- **TITLE PAGE**

PROJECT REPORT

(Project Term January-May 2021)

(Share Market Prediction using Machine Learning)

Submitted by

(Suman Kumar Das) **Registration Number : 11907812**

Project Group Number 15

Course Code : INT246

Under the Guidance of

(Dr. Sagar Pande)

School of Computer Science and Engineering

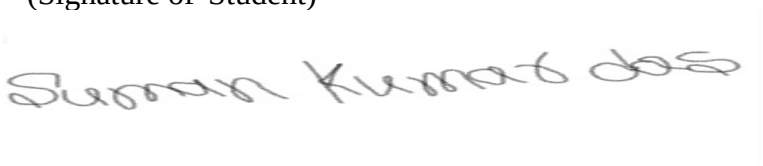
DECLARATION

We hereby declare that the project work entitled ("Share Market Prediction using Machine Learning") is an authentic record of our own work carried out as requirements of Project for the award of B.Tech degree in (Computer Science and Engineering) from Lovely Professional University, Phagwara, under the guidance of (Name of Faculty Mentor), during August to November 2020. All the information furnished in this project report is based on our own intensive work and is genuine.

Project Group Number: 15

Name of Student : Suman Kumar Das
Registration no. : 11907812

(Signature of Student)



Date: 20/112021

TABLE OF CONTENTS

1. INTRODUCTION

The first-level subheading uses the formatting style Heading 1. It is centered, boldface, single line spaced, and it advances the text after it by two lines (24pt). No extra carriage returns are needed to correctly space the text that follows. First-level subheadings should be in all capital letters. You must capitalize the first-level subheadings yourself. MS Word can simulate Title Case capitalization but it will capitalize the first letter of EVERY word including articles and prepositions. First-level subheadings must not have more than a single blank line space before or after the heading.

1.1. SECOND-LEVEL SUBHEADING

This is the second-level subheading of the first section. The second-level subheading uses the formatting style Heading 2. It is left aligned, boldface and single spaced, and it advances the text after it by one line. Second level subheadings are in Title Case (The first letters of principal words must be capitalized).

1.2. ANOTHER SECOND-LEVEL SUBHEADING

The format of this subheading is the same with the first one. The purpose of this subheading is to show you that if you have a subheading of a certain level, you must have more than one. The rationale is that you cannot have a list of only one item.

1.2.1. THIRD-LEVEL SUBHEADING

The third-level subheading uses the formatting style Heading 3. It uses the same formatting with the second-level subheading except that for the third-level subheading, only the first letter of the first word and proper nouns are capitalized (Sentence case).

1.2.2. ANOTHER THIRD-LEVEL SUBHEADING

The heading above shows that if you have a subheading of a certain level, you must have more than one. The rationale is that you cannot have a list of only one item.

CERTIFICATE

This is to certify that the declaration statement made by this group of students is correct to the best of my knowledge and belief. They have completed this Project under my guidance and supervision. The present work is the result of their original investigation, effort and study. No part of the work has ever been submitted for any other degree at any University. The Project is fit for the submission and partial fulfillment of the conditions for the award of B.Tech degree in (Computer Science and Engineering) from Lovely Professional University, Phagwara.

Signature and Name of the Mentor

Designation

School of Computer Science and Engineering,
Lovely Professional University,
Phagwara, Punjab.

Date : 20/11/2021

CONTENTS

1. INTRODUCTION

INFORMATION ON SHARE

We all have heard the word share one way or the other. Particularly share is related with the associates and companies which are commercialized and are to settling in the world of marketization. The other word used for share is share which is prominently used in day to day life. People even term is as an investment plan and its something people see as a long term investment that secures and provides an abundant funds during the retirement age.

Buying a company share is purchasing a small share of it. People invest on the same to get a long term benefit which they think is less value for now but has to potential to grow with the time. Its an investment that provides the long time run and deals with long time goals with the fair objectives. The value of share you invest today has to give you an yield of best tomorrow but its not the same.

Market is unpredictable so are the resources and the factors that are taken to drive it off or on the set. Its never been on the same level and the pattern of the same is still unpredictable till the time. Some closeness and prediction method had been derived and approximates values and the rough figures are generated hoping for the best but all of the resource can’t be trusted and are still unpredictable in nature.

Knowing the market situation and researching on the same is the best way to find the reliability for which there are many agents who have taken the same as a profession and are making a fortune out of it. They predict and advise but the advisory

cost and the charge is higher and the stock evaluation is never less the same.

2. PROBLEM DEFINITION

Share is an unpredictable curve that had been in picture ever since. Its essence had been ever long living and indulging. It had grown its popularity with respect to time.

People are more fascinating and interested on the same then before times. Same for the case for the organization. Organization had created it as a better source of revenue generation rather than investing and taking a loan approval from the bank It's way efficient and less hectic from the firm point of view.

Share is unpredictable and its been the same from the start. Its way of escalating and deescalating had been phenomenon and experiencing the same is the best integral part of it. It has its upper hand and flexibility with the changes that has the chances of uprising as well as crashing the whole market. Its easily defined in few words but making an essence and understanding the same is way more hectic and time consuming.

Simpler it sound complex are its phenomenon and integrating the same. Its has its whole different sets of dependencies and integration from different agents which fluctuate the same in the market. Finding an accurate and getting the exact values out of the same is still unaligned and no particular model of the same is seen in the market value.

Finding the closest and getting an accurate proximate value out of such an

unpredictability is a problem in itself. Merging of the data getting the best prediction to

increase the efficiency alongside considering the different expects of the moderator is tough and we took the same in consideration and implemented with every aspect to generate the best out of the same and get a result that can be better interrupted and the efficiency remains the same with the value of different aspects of creating an impact

of reducing the risk and influencing the same over the time period to gain the most out of it.

This is totally based on Machine Learning Algorithm to proceed and provide an effective result. Getting the data and processing it and generating a forecast for three days is the problem statement that we worked on.

3. Data Flow Diagram (DFD)

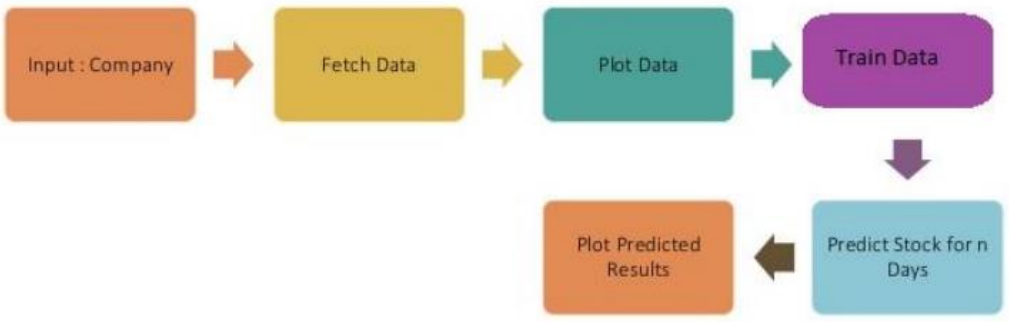


Fig. Data Flow Diagram

In the above fig. we are taking a company fetching the data of the company from the panda's data-reader library then we are plotting the data, then we train the data to

predict the share for certain number of days. In this way data is flowing in our system.

4. PROJECT PURPOSE

Share market prediction is a prediction system software that illuminate the risk that undergoes during the investment in share market. It predicts the share rates and its rate of exchange acknowledging the basic understanding and the statistical analysis in front of users.

Data is considered as the digital fuel that gives the possibilities of higher yearn and gives the upcoming terms. Knowledge is power and same holds correct with the share. Share is unpredictable and over-changing its dynamic in nature. The rise and fall of the same is uneven and can't be classified so easily. Dependencies of the same deals with flexible resources and the agents behind it.

Investment during a fiscal day determines the opening share market for the next day. It has its dependencies and is total integration with the level of finances and revenue generation. The share is tremendous and hectic in nature. The main theme of the project is to predict the turning curves and bring the predictability method and undergo the process and algorithms to conclude to a viable resource source.

Everything flows a pattern. Pattern is the way of derivation and so holds true for the share too. Share in day to day life follows a pattern movement. Increase in some resource can increase the price of some whereas decrease the price rate for the others, The source and the outcome are derived on the polarity basis which can either be

positive, neutral or an negative flow. Correlation of the given polarity is determined and

an effective source and reliability is established.

This project helps in bridging the resources and empowering the people to know

and trade the most out of share and understand the generation and the vulnerabilities

that has to be seen and predicted.

5. SOFTWARE REQUIREMENT ANALYSIS

FUNCTIONAL REQUIREMENTS

Functional requirements deals with the functionality of the software in the

engineering view. The component flow and the structural flow of the same is enhanced

and described by it.

The functional statement deals with the raw datasets that are categorized and

learning from the same dataset. Later the datasets are categorized into clusters and the

impairment of the same is checked for the efficiency purpose. After the dataset cleaning

the data are cleansed and the machine learns and finds the pattern set for the same it

undergoes various iteration and produce output.

NON-FUNCTIONAL REQUIREMENTS

Non functional requirement deals with the external factors which are non-

functional in nature It is used for analysis purpose. Under the same the judgment of the

operations are carried out for its performance. Share is feasible and is ever changing so

these extra effects and the requirements helps it to get the latest updates and integrate in a one go where the technicians can work on and solve a bug or a draft if any.

The non-functional requirements followed are it's efficiency and hit gain ratio.

The usability of the code for the further effectiveness and to implement and look for the

security console. The System is reliable and the performance is maintained with the support of integration and portability of the same.

SOFTWARE REQUIREMENTS

- Operating system : Windows
- IDE : Jupiter Notebook
- Data Set : .csv file
- Visualization : mat plot lib, pandas.
- Server : Web Server with HTTP process.

JUPYTER NOTEBOOK

Jupiter Notebook or so called IPython Notebook is an interactive web based computational mean for starting with Jupiter Notebook documents. The term notebook

itself is a huge entity to represent the integration with different entity sets. JSON is the main document form from the same for the execution which follows the brief on the schema and the input and output means. It has high integration with several language

set and has various flexibilities with the choices.

6. DESIGN

DESIGN GOALS

To make the project runs smoothly it's required that we make plan and design some accepts like flowcharts and system architecture which are defined below.

Data Collection

Data collection is one of the important and basic thing in our project. The right dataset must be provided to get robust results. Our data mainly consists of previous year or weeks share prices. We will be taking and analyzing data from Kaggle. After that seeing the accuracy we will use the data in our model.

Data Preprocessing

Human can understand any type of data but machine can't our model will also learn from scratch so it's better to make the data more machine readable. Raw data is usually inconsistent or incomplete .Data preprocessing involves checking missing values, splitting the dataset and training the machine etc.

Training Model

Similar to feeding somethings, machine/model should also learn by feeding and learning on data. The data set extracted from Kaggle will be used to train the model. The training model uses a raw set of data as the undefined dataset which is collected from

the previous fiscal year and from the same dataset a refine view is presented which is seen as the desired output. For the refining of the dataset various algorithms are implemented to show the desired output

SYSTEM ARCHITECTURE

The dataset we use for the proposed project is been taken from Kaggle. But, this data set is in raw format. The data set is a collection of valuation of share market information about some companies. The initial step is to convert raw data into processed data. Which is done by feature extraction, since the raw data collected have multiple attributes but only some of those attributes are needed for the prediction. Feature extraction is a reduction process.

The structure, behavior and views of a system is given by structural model.

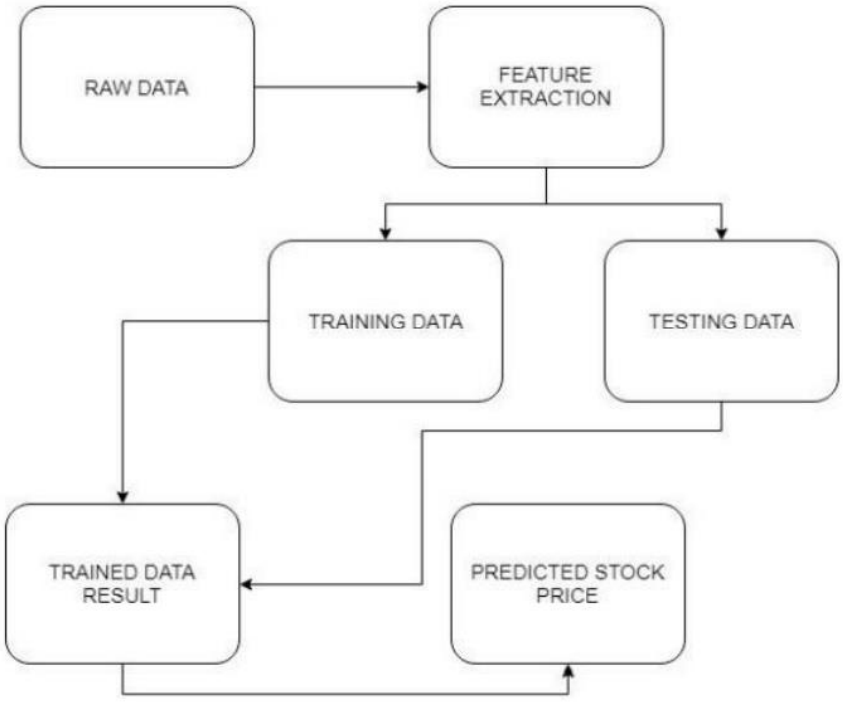


Fig. System Architecture

The above figure gives the demonstration on the dataset extraction and refining the raw dataset by categorizing into two phase of training and testing data. From the given dataset a well modified categorization is extracted and a graph set is plotted to gain the required output which gives the share prediction range.

Use case Diagram

A dynamic and behavioral diagram in UML is use case diagram. Use cases are basically set of actions, services which are used by system. To visualize the functionality

requirement of the system this use case diagram are used. The internal and external events or party that may influence the system are also picturized. Use case diagram specify how the system acts on any action without worrying to know about the details how that functionality is achieved.

For the project we have created the below mentioned use case diagram.

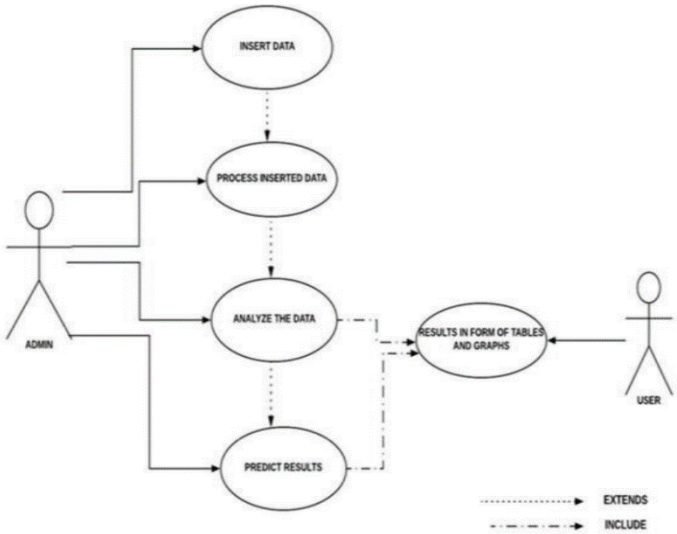


Fig. Use Case Diagram

The above figure shows the use-case diagram of the entitled project and it’s flow. From the diagram it’s seen that the user gives the raw dataset as input and with the flow of the input in the system.

The system evaluates and process the dataset train itself with the provided dataset and extract the meaningful dataset to process and refine the cluster data and from the given cluster of the data, the plotting of the data values are shown and with

7. TESTING

The purpose of testing is to get errors. Testing is that the process of trying to get every conceivable fault or weakness during a work product. It provides how to see the functionality of components, sub assemblies, assemblies and/or a finished product it's the method of exercising software with the intent of ensuring that the software meets its requirements and user expectations and doesn't fail in an unacceptable manner. There are various sorts of test. Each test type addresses a selected testing requirement. The various types of testing that follows are listed as below.

UNIT TESTING

Unit testing involves the planning of test cases that validate that the interior program logic is functioning properly, which program inputs produce valid outputs. All decision branches and internal code flow should be validated. it's the testing of individual software units of the appliance.

It is done after the completion of a private unit before integration. this is often a

structural testing, that relies on knowledge of its construction and is invasive. Unit tests

perform basic tests at component level and test a selected business process, application,

and/or system configuration.

Unit tests make sure that each unique path of a business process performs

accurately to the documented specifications and contains clearly defined inputs and

expected results.

INTEGRATION TESTING

Integration tests are designed to check integrated software components to work

out if they really run together program. Testing is event driven and is more concerned

with the essential outcome of screens or fields.

Integration tests demonstrate that although the components were individually

satisfaction, as shown by successfully unit testing, the mixture of components is correct

and consistent. Integration testing is specifically aimed toward exposing the issues that

arise from the mixture of components.

VALIDATION TESTING

Validation testing is that the process of ensuring if the tested and developed

software satisfies the client /user needs. The business requirement logic or scenarios

need to be tested intimately . All the critical functionalities of an application must be

tested here.

As a tester, it's always important to understand the way to verify the business

logic or scenarios that are given to you. One such method that helps intimately evaluation of the functionalities is that the Validation Process.

SYSTEM TESTING

System testing of software or hardware is testing conducted on an entire , integrated system to guage the system's compliance with its specified requirements. System testing falls within the scope of recorder testing, and intrinsically , should require

no knowledge of the inner design of the code or logic.

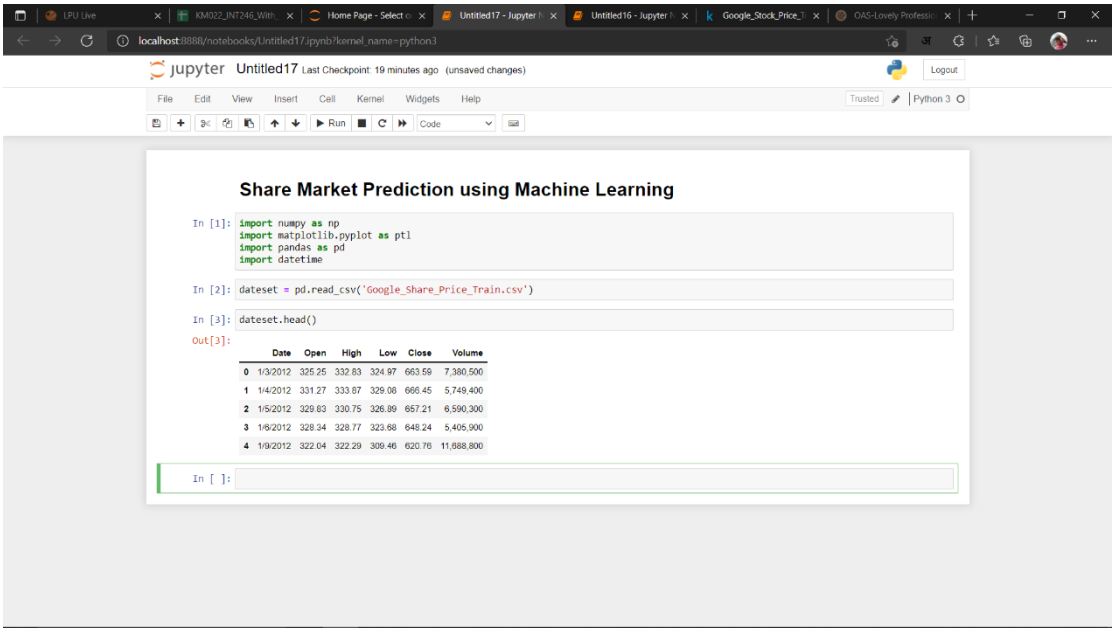
As a rule, system testing takes, as its input, all of the "integrated" software components that have successfully passed integration testing and also the software itself integrated with any applicable hardware system(s).

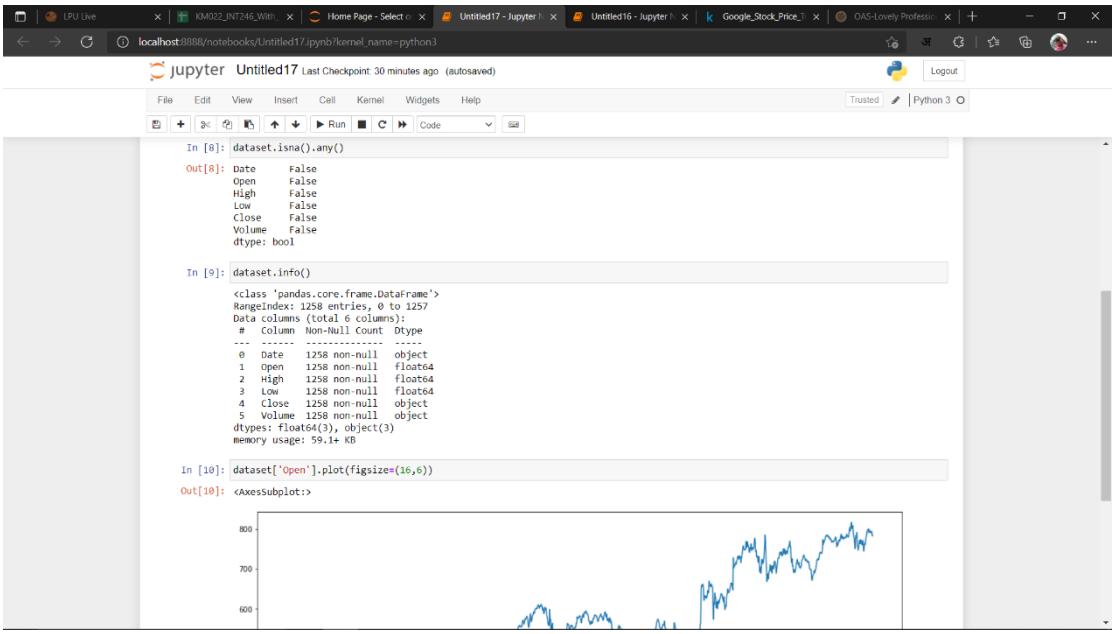
System testing may be a more limited sort of testing; it seeks to detect defects both within the "inter-assemblages" and also within the system as an entire .

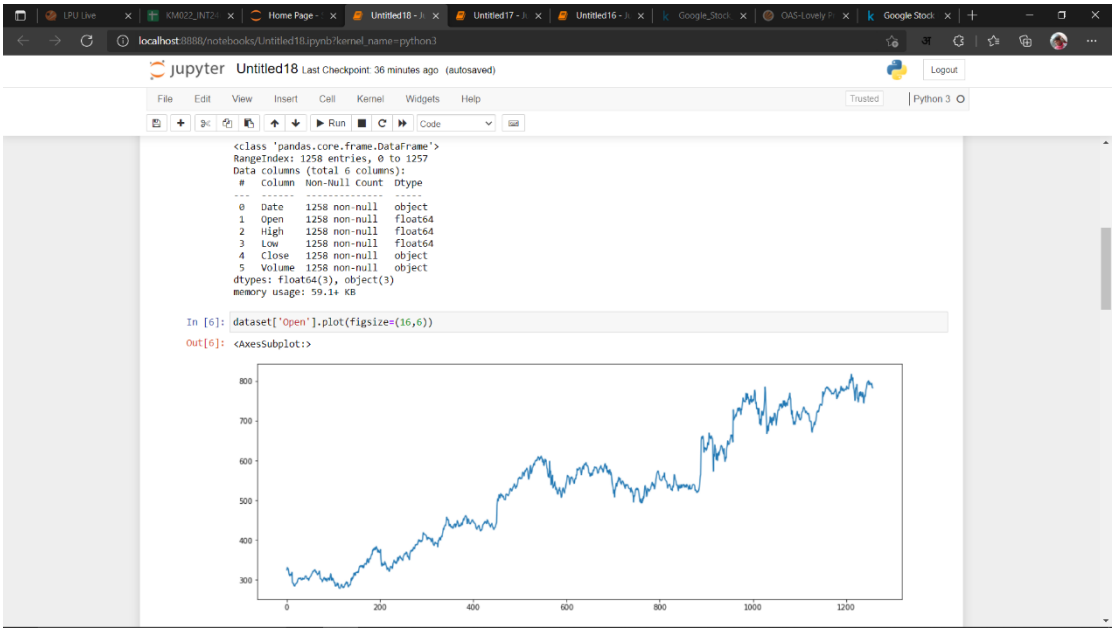
System testing is performed on the whole system within the context of a Functional Requirement Specification(s) (FRS) and/or a System Requirement Specification (SRS).

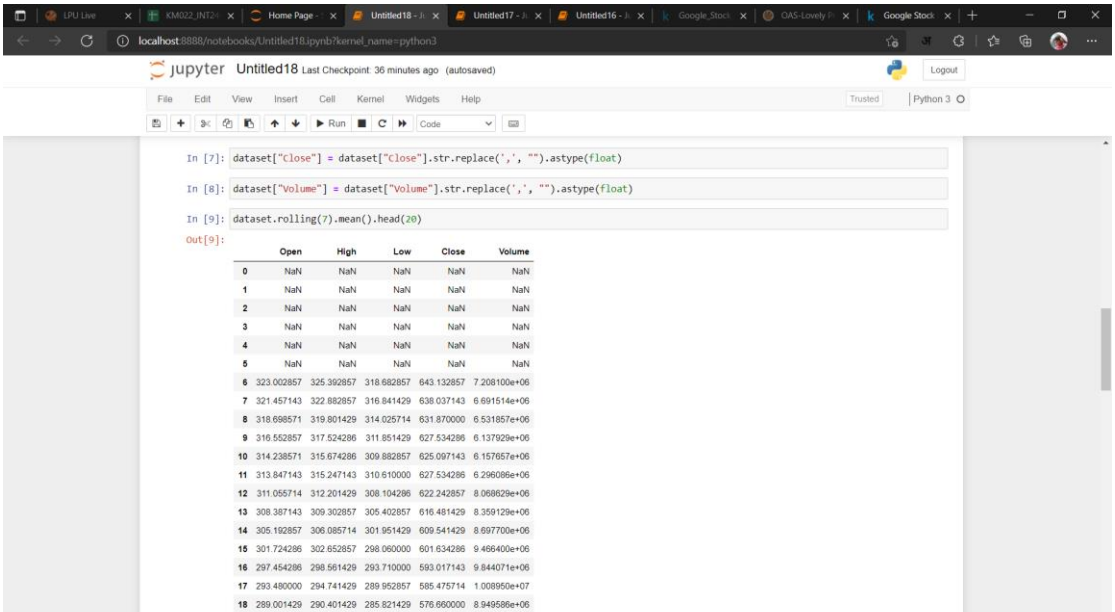
System testing tests not only the planning , but also the behavior and even the believed expectations of the customer. it's also intended to check up to and beyond the bounds defined within the software/hardware requirements specification.

8. Source Code (where ever applicable) or System Snapshots

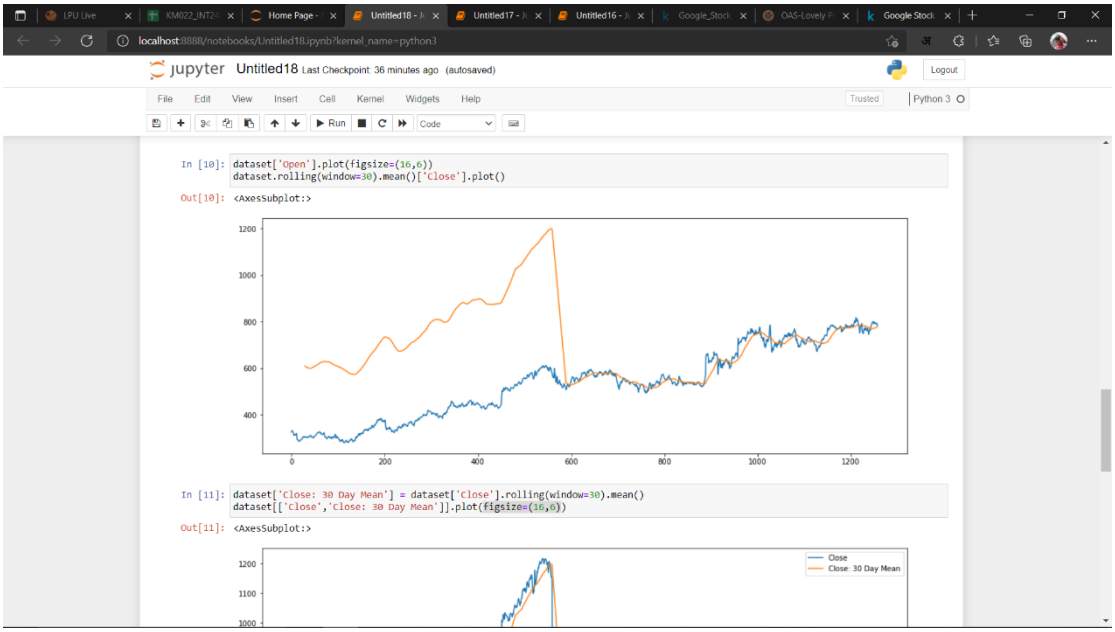

jupyter Untitled17 Last Checkpoint: 19 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3
Run Code
Share Market Prediction using Machine Learning
In [1]:
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import datetime
In [2]:
dateset = pd.read_csv('Google_Share_Price_Train.csv')
In [3]:
dateset.head()
Out[3]:
Date Open High Low Close Volume
0 1/3/2012 325.25 332.83 324.97 663.59 7,380,500
1 1/4/2012 331.27 333.87 329.08 666.45 5,749,400
2 1/5/2012 329.83 330.75 326.89 657.21 6,590,300
3 1/6/2012 328.34 328.77 323.68 648.24 5,405,900
4 1/9/2012 322.04 322.29 309.46 620.76 11,688,800
In [ ]:

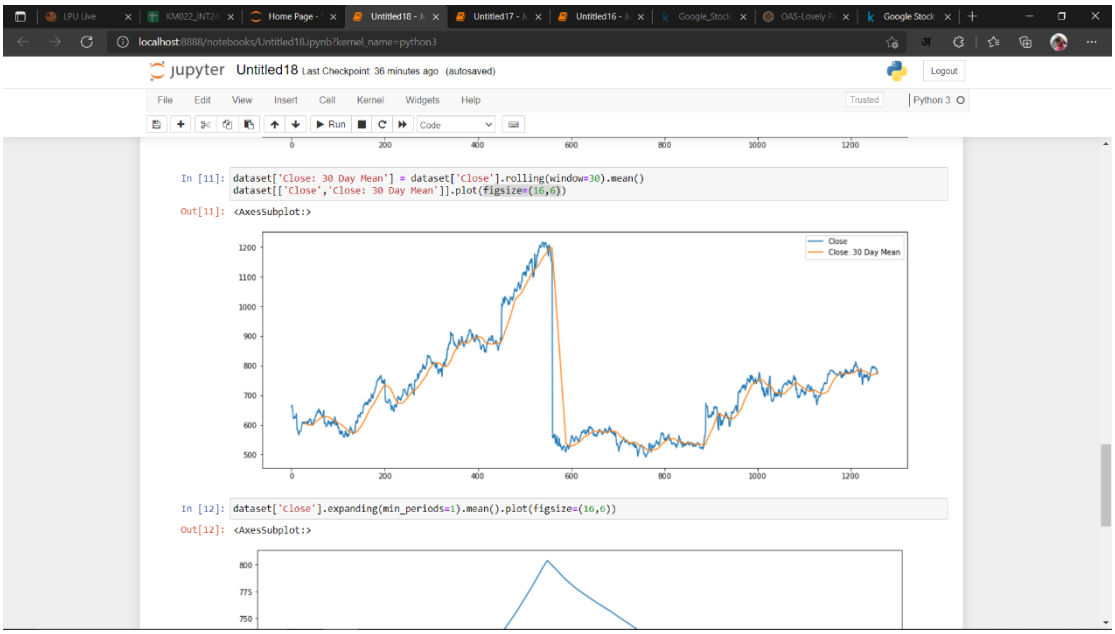

jupyter Untitled17 Last Checkpoint: 30 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3
Run Code
In [8]:
dataset.isna().any()
Out[8]:
Date False
Open False
High False
Low False
Close False
Volume False
dtype: bool
In [9]:
dataset.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1258 entries, 0 to 1257
Data columns (total 6 columns):
Column Non-Null Count Dtype
0 Date 1258 non-null object
1 Open 1258 non-null float64
2 High 1258 non-null float64
3 Low 1258 non-null float64
4 Close 1258 non-null object
5 Volume 1258 non-null object
dtypes: float64(3), object(3)
memory usage: 59.1+ KB
In [10]:
dataset['Open'].plot(figsize=(16,6))
Out[10]:
<AxesSubplot:>
800
700
600

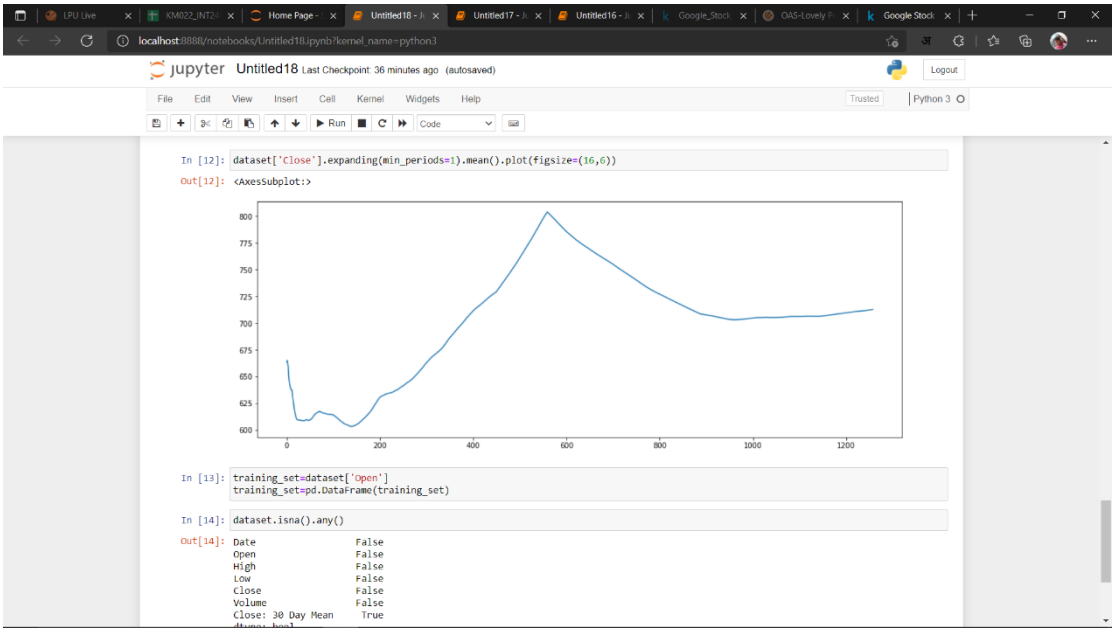




	Open	High	Low	Close	Volume
0	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN
5	NaN	NaN	NaN	NaN	NaN
6	323.002857	325.392857	318.682857	643.132857	7.208100e+06
7	321.457143	322.882857	316.841429	638.037143	6.691514e+06
8	318.698571	319.801429	314.025714	631.870000	6.531857e+06
9	316.552857	317.524286	311.851429	627.534286	6.137929e+06
10	314.238571	315.674286	309.882857	625.097143	6.157657e+06
11	313.847143	315.247143	310.610000	627.534286	6.296086e+06
12	311.055714	312.201429	308.104286	622.242857	8.068629e+06
13	308.387143	309.302857	305.402857	616.481429	8.359129e+06
14	305.192857	306.085714	301.951429	609.541429	8.697700e+06
15	301.724286	302.652857	298.060000	601.634286	9.466400e+06
16	297.454286	298.561429	293.710000	593.017143	9.844071e+06
17	293.480000	294.741429	289.952857	585.475714	1.008950e+07
18	289.001429	290.401429	285.821429	576.660000	8.949586e+06
19	288.465714	289.902857	285.355714	575.821429	6.530857e+06


jupyter Untitled18 Last Checkpoint: 36 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [10]:
dataset['Open'].plot(figsize=(16,6))
dataset.rolling(window=30).mean()['Close'].plot()
Out[10]:
<AxesSubplot:>
In [11]:
dataset['Close: 30 Day Mean'] = dataset['Close'].rolling(window=30).mean()
dataset[['Close','Close: 30 Day Mean']].plot(figsize=(16,6))
Out[11]:
<AxesSubplot:>
Close
Close: 30 Day Mean


jupyter Untitled18 Last Checkpoint: 36 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [11]:
dataset['Close: 30 Day Mean'] = dataset['Close'].rolling(window=30).mean()
dataset[['Close','Close: 30 Day Mean']].plot(figsize=(16,6))
Out[11]:
<AxesSubplot:>
Close
Close: 30 Day Mean
In [12]:
dataset['Close'].expanding(min_periods=1).mean().plot(figsize=(16,6))
Out[12]:
<AxesSubplot:>


jupyter Untitled18 Last Checkpoint: 36 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Code
In [12]: dataset['Close'].expanding(min_periods=1).mean().plot(figsize=(16,6))
Out[12]: <AxesSubplot:>
800
775
750
725
700
675
650
625
600
0
200
400
600
800
1000
1200
In [13]: training_set=dataset['Open']
training_set=pd.DataFrame(training_set)
In [14]: dataset.isna().any()
Out[14]: Date False
Open False
High False
Low False
Close False
Volume False
Close: 30 Day Mean True