. Once the right data is collected, it then can be used to train a machine and to generate a predictive result

Chapter No: 5

Technological Stack:

5.1 Hardware Requirements:

Processor: Intel Core 4 or above

RAM: 4GB or above

5.2 Software Requirements:

Backend: Jupyter/colab

JupyterLab is the latest web-based interactive development environment for notebooks, code, and data. Its flexible interface allows users to configure and arrange workflows in data science, scientific computing, computational journalism, and machine learning. A modular design invites extensions to expand and enrich functionality.

Front end: Streamlit

Streamlit is a free and open-source framework to rapidly build and share beautiful machine learning and data science web apps. It is a Python-based library specifically designed for machine learning engineers.

Dataset:

For dynamic dataset for the user to choose from, we have used yfinance i.e., yahoo finance library so that there is no use of particular csv file as a dataset (static dataset) as yfinance will have every stock's data in the server itself. We are just fetching the data from there to represent the graph provided for the stock by the user.

CHAPTER 6:

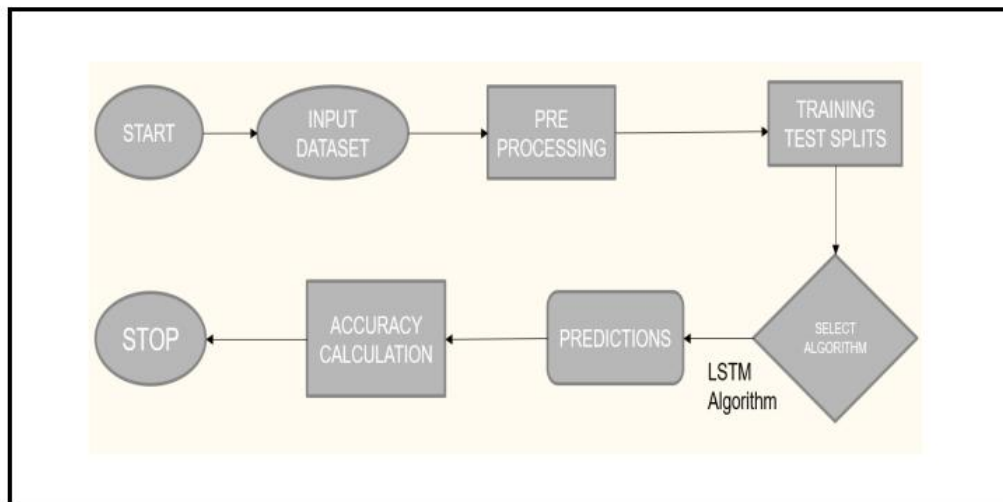# Design and Architecture

6.1 Project Architecture

Figure 1: Project Architecture

The data from the dataset is acquired by the model by splitting the data in certain percentage for training and testing purpose and predictive model is made using prediction algorithm, then make the prediction using different dataset for the accuracy of the predictive model and evaluate the accuracy of the model prediction for the required stock.

LSTM

LSTM follow this chain-like structure, however the repeating module has a different structure. Instead of having a single neural network layer, there are four layers, interacting in a very special way as shown
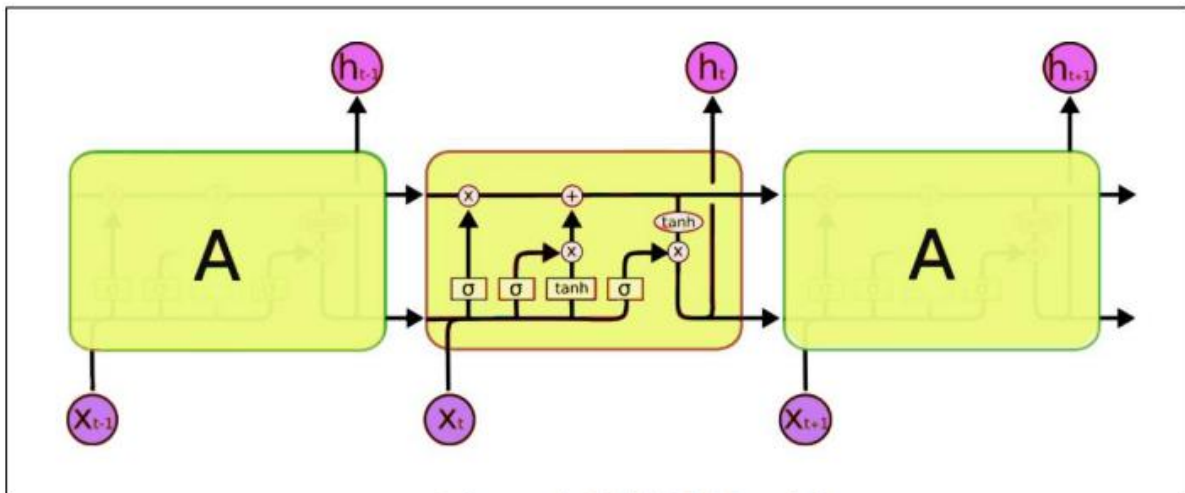

Figure 2: LSTM Algorithm

Purpose:

- The proposed algorithm using the market data to predict the share price using machine learning techniques like recurrent neural network named as Long Short-Term Memory (LSTM).

- This system will provide accurate outcomes in comparison to currently available stock price predictor algorithms.

- The network is trained and evaluated with various sizes of input data to urge the graphical outcomes.
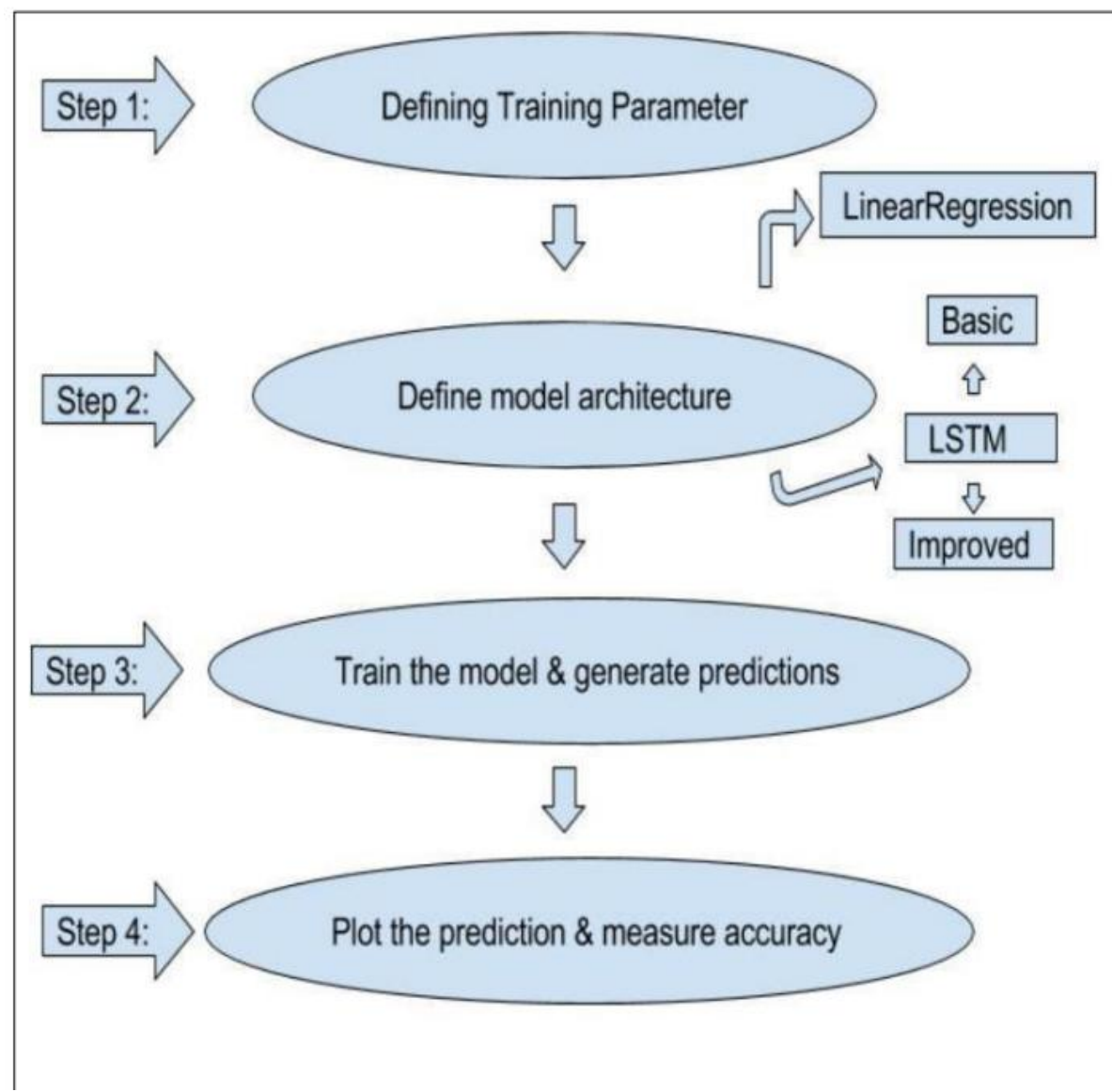
LSTM Model

Figure 3: LSTM Model

In LSTM Model, defining the training parameter is required and model architecture is defined (here it is LSTM) to train the model and generate predictions and to plot the prediction and measure the accuracy of the predictive model.

6.2 PROJECT DESIGN

This project will be implemented through the keras flow library using LSTM Neural Networks. Development workflow will follow the below sequence:

1) Set Up Infrastructure :

  • Incorporate required Libraries (Keras, Pandas, Matplotlib, NumPy).

2) Prepare Dataset :

  • Process the requested data into pandas Data frame

  • Develop function for normalizing data

3) Develop Basic LSTM Model   :

  •Set up basic LSTM model with Keras utilizing some parameters

4) Train LSTM Model :

  • Develop, document, and Compare results using additional labels for the LSTM model
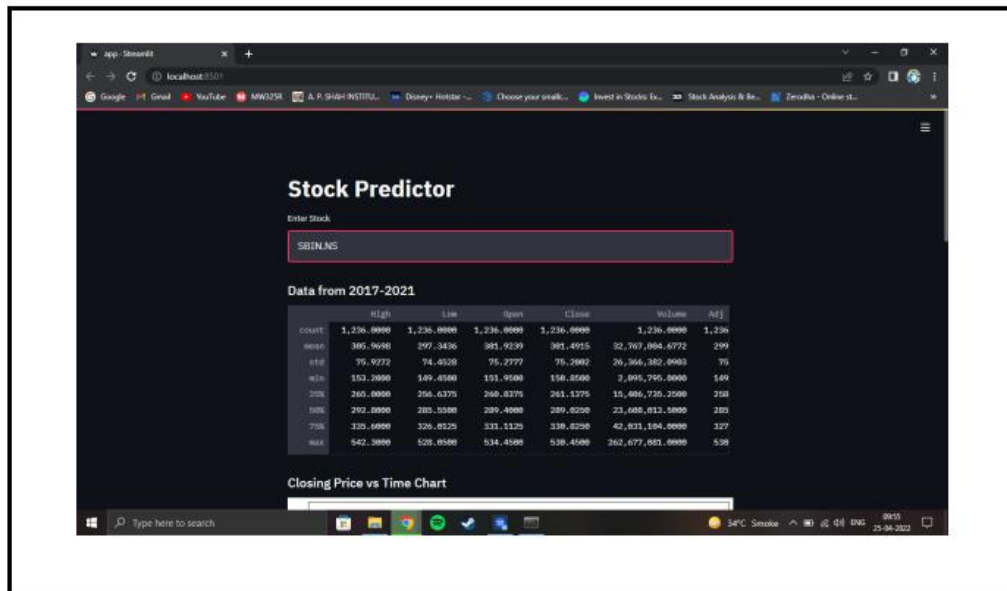
Chapter No.7:

Project Implementation:

Figure 4: Stock Data Description

Dashboard Page: A dashboard is a visual representation of the most important information needed to achieve one or more goals, consolidated and arranged on a single screen so that information can be controlled at a glance.

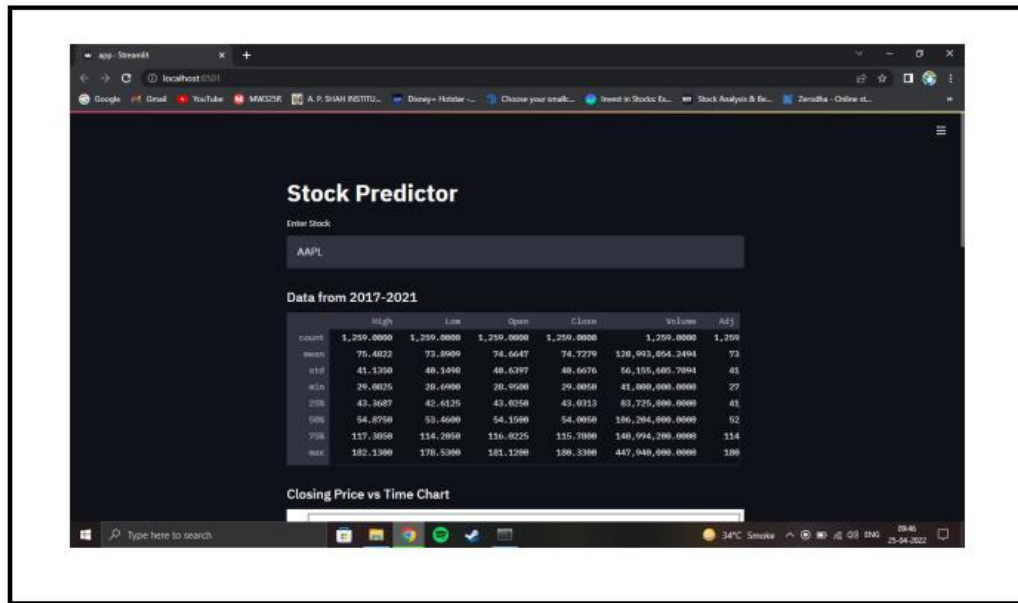(The following page is same as before for different stock)
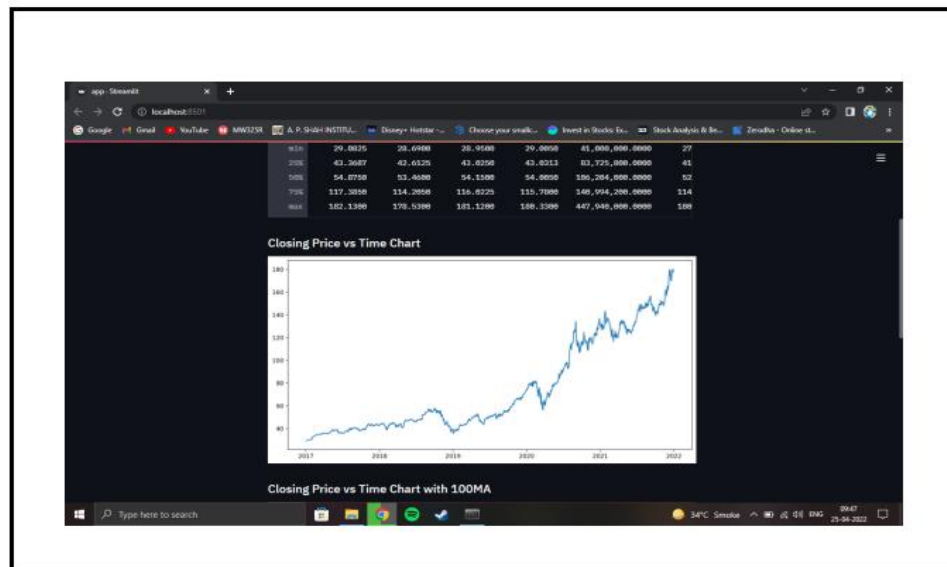

Figure 5: AAPL Stock

Figure 6: Closing vs Time Chart

Visualization of Graph: Graph of the data provided from start date to end date table above is represented in terms of Closing Price vs Time Chart.

Chapter No: 8

PROJECT SCHEDULING

| Group Member	Time	Work to be done
Pushkar Telavane	2nd week of February	Literature Review, reading papers and choosing algorithm.
Shubhangi Tripathi	4rd week of February	Training the dataset on the Colab notebook.
Vatsal Singh	1st week of March	Learning and testing the functionalities and properties of the LSTM.
Everyone	3rd week of March	Creating a file uploading option with python.
Everyone	4th week of March	Connecting the improved UI with python engine on Streamlit.
Everyone	2nd week of April	Testing and fixing errors and making required changes. Making of Report and Presentation.

GANTT CHART
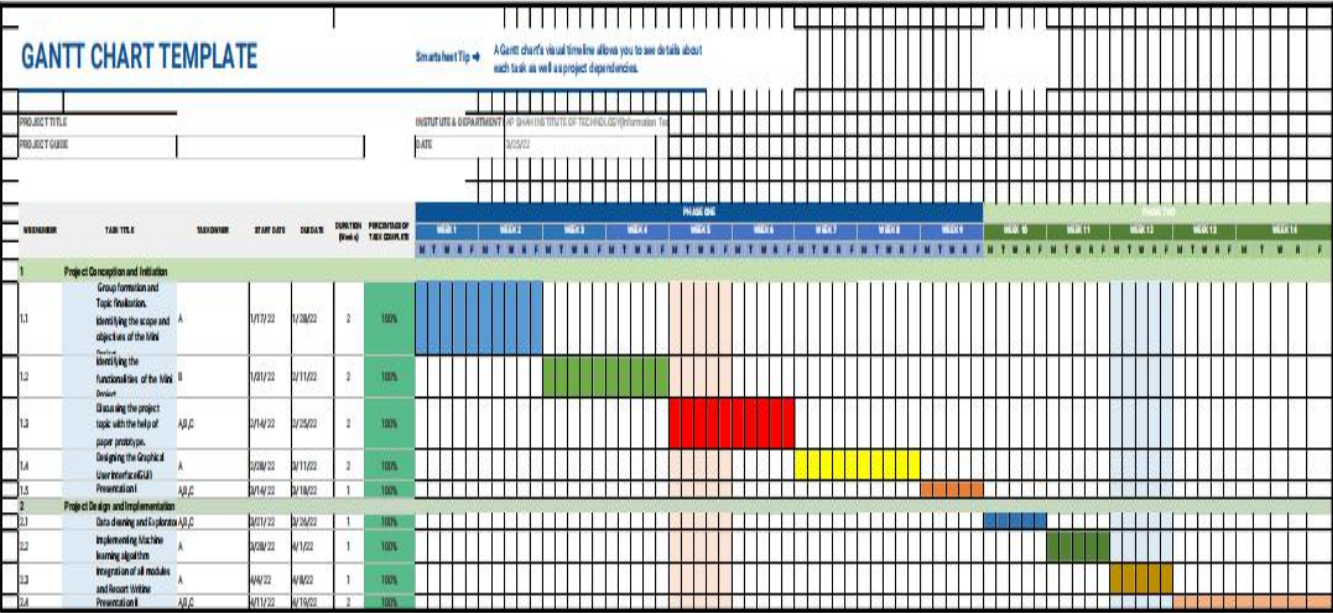
Figure 7 : Gantt Chart

Gantt chart help teams to plan work around the deadlines and properly allocate resources. Projects planners also use Gantt charts to maintain a bird's eye view of projects. They depict, among other things, the relationship between the start and end dates of tasks, milestones, and dependent tasks. Modern Gantt chart programs such as Jira Software with Roadmaps and Advanced Roadmaps synthesize information and illustrate how choices impact deadlines.

Chapter No: 9

CONCLUSION

In this project, we analyze the growth of the companies from different sector and try to find out which is the best time span for predicting the future price of the share. So, this draws an important conclusion that companies from a certain sector have the same dependencies as well as the same growth rate. The prediction can be more accurate if the model will train with a greater number of data set. Moreover, in the case of prediction of various shares, there may be some scope of specific business analysis. We can study the different pattern of the share price of different sectors and can analyze a graph with more different time span to fine tune the accuracy. This framework broadly helps in market analysis and prediction of growth of different companies in different time spans. Incorporating other parameters (e.g., investor sentiment, election outcome, geopolitical stability) that are not directly correlated with the closing price may improve the prediction accuracy. The study indicates that varying time steps on this particular LSTM model does not impact the predictive powers of the model to a large extent, however 10-time steps seem to be the optimal amount of time steps among the different settings used in this model.

REFERENCES

[1]Neuro-fuzzy modeling and control by J.-S.R. Jang; Chuen-Tsai Sun from page 378 - 406 on March 1995

[2]A comparative study of recurrent neural network models for lexical domain classification by Suman Ravuri; Andreas Stolcke published in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

[3]Author Raimi Karim " Animated RNN , LSTM and GNU " published in Year 2018 ef124d06cf45.

[4] R.V.Argiddi, S.S.Apte "a study of association rule mining and in fragmented item-sets for prediction of transactions outcome in stock trading systems" IJCET-2012.

[5] Wei Huang, Yoshiteru Nakamori, Shou-Yang Wang, "Forecasting stock market movement direction with support vector machine", Computers & Operations Research, Volume 32, Issue 10, October 2005, Pages 2513–2522.

[6] Based on Time Sequence of ARIMA Model in the Application of Short-Term Electricity Load Forecasting by Li Wei; Zhang Zhen-gang 29 January 2010.

[8] Jen-Zhung Chien ,Siyuan Liu, Guangzhong Liao, Yifan Ding, " Stock Transaction Prediction Modeling and Analysis Based on LSTM ", 2019 13th IEEE Conference on Industrial Electronics and Applications(ICIEA), DOI:10.1109/ICIEA.2018.8398183.