Chandigarh University

NOVEMBER 2022

Stock Price Prediction

A PROJECT REPORT

Submitted by

DEVANSH SINGH 20BCS9102

JAYANT AGGARWAL 20BCS9143

ANKIT KUMAR SINGH 20BCS7529

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

Certified that this project report **"STOCK PRICE PREDICTION"** is the bonafide work of "DEVANSH, JAYANT, ANKIT**"** who carried out the project work under my supervision.

Signature of HOD   Signature of
supervisor

Submitted for the project viva-voce examination held.

INTERNAL EXAMINER   **EXTERNAL EXAMINER**

TABLE OF CONTENT

ABSTRACT

The prediction of stock value is a complex task which needs a robust algorithm background in order to compute the longer term share prices. Stock prices are correlated within the nature of market; hence it will be difficult to predict the costs. 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, in that process weights are corrected for each data points using stochastic gradient descent. 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. Keywords: Machine Learning, Stock Price Prediction, Long Short-Term Memory, Stock Market, Artificial neural Networks, National Stock Exchange.

CHAPTER 1.
INTRODUCTION

The share market is a place where the shares of a public company are traded. As discussed in [7] 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 o 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. Among pilgrim investors, few are strategic holdings of the city party, Mauritius limited, Tiger Ecumenical five holdings.

1.1. Stock Market:

A stock market is a public market where we can buy and sell shares for the listed companies. This share is known as equities and represents the ownership of the companies. The stock exchange is a mediator who allows us to buy and sell stock.

Importance of stock market:

- Stock markets help companies to raise capital.

- It helps generate personal wealth.

- Stock markets serve as an indicator of the state of the economy.

- It is a widely used source for people to invest money in companies with high growth potential.

1.2. Stock Price Prediction:

- Stock price prediction using machine learning helps us to discover the future values of the stock. The entire idea of predicting stock prices is to gain important benefits. Predicting the stock price is a challenging task to do. There are many factors involved in the prediction such as physical and psychological factors, rational and irrational behaviour, and so on. All the factors combine to make share prices dynamic and volatile. This is exceedingly difficult to predict stock prices with high accuracy.

1.3. Machine Learning:

- Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

- Recommendation engines are a common use case of machine learning. Popular use case of machine learning. Popular use case of machine learning such as fraud detection, spam filtering, malware threat detection, etc.

1.4. Long Sort Term Memory Network:

- Here, we use Long Short Term Memory Network (LSTM) to develop a model to predict the stock prices of Google.

- LSTM are the type of Recurrent Neural Network which is used to predict long term dependencies. It is commonly used for processing and predicting time-series data.
- LSTM has a chain like structure. LSTM has four interacting layers communicating extraordinarily
- LSTMs (Long Short-Term Memory) work in a three-step process.
    - The first step in LSTM is to decide which information to be omitted from the cell in that time step. It is decided with the help of a sigmoid function. It looks at the previous state (ht-1) and the current input xt and computes the function.
    - There are two functions in the second layer. The first is the sigmoid function, and the second is the tanh function. The sigmoid function decides which values to let through (0 or 1). The tanh function gives the weightage to the values passed, deciding their level of importance from -1 to 1.
    - The third step is to decide what will be the final output. First, you need to run a sigmoid layer which determines what parts of the cell state make it to the output. Then, you must put the cell state through the tanh function to push the values between -1 and 1 and multiply it by the output of the sigmoid gate

- **1.5. Recurrent Neural Network (RNN)**
- A Recurrent Neural Network (RNN) is a type of artificial neural network (ANN), which uses sequential data or time series data. This deep learning algorithm is commonly used for ordinal or temporal problems, such as natural language processing (nlp), image processing and speech recognition. They are used in some popular applications such as Siri, Google Assistance, Alexa, and Google Translate.
- Like feed forward and convolutional neural networks (CNNs), recurrent neural networks utilize training data to learn. They are distinguished by their "memory" as they take information from prior inputs to influence the current input and output.

Scope

1. Justification

Existing Problem :

1. The prediction model was run on one stock set only and not on the entire market. This creates a certain degree of short sightedness on the entire evaluation process.
2. The method of prediction using Linear Regression is comparatively less accurate, when compared to other methods.
3. Does not consider Random Forest prediction model for prediction, which theoretically gives a higher accuracy when run on a small dataset.

How application will solve the problem?

1. This is the first model to solely use, Long Short-Term Memory (LSTM) as a prediction technique which gives a more accurate result.
2. This paper provides a good explanation of the LSTM model and how it is trained. This provides both the graphical and mathematical implementations.
3. The Long Short Term Memory (LSTM) is a counterfeit intermittent neural system (RNN) design[1] used in the field of deep learning, Unlike standard feed forward neural systems, LSTM has input associations.

Scope Description

In Scope

- Application would provide an interface to user for predicting stock prices.
- Application would provide an interface to user for accurate prediction of stock prices by using LSTM.

Out of Scope

- Application would provides a learning example for investors who believe in   buying the dip.
- Application can not be used to perform multiple sharing at a single time .

2 . Planning and Task Definition

Task Definitions

The project is basically divided into four major tasks:

Project Planning and Research

Project planning involves laying out all project details like project features, design, layout and schedule. All the details regarding project are discussed and written record is made to later check it if required.

Before starting up the project it is important to make research regarding project that what existing solution is available to the problem which project is going to solve and how the project to be made differs and better from them. Also, research regarding technology that is to be used for designing of the application is to be made beforehand.

Frontend Design

Frontend design include designing interface for cross platform application that will provide user interface (UI) for the application. The Frontend Task will contain several small subtasks which will together make up the application frontend using ploty dash.

Testing

Like every other application this application also need to be tested before deployed. The application needs to be tested in real environment to make sure every thing works properly and there is no bug in the application.

CHAPTER 2.
LITERATURE REVIEW

Studies using Long Sort Term Neural Network (LSTM) to predict stock price.

The first set of articles includes studies that primarily focus on stock market prediction using Long

Short Term Neural Network (LSTM). Long Short-Term Memory networks – usually just called "LSTMs" – are a special kind of RNN, capable of learning long-term dependencies. They were introduced by Hochreiter & Schmidhuber (1997), and were refined and popularized by many people in the following work. They work tremendously well on a large variety of problems, and are now widely used.

LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for extended periods of time is their default behavior, not something they struggle to learn!

All recurrent neural networks have the form of a chain of repeating modules of neural network. In standard RNNs, this repeating module will have a quite simple structure, such as a single tanh layer. LSTMs also have this chain like structure, but the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a distinct way. LSTM is a type of Recurrent Neural Network in Deep Learning that has been specifically developed for the use of handling sequential prediction problems. For example:

- Weather Forecasting
- Stock Market Prediction
- Product Recommendation
- Text/Image/Handwriting Generation

- Text Translation

Since LSTMs are effective at capturing long-term temporal dependencies without suffering from the optimization hurdles that plague simple recurrent networks (SRNs), they have been used to advance the state of the art for many difficult problems. This includes handwriting recognition and generation, language modeling and translation, acoustic modeling of speech, speech synthesis, protein secondary structure prediction, analysis of audio, and video data, among others.

Like other Neural Networks, they contain neurons to perform computation, however for LSTM, they are often referred to as memory cells or simply cells. These cells contain weights and gates; the gates being the distinguishing feature of LSTM models. There are 3 gates inside of every cell. The input gate, the forget gate, and the output gate.

Developing a dashboard using Plotly Dash.

The second set of articles includes the development of a dashboard for stock price prediction. Dash is a python framework created by plotly for creating interactive web applications. Dash is written on the top of Flask, Plotly.js and React.js. With Dash, you do not have to learn HTML, CSS, and JavaScript to create interactive dashboards, you only need python. Dash is an open source and the application build using this framework are viewed on the web browser.

Dash is a user interface library for creating analytical web applications. Those who use Python for data analysis, data exploration, visualization, modelling, instrument control, and reporting will find immediate use for Dash. In June of 2017, plotly formally released Dash as an open- source library

Dash makes it dead-simple to build a GUI around your data analysis code. Dash app code is declarative and reactive, which makes it easy to build complex apps that contain many interactive elements. Every aesthetic element of the app is customizable: The sizing, the positioning, the colors, the fonts. Dash apps are built and published on the Web, so the full power of CSS is available. While Dash apps are viewed in the web browser, you do not have to write any JavaScript or HTML. Dash provides a Python interface to a rich set of interactive web- based components.

3. Problem Identification :

The stock market appears in the news every day. You hear about it every time it reaches a new high or a new low. The rate of investment and business opportunities in the Stock market can increase if an efficient algorithm could be devised to predict the short-term price of an individual stock. We are given Tesla stock price from 2010 to 2020. The task is to predict the trend of the stock price for 2018. Note that, based on Brownian Motion, the future variations of stock price are independent of the past. So, it is impossible to predict the exact stock price, but possible to predict and capture the upward and downward trends.

Best Solution Available in Market

Stock Market Prediction using Simple Linear Regression

- By using Linear Regression, the project forecasts the TCS datasets behaviour and the net outcome of other approaches.

- The model incorporates techniques for real world machine learning applications including acquiring and analysing a large dataset; using a variety of techniques to train the model and predict potential outcomes.

Problems With Existing Solutions

- The prediction model was run on one stock set only and not on the entire market. This creates a certain degree of short sightedness on the entire evaluation process.

- The method of prediction using Linear Regression is comparatively less accurate, when compared to other methods.

- Does not consider Random Forest prediction model for prediction, which theoretically gives a higher accuracy when run on a small dataset.

How our solution will Solve it?

This is the first model to solely use, Long Short-Term Memory (LSTM) as a prediction technique which gives a more accurate result.

This paper provides a good explanation of the LSTM model and how it is trained. This provides both the graphical and mathematical implementations.

4. Objectives:

The objective is to predict the stock prices to make more informed and accurate investment decisions.

Technical Objective:

The technical objectives will be implemented in Python using LSTM model by Pyplot and matplotlib. The system must be able to access a list of past prices. It must calculate the estimated price of stock based on past data. It must also provide an instantaneous visualization of the market indexes. We use Plotly Dash for developing a dashboard for the prediction of stock. It helps to show ups and downs in stock.

Experimental Objective:

The prediction system will be implemented one using Long Sort Term Neural Network (LSTM)M to find more accuracy in the prediction of stock price.

CHAPTER 3.
PRELIMINARY DESIGN

3.1 Features & Characteristics of dashboard:-

- The dashboard is very interactive and it is based on the LSTM technology of machine learning algorithm which able to give the efficient result and accurate than others type of machine learning algorithm technology.

- Dashboard is basically designed and gathers data from the yahoo finance limited which is one of the biggest data repository in the field of data analytics.

- We are using plotly Dash framework for build our dashboard and perform action in an interactive manner.

- Our dashboard is fully based on user requirements and it is too familiar with the user. That means anyone who has not deep knowledge about this field can also handle and run it easily.

- It helps the user to show the closing value of the stock, It also shows some important information and show the following plotting diagram that helps user to predict easily the future price of the stock.

- On the search bar of the dashboard, we just write the keyword of the stock and click on the search and it will show all the data.

3.2 Constraint Identification:-

Time Constraint:

During the project It takes about 3 months to complete. 1st we take 1 and half months to learn about the machine learning module and we learn some research paper on the stock price prediction, we learn the terms about the stock and after that we create a team of 4 members and divide our works and then we jump into the project and 1st we make a documentation that what we should analyse, which is necessary for the prediction of a stock price it took about 1012 days.

After completed all of the above-mentioned work we move on the coding part to do the coding part first we prepare our environment (that means we install the required libraries, code editor

and packages) to compile our program. It took about 1-2 days to prepare an environment. After that we started writing a code to analyse all our required term like import libraries, import data, reading data, writing data, analysis data like (mean of the stock price, mode of the stock price, average of the stock price, and plotting of the required data). It took lot of the time to complete all these things.

At later we decide to create a dashboard first we decide to create a mock-up design for our dashboard, it took time about 2 days to complete it. After completing our mock-up design, we need to write some codes to develop a dashboard and at that time we are on the dashboard according to me it should take 2 days more to complete it.

Scope Constraint:

The scope of this project is to predict the stock price in order to make more informed and accurate investment decision, and we are creating a dashboard because every one can not write codes to analyse the stock price that's why we creating a dashboard that everyone who need to analyse the price trend of the stock that they can simply do that. And we are trying our best to crate our dashboard user friendly.

Cost Constraint:

The cost constraint of our project is null, because here we use python which is an open-source programming language and we are using VS code editor which is also an open -source software and rest of all the libraries and packages are open-source. That why our cost constraint is null.
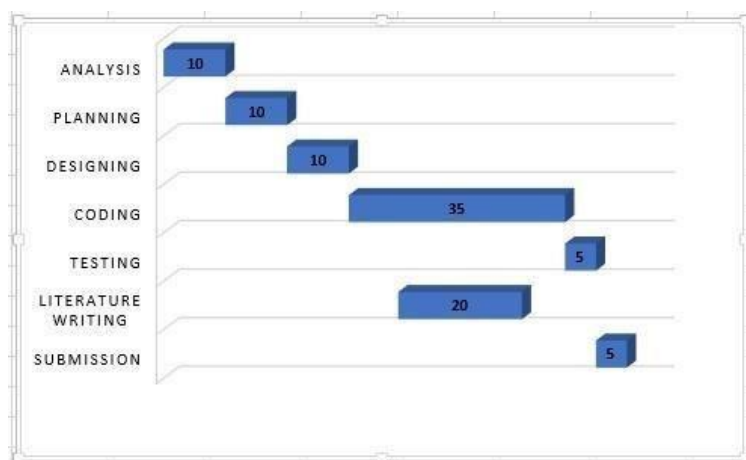
Analysis of features and finalization subjects to constraints:

As we know the stock market appears in the news every day. You hear about it every time it reaches a new high or a new low. The rate of investment and business opportunities in the Stock market can increase if an efficient algorithm could be devised to predict the short-term price of an individual stock. We are given Tesla stock price from 2014 to 2021. The task is to predict the trend of the Stock price for any year. Note that, based on Brownian motion, the future variation of stock price are independent of the past. So, it is impossible to predict exact stock price, but possible to predict and capture the upward and downward trends.

During the project It takes about 3 months to complete. 1st we take 1 and half months to learn about the machine learning module and we learn some research paper on the stock price prediction, we learn the terms about the stock and after that we create a team of 4 members and divide our works and then we jump into the project and 1st we make a documentation that what we should analyse, which is necessary for the prediction of a stock price it took about 1012 days.

After completed all of the above-mentioned work we move on the coding part to do the coding part first we prepare our environment (that means we install the required libraries, code editor and packages) to compile our program. It took about 1-2 days to prepare an environment. After that we started writing a code to analyse all our required term like import libraries, import data, reading data, writing data, analysis data like (mean of the stock price, mode of the stock price, average of the stock price, and plotting of the required data). It took lot of the time to complete all these things.
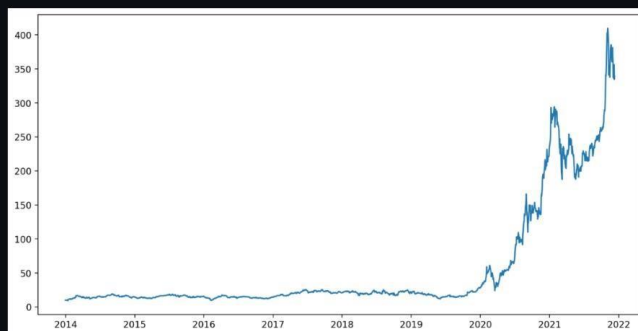
At later we decide to create a dashboard first we decide to create a mock-up design for our dashboard, it took time about 2 days to complete it. After completing our mock-up design, we need to write some codes to develop a dashboard and at that time we are on the dashboard according to me it should take 2 days more to complete it.
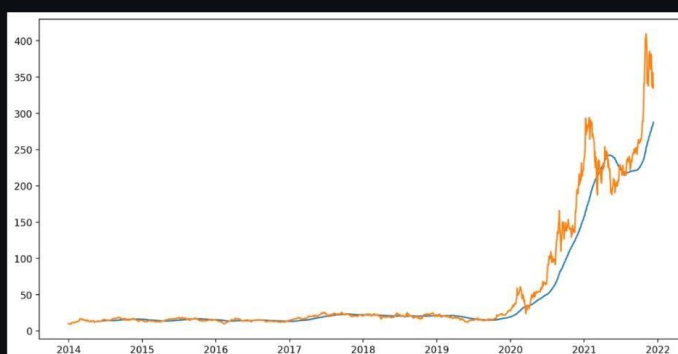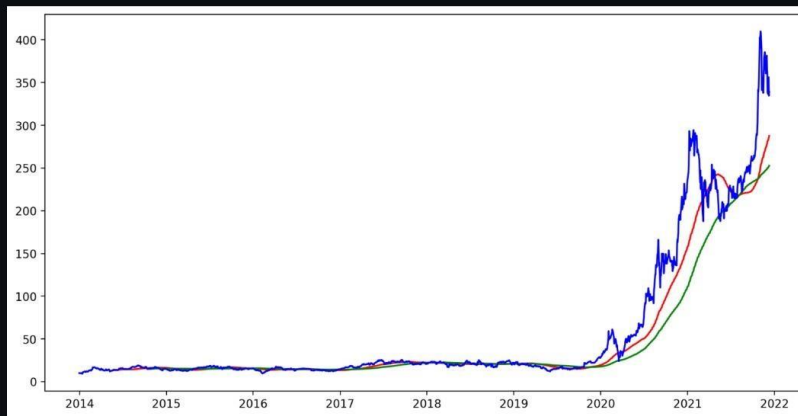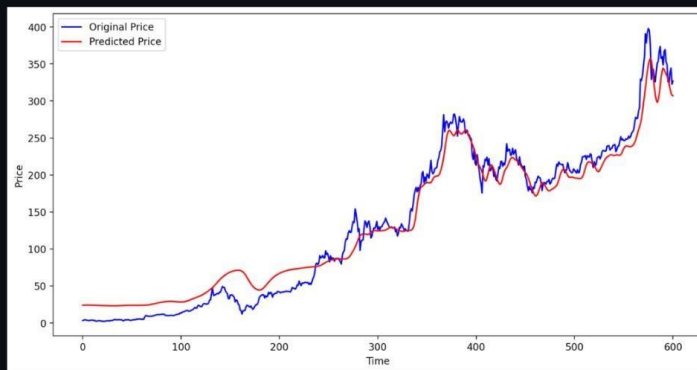
Design Selection:
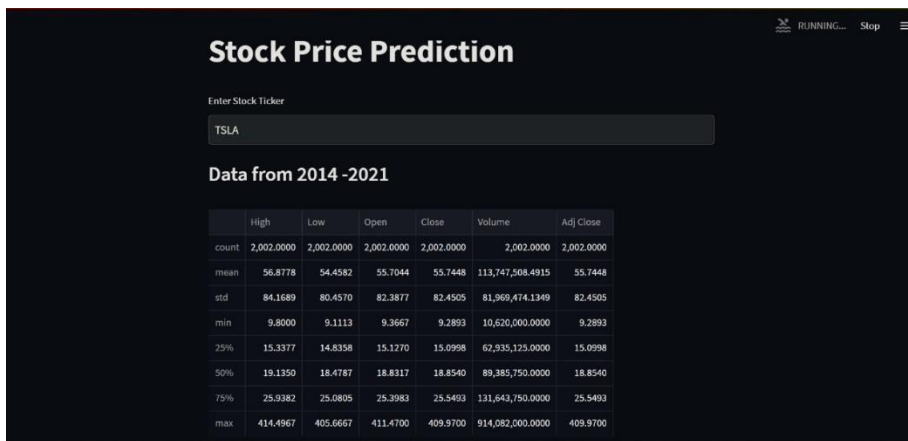
Closing Price Vs Time Chart with 200MA


Prediction Vs Original

DFD: A data flow diagram (DFD) maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination. Data flowcharts can range from simple, even hand-drawn process overviews, to in-depth, multi- level DFDs that dig progressively deeper into how the data is handled. They can be used to analyse an existing system or model a new one. Like all the best diagrams and charts, a DFD can often visually "say" things that would be hard to explain in words, and they work for both technical and nontechnical audiences, from developer to CEO.
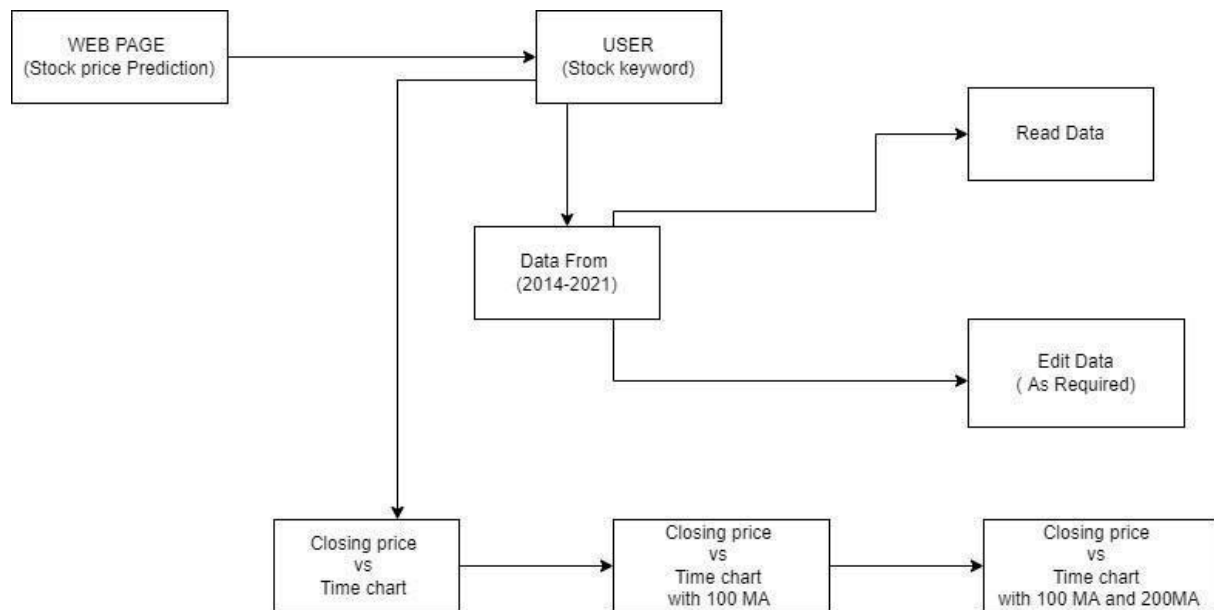
ER Diagram: ER Diagram stands for Entity Relationship Diagram, also known as ERD is a diagram that displays the relationship of entity sets stored in a database. In other words, ER diagrams help to explain the logical structure of databases. ER diagrams are created based on three basic concepts: entities, attributes and relationships.

ER Diagrams contain different symbols that use rectangles to represent entities, ovals to define attributes and diamond shapes to represent relationships.
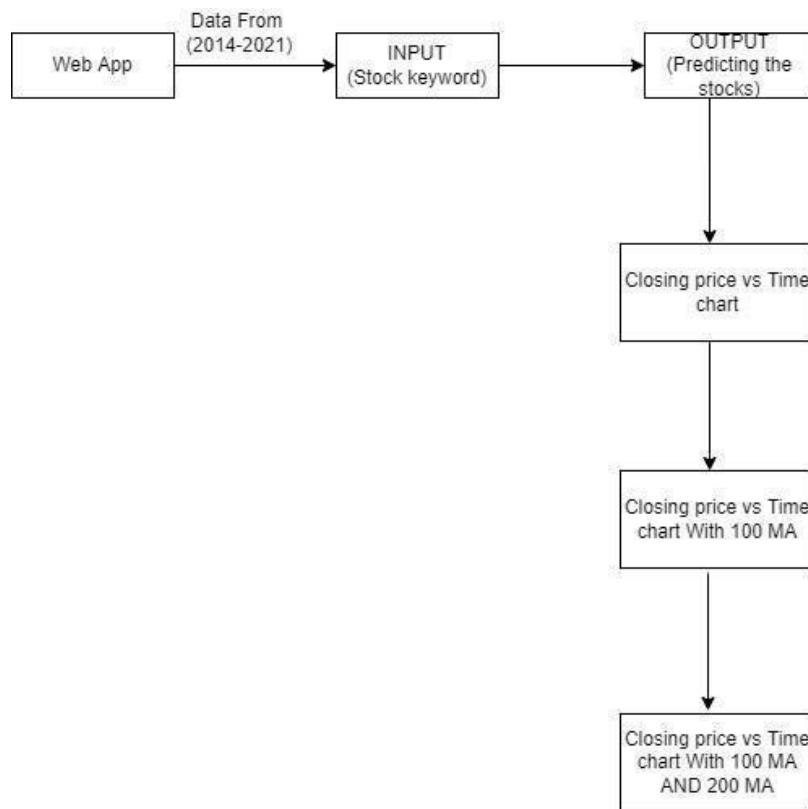
Sequence Diagram: A sequence diagram is a type of interaction diagram because it describes how—and in what order—a group of objects works together. These diagrams are used by software developers and business professionals to understand requirements for a new system or to document an existing process. Sequence diagrams are sometimes known as event diagrams or event scenarios.
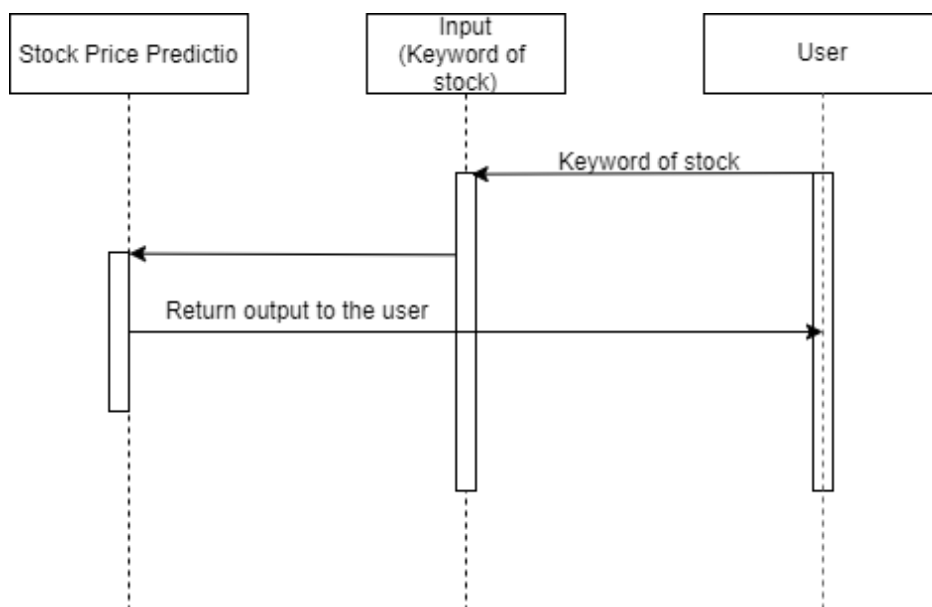
CHAPTER 4
RESULT ANALYSIS AND VALIDATION

The implementation of proposed LSTM model using python which predicts the future price of TATAMOTORS share based on its historical data. The below visualization figure shows the visualization of TATASHARE prediction. In our paper the implementation of an algorithm which predicts the stock price of a share for given period of time, the below graph from our algorithm will show the predicted price of TATAMOTORS share. In the result shown in the below graph is the plotted form our algorithm outcome by applying 96 LSTM units for achieving the accuracy. The Fig 2 is drawn from original dataset and also shown the result by comparing its correctness with the trained model from algorithm that is defined in the previous section. the "x" axis is share price. The "y" axis is days. The data is slot of 1500 days is shown in the Fig 3.
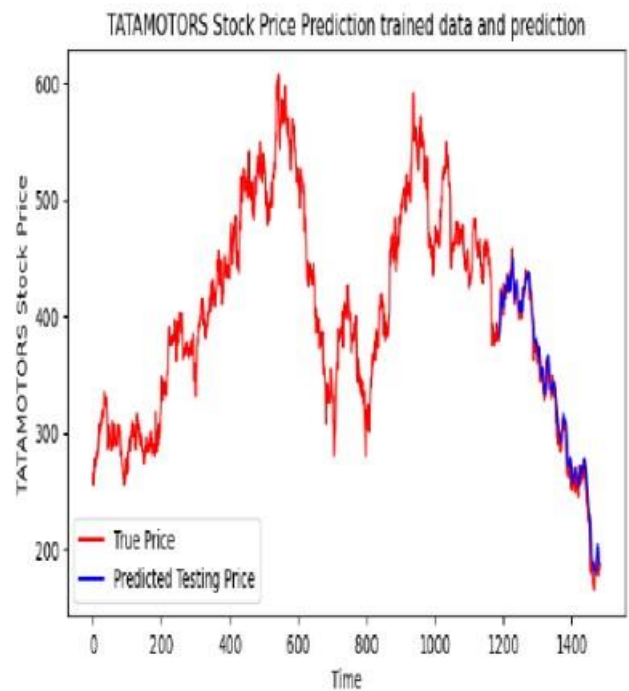

Figure 2: predicted testing stock price

The Fig 3 is drawn from original dataset also shown the result by comparing its correctness with the trained model from algorithm which that is defined in the previous section. the "x" axis is share price. The "y" axis is days. The data is slot of 300 days is shown in the Fig 3.
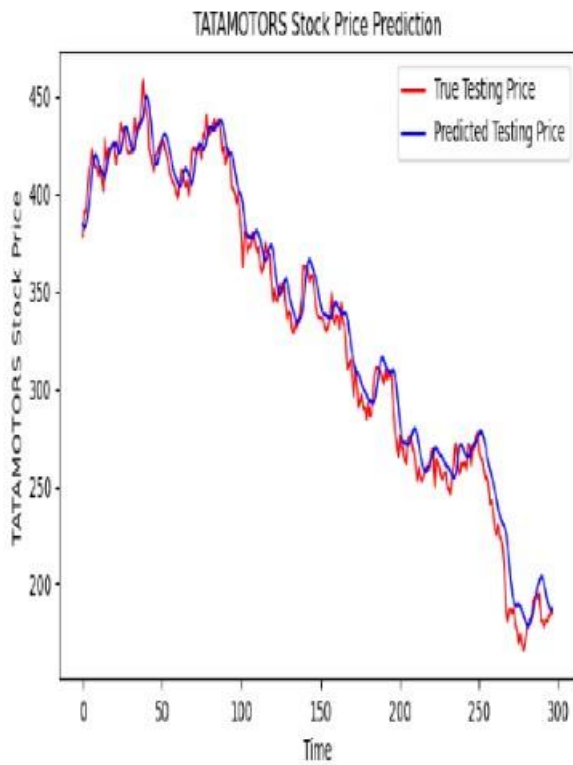
Figure 3: predicted stock price

In the Fig 2, the graph has been plot for whole data set along with some part of trained data. the graph is showing the open price of TATAMOTORS share for 1484th day's opening price with very minimal loss. the algorithm has plotted the graph successfully along with the predicted price testing price (blue) and true price (red), there is a slight difference in predicting the price between the predicted price testing price (blue) and true price (red), which proves that our algorithm is able to predict the with minimum loss rate for the given complete data set of a particular share. In the Fig 3, the graph is showing the open price of TATAMOTORS share for 300th day's opening price with very minimal loss. the algorithm has plotted the graph successfully along with the predicted price testing price (blue) and actual testing price (red), there is a slight difference in predicting the price between the predicted price testing price (blue) and actual testing price (red), which proves that our algorithm is able to predict the with minimum loss rate of 0.0024. The proposed algorithm is able to predict the share price with very low loss and error rate, if increase the epoch batch rates the training will be more efficient, in the above section we have used epoch batch size of 50 to predict the stock prices. The figures shown in the previous section (fig 2 and fig 3) of the proposed algorithm is able to predict the price, with loss: 0.0024 300th days open price was 172 rupees INR and our predicted price is 166 rupees per share.

CHAPTER 5.
CONCLUSION

The study of the share is carried out in this paper and it can be carried out for several shares in the future. Prediction could be more reliable if the model trains a greater number of data sets using higher computing capacities, an increased number of layers, and LSTM modules. In future enhancement the inclusion of sentiment analysis from social media to understand what the market thinks about the price variation for a particular share and it can be implement this by adding twitter and Facebook API to our program as Facebook is a leading social media which has lots of market trend information posted by users.

CHAPTER 6.
REFERENCES

[1] Hiba Sadia, Aditya Sharma, Adarrsh Paul, SarmisthaPadhi, Saurav Sanyal- "Stock Market Prediction Using Machine Learning Algorithms", IJEAT, 2019.

[2] Raut Sushrut Deepak, Shinde Isha Uday, Dr. D. Malathi, "Machine Learning Approach In Stock Market Prediction", IJPAM 2017

[3] M. S. Hegde, G. Krishna and R. Srinath, "An Ensemble Stock Predictor and Recommender System," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, 2018, pp. 1981-1985.

[4] M. Roondiwala, H. Patel and S. Varma, "Predicting stock prices using LSTM," International Journal of Science and Research (IJSR), vol. 6, no. 4, pp. 1754-1756, 2017.

[5] T. Kim and H. Y. Kim, "Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data," PloS one, vol. 14, no. 2, p. e0212320, April 2019.

[6] S. Selvin, R. Vinayakumar, E. A. Gopalkrishnan, V. K. Menon and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in International Conference on Advances in Computing, Communications and Informatics, 2017.

[7] Loke.K.S. "Impact Of Financial Ratios And Technical Analysis On Stock Price Prediction Using Random Forests", IEEE, 2017.

[8] . Xi Zhang1, Siyu Qu1, Jieyun Huang1, Binxing Fang1, Philip Yu2, "Stock Market Prediction via Multi-Source Multiple Instance Learning." IEEE 2018.