

STOCK PRICE PREDICTION

A PROJECT REPORT

Submitted by

RITIKA PANDEY -201500579

VISHESH SRIVASTAVA -201500800

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

Computer Engineering and Application



November 2022

BONAFIDE CERTIFICATE

Certified that this project report “**STOCK PRICE PREDICTION**” is the bonafide work of “**RITIKA PANDEY and VISHESH SRIVASTAVA**” who carried out the project work under my/our supervision.

SIGNATURE

Rohit Agarwal

HEAD OF THE DEPARTMENT

Department of Computer Engineering
& Applications

SIGNATURE

Md Farmanuel Haque

SUPERVISOR

(Technical Trainer)

Department of Computer Engineering
& Applications

Submitted for the project viva-voice examination held on_____

INTERNAL EXAMINER

EXTERNAL EXAMINER

ABSTRACT

The goal of Stock Market Prediction is to forecast the future value of a firm's earnings through stocks. Machine learning, which produces forecasts depends on the principles of existing stock markets indexes by training on its prior values, is the latest phenomenon in stock market forecasting techniques. Machine learning makes use of a variety of models to create accurate predictions. The research is focused on share value prediction using regression and other tools of machine learning.

TABLE OF CONTENT

<u>CHAPTER</u>	<u>PAGE No.</u>
Bonafide Certificate	i
Abstract	ii
Chapter 1 -- Introduction	3-5
Chapter 2 –Project Description	6-9
2.1. Purpose	6-7
2.2. Motivation	8
2.3. Problem statement	8
2.4. Special Features	9
Chapter 3 –Tools and Technologies	10-19
Chapter 4- Implementation Modules and Screen Shots	20-24
Chapter 5 – Conclusion and Future Scope	25-26
Bibliography	27

LIST OF FIGURES

FIGURE NAME	PAGE
Fig 1- Machine Learning Processes	5
Fig 2- Python Libraries	11
Fig 3- Uses of NumPy	12
Fig 4- Features of Pandas	13
Fig 5- Types of Matplotlib	14
Fig 6- TensorFlow	15
Fig 7- Kaggle dataset	15
Fig 8- Waterfall method	17
Fig 9- Data flow chart	18
Fig 10- System Architecture	19
Fig 11- ML code screenshot-1	20
Fig 12- ML code screenshot-2	20
Fig 13- ML code screenshot -3	21
Fig 14- ML code screenshot -4	21
Fig 15- UI-1	22
Fig 16- UI-2	22
Fig 17- UI-3	23
Fig 18- UI-4	23
Fig 19- UI-5	24
Fig 20-UI-5	24

CHAPTER -1 INTRODUCTION

- Stock Market Prediction's purpose is to estimate the future worth of a company's earnings via stocks. Machine learning is a new occurrence in stock market forecasting approaches that provides forecasts founded on the principles of current stock indexes by training on their prior data. To develop correct estimates, machine learning employs a range of products. The study focuses on predicting stock prices using depreciation and other machine learning technologies. The purpose of the Stock Market Prediction is to predict the future value of a company's stock. Machine learning is a novel event in sharemarket forecasting methodologies that deliver projections based on actual stock index principles by training on their previous data. Machine learning uses a variety of materials to produce accurate predictions. The study focused on using downward and other machine learning methods to predict stock prices. In other words, the stock is a disproportionate amount or ownership symbol that explains the security protocols and formal agreement, an individual and a corporation. The stock has been around since the beginning, and it has been a nice word owing to its inclination for ambiguity. People who were investigating the topic and putting it into practice regularly have made a fortune. There is a slew of brokers in the market who can help you comprehend and engage in the subject, but their fees are arduous and exorbitant. For an investor, it is a risky situation in which they put their money into something in the hopes of getting a higher profit. If the value of the asset rises, so does the value of the shares and its cost, resulting in a financial benefit for both sides. It is even seen as a

sideline business in society, with individuals believing it to be a sign of good fortune. When a person buys a company shares, they are known as a stockholder, and they will receive a portion of the profits or benefit which they have invested in. The stock can be sold and bought as needed by the investor. They can sell their shares with their particular or some other customers, and several investment bankers in the industry are interested in doing so.

- **MACHINE LEARNING:**
- Machine learning is a subset of artificial intelligence. It is a crucial part for making a machine intelligent. So, basically it works with dataset which can be stored in form of database (i.e., rows and columns). It makes machine application to work intelligently and rationally. It works as it takes some input from the past and then learn from them.

What are the different types of machine learning?

- There are four basic approaches:
- **Supervised learning:** when the learning is done as inputs and output both are specified then the learning is supervised. In this type of machine learning, Data scientists provide algorithms with labeled training data and define variables that they want the algorithm to test for relevance.

- **Unsupervised learning:** In this type of ML machine itself tries to find out some meaningful connection on the dataset it is being provided. Data trained by algorithms and predictions or outgoing recommendations are predetermined.
- **Semi-supervised learning:** When we combine supervised and supervised reading it is called less supervised reading. In this machine do have some previously data set in which it was trained but it also does have freedom to explore new things and develop its understanding more and more.
- **Reinforcement learning:** reinforcement learning is used when we have to teach a machine a process which itself is a multi-stage but do have some sort of clearly defined rules. Data scientists devise an algorithm on how to complete a task. But for the most part, the algorithm decides for itself (it works in real time) what steps to take on the path. Used in playing chess algorithms, satellite launch algorithms. Here, accuracy and efficiency are must because it has to act in real world environments.

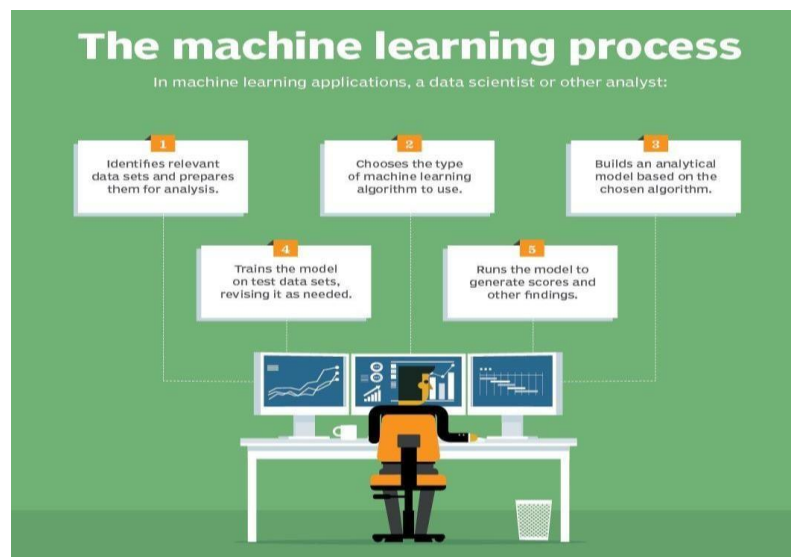


Fig 1- Machine Learning Processes

Chapter 2 –Project Description

2.1. Purpose

- Computers are good at processing large amounts of data, data that we are creating more and more of machine learning lets computers make decisions about data. It lets a computer learn from data and make predictions and decisions. It allows computers to do tasks that in the past required humans and make our lives hopefully easier.
- Stock market prediction is a computerized prediction method that highlights the danger that investors face while investing in stocks. It forecasts stock prices and currency rates while considering the user's basic knowledge and statistical analysis.
- Data is regarded as a digital propellant that provides higher yearning prospects as well as futurephrases. Knowledge is power, and so is the stock market. In fact, stocks fluctuate unexpectedly and change frequently. The same fall and rise are rare and difficult to distinguish. Requirements on the same dealing with greater agility and the people that work behind the scenes.
- The beginning stock market until the next day is determined by investments made throughout a financial day. It has its own set of requirements and is completely integrated with the financial and income-producing levels. The stock market is vast and frenetic. The project's

major goal isto forecast bending curves, develop a prediction approach, and go through the process and algorithms to arrive at a sustainable resource source.

- Anything follows a predictable pattern. The way of origination is the pattern, and the same is valid for the stock.
- This initiative assists in connecting information and equipping people to comprehend and tradethe most of inventory, as well as the production and weaknesses that must be now seen anticipated. The improvement of the very same is done with the resources graph, which allowsa user or customer to evaluate the same and consider the wants and key facts before dealing andconsidering those factors for the return that the person is prepared to invest on. The stock projection is forecasted using the current data source, and the projection is for the following week. The pattern is a hurdle, which is why the report exists.

2.2 Motivation

- The issue of stock price prediction is a well-known and essential one. We can get insight into market behaviour over time with an effective stock forecasting model, recognising tendencies that might otherwise go unnoticed. With increasing computer capacity, machine learning will be an effective way to overcome this challenge. Unfortunately, many machine learning algorithms cannot work with the stock offering information and adding extra characteristics can cost a lot of money per day. In this project, we will present an approach for incorporating user forecasts into the current machine learning while utilising publicly available historical data to improve our outcomes.

2.3 Problem statement

- The stock market is published daily in the news. Whenever it achieves a new high or a new low, you hear about it. If an effective method could be created to forecast the short-term price of an individual stock, the investment rate and business prospects in the Stock market can grow. Previous stock forecasting approaches include the use of Ai Cognitive and Convolution NeuralNetworks with a loss of inaccuracy of 20 per cent on average.

2.4 Special Features

- This project aims for a collected data set, analysing it with Machine Learning Algorithms and generating output. Visualization and the forecast that delivers a lift are two key aspects incorporated. Various prediction algorithms are used to predict that are accurate and sufficient in the form to yield to the source of a favourable resource. Digging and initialising the expectations that must be considered. To connect and enhance the expenditure, risks are being mitigated.
- Evaluating and implementing the results to support the actual situation. Keep track of progress and analysis of the results daily to determine the flow and degree of integration

Chapter 3- Tools and Technologies

Minimum Hardware Requirement:

- Intel core i3 or i5
- Graphics Card
- RAM 8 GB

Software Requirement:

- Windows 8 and above
- Python 3.5
- Browser(any)
- VS code
- IDE PyCharm
- Jupyter notebook
- Streamlit Framework

- **Data Collection: -**

The data entered in the previous phase is collected here collected online from yahoo.com

- **Data Preprocessing: -**

The collected data is converted into our desired file format that is “.csv” from yahoo.com etc.

- **ML Algorithms and Modeling: -**

These are the various Python libraries: -

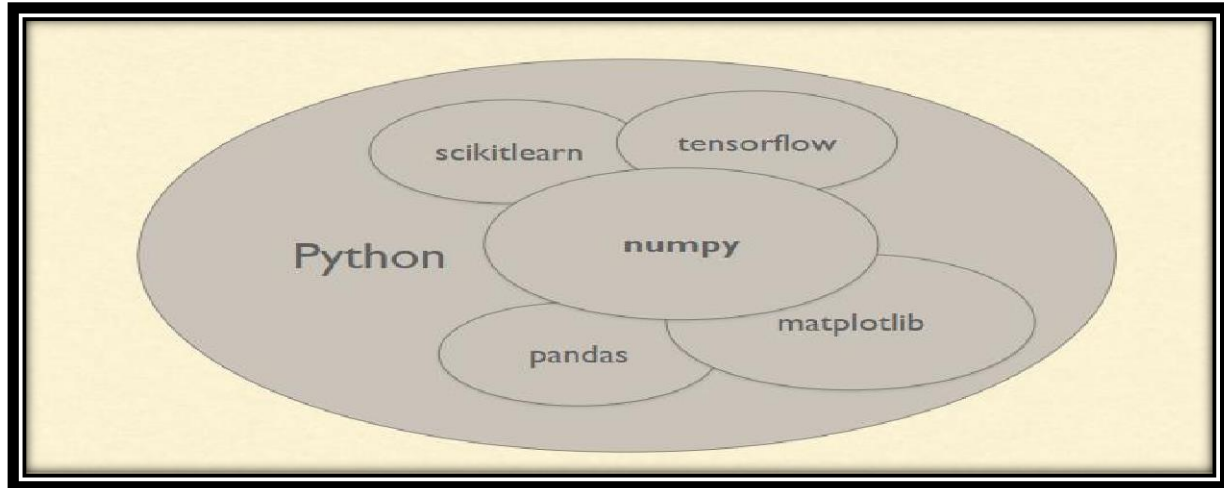


Fig 2-Python Libraries

- **NumPy:** NumPy stands for Numerical Python. This Python library gives the numerical capacity to deal with enormous measurement clusters. Through NumPy, many mathematical operations can be easily used widely in fast scientific computing.
- The NumPy library's most significant characteristic is this. The homogenous array object is what it is called. On the array, we conduct all procedures. NumPy arrays can be either one-dimensional or multifunctional.
- Same-sided members consist of a single column or row of data. Every column is considered a size. Format compared to excel spreadsheet. The parts are the same.

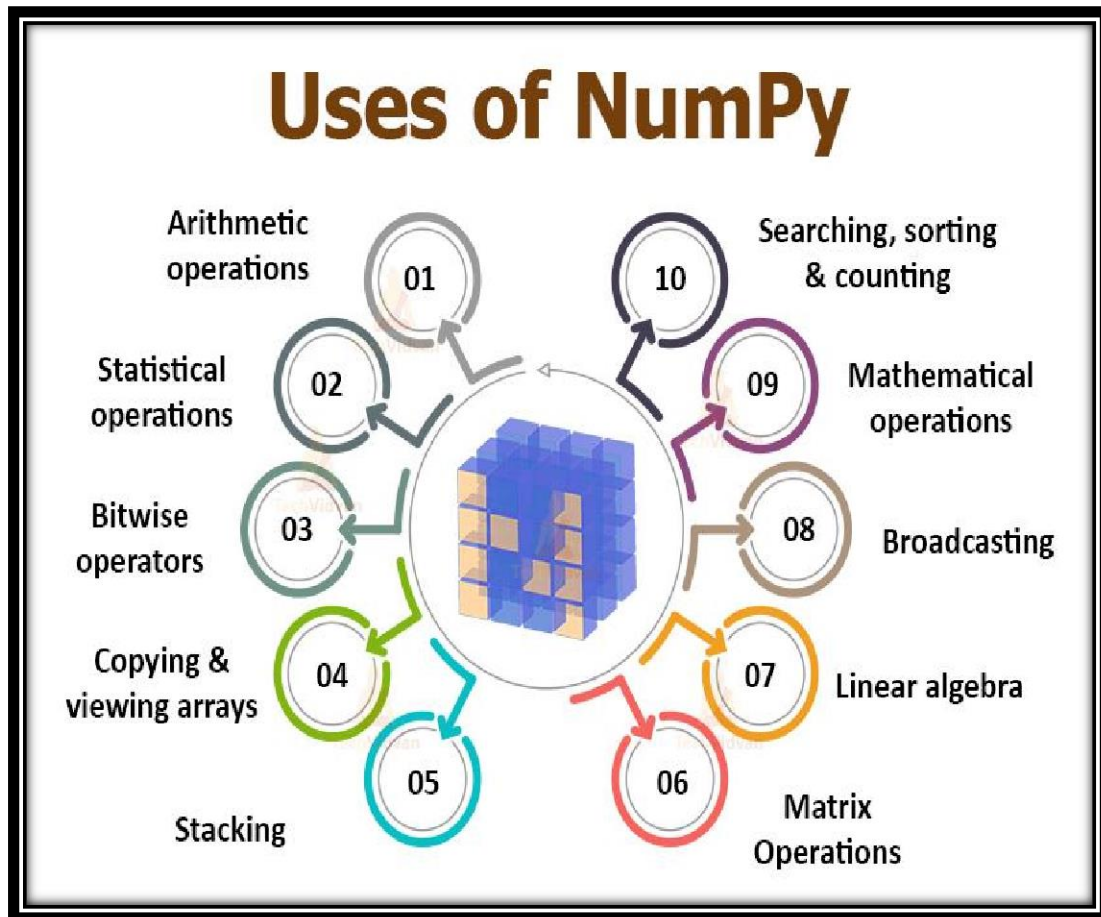


Fig 3-Uses of NumPy

- **Pandas:** This is the most mainstream Python library for information control and investigation. Pandas is the helpful capacity to control enormous measures of organized information and give the least demanding strategy to perform examination.

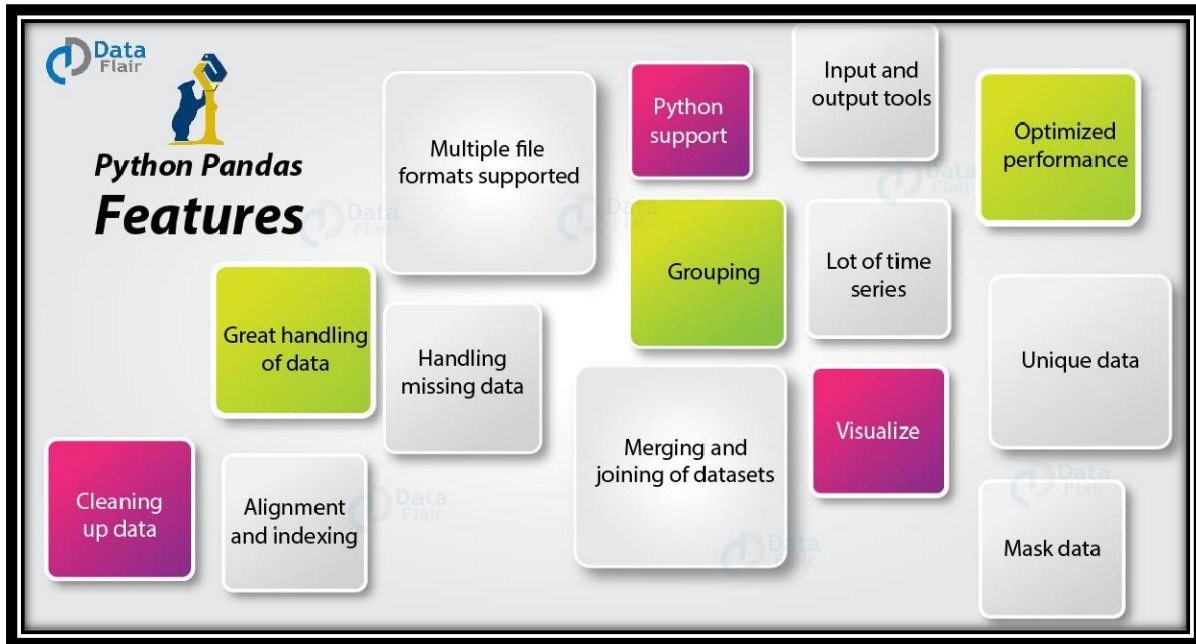


Fig 4- Features of Pandas

- With default and customizable indexing, the Data Frame objects are effective and easy. Tools for extracting data from various file types into in-memory data items.
- Data sets can be reshaped and pivoted.
- Slicing, indexing, and resampling big data sets depending on labels. A data structure's components can be removed or added.
- For aggregate and modifications, arrange by data. Information merging and joining with superior efficiency. Time-series data is a useful feature.

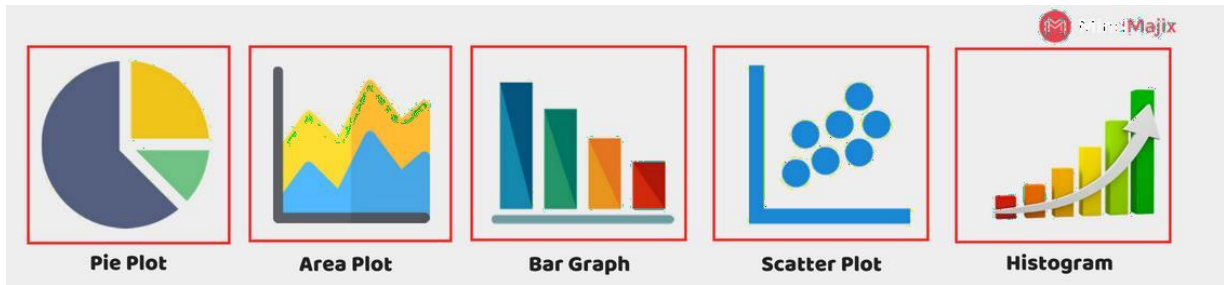


Fig 5- Features of Pandas

- **TENSERFLOW**

- TensorFlow is an open source and open source software used for ML & AI. It can be blamed for a variety of activities but focuses on the training and disruption of deep neural networks. It is a symbolic mathematical library based on data movement and segmentation. It is used for research and production.
- It is a rich way to manage all aspects of a machine learning program, but this class focuses on using training and machine learning learning models. TensorFlow APIs are organized sequentially, with high-level APIs built into low-level APIs.
- Features:
- Includes features that explain, make good use of, and calculate mathematical expressions easily with the help of multi-case planning called tensors.
- Includes the highest measurable computation feature for various data sets.
- TensorFlow uses a computer GPU, automatic management. It also features a unique feature to increase memory and used data.

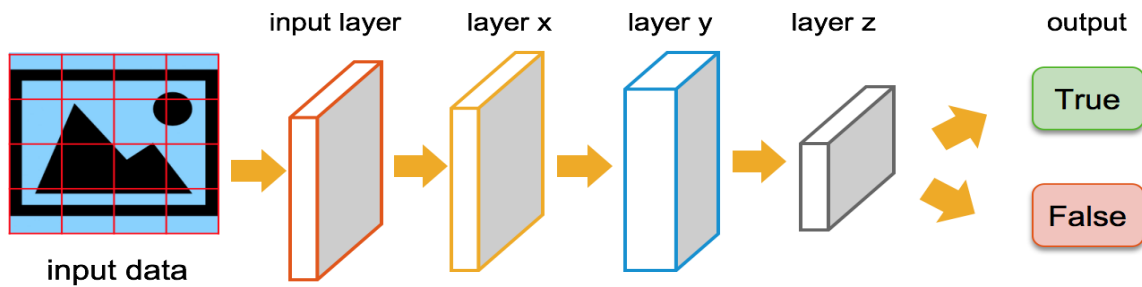


Fig 6- TensorFlow

• KAGGLE

- Kaggle is an online community of data scientists. Provides user with resources such as publishing data sets, testing and building models in a web-based environment. It helps the users to solve complex problems by communicating and working with the other users. It is a platform to compete with others in machine learning and AI based competitions. It is a great platform to learn new techniques, algorithms and trends in machine learning, data science and AI related things.
- It provides scholars with a platform for solving real-life problems. It is also a great site to download and train your data sets for your projects. It has millions of data sets which you can use and modify as per your needs.

House	Sale price (100\$)	Size (sqft)	Age (years)
Avalon	2050	2650	13
Cross Winds	2080	2600	*
The White House	2150	2554	6
The Rectory	2150	2921	3
Larchwood	1999	2580	4
Orchard House	1900	2580	4
Shangri-La	1800	2774	2
The Stables	1560	1920	1
Cobweb Cottage	1450	2150	*
Nairn House	1449	1710	1

Fig 7- Kaggle is used to provide the datasets that are used to generate the data

- **CMD**

- CMD is the default command line interpreter. It is usually called cmd or command prompt. The Cmd section provides a simple framework for writing online targeted translators. This is often the case with test equipment, management tools, and prototypes that will be rolled out over a relatively long period of time. It looks like Windows Explorer on windows , but without mouse cursor. Other command line names are: cmd, CLI, prompt, console or terminal. The part that comes and includes \$ or > is called Command line Prompt, or prompt for short. It encourages you to put something in there. In the lesson, if we want you to type an order, we will enter \$ or >, and sometimes more than the left.

- **JUPYTER NOTEBOOK**

- JupyterLab is a web-based interactive development environment for Jupyter notebooks, code, and data. JupyterLab is flexible: configure and arrange the user interface to support a wide range of workflows in data science, scientific computing, and machine learning. JupyterLab is extensible and modular: write plugins that add new components and integrate with existing ones.
- The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

- **Waterfall methodology**

- In this case, the development process looks like a stream, passing stages in sequence. But the transition from one phase to the next takes place only after the previous one has been completed, While Agile is a time-focused philosophy that allows creating a project incrementally, dividing it into small pieces. One of its main benefits is the ability to adapt and change at any step.

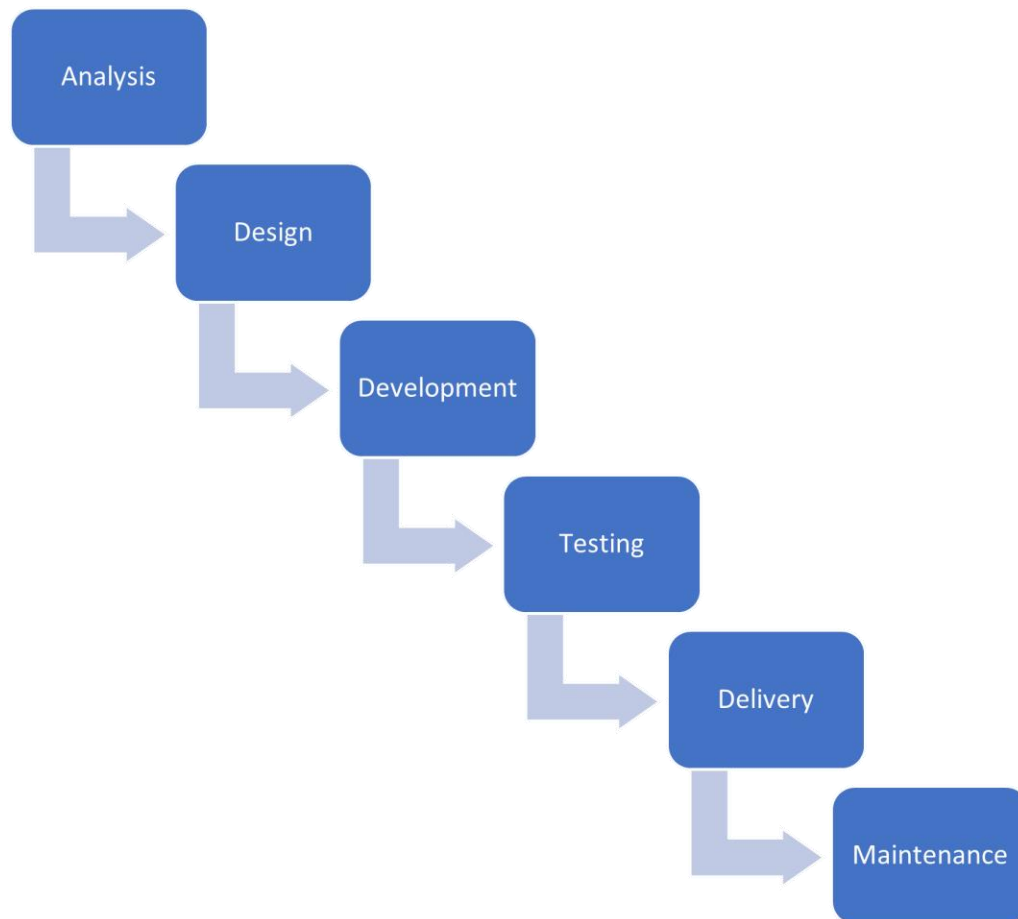


Fig 8- Waterfall method

Data Flow Diagram

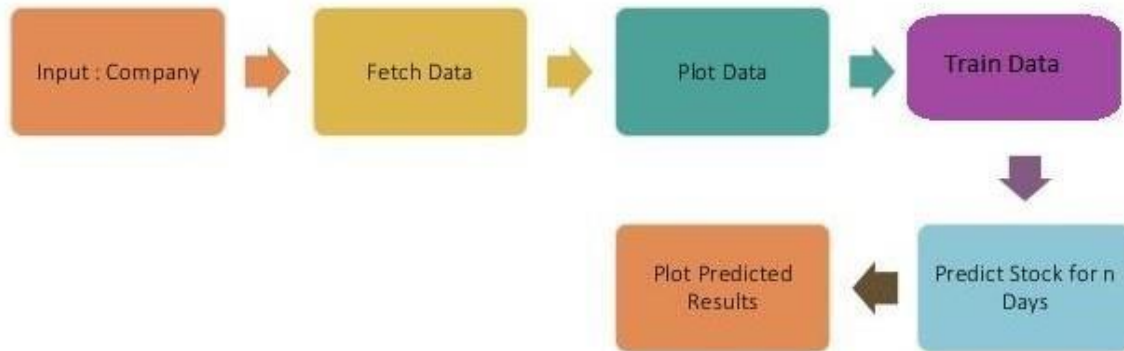


Fig 9– DATA FLOW CHART

In fig., we are selecting a corporation and retrieving its data using Panda's data-reader library, charting the data, and then training the information to forecast the stocks for a set time frame.

- **SYSTEM ARCHITECTURE**

- We obtained the information for the suggested research from Yahoo. This data set, however, is in plaintext form. The set of data is a compilation of Share Market evaluation statistics for a few corporations. The first step is to transform unclassified data into classified data. Because the basic data obtained has various qualities, but few of those characteristics are required for the forecast, this is done through extracting features. A reducing technique is feature extraction.

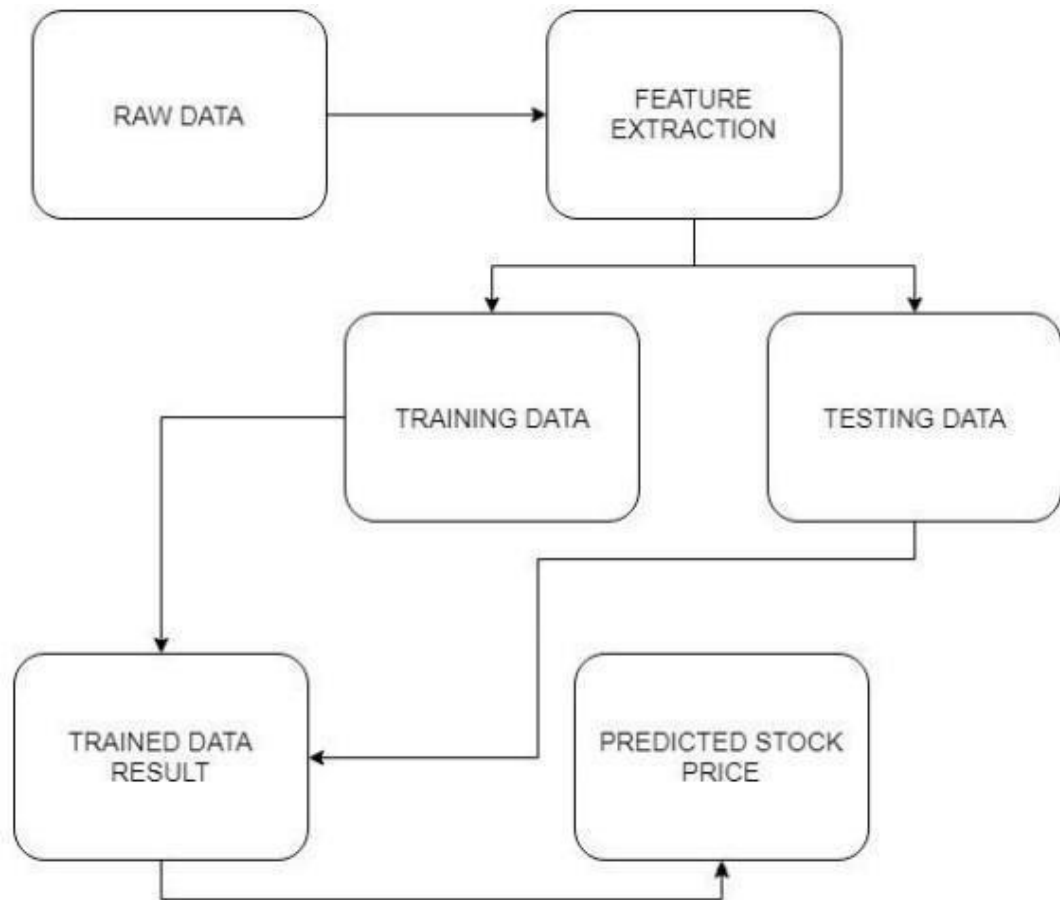


Fig 10- SYSTEM ARCHITECTURE

Chapter 4- Implementation Modules and Screen Shots

CODE FOR THE PROGRAM:

```
jupyter predict.py a few seconds ago Logout
File Edit View Language Python

2 import numpy as np
3 import pandas as pd
4 import matplotlib.pyplot as plt
5 import pandas_datareader as data
6 from keras.models import load_model
7 import streamlit as st
8 from sklearn import datasets
9 from datetime import date
10 from plotly import graph_objs as go
11
12 start = "2015-01-01"
13 end = date.today().strftime("%Y-%m-%d")
14 # In[6]:
15
16 def app():
17     start = "2015-01-01"
18     end = date.today().strftime("%Y-%m-%d")
19
20     st.title('Market Prediction Zone:-')
21
22     user_input=st.text_input('Enter Stock Ticker:', 'AAPL')
23     # In[7]:
24
25     data_load_state = st.text('Loading data...')
26     data = load_data(user_input)
27     data_load_state.text('Loading data... Done!!!!')
28
29     st.subheader('Raw Data Till Date:')
30     st.subheader(end)
31     #data=data.reset_index()
32     #data=data.drop(['Adj Close'], axis=1)
33     st.write(data.head())
34     st.write("----")
35
36     fig = go.Figure()
37     fig.add_trace(go.Scatter(x=data['Date'], y=data['Open'], name="stock_open"))
38     fig.add_trace(go.Scatter(x=data['Date'], y=data['Close'], name="stock_close"))
39     fig.layout.update(title_text='Time Series data', xaxis_rangeslider_visible=True)
40     st.plotly_chart(fig)
41 # In[8]:
```

Fig 11-ML Code

```
jupyter predict.py a few seconds ago Logout
File Edit View Language Python

41 # In[8]:
42
43
44 #visuals
45 st.subheader('Closing Price vs Time Chart')
46 fig=plt.figure(figsize=(16,6))
47 plt.plot(data.Close)
48 st.pyplot(fig)
49
50
51 # In[11]:
52
53
54
55 st.subheader('Closing Price vs Time Chart with 100MA')
56 ma100=data.Close.rolling(100).mean()
57 fig=plt.figure(figsize=(16,6))
58 plt.plot(ma100)
59 plt.plot(data.Close)
60 st.pyplot(fig)
61
62
63 # In[12]:
64
65
66 st.subheader('Closing Price vs Time Chart with 200MA')
67 ma200=data.Close.rolling(200).mean()
68 fig=plt.figure(figsize=(16,6))
69 plt.plot(ma200,'r')
70 plt.plot(ma100,'g')
71 plt.plot(data.Close,'b')
72 st.pyplot(fig)
73
74
75 data_training=pd.DataFrame(data['Close'][0:int(len(data)*0.70)]) #we take only 70% values of close column
76 data_testing=pd.DataFrame(data['Close'][int(len(data)*0.70): int(len(data))])#we take remaining 30% values of close column
77
78 print(data_training.shape)
79
80 print(data_testing.shape) # addition of both values will be equal to df.shape
```

Fig 12-ML Code

```

75 data_training=pd.DataFrame(data['Close'][0:int(len(data)*0.70)]) #we take only 70% values of close column
76 data_testing=pd.DataFrame(data['Close'][int(len(data)*0.70): int(len(data))])#we take remaining 30% values of close column
77
78 print(data_training.shape)
79
80 print(data_testing.shape) # addition of both values will be equal to df.shape
81
82 # In[14]:
83
84
85
86 from sklearn.preprocessing import MinMaxScaler #to convert datas into scaled data
87 scaler=MinMaxScaler(feature_range=(0,1))# each value of closing price will be scaled between 0 and 1
88
89
90 # In[15]:
91
92
93 data_training_array=scaler.fit_transform(data_training)#converts values to array
94
95
96 # In[17]:
97
98
99 # Load our model
100 model=load_model('keras_model.h5')
101
102 #testing part
103 past_100_days=data_training.tail(100)
104 final_df = past_100_days.append(data_testing, ignore_index=True)
105 input_data=scaler.fit_transform(final_df)
106
107 x_test=[]
108 y_test=[]
109
110 for i in range(100,input_data.shape[0]):
111     x_test.append(input_data[i-100:i])
112     y_test.append(input_data[i,0])
113
114

```

Fig 13-ML Code

```

101
102 #testing part
103 past_100_days=data_training.tail(100)
104 final_df = past_100_days.append(data_testing, ignore_index=True)
105 input_data=scaler.fit_transform(final_df)
106
107 x_test=[]
108 y_test=[]
109
110 for i in range(100,input_data.shape[0]):
111     x_test.append(input_data[i-100:i])
112     y_test.append(input_data[i,0])
113
114 x_test,y_test=np.array(x_test),np.array(y_test)
115 y_predicted=model.predict(x_test)
116
117 scaler=scaler.scale_
118
119 scale_factor=1/scaler[0]
120 y_predicted=y_predicted*scale_factor
121 y_test=y_test*scale_factor
122
123
124 # In[18]:
125
126
127 st.subheader('Prediction vs Original')
128 fig2=plt.figure(figsize=(16,6))
129 plt.plot(y_test,'b',label= 'Original Price')
130 plt.plot(y_predicted,'r',label= 'Predicted Price')
131 plt.xlabel('Time')
132 plt.ylabel('Price')
133 plt.legend()
134 st.pyplot(fig2)
135
136 def load_data(ticker):
137     df = data.DataReader(ticker, 'yahoo', start,end)
138     df.reset_index(inplace=True)
139     return df

```

Fig 14-ML Code

UI: -

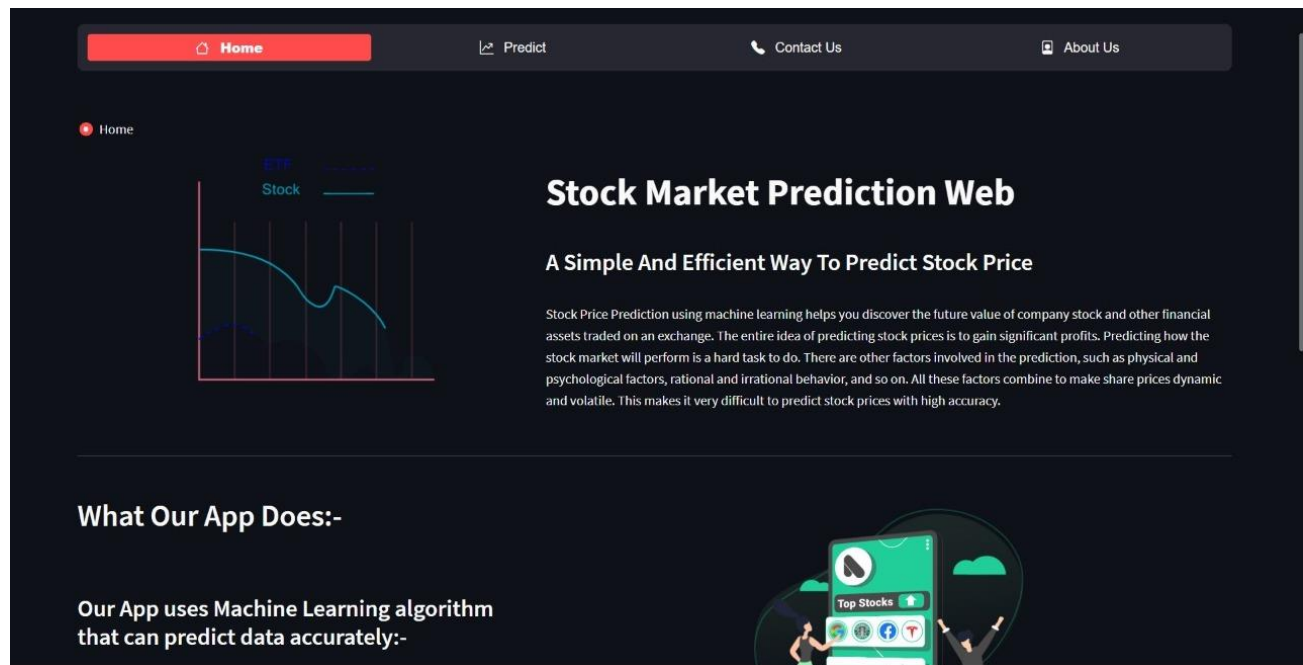


Fig 15- UI-1

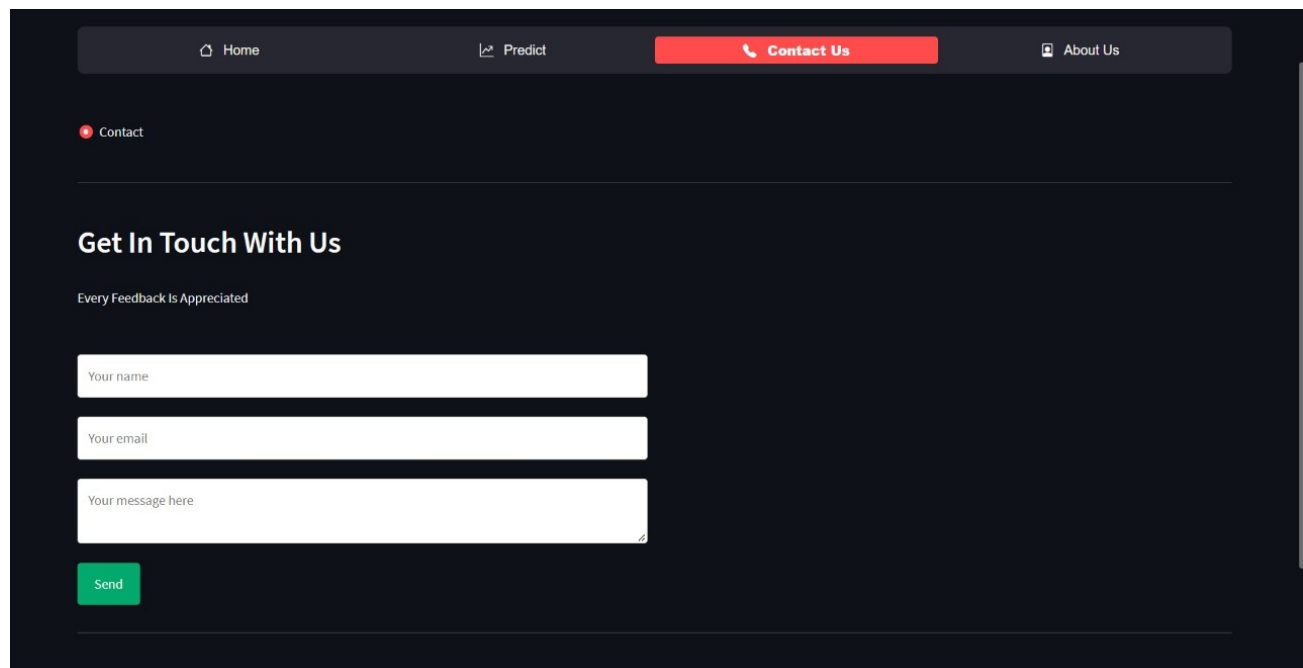


Fig 16- UI-2

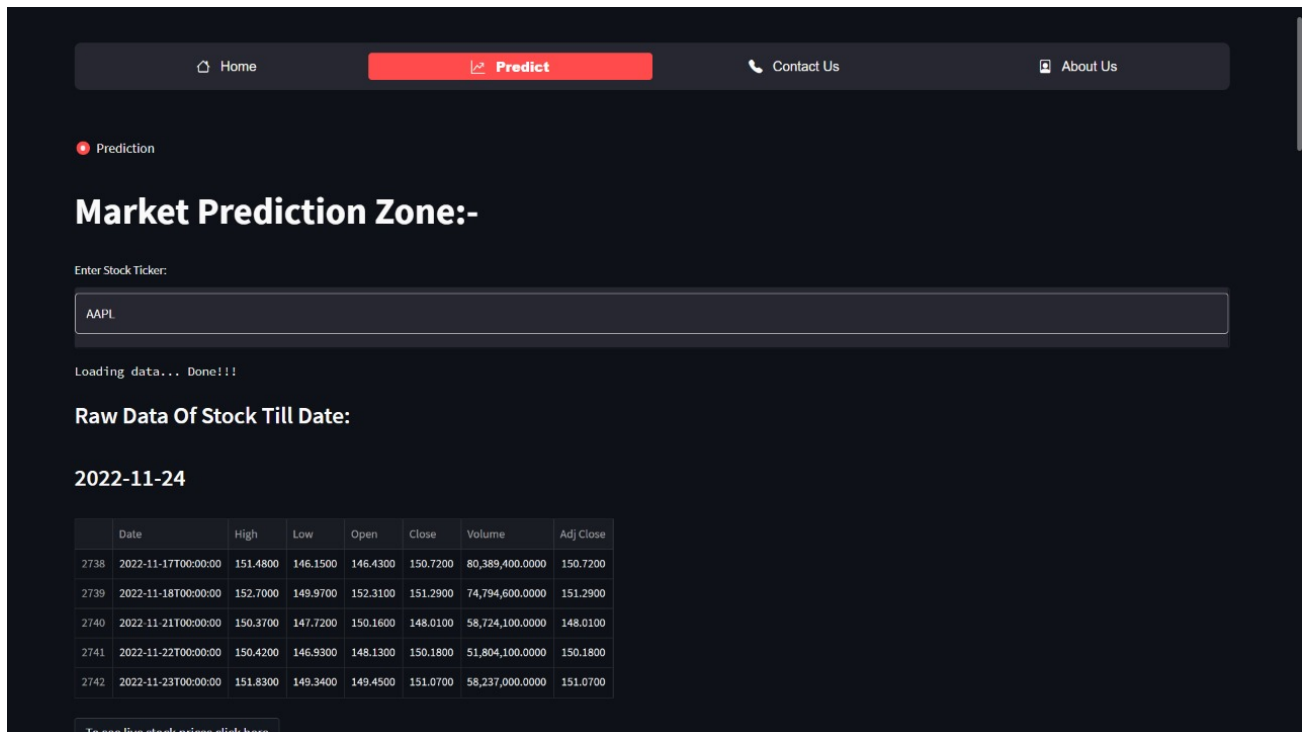


Fig 17- UI-3

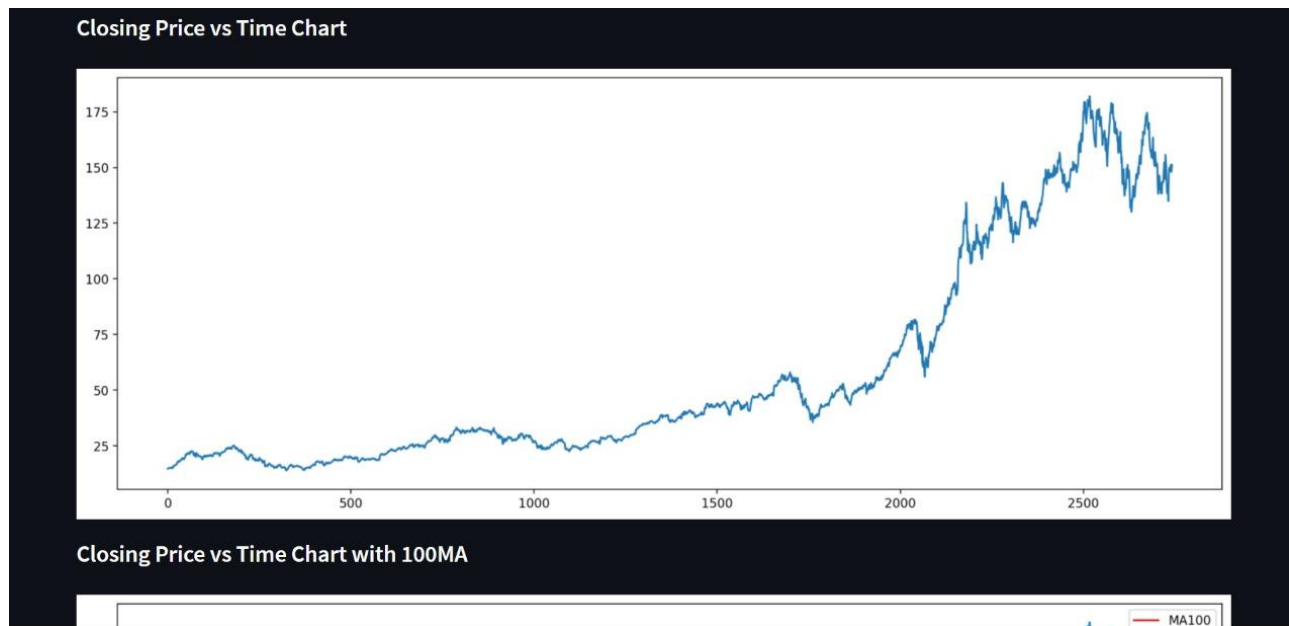


Fig 18- UI-4

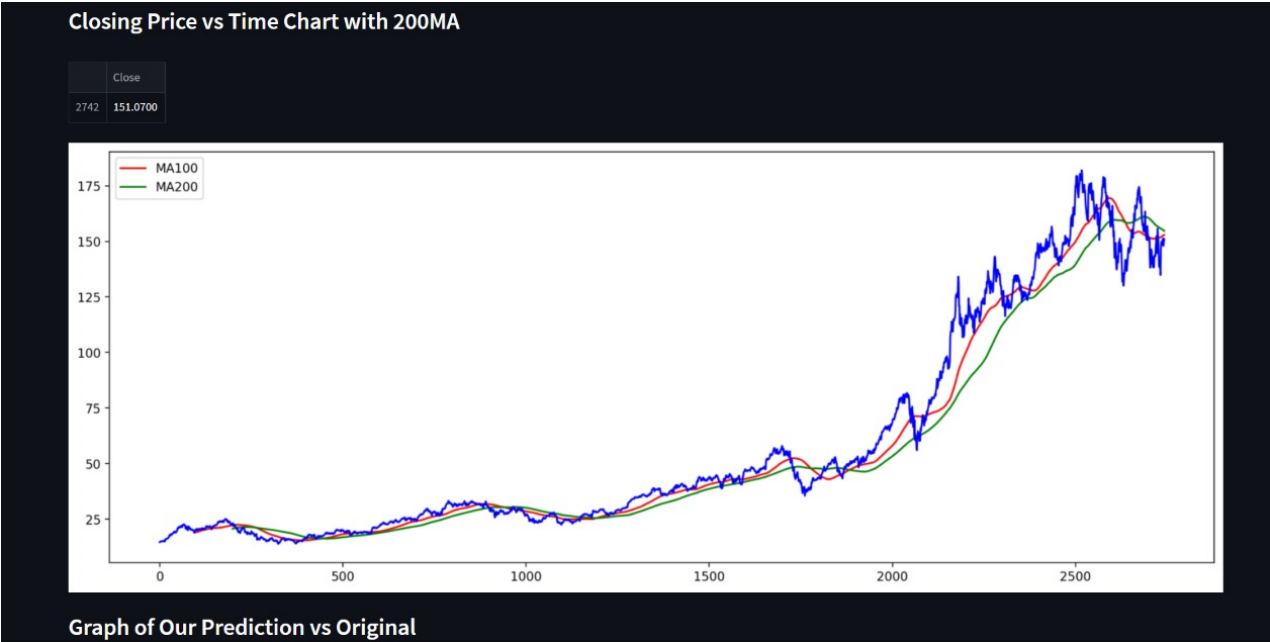


Fig 19- UI-5

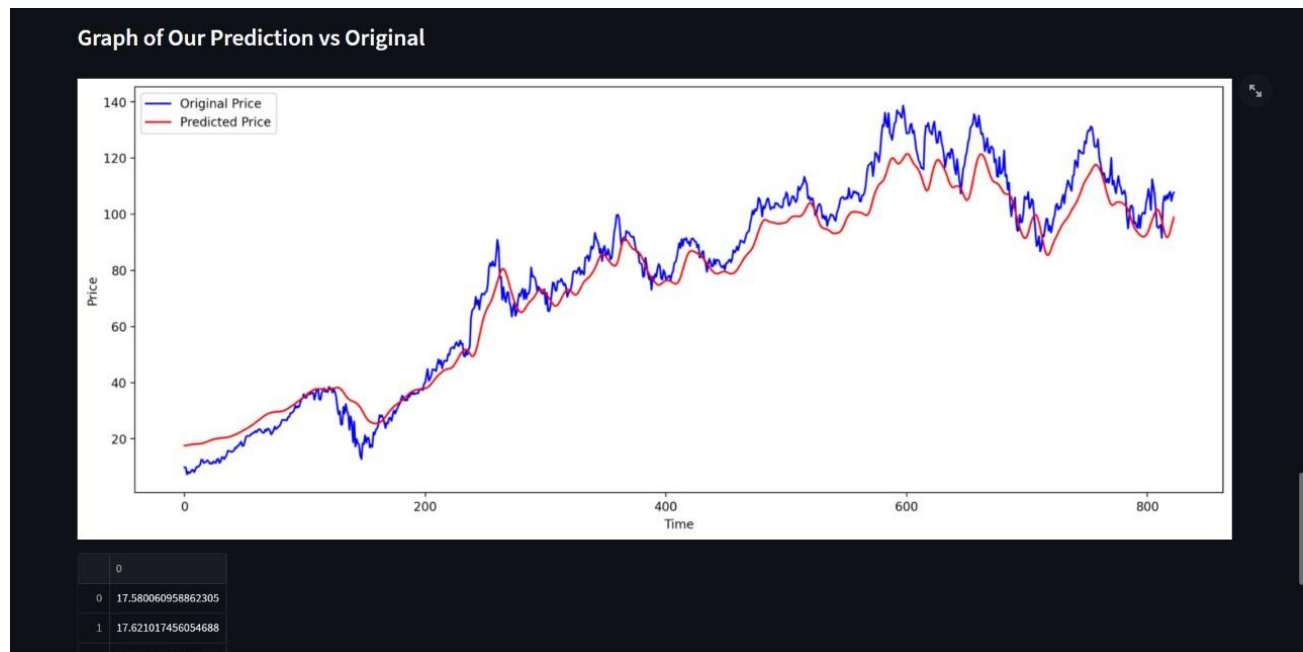


Fig 20-UI-6

Chapter 5 – Conclusion & Future Scope

- To summarise, stocks is an unpredictably unexpected system that monitors chain parts and their interdependencies. It is characterised as a curve that changes over time, shifting the value from bottom to top and vice versa.
- Because the degree of coordination is greater with the other dependencies, keeping one out compromises the degree of precision. Accuracy is not really a phrase used for the share markets since accurate forecasting for any financial day has been impossible because the market is always shifting and shifting the tables day and night. With more constituent components and interconnections, it becomes more possible and adaptable, making it even more difficult to anticipate.
- The estimated amount is factored, and the hit, gain, or benefit rate for the very same is computed.
- Many high-level machine learning algorithms have been implemented and integrated into the project and the signal is taken from it, which tries to make the user recognizable with the output in graphical form so that they are very easy to see and understand. No matter what the situation it will be easier, and then they will decide to engage in the same work and then benefit from everything. The data is filtered and cleaned before it can be analyzed to get the desired result. The stock market is the finest way for businesses to expand, so it provides an extra profit for those who are interested in investing and getting profit from it. Since then, the terminology "stock" is still in use, and it continues to expand in

popularity. Investors and traders have put their money into it and profited handsomely. Intermediate agents and stocks sellers as well as others who study and engage in the very same. The expense of the stock consultation is substantial and costly. So, whenever it comes to individuals, they think lots and invest heavily, but there is no guarantee that the same would deliver a term fiscal.

- As a result, stocks are more volatile than ever before, and the likelihood of their rise is greater than before. If the stock market and its forecasting can be done properly, it will be beneficial to both people and organizations. The potential risk must be reduced so that the game's effectiveness may be maximised, and individuals can be confident in their big investment.
- The accuracy can be increased by many methods, such as real - world scenarios and the methods for performing and analysing real life situations

REFERENCE

- Downey, A., 2008. How to think like a computer scientist: learning with python. Green Tea Press.
- Nixon, R., Learning, P.H.P. and MySQL, J., 2016. CSS & HTML5: Guide to creating powerful websites, St. Petersburg: Piter, (4), pp.86-89.
- Forouzan, B.A. and Fegan, S.C., 2007. Data Communication and Networking. McGraw-Hill Companies. Inc. New York.
- Robey, A., Shukla, K., Agarwal, K., Joshi, K. and Joshi, S., 2019. Personality prediction system through CV analysis. Int. Res. J. Eng. Techno, 6(2), pp.2395-0056.
- Arora, A. and Arora, N.K., 2020. Personality Prediction System with CV Analysis. Ambient Communications and Computer Systems, pp.337-343.
- Dheeraj, K.N. and RJ, G., 2021. Personal Prediction Using CV Analysis. International Journal on Cybernetics & Informatics (IJCI), 10 (1), p.2.
- YouTube link- Machine Learning full course-10 Hours | Learn Machine Learning from Scratch | Eureka , 2020