**A Project Report**

On

**“STOCK PREDICTION USING MACHINE LEARNING”**

Submitted to the

Savitribai Phule Pune University

In Partial fulfillment for the award of the Degree of

Bachelor of Technology

in

Information Technology

By

Aditya Munjal [BITA04]

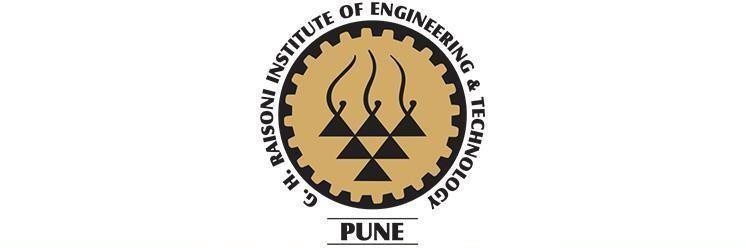
Atharva Chavan [BITA08]

Fazal Azizi [BITA10]

Ganesh Shinde [BITA12]

Under the guidance of

Ms. Vaishali Anaspure



**DEPARTMENT OF INFORMATION TECHNOLOGY**

**G. H. RAISONI COLLEGE OF ENGINEERING AND MANAGEMENT**

(An Autonomous Institute affiliated to SPPU)

WAGHOLI, PUNE – 412207

**SAVITRIBAI PHULE PUNE UNIVERSITY**

**2023-24**

# CERTIFICATE

This is to certify that the project-based seminar report entitled **“STOCK PREDICTION USING MACHINE LEARNING”** being submitted by **Aditya Munjal, Atharva Chavan, Fazal Azizi, Ganesh Shinde** is a record of bonafide work carried out by him/her under the supervision and guidance of **Ms. Vaishali Anaspure** in partial fulfillment of the requirement for **B. Tech (Information Technology Engineering) – 2020 course** of Savitribai Phule Pune University, Pune in the academic year 2023-2024.

Date:

Place: Wagholi, Pune.

|  |  |  |
| --- | --- | --- |
| **Ms. Vaishali Anaspure** | **Dr. Shaikh Abdul Waheed** | **Dr. Poonam Gupta** |
| (Guide Name) | (Project Coordinator) | (HOD) |

**External Name & Sign Dr . R . D. Kharadkar**

(Director)

# ACKNOWLEDGEMENT

We here wish to take this opportunity to express our gratitude to our Project Guide Ms. Vaishali Anaspure, Project Review Committee Members and Head of Department Dr. Poonam Gupta for their consistent guidance and motivation toward the completion of our project.

We take a great honor in presenting this Project Report to our Director, Dr. R. D. Kharadkar.

We are very grateful to our teaching staff for guiding us all over the duration of the degree. They were very helpful to us as and when we required their help. We are also very grateful to non-teaching staff to help us in the laboratory in various ways.

We would also like to extend our gratitude to those friends whose knowledge and time helped us in many ways.

|  |  |  |
| --- | --- | --- |
| **Name** | **Roll No** | **Sign** |
| Aditya Munjal | BITA04 |  |
| Atharva Chavan  Fazal Azizi  Ganesh Shinde | BITA08  BITA10  BITA12 |  |

# ABSTRACT

Stock is an unpredictable curve. Prediction in stock market is covered with the complexity and instability. The main aim for the persuasion of the topic is to predict the stability in the future market stocks. Many researchers have performed their research on the movement of future market evolution. Stock consists of fluctuating data which makes data as an integral source of efficiency.

Impact on the same chances the efficiency of the prediction. In the recent trend of Stock Market Prediction Technologies machine learning has integrated itself in the picture for deployment and prediction of training sets and data models.

Machine Learning employs different predictive models and algorithms to predict and automate things of requirement. The Paper focuses on the use of Regression and LSTM to predict stock values.

# CONTENTS PAGE NO

### ACKNOWLEDGEMENT [ABSTRACT](#_bookmark0)

**LIST OF FIGURES LIST OF TABLES CHAPTER 1**

1. INTRODUCTION 1
   1. [INFORMATION ON STOCK MARKET 1](#_bookmark1)
   2. [MODULES DESCRIPTION](#_bookmark2) 2
   3. [MOTIVATION & OBJECTIVE](#_bookmark3) 3

[CHAPTER 2](#_bookmark5)

1. LITERATURE SURVEY… 5
   1. [EXISTING SYSTEM 6](#_TOC_250013)
   2. [PROPOSED SYSTEM 7](#_TOC_250012)

[CHAPTER 3](#_bookmark4)

1. PROBLEM DEFINITION 9
   1. [PROBLEM TO BE SLOVED](#_bookmark7) 9
   2. [PROJECT PURPOSE 9](#_TOC_250011)
   3. [METHOD USED 10](#_TOC_250010)
   4. [TECHNOLOGIES USED 11](#_TOC_250009)

CHAPTER 4

1. [REQUIREMENT ANALYSIS 1](#_bookmark6)4
   1. FUNCTIONAL REQUREMENTS 14
   2. NON- FUNCTIONAL REQUREMENTS 14
   3. HARDWARE REQUREMENTS 14
   4. SOFTWARE REQUREMENTS 15
   5. [SOFTWARE DESCRIPTION 15](#_TOC_250008)

[CHAPTER 5](#_TOC_250007)

1. HIGH LEVEL DESIGN… 16
   1. DESIGN GOALS 16
   2. [SYSTEM ARCHITECTURE](#_bookmark8) 17
   3. [USE CASE DIAGRAM](#_bookmark9) 18

i

* 1. [DATA FLOW DIAGRAM](#_bookmark10) 19

CHAPTER 6

1. IMPLEMENTION 21
   1. [LINEAR REGRESSION 2](#_bookmark11)1
   2. [POLYNOMIAL REGRESSION 2](#_bookmark12)1
   3. [ARIMA MODEL 22](#_TOC_250006)
   4. [NESPY 23](#_TOC_250005)
   5. [CELERY 23](#_TOC_250004)
   6. [RADIS SERVER 23](#_TOC_250003)
   7. [EDA ALGORITHM 24](#_TOC_250002)
   8. [THRESHOLDING WITH SHORT TERM MEMORY 24](#_TOC_250001)
   9. [COMPARITIVE STUDY OF IMPLEMENTED MODULES 25](#_TOC_250000)

[CHAPTER 7](#_bookmark13)

1. REASULT 26
   1. LOGIN PAGE 26
   2. REGISTRATION PAGE 27
   3. IBM STOCK DISPLAYING CHART 28
   4. CHART OF TATA POWE STOCK 29
   5. LIST OF STOCKS 30
   6. ACCURACY OF LINER REGRESSION 31
   7. PREDICTION 32

[CHAPTER 8](#_bookmark14)

1. CONCLUSION AND FUTURE WORK 33
   1. [CONCLUSION](#_bookmark15) 33
   2. [FUTURE WORK 3](#_bookmark16)3

[CHAPTER 9](#_bookmark17)

1. REFERENCE 35
2. PUBLICATION 36

ii

# LIST OF FIGURES

**Fig no. Figure Name Page no.**

3.2.1 System Flow 9

3.3.2 LSTM diagram 12

* 1. System Architecture 20
  2. Use Case Diagram 22
  3. Data Flow Level 0 Diagram 23
  4. Data Flow Level 1 Diagram 24
  5. Screenshots 30
     1. Login Page 31
     2. Registration Page 32
     3. IBM stock displaying chart 33
     4. Chart of Tata Power Stock 34
     5. List of Stocks 35
     6. Accuracy of liner Regression 36
     7. Prediction 37

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Fig.no** | **Table Name** | **Page No.** |
| 1 | LITEATURE SURVEY TABLE | 5, 6 |
| 2 | TECHNOLOGY USED TABLE | 11 |
| 3 | SOFTWARE REQUIREMENT TABLE | 15 |
| 4 | COMPARITIVE STUDY OF IMPLEMENTED  MODULES | 25 |

**CHAPTER 1**

**INTRODUCTION**

# INFORMATION ON STOCK MARKET

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

Buying a company stock 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. It’s 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 a yield of best tomorrow but it’s not the same.

Market is unpredictable so are the resources and the factors that are taken to drive it off or on the set. It’s 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 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.

Market is changing in an instantaneous rate even in a day there are many highs and lows in the market and having said the resources and the timing the external and internal agent. Stock is a fascinating resource to start with. The main resources for the company are the fund to carry out the daily work and create a profit out of it. In time of need for a higher budget estimation and to overgrow from the resources they need the finance and undergoing a finance loan for approval, passing and having one is hectic and the banks are vultures for which the interest rate is higher than the other form of investment hence limiting the margin of the product. Stock is another way for company to collect revenue and boost up the production for the upper yield to gain the cost out of the business plan for the bigger picture. This is found to be effective way to invest

and grow in the commercial field and a better alternative to tackle the financial crisis during the requirement.

For an investor it’s a risk phenomenon where they invest their saving and hope it brings back the return in higher yield. If the evaluation of the same increases, then the stock evaluation and its price increases causing the financial gain to both the parties. In Indian Society it is even consider as a side point business and people believe it as a hand of luck. When an individual purchases a company stock then they’re referred as a shareholder and they will get a share out of the same as they have invested in their profit or the gain. An investor can sell and buy the stock as per their needs. They can share their stock to their respective or the other individuals where as there are many stock brokers available out in the firm playing with the same.

# MODULES DESCRIPTION

* + 1. **DATA SET**

This is the fundamental module before starting of the project. The dataset is a group of data that are mended together to show the data variations in a time span to undergo further estimation and the source of the resources and its outcome for the later time of evaluation. It generates the result optimization and gives a feasible time period to customize and get the flow to the derivation.

This increases and are used in the level of research and finding the best suitable resource of the same the resources have to be finely estimated and derived for the best possible outcome and the finest the value become the better is the level of extraction and closure is the best yield values that needs to be considered.

# DATA ABSTRACTION

Abstraction is the finding of the resource to its best to categorize the above dataset and learning the best out of it. Abstraction of the data is an integral part to the flow. All the data are a huge set of chunks which on processing can limited the yield result and the computational mean too. Thus, with the available resources the data yield had to be derivative.

Abstraction of the dataset is to customize the data set and finding the best suitable constraints to take into consideration and the unwanted resources are the dump which will be dumped and the supreme cluster is created with the valuable constrains and a pattern is needed to be derived from the same. Data are cleared on this level for the beginning of the process. The

valuable data are the set that brings the value to the data set for a better understanding and giving a better yield and production by evaluating the same.

This is a feature abstraction module to extract the featuring of the dataset. This is a feature model process where all the feasible resources are categorized and the same will be in use for the featuring.

# TRAINING DATASET

After the abstraction of the data and clustering of the same. The machine had to be trained for which the training data plays the important role. There are thousands of machine learning algorithms that are into place and evolving with the same. The best to the practice of machine learning is to yield the result and the content to derive what’s needed with the time frame.

This is a supervised learning form where the input is passed so that the system learns from the same. Various variants of inputs are passed which were stored in the dataset. Every resource is considered and taken into consideration. After considering the whole set of information and the resource the machine tries to learn from the passed dataset. The dataset has to be wide and versatile. After considering the learning it tries to integrate with the same type and flow like the same as the human mind and creates a pattern.

# MOTIVATION & OBJECTIVE

* + 1. **MOTIVATION:**
       - To determine the data fluctuating or stable entities for the learning and implement for the real-world use of system.
       - decipher real time changes
       - characteristics of stock price, increasing interest in investments in stalks, and awareness about means of studies for stalks.
       - Analysis is given on the trend in given for period of time

## Objective:

* Predict the future stalk rates and the approximate growth factor of company using

Machine Learning

* Use of Sentimental Analysis to study and give correct assumption.
* Use of TF-IDF
* improve stock price prediction

# CHAPTER 2

**LITERATURE SURVEY**

One of the integral parts to maintain the consistency is the literature survey. It’s the crucial steps to be followed in the development process. The Software Development needs authenticity of the resources and the availability of the same. This part helps in discovering the content that been worked on and find the utilization and the implementation of the same in today’s time. The key factor to the development is the economy and the strength of the product. Once the innovation of the same undergoes through the building phase the support and the resource flow is to be monitored and computed. This is also known as the Research phase where all the research is embedded and done to carry the flow.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sr. No | Paper Name | Author Name | Algorithm Used | Accuracy Achieve |
| 1 | Analyzing Stock Market Trend Prediction using Machine & Deep Learning Models 978-1-7281-9437-  0/20/$31.00 ©2020  IEEE.DOI: 10.1109/CITISIA50690.2020.93  71852 | Doan Yen Nhi Le1, Angelika Maag2 | SVM, LSTM,  RNN, DNN, LSTM, | 0.756 |
| 2 | Stock Market Prediction using Ensemble of Deep Neural Networks. 978-1-7281-6946- 0/20/$31.00 ©2020IEEE | Lu Sin Chong, Kian Ming Lim, Chin Poo Lee | CNN, LSTM, RNN, ANN | 59% or 65% |
| 3 | Predicting Stock Market trends using machine learning and deep  learning.DOI:10.1109/ACCE SS.2020.2012966, IEEE | Mojtaba Nabipour Pooyan Nayyeri  , Hamed Jabani | SVM, KNN, ANN, LSTM | 56 or 57% |
| 4 | Prediction of TCS Stock Prices Using Deep Learning Models. 978-1-6654- 0521- 8/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICACCS51430.2021.94  4185 | Srilakshmi.K Sai Sruthi.Ch | LSTM, CNN, SARIMA,RMS | 65% or 75% |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 5 | Predicting Stock Market Price: ALogical Strategy | Milon Biswas, Milon Biswas, Md. Ashraful Islam | LSTM ,  Linear Regression, XG-Boost | 60% or 70% |
| 6 | [Analysis of Investor](https://ieeexplore.ieee.org/document/8806504/) [Sentiment and](https://ieeexplore.ieee.org/document/8806504/) [Stock Market](https://ieeexplore.ieee.org/document/8806504/) [Volatility Trend Based on](https://ieeexplore.ieee.org/document/8806504/) [Big Data Strategy](https://ieeexplore.ieee.org/document/8806504/)[2020](https://ieeexplore.ieee.org/xpl/conhome/8790374/proceeding) [International](https://ieeexplore.ieee.org/xpl/conhome/8790374/proceeding) [Conference on](https://ieeexplore.ieee.org/xpl/conhome/8790374/proceeding) [Robots & Intelligent](https://ieeexplore.ieee.org/xpl/conhome/8790374/proceeding) [System](https://ieeexplore.ieee.org/xpl/conhome/8790374/proceeding)  Year:2020 IEEE | Du Peng. | iVIX, BigData | 70% or 85% |
| 7 | [Sentiment Analysis for](https://ieeexplore.ieee.org/document/9175549/) [Stock Price](https://ieeexplore.ieee.org/document/9175549/) [Prediction](https://ieeexplore.ieee.org/document/9175549/) [IEEE](https://ieeexplore.ieee.org/xpl/conhome/9166909/proceeding) [Conference on](https://ieeexplore.ieee.org/xpl/conhome/9166909/proceeding) [Multimedia](https://ieeexplore.ieee.org/xpl/conhome/9166909/proceeding) [Information Processing and](https://ieeexplore.ieee.org/xpl/conhome/9166909/proceeding) [Retrieval](https://ieeexplore.ieee.org/xpl/conhome/9166909/proceeding) Year: 2020 | Conference Paper  | Publisher: IEEE | [Rubi](https://ieeexplore.ieee.org/document/9175549/) Gupta, MinChen | Stock Twits | 65 or 70% |
| 8 | Analysis of machine learning algorithm for stock market prediction. 978-1- 7281-7213-2/20/$31.00 c  2020 IEEE | Sahil vazirani, Abhishek sharma and Pavika sharma | KNN, SVM, MAE, MSE,  RMSE | 85% or 90% |

# EXISTING SYSTEM

As many have invested their time and effort in this world trade for getting it closer and more reliable to the people for carrying out the resources and make their lifestyle more deliberate than the previous. In the past few years’ various strategies and the plans had been derived and deployed ever since it’s continuation and the topic are still a point of research where people are coming up with ideas to solve.

Intelligence fascinates mankind and having one in machine and integrating on the same is the hot key of research. There are various people contributing on the same research. A Shweta tried its invention on two nonlinear process and had come up with TS which is used as a model for fuzzy sets.

All the learning system from the past are limited and are simplest in nature where learning of the simple algorithm for a computational mean is not enough which can even be done by human brain itself. The main motto of learning was limited and learning model was not efficient.

The existing models can’t cope up with the vulnerabilities and remove the rarest information

that they can’t process causing it a major data loss which creates a problem in forecasting. Observation is the integral part in the resource and prediction management. If the outcome can’t be observed its point of time estimation is compromised causing it less liable in market. Monitoring of the same is not possible in the existing system.

The existing system in stock market predictions are apparently biased because it considers an only source point for data source. Before the prediction of the data set a simple data retrieval should be generated and tested on the training data set which are more flexible and versatile in nature.

Loss of sights is a major problem in the existing system as the stock varies each day and the loss margin can be higher with respect to time. An initial instance is taken for prediction.

# PROPOSED SYSTEM

Stock is unpredicted and liberal in nature. The follow of the same is impressive and reluctant in nature. Finding the predictability and getting the nearest is the best hit goal for the same. The exact and accurate estimation of the same is never-less possible.

There are various constrains that in-fluctuate the pricing and the rate of stock. Those constrains had to be taken in consideration before jumping to the conclusion and report derivation.

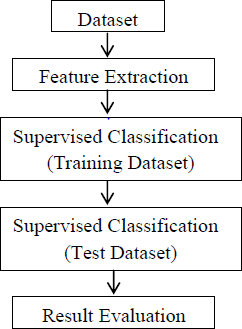


Fig. 2.2.1 : System flow

Here as described in the figure above, the proposed system will have an input fromthe dataset

which will be extracted featured wise and Classified underneath. The classification technique used is supervised and the various techniques of machine level algorithms are implemented on the same.

Training Dataset are created for training the machine and the test cases are derived and implemented to carry out the visualization and the plotting’s.

# CHAPTER 3 PROBLEM DEFINITION

# PROBLEM TO BE SLOVED

Stock 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.

Stock is unpredictable and it’s been the same from the start. It’s 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 its 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.

# PROJECT PURPOSE

Stock market prediction is a prediction system software that illuminate the risk that undergoes during the investment in stock market. It predicts the stock 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 earn and gives

the upcoming terms. Knowledge is power and same holds correct with the stock. Stock 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 physical day determines the opening stock market for the next day. It has its dependencies and is total integration with the level of finances and revenue generation. The stock 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 stock too. Stock 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 a 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 stock and understand the generation and the vulnerabilities that has to be seen and predicted. The enhancement of the same is done with the resource graph which makes a user or the customer to analyses the same and take the needs and important details before dealing and consider those things for the yield that the person is willing to invest on. Forecasting of the stock prediction is done by the available data source and the prediction is done for the upcoming week. The predictability itself is a challenge and that’s the main purpose of the report.

# METHOD USED

In stock prediction using Machine learning, we mostly focus on Linear regression as an algorithm which will work on preferred languages that are python.

# LONG SHORT-TERM MEMORY

Sequence prediction problems have been around for a long time. They are considered as one of the hardest problems to solve in the data science industry. These include a wide range of problems; from predicting sales to finding patterns in stock markets’ data, from understanding movie plots to recognizing your way of speech, from language translations to predicting your

next word on your iPhone’s keyboard.

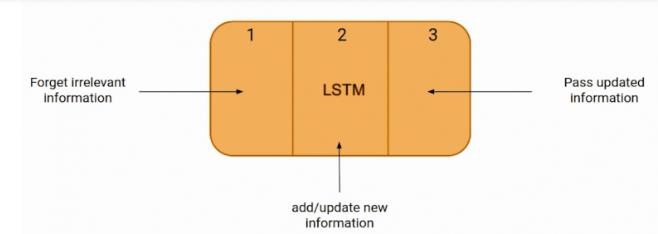


Fig 3.3.2: LSTM diagram

With the recent breakthroughs that have been happening in data science, it is found that for almost all of these sequence prediction problems, long short-Term Memory networks, LSTMs have been observed as the most effective solution.

LSTMs on the other hand, make small modifications to the information by multiplications and additions. With LSTMs, the information flows through a mechanism known as cell states. This way, LSTMs can selectively remember or forget things. The information at a particular cell state has three different dependencies. Industries use them to move products around for different processes. LSTMs use this mechanism to move information.

# TECHNOLOGIES USED

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SR.**  **NO** | **GRAPHICS USER**  **INTERFACE** | **MIDDLEWARE** | **DATASET** | **IDE** |
| 1 | HTML | PYTHON, MACHINE LEARNING | YAHOO FINANCE, NSE AND BSE | JUPYTER NOTEBOOK, |
| 2 | CSS | VS Code |
| 3 | JAVA SCRPIT |  |

* + 1. **GRAPHICS USER INTERFACE**
       1. **HTML**

Hypertext Markup Language, a standardized system for tagging text files to achieve font, color, graphic, and hyperlink effects on World Wide Web pages.

* + - 1. **CSS**

CSS stands for Cascading Style Sheets. CSS describes how HTML elements are to be displayed on screen, paper, or in other media.CSS saves a lot of work. It can control the layout of multiple web pages all at once. External stylesheets are stored in CSS files.

**3.4.1.2 JAVA SCRIPT**

An object-oriented computer programming language commonly used to create interactive effects within web browsers.

* + 1. **MIDDLEWARE**
       1. **PYTHON**

Python is a high-level, general-purpose and a very popular programming language. Python programming language (latest Python 3) is being used in web development, Machine Learning applications, along with all cutting-edge technology in Software Industry. Python is currently the most widely used multi-purpose, high-level programming language.

Python allows programming in Object-Oriented and Procedural paradigms. Python programs generally are smaller than other programming languages like Java. Programmers have to type relatively less and indentation requirement of the language, makes them readable all the time.

# MACHINE LEARNING

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that makes it more similar to humans Machine learning is actively being used today, perhaps in many more places than one would expect. Many machine learning networks trained on images have a curious phenomenon in common: in early layers of the network, a machine learning model tries to learn a low level of features, like

detecting edges, colors, variations of intensities, etc. such kind of features appears not to be specific to a particular dataset or a task because of no matter what type of image we are processing either for detecting a lion or cars. in both cases, we have to detect these low-level features. all these features occur regardless of the exact cost function or image dataset. thus, learning these features in one task of detecting lion can be used in other tasks like detecting humans.

* + 1. **DATASETS**

# YAHOO FINANCE

Finance is a media property that is part of the Yahoo! network. It provides financial news, data and commentary including stock quotes, press releases, financial reports, and original content. It also offers some online tools for personal finance management.

Yahoo Finance search - Find ticker symbol quotes, company names, and market info in the search designed to return financial results.

Interactive charts - Once you find a quote, you can graph its performance over varying time ranges, indicators, and parameters.

* + - 1. **NSE AND BSE**

India’s two main stock exchanges are the Bombay Stock Exchange (BSE) and National Stock Exchange (NSE). These two are among Asia’s largest stock exchange surpassed only by the stock exchanges of Japan and China

The Bombay Stock Exchange is one of Asia’s oldest stock exchange, beginning operations on [July 9,](https://byjus.com/free-ias-prep/this-day-in-history-jul09/) [1875,](https://byjus.com/free-ias-prep/this-day-in-history-jul09/) as “The Native Share & Stock Brokers Association”.

The National Stock Exchange is India’s biggest stock exchange in terms of market capitalization. Its operation beginning in 1992, it was the first exchange to bring in fully automated trading to India.

# CHAPTER 4 REQUIREMENT ANALYSIS

* 1. **FUNCTIONAL REQUIREMENTS**

Functional requirements deal 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 it undergoes various iteration and produce output.

# NON-FUNCTIONAL REQUIREMENTS

Nonfunctional 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 is carried out for its performance. Stock 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 goes where the technicians can work on and solve a bug or a draft if any.

The non-functional requirements followed are its 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.

# HARDWARE REQUIREMENTS

Processor: Intel i3 or above

RAM: Minimum 225MB or more.

Hard Disk: Minimum 2 GB of space Input Device: Keyboard

Output Device: Screens of Monitor or a Laptop.

* 1. **SOFTWARE REQUIREMENTS**

|  |  |
| --- | --- |
| **Requirement** | **Description** |
| 1. Operating System | Windows and Linux |
| 2. IDE | Visual Studio, Jupiter notebook, anaconda navigator |
| 3. Data Set | CSV File |
| 4. Visualization library | Math plot lib, pandas |
| 5. Server | Webserver with HTTP Process, Radis server |

# SOFTWARE DESCRIPTION

* + 1. **NSE TOOLS**

The NSE tools library collects real-time data from India's National Stock Exchange. It can be utilized in a variety of projects that require live quotes for a specific stock or index, as well as the creation of massive data sets for additional data analytics. We may also create command line interface applications that deliver real-time market data at considerably higher speeds than any browser.

Main features of NSE Tools:

* Getting live quotes for stocks using stock codes.
* Return data in both json and python dict and list formats.
* Getting quotes for all the indices traded in NSE, e.g., CNX NIFTY, BANKNIFTY
* Getting list of top losers.
* Getting list of top gainers.
* Helper APIs to check whether a given stock code or index code is correct.
* Getting list of all indices and stocks.
* Cent: percent unit test coverage.
  1. **DESIGN GOALS**

# CHAPTER 5

**HIGH LEVEL DESIGN**

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 things in our project. The right dataset must be provided to get robust results. Our data mainly consists of previous year or weeks stock prices. We will be taking and analyzing data from Cagle. 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 Cagle 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 BSE and NSE. But this data set is in raw format. The data set is a collection of valuation of stock 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.

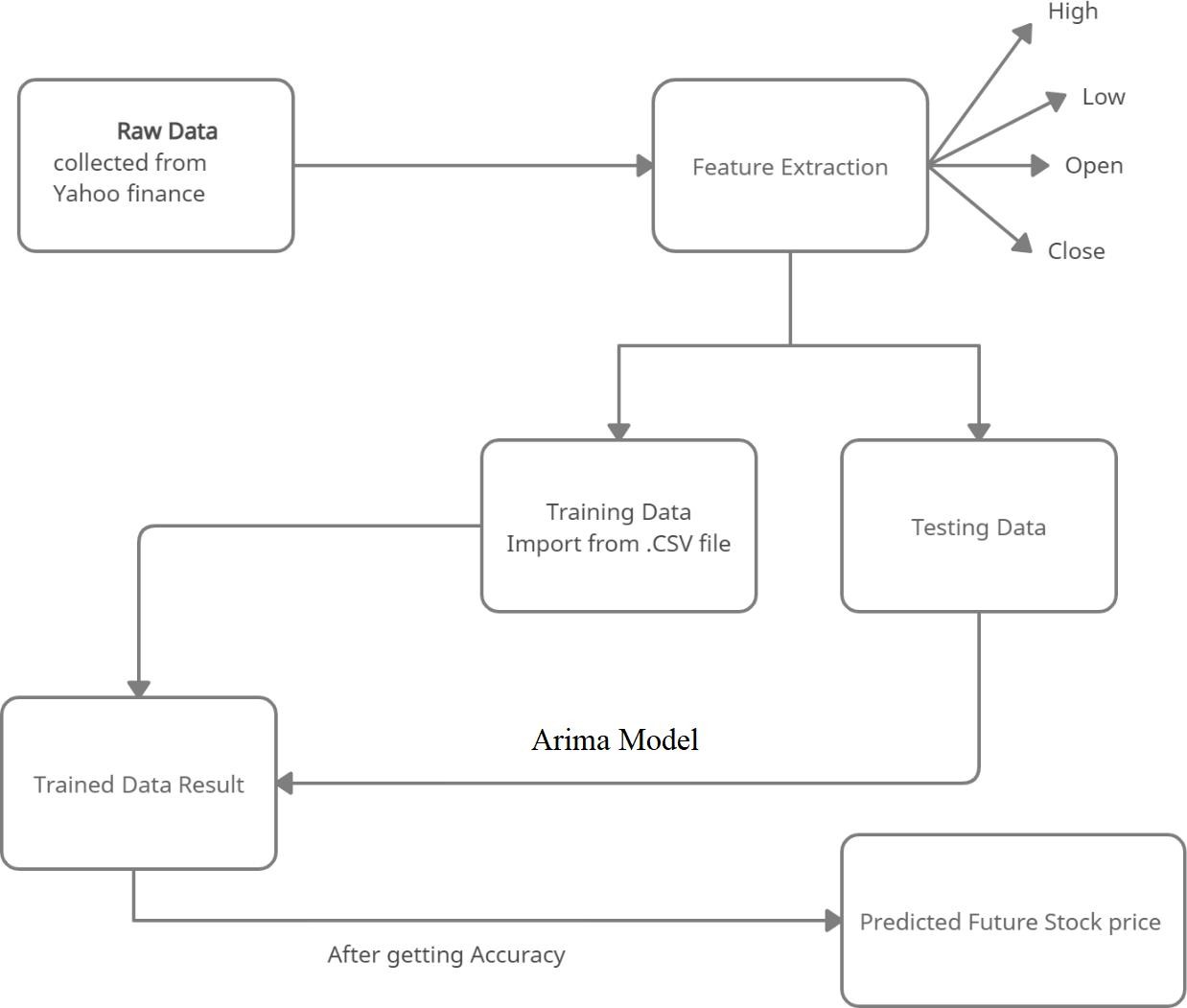


Fig 5.2: System Architecture

The above figure 4.1 gives the demonstration on the dataset extraction and data from dataset a well modified categorization is extracted.

Architecture model defines the way system works, in this architecture the system reads the data in first section where the data of past five years was implemented and the preprocessing and feature extraction is done with the parameters like high, low, open, close

are given. After words we trained the data and tested the data by using the Arima model and predicted the future stock prices and also additional feature is added to it, i.e., it displays the user whether user need to buy the stock or not and vice-versa for the cells.

# 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.

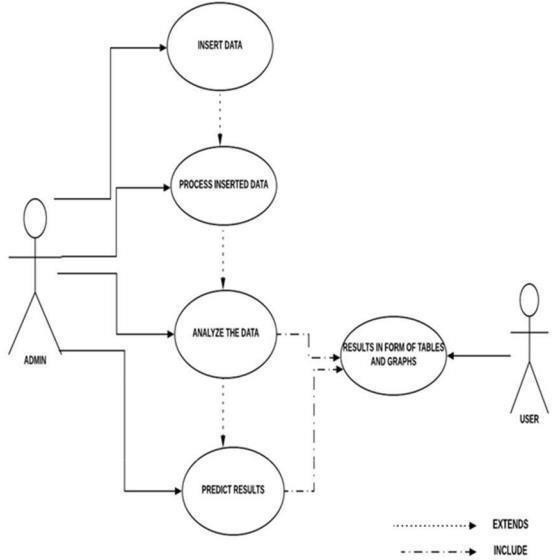


Fig 5.3: Use Case Diagram

The above figure 4.2 shows the use-case diagram of the entitled project and its 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.

# DATA FLOW DIAGRAM

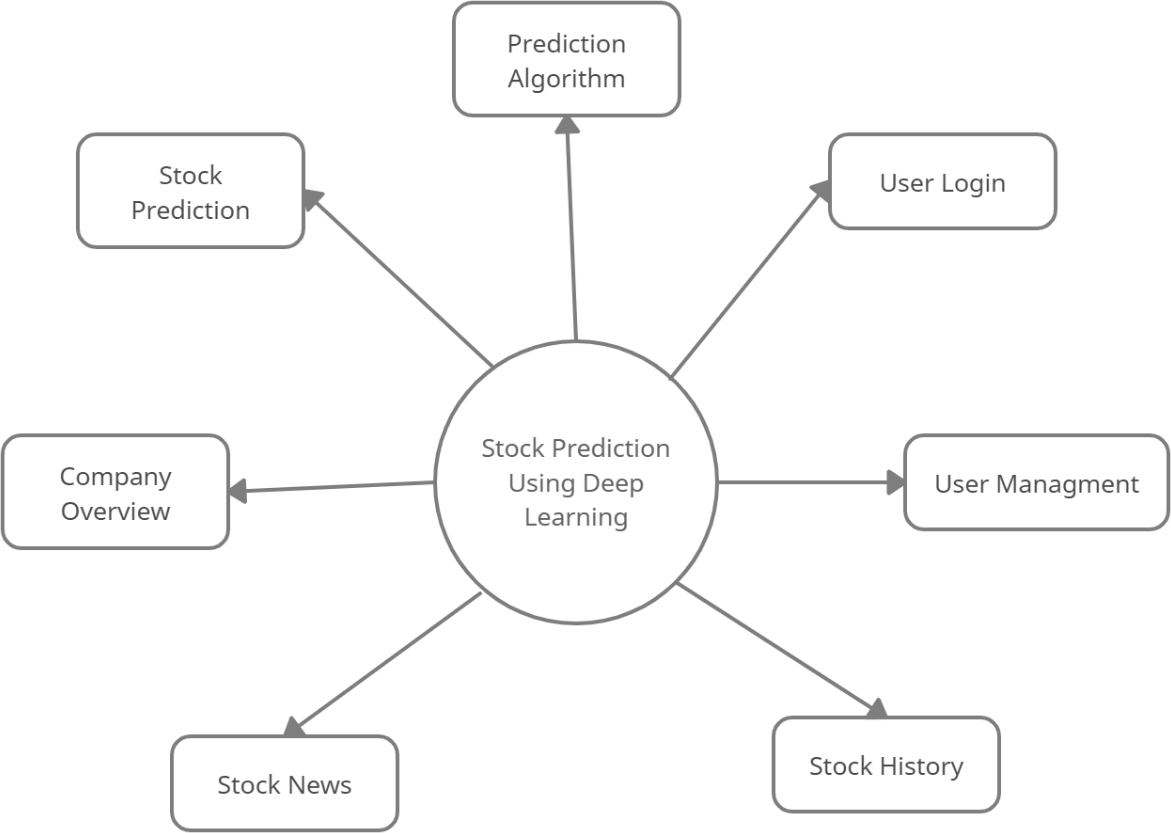


Fig 5.4: Data Flow Level 0 Diagram

In the above fig 5.4 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 stock for certain number of days. In this way data is flowing in our system.

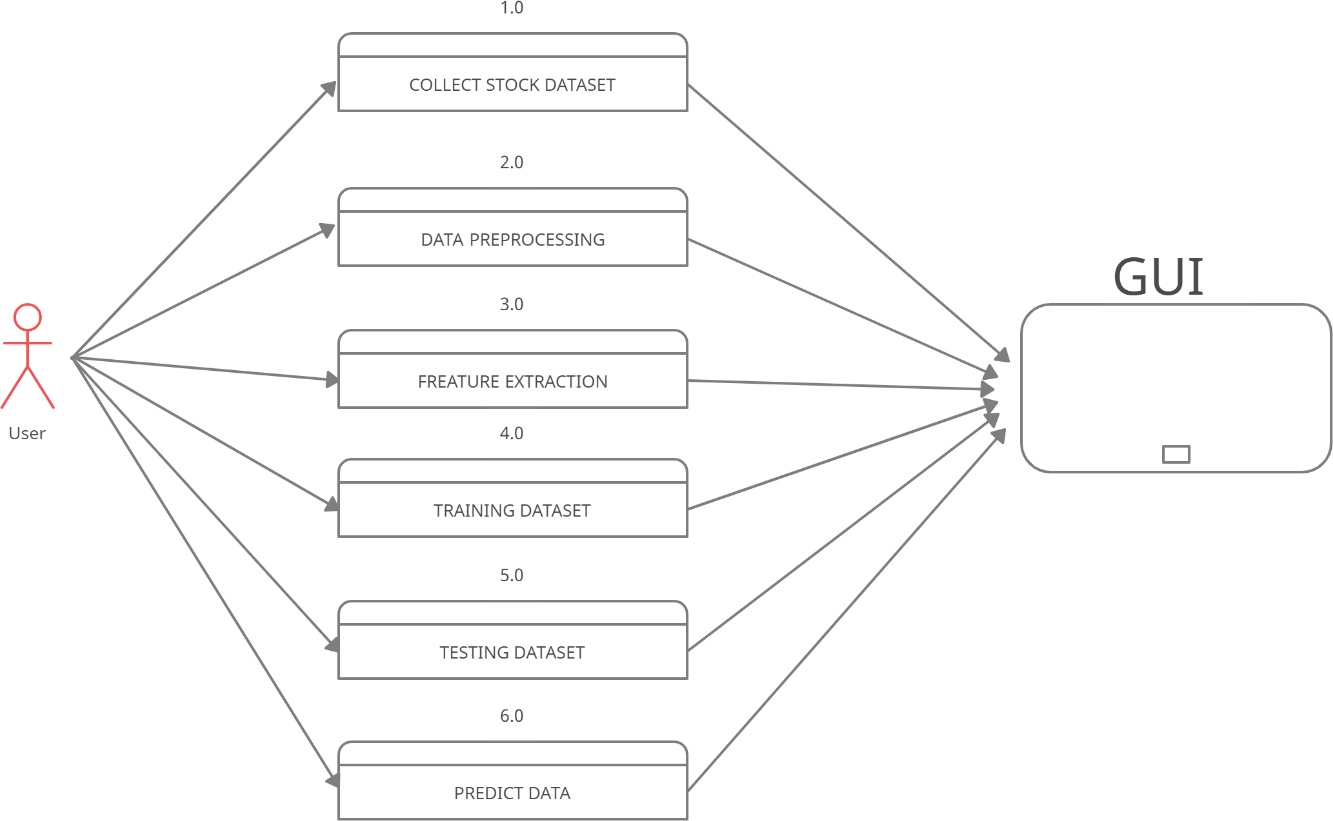


Fig 5.5: Data Flow Level 1 Diagram

# CHAPTER 6 IMPLEMENTATION

These are the Machine Learning Algorithms implemented during the building of the

project.

# LINEAR REGRESSION

One of the well-known algorithms used in machine learning is the linear regression. It is covered under both statistical as well as in machine learning. It is used for analyzing the dependency between two variables one is known dependency which value is known and the other is unknown. The value of the unknown dependency is checked with the known dependencies and the result is found as derived on its basis.

The dependency of the variable chances and are categorized into two types. Positive Linear Regression is the regression flow when both the dependencies show the growth rate and both are totally depended and supportive with the changes flow. Negative Regression is the regression flow where one dependency cancels the growth of the other. If one dependency shows the tendency to grow whereas the other one is decreasing, then this graph flow comes in picture.

They are Single Linear Regression (SLR), it’s the fundamental block of linear regression. It assumes that the two dependencies are linearly aligned and changing the values on the same will affect the other equally.

Multi Linear Regression is an extension of the SLR algorithm here different fundaments are considered with regards to the dependencies. It even deals with residual errors.

# POLYNOMIAL REGRESSION

It is a form of Non-Linear Regression. In this form of regression, the two constraints one having known dependencies whereas the other part is unknown and is generalized with the help of nth polynomial value. Research is a wide level of scope that one’s involved. There are various curves and lines estimation which can’t be normally fixed and plotted with the limitation of linear regression if trying to do so there will be the higher error ratio which will bring down the integrity and the reliability of the system in itself.

Thus, to overcome this barrier and to represent the most of the curve in every way possible either it be a straight line, or hysteresis curves. This regression helps to analyze the

curve in every possible format and helps to reduce the redundancy and the trial points of errors alongside with optimized cost factor which is a great boost to the algorithm itself. It’s widely used for the complexity to solve and takes the particular values which are unique in nature and peculiar values that needs to be considered before to set the outcome. The natural uses of the same is found in epidemics growth and to see the growth ratio of the tissue.

Including all the values of peculiarity increases the effectiveness and the efficiency of the same so it is more reliable than a Linear Regression. It has its wide range coverage so no distortion of information. No data is loss during the processing and cleansing of the dataset.

**y = m x + b**

(Where m is the slope, and b is the y-intercept. It could just as easily be written f (x) = c0 + c1 x with c1 being the slope and c0 the y-intercept.)

A prediction model is generated from the high dependencies set that increases the expectancy and get ones closer to the proximate values. It provides the best figure of constraint dependencies with one another making it easy for the user to understand and seethe conversion of the same.

# ARIMA MODEL

When we want to predict a company’s stalk prize with an ARIMA model. We have to collect data and pre-process it and train ARIMA model, based on trends of data we will have to choose the order of differencing required for this model, based on auto co-relations and partial auto co-relations we can determine the order of regression order of moving average.

#### Working of ARIMA model:

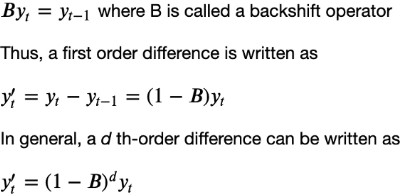
ARIMA is one of the most popular economical models and used to work with time series with the stock prices. An ARIMA model is basically ARMA model fitted on d-the order differenced time series such that the final differenced time series is stationary.

#### Auto-Regression

AR is a regression model with lagged values of Y, until P-th time in the past, as predictors. Here p= the number of lagged observations in the model, is white noise at time t, c is a constant and pi-s parameter.

* 1. Integrated I(d)- The difference is taken d times until the original series becomes stationary. A stationary time series is one whose properties do not depend on the time at which the series

is observed.



#### Moving Average MA(q)

A moving average model uses a regression-like model on past forecast errors. Here, fi is white noise at time t, c is a constant, and 0, s are parameters.

Combining all of the three types of models above gives the resulting ARIMA (p, d, i) model.

# NESPY

NSEpy is a library for extracting historical and real-time data from the National Stock Exchange's website. The API for this library is designed to be as basic as possible. Python and the Scipy stack are excellent tools for data analysis, and NSEpy's major goal is to provide analysis-ready data-series for use with the Scipy stack. NSEpy is compatible with the Technical Analysis library like MACD, RSI.

# CELERY

Task queues are a way to distribute work across multiple threads or machines. A task is a unit of work that goes into a task queue. Task queues are regularly monitored by dedicated worker processes for new work to accomplish. Celery communicates with clients and workers via messages, with a broker acting as a middleman. To begin a job, the client adds a message to the queue, which is then delivered to a worker by the broker. Multiple workers and brokers can be used in a Celery system, allowing for high availability and horizontal scaling.

# RADIS SERVER

Redis is a powerful key-value store that is free source and ideal for developing high- performance, scalable web applications.

Redis is distinguished by three primary characteristics.

1. Redis stores its database entirely in memory, with the disc serving only as a backup.
2. When compared to other key-value data stores, Redis has a large number of data types.
3. Redis is capable of replicating data to an unlimited number of slaves.

# EDA ALGORITHM

The four fundamental components of the EDA framework are selection, learning, sampling, and replacement. Each of the four components can be considered for enhancement using fascinating AMP optimization domains techniques. The Cyber-EDA in this work, which summarizes the new features for constructing each EDA component.

### EDA SELECTION COMPONENT

Diversification Generation: For constructing the initial or rebuilt populations, as well as updating the reference set, the diversification generation approach is used. The goal is to keep contrasting alternatives in the hopes of producing promising answers that would be impossible to get through greedy selection. Similarly, for the EDA selection component, we use the diversification generation method to produce different individuals from the initial or rebuilt populations. It's worth mentioning that the diversification generation method isn't used on generational populations in order to preserve evolutionary information. Diversification control can be aided by a variety of methods. Using the Taguchi approach, which is widely used in the experimental design industry, we may directly construct a varied population.

# THRESHOLDING WITH SHORT TERM MEMORY

The thresholding strategy to manage the short-term memory, which records the frequencies of the solutions selected from the present population, to lessen the sampling bias caused by the population's limited size and the sampling process. The thresholding technique applies an adaptive threshold to the highest frequency and evenly distributes the excess counts across the remaining frequency beams. As a result, the solution with the highest frequency will not dominate the probability model, and we will be able to avoid oversampling. The adaptive threshold is set based on the number of executed function evaluations, which represents the proportion of the projected evolution that has been completed.



Tt the adaptive threshold's current value is determined by adding an increment to the prior

threshold value Tt-1. This increase is proportional to the percentage of the total evolution that has been completed, where n is a constant that represents the relative scaling between the number of function evaluations and the threshold value.

# COMPARATIVE STUDY OF IMPLEMENTED MODULES

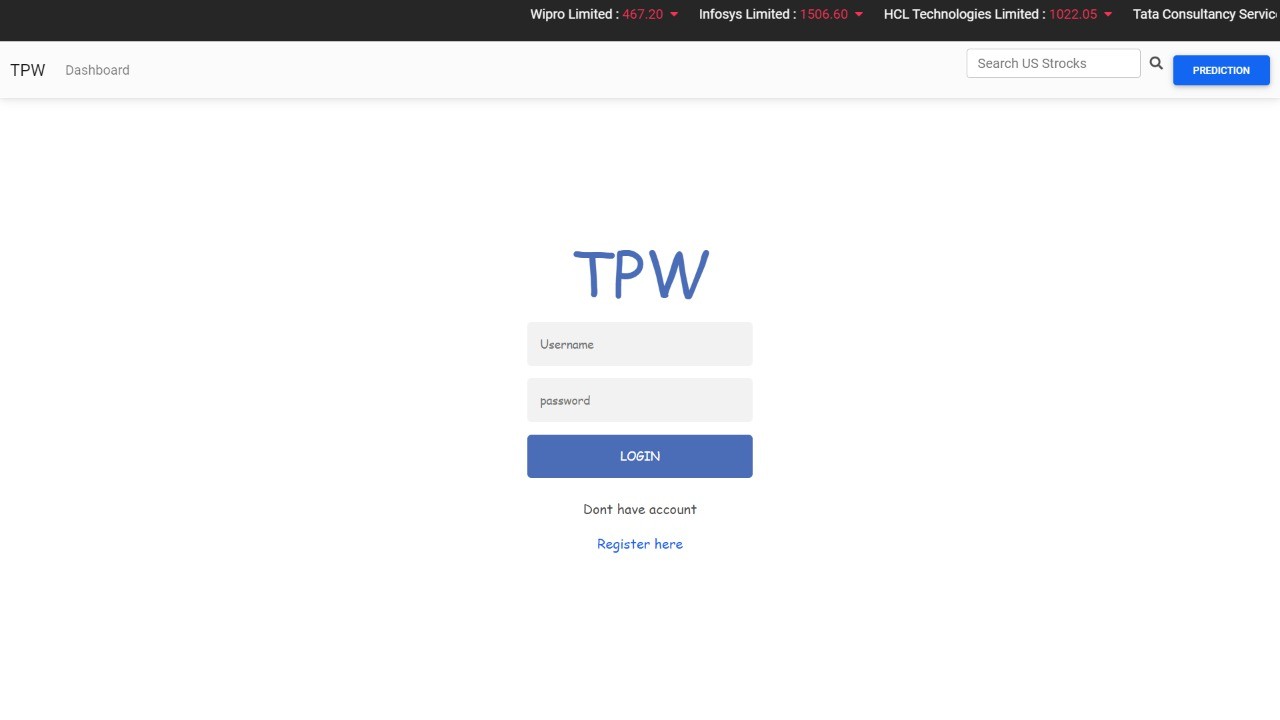
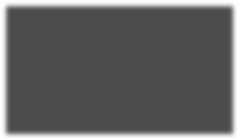
|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. no** | **Implemented**  **Algorithms and Model** | **Description** | **Accuracy Percentage (%)** |
| **1.** | Linear Regression | Performed analysis on data elements to define a single entity, i.e., result value of stock  predicted. | 79 % to 81 % |
| **2.** | Arima Model | Variable difference ordering and algorithm  implemented to increase the accuracy rates | 91 % to 93 % |
| **3.** | EDA Algorithm | Diversification is generated and thresh  holding of short-term memory is done | 89% to 92 % |

**CHAPTER 7**

**RESULT**

## Login Page

Fig. 7.1.1 Login Page



In this figure login page are there. This page is used to login existing user, if user is not available in database it won’t login. If user is valid then it allowed user to login to dashboard page. There is “Register Here” button which redirect user to registration form where user can register themselves.

## Registration Page

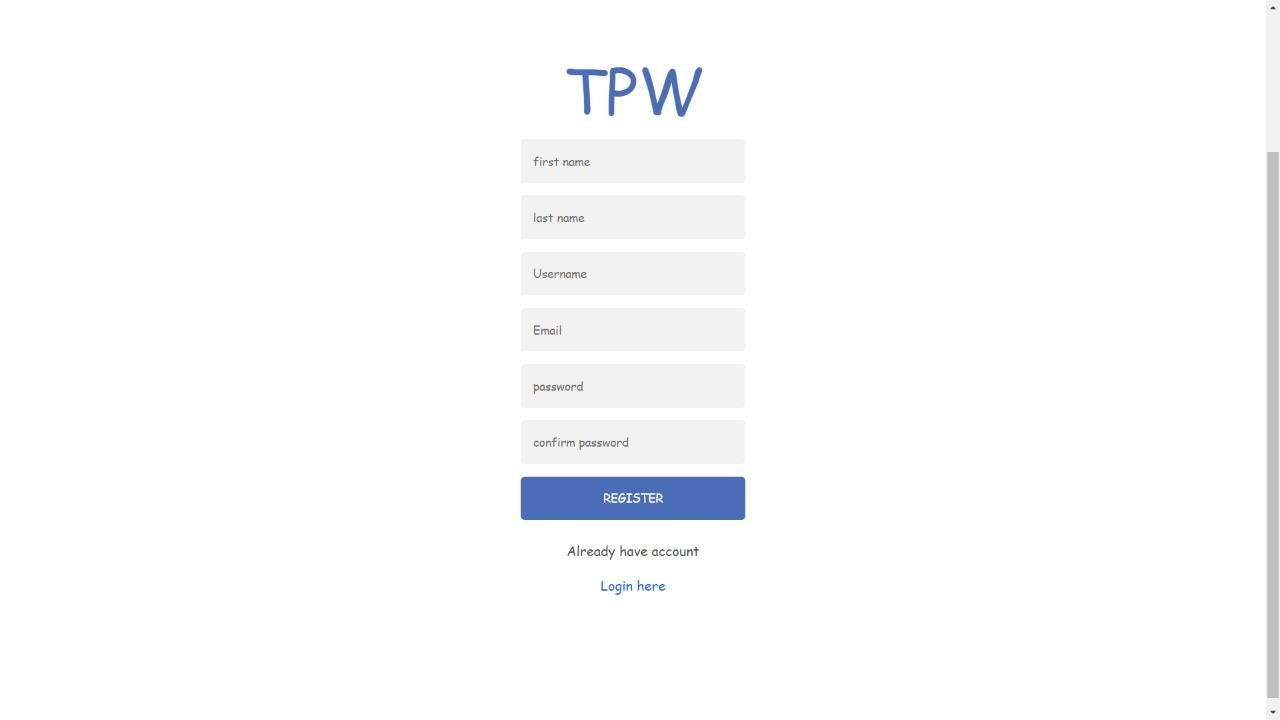
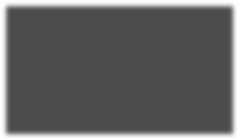


Fig. 7.1.2 Registration Page

This is Registration Page, where user need to choose valid Username, valid Email and valid Password. After entering valid information user can register in to database.

After successfully register, user will be redirected to login page, where user need to login using the credential.

## IBM stock displaying chart

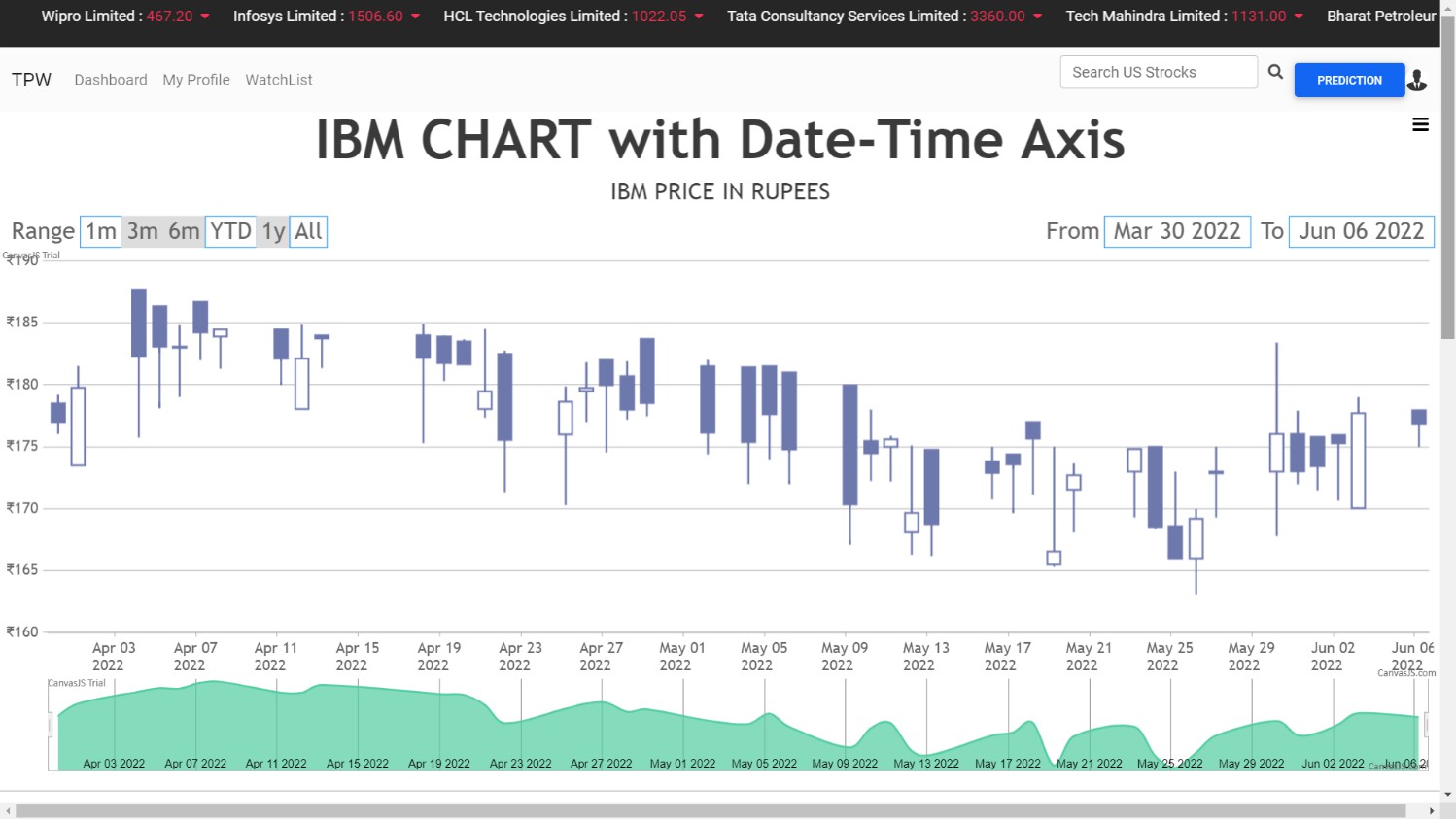
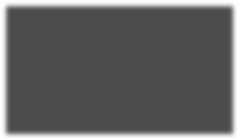


Fig 7.1.3: IBM stock displaying chart

In above figure we are displaying the chart showing fluctuations in costs of IBM shares, there are other options shown (1m, 3m, 6m, YTD, 1y, All) these options display the stocks of different timeline with candlestick and volume on that specific day. On right side user can choose specific date using “From – To” option, using month and date.

## Chart of Tata Powe Stock

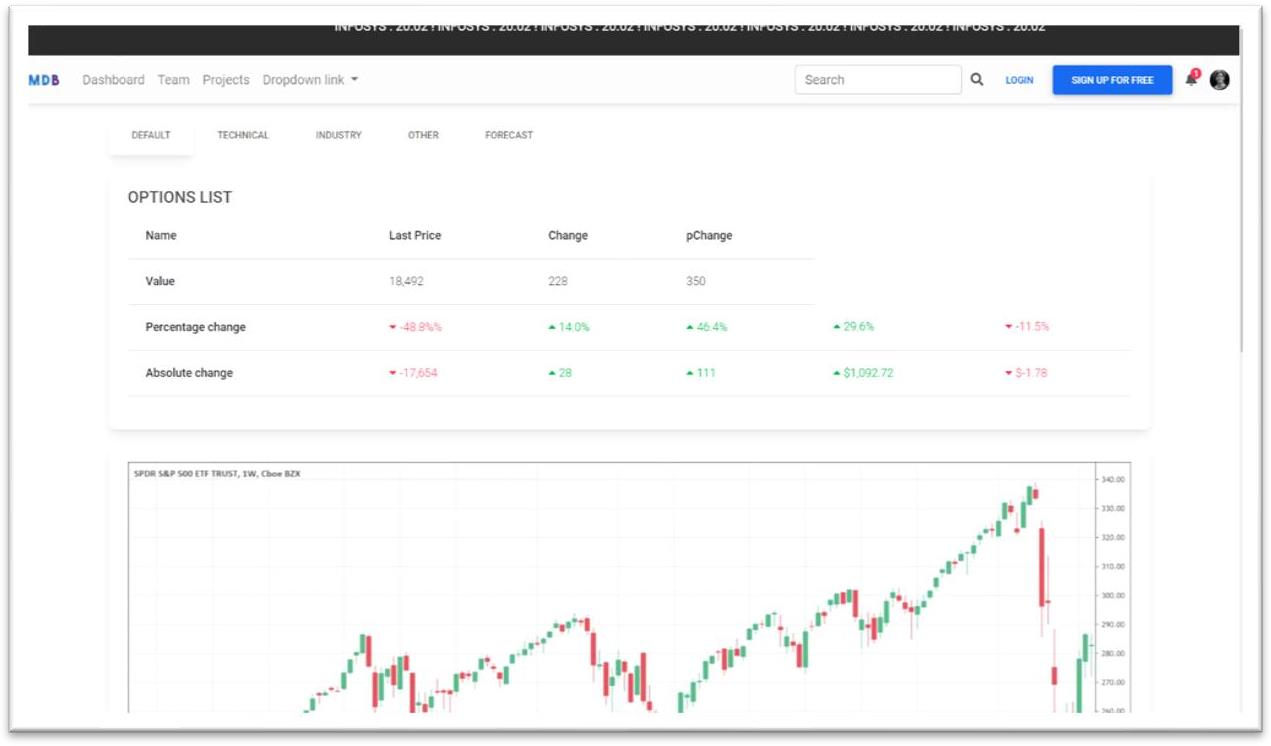


Fig 7.1.4: Tata power Chart

In above figure the information about Tata Power is showing. Per day candle with timeline and percentage change of the stock and Absolute change in stock form previous day and current value of stock.

## List of stock

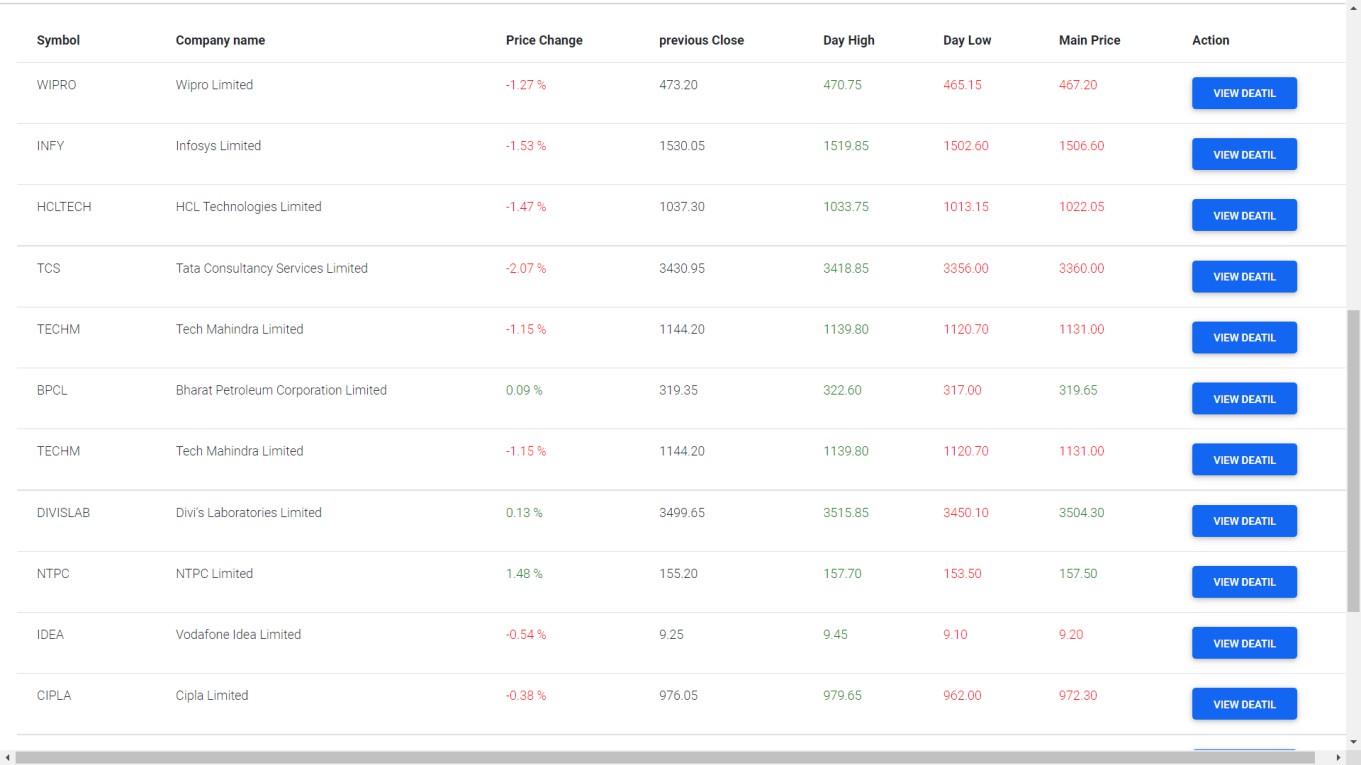
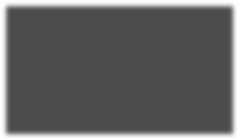


Fig 7.1.5: List of Stock

In above figure there are list of various stock where user can see details about specific stock.

It also shows price change, previous close, Day high, Day Low, Market Price.

## Accuracy of liner Regression

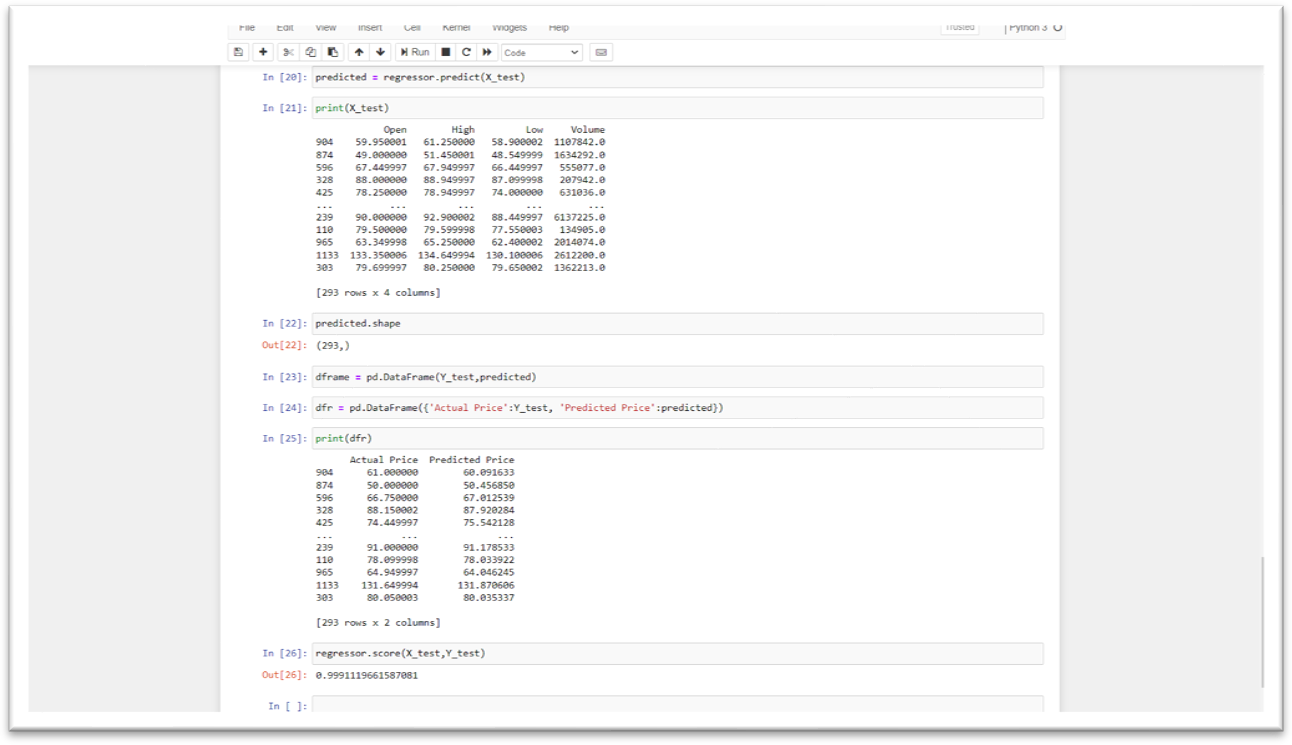


Fig 7.1.6: Accuracy of Liner Regression

This above figure the accuracy of liner regression for stock, with actual price and predicated price. It gives around 0.99 percentage.

## Prediction

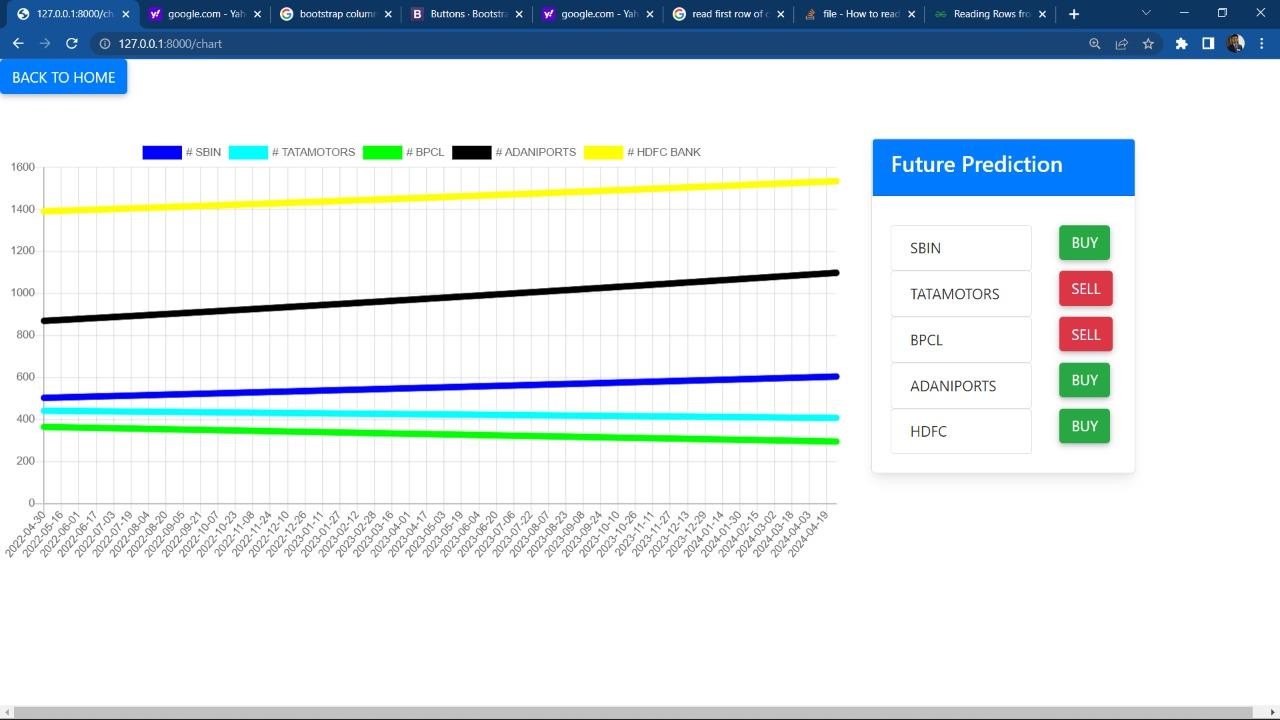
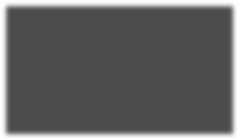


Fig 7.1.7: Prediction

In above figure the actual prediction is showing. Different stock indicates in different color line. The line shows that where stock goes in upcoming year, and according to future prediction it shows whether user need to buy it sell the stock.

# CHAPTER 8 CONCLUSION AND FUTURE WORK

# CONCLUSION

To conclude stock is an unpredictable mechanism which follows the segments of chain and the dependencies of the same are unpredictable. It is defined to be a curve which keeps on changing and turning the price from low to high and vice-versa.

As the integration of the same is higher with other dependencies so leaving one dependency compromises the level of accuracy. Accuracy is not the term used over in stock as the actual prediction is not possible for any fiscal days it keeps on changing and turning the tables day and night. Having higher component assets and the dependencies makes it more feasible and flexible in nature causing it even harder to predict. The approx. value is taken into consideration and the hit or profit or the gain rate is calculated for the same.

In the project various high level machine learning algorithms are implemented and integrated and the output is generated from the same making a user visible with the outputs in the form of graph which makes it easier for them to see and interpret what’s the scenario and they can decide on the same to invest and get the benefit out of it,

The proposed software takes the raw set of data from the dataset or the .csv file and process it. The cleaning and cleansing of data are done and then further processed to gain effective outcomes. After the computational mean the output is displayed on the screen in the form of a graph.

# FUTURE WORK

Stock markets are the best alternative for businesses to grow and it’s a side-way income for individuals who are ready to invest and earn from the same. The term stock has been in the picture ever since and it’s growing in bulk every day. There are thousands of investors investing in the same and making a fortune out of it.

There are middle-level agents and stock vendors who learn and invest in the same. The cost for the consultation on the stock is bulky and expensive. So, when it comes to people, they think a lot and invest and there’s no chance and certainty for the same to produce a yield result.

So, stock is unpredictable and the tendency of its growth is higher than ever. If the stock market and its prediction can be done accurately then it’s going to be a gain for both the individuals and the organization. The risk factor has to be mitigated so the efficiency of the system should be high and people can be certain about their investment in time.

The project can be further continued to gain the effectiveness of the prediction with additional implementations of the content that can involve real-time scenario and the way of executing and processing the real-time scenario. Various constraints has to be added and performance of the same can be acrylate in the future time for effective results. The expected form of the display is a graph whereas from the same the appearance and setting of the display can be integrated and a pie chart and a custom graph can be further implemented the same.

# CHAPTER 9 REFERENCES

1. D. P. Gandhmal and K. Kumar, "Systematic analysis and review of stock market prediction techniqe s," Computer Science Review, vol. 34, 2019, doi: 10.1016/j.cosrev.2019.08.001.
2. Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index ovemet using Trend Deterministic Data Preparation and machine learning techniques. Expert Systems with Applications, 42(1), 259-268.

[3]E.Cambria, D. Olsher, and D. Rajagopal, ³Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis, ́ AAAI, pp. 1515±1521, 2014.

1. M. Ballings, D. V. D. Poel, N. Hespeels, and R. Gryp, “Evaluating multiple classifiers for stock price direction prediction,” Expert Systems with Applications, vol. 42, no. 20, pp. 7046–7056 2015.
2. L. Nassar, I. E. Okwuchi, M. Saad, F. Karray and K. Ponnambalam, "Deep Learning Based Approach for Fresh Produce Market Price Prediction," 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, United Kingdom, 2020, pp. 1-7, doi: 10.1109/IJCNN48605.2020.9207537.
3. V. Unadkat, P. Sayani, P. Kanani and P. Doshi, "Deep Learning for Financial Prediction," 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), Kottayam, India, 2018, pp. 1-6, doi:10.1109/ICCSDET.2018.8821178.
4. Convolutional LSTM, LSTM, Fully Connected doc- X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, “Convolutional LSTM network: A machine learning approach for precipitation nowcasting,” CoRR, vol. abs/1506.04214,2015.
5. <https://ieeexplore.ieee.org/>
6. <https://finance.yahoo.com/>