

# **STOCK MARKET ANALYSIS**

Submitted in partial fulfillment of the requirements of the degree of

## **BACHELOR OF COMPUTER ENGINEERING**

by

**Rakshit Shah (19102008)**

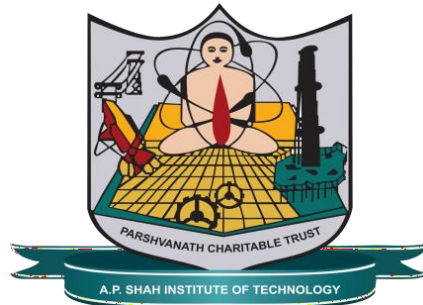
**Catherine Joshi (19102007)**

**Poojan Shah (20202001)**

**Rutuja Jain (19102044)**

Guide

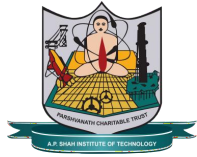
**Prof. Rushikesh Nikam**



**Department of Computer Engineering**

**A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE**

**(2021-2022)**



## A.P SHAH INSTITUTE OF TECHNOLOGY

### CERTIFICATE

This is to certify that the Mini Project 2B entitled “Stock Market Analysis” is a bonafide work of “**Rakshit Shah (19102008), Catherene Joshi (19102007), Poojan Shah (20202001), Rutuja Jain (19102044)**” submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **Bachelor of Engineering in Computer Engineering**

---

Guide

Prof. Rushikesh Nikam

---

Project Coordinator

Prof. Rushikesh Nikam

---

Head of Department

Prof. Sachin Malave



# A.P SHAH INSTITUTE OF TECHNOLOGY

## Project Report Approval

This Mini project report entitled “**Stock Market Analysis**” by “**Rakshit Shah (19102008), Catherine Joshi (19102007), Poojan Shah (20202001), Rutuja Jain (19102034)**” is approved for the degree of *Bachelor of Engineering* in *Computer Engineering*, 2021-22.

Examiner Name Signature

1. \_\_\_\_\_

2. \_\_\_\_\_

Date:

Place:

## Declaration

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

-----  
(Signature)

-----  
Rakshit Shah (19102008)

-----  
(Signature)

-----  
Catherine Joshi (19102007)

-----  
(Signature)

-----  
Rutuja Jain (19102044)

-----  
(Signature)

-----  
Poojan Shah (20202001)

Date:

## **Abstract**

As the stock market is extremely variable and indeterministic, predicting the uptrend and downtrend may be a complicated process. The stock market features a limitless number of aspects that regulate these directions and trends. The system aims to demonstrate the utilization of Machine Learning and other algorithms in finance to analyze the past performance of the Index and thus predict the longer-term possibility. Using BeautifulSoup, a Web Scraping tool, news articles would be extracted from reliable websites such as India Infoline, EquityBulls, etc. The web interface would also feature a News Section Panel. The trending news related to the stock market could aid the investor to develop trading strategies. Stock markets tend to react very quickly to a lot of factors such as the news, earnings reports, etc. With the help of TextBlob a Python Library and twitter API, Sentiment Analysis would be implemented to derive the opinion or attitude of news i.e the investor could determine whether a piece of writing is positive, negative, or neutral. To analyze the price action of a the stocks an interactive candlestick chart pattern for one year time frame is displayed with the help of Bokeh library which provides high-performance interactive charts and plots.

# CONTENTS

1 Introduction .....	9
2 Literature Survey .....	10
3 Problem Statement.....	15
4 Objective and Scope.....	16
5 Experimental Setup.....	17
5.1 Hardware Requirement	
5.2 Software Requirement	
6 System Design.....	18
6.1 UML Diagram	
6.2 Algorithm/Process (with Expected input and outputs)	
7 Implementation.....	22
7.1 Code	
7.2 Screenshots	
8 Result.....	0
9 Conclusion.....	31
References .....	0
Annexture 1- Project Planning (Using Gantt chart) .....	0

# List of Figures

6.1 Class Diagram.....	19
6.2.1 Data Flow Diagram Level 0 .....	20
6.2.2 Data Flow Diagram Level 1 .....	21
6.3 Sequence Diagram .....	22

# List of Graphs

Graph 10.1: Bar Graph Depicting the Difference Between Both the RMSE .....



# 1. Introduction

The stock market refers to public markets that exist for issuing, buying, and selling stocks that trade on a stock exchange or over-the-counter. Stocks, also known as equities, represent fractional ownership in a company, and the stock market is a place where investors can buy and sell ownership of such investable assets. An efficiently functioning stock market is considered critical to economic development, as it gives companies the ability to quickly access capital from the public.

Predicting stock market movements is a problem of huge interest. Every grown up out there wants to make more money out of the existing money. While doing so, no one would like to put themselves at risk i.e., no one would want to see their amount reducing. Utmost care and alertness are the keys to a safe stock market investment. As the stock market is extremely variable and indeterministic, predicting the uptrend and downtrend may be a complicated process. The stock market features a limitless number of aspects that regulate these directions and trends.

This project helps an investor to be alert of the stock market related news and charts from different websites by showing their gist under one roof. The public talks regarding stock markets are the most usable and reliable words for any investor to analyze the market. Sentiment analysis of the tweets on Twitter helps to get an overview of the current stock market.

## 2. Literature Survey

Stock price forecasting may be a popular and important topic in financial and academic studies. A share Market is an untidy place for predicting since there are not any significant rules to estimate or predict the worth of a share within the share market. Ashutosh S. et al [9], predict the stock prices using previous closing price and trading volumes to visualize both the anticipated price as well as values over time as well as the optimal parameters for the model. The model which we have used predicts 60 data points i.e the past two months' data based on the test data set and the last data point is pushed as the output. In order to add new information, it transforms the existing information completely by applying a function. Because of this, the entire information is modified, on the whole. LSTMs, a special kind of Recurrent Neural Nets (RNN), are capable of learning long-term dependencies. These are used specifically for learning the sequences and patterns of the time-series-based data. 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. The paper shows a comparative study of the models using both Linear Regression as well as LSTM. The mean balancing done over the processed LSTM helped us get better results and more accurate patterns over hysterical data sets. LSTM is capable of learning order dependence in sequence prediction problems. This is a behavior required in complex problem domains like machine translation, speech recognition, and more.

Yoshua B. et al [6]., defines 3 basic requirements of a recurrent neural network. This paper mentions that the system must be able to store information for an arbitrary duration, be resistant to noise i.e. fluctuations of the inputs that are random or irrelevant to predicting a correct output. and that the system parameters be trainable in reasonable time. The paper also describes the minimal task for demonstrating recurrent neural networks.

Polamuri S.R et al [5]. mentions implementing a linear statistical model (LSTM). In an LSTM Model, a linear model that is more or less ideal is primarily created and data is then added to it so that the linear model reflects the properties of the actual data. The main advantage of this linear model of the statistic is that the particular data are incorporated into the ideal linear model. All the required pre-calculations for the prediction are used as an input for the prediction function.

Predicting stock market movements is a problem of huge interest. Every grown up out there wants to make more money out of the existing money. While doing so, no one would like to put themselves at risk i.e., no one would want to see their amount reducing. Utmost care and alertness are the keys to a safe stock market investment. Before the large expansion of social media, talking with each other, newspapers and new channels on television were the only few ways available to be alert. The humongous growth in use of social media has made the platform suitable for discussing a huge range of topics. Stock market is an ever-hot topic for such discussions, especially on the Twitter platform. Positive news and tweets in social media about a company would definitely encourage people to invest in the stocks of that company and as a result the stock price of that company would increase. The author describes a method for predicting stock prices using Twitter tweets about the various companies. Sentiment analysis of the collected tweets is used for the prediction model for finding and analyzing the correlation between contents

of news articles and stock prices and then making predictions for future prices will be developed by using machine learning.

V. S. Pagolu et al [10] describes how well the changes in stock prices of a company, the rises and falls, are correlated with the public opinions being expressed in tweets about that company. According to this research paper, understanding the author's (twitter user who has posted) from a piece of text is the objective of sentiment analysis. The paper has employed two different textual representations, Word2vec and N-gram, for analyzing the public sentiments in tweets.

Sentiment classification is the task of judging opinion in a piece of text as positive, negative or neutral. Twitter is a valuable source and a powerful tool for conducting studies and making predictions. The tweets regarding Microsoft were extracted. These tweets, in this project, were collected using Twitter API and filtered using keywords like \$MSFT, #Microsoft, #Windows etc. Tweets consisted of many acronyms, emotions and unnecessary data like pictures and URLs. So the tweets were preprocessed to represent the correct emotions of the public.

The classifier with features as Word2vec representations of human annotated tweets trained on Random Forest algorithm with a split percentage of 90 for training the model and remaining for testing the model showed an accuracy of 70.5%. Total data was split into two parts, 80 percent to train the model and remaining for testing operations. The classifier results show an accuracy value of 69.01% when trained using the Logistic regression algorithm and the accuracy rate varied with the training set.

This paper has shown a strong correlation that exists between rise/fall in stock prices of a company to the public opinions or emotions about that company expressed on twitter through tweets. The main contribution of the project was the development of a sentiment

analyzer that can judge the type of sentiment present in the tweet. The tweets were classified into three types: positive, negative and neutral. The claim That sentiments of the public in twitter about a company would reflect in its stock price was well supported by the results achieved.

Vangala R. V. et al [7], suggests that the web framework flask provides the essential tools, libraries and technologies to build an internet application. The web applications are often web pages, a blog, a wiki, or go as web-based calendar applications, or billboard websites. Flask is a product of Python. It is a lightweight web application framework and baseband on the WSGI toolkit and Jinja2 template engine. Flask is a component of the Categories of the micro framework i.e these frameworks have little to no dependencies on external libraries. The framework ensures Integrated Support for Unit Testing. Flask is compatible with the Google app engine. Flask is sufficient for almost all the machine learning models.

P. Ashiwal et al [8], states that there is a huge and rapid growth in unstructured data every moment. The production and generation of data are predicted to be 44 times more as compared to the data back then. All these figures and statistical data are amazing and growing in an exponential pattern. Such data is unstructured in nature which means the data of different and heterogeneous formats. There are a number of approaches by which the live data can be obtained for research and development. One of these approaches is getting data from Open Data Portals. The open data portals provide authentic datasets for research and development in multiple domains. The data sets can be downloaded from these portals in multiple formats including XML, CSV, JSON and many others. Many times data is not easily accessible – although it does exist. As much as we wish everything was available in CSV or the format of our choice – most data is published in different forms on the web. To combine it with other datasets and explore it independently, one could opt for Screen Scraping. Screen Scraping is the technique to capture the data that is

being displayed in a human-readable format on the destination terminal and replicate it at the source terminal for further processing. Screen scraping is sometimes referred to as terminal emulation. Though there are other ways to get the data out of the web i.e., from web-based APIs, such as interfaces provided by online databases and many modern web applications including Twitter, Facebook, etc. This is a fantastic way to access government or commercial data, as well as data from social media sites. Extracting information from PDFs is beyond the scope of this paper, but there are some tools and tutorials that may help you do it. But the advantage of scraping is that you can do it with virtually any website — from weather forecasts to government spending, even if that site does not have an API for raw data access. However, screen scraping is not an independent process. Before scraping the output, Crawlers are responsible to navigate to the destination terminal. The search key entered at the source machine engages the crawlers to navigate through the links on the web. Once the crawlers successfully reach the correct page that matches up with the search string, the scraping process starts.

B. Lindemann et al [11] gives an overview of the LSTM Model used on n time series data or data sequences to predict nonlinear time-variant system outputs. Wide range of network types available, but the modeling accuracy is strongly dependent on the fit of network architecture and considered a problem. LSTM with an optimized cell state representation, such as hierarchical and attention-based LSTM, shows an improved ability to process multidimensional data whereas LSTM with interacting cell states, such as Grid and cross-modal LSTM, are able to cooperatively predict multiple quantities with high precision. The Seq2Seq LSTM Model can predict multiple quantities in terms of a multi-step ahead prediction with minimized error propagation. Moreover, partially conditioned Seq2Seq LSTM show the best suitability to model both short-term and long term dependencies

## **3.Problem Statement**

Analyzing the Stock Market News and predicting the Index using LSTM and checking the sentiment of the Market.

## **4.Objectives and Scope**

### **4.1 Objectives**

- To display the trending news related to the stock market.
- To analyze the past performance of the Index and thus predicting the future possibility.
- To show the sentiments of the market.
- To plot an interactive candlestick chart of a stock.

### **4.2 Scope**

The main idea behind the project is to analyze the major indices in the Indian markets and predict the markets accordingly using Regression. Analysis of stocks using data mining will be useful for new investors to invest in the stock market based on the various factors considered by the software. The exchange provides an efficient and transparent market for trading in equity, debt instruments and derivatives. Our web application will be analyzing sensex based on the company's stock value. Our web application also provides stock related news by scraping news from different websites using python and its libraries which can keep a user aware about the live market. A market mood index (sentiment) is displayed which can help an investor to get an overview of the greed or fear amongst the investors in the market. An interactive candlestick chart is plotted to analyze the price action of a particular stock.



# **5.Experimental Setup**

## **5.1 Hardware requirements**

- Operating Systems
- High Speed RAM
- Fast Processor

## **5.2 Software requirements**

- VSCode
- Python 3.8 and above.
- HTML, CSS, JavaScript, Bootstrap

## 6. System Design

### 6.1 Class diagram

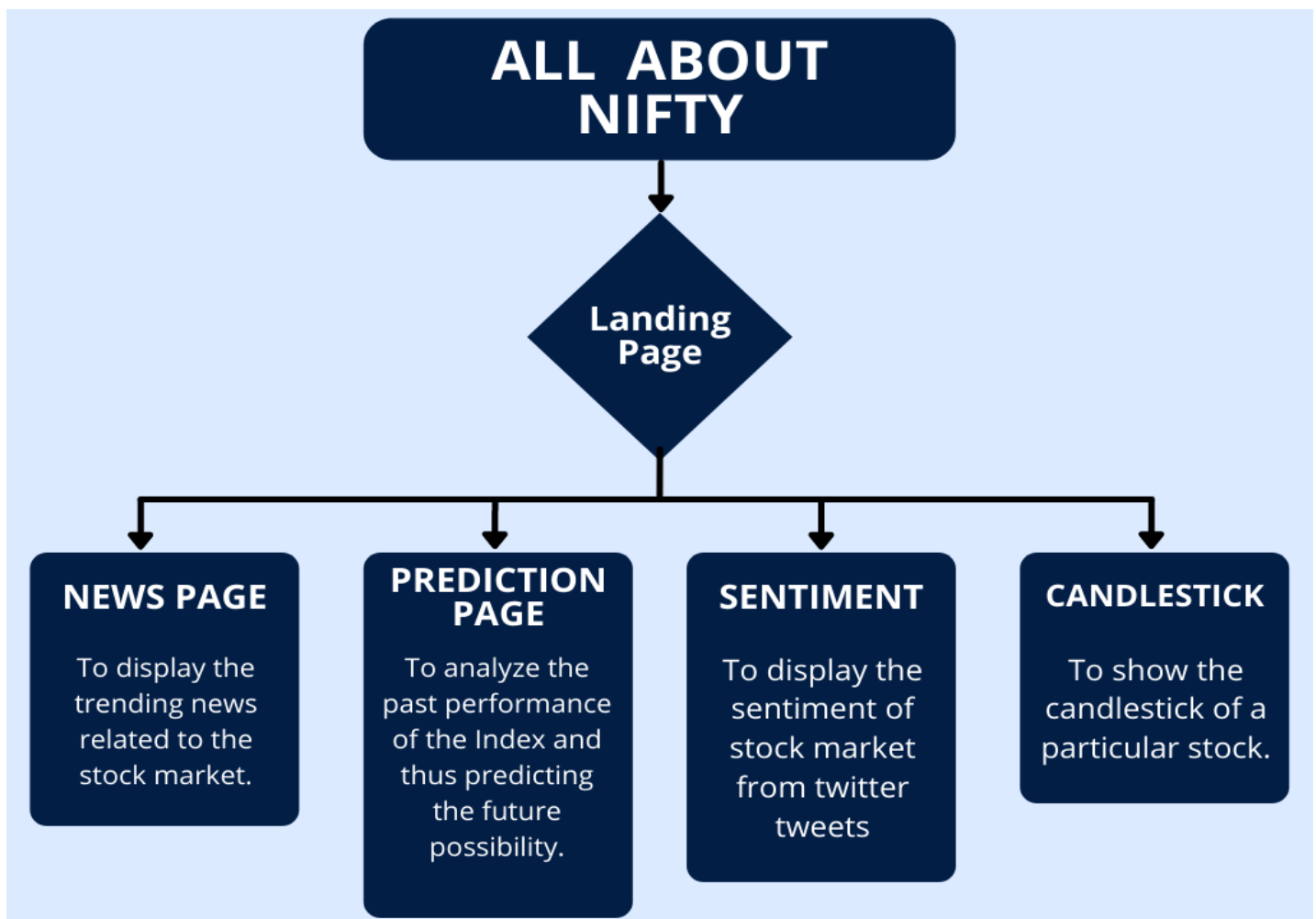


Fig 6.1: Class Diagram

## 6.2 Data Flow Diagrams

### 6.2.1 Data Flow Diagram Level 0

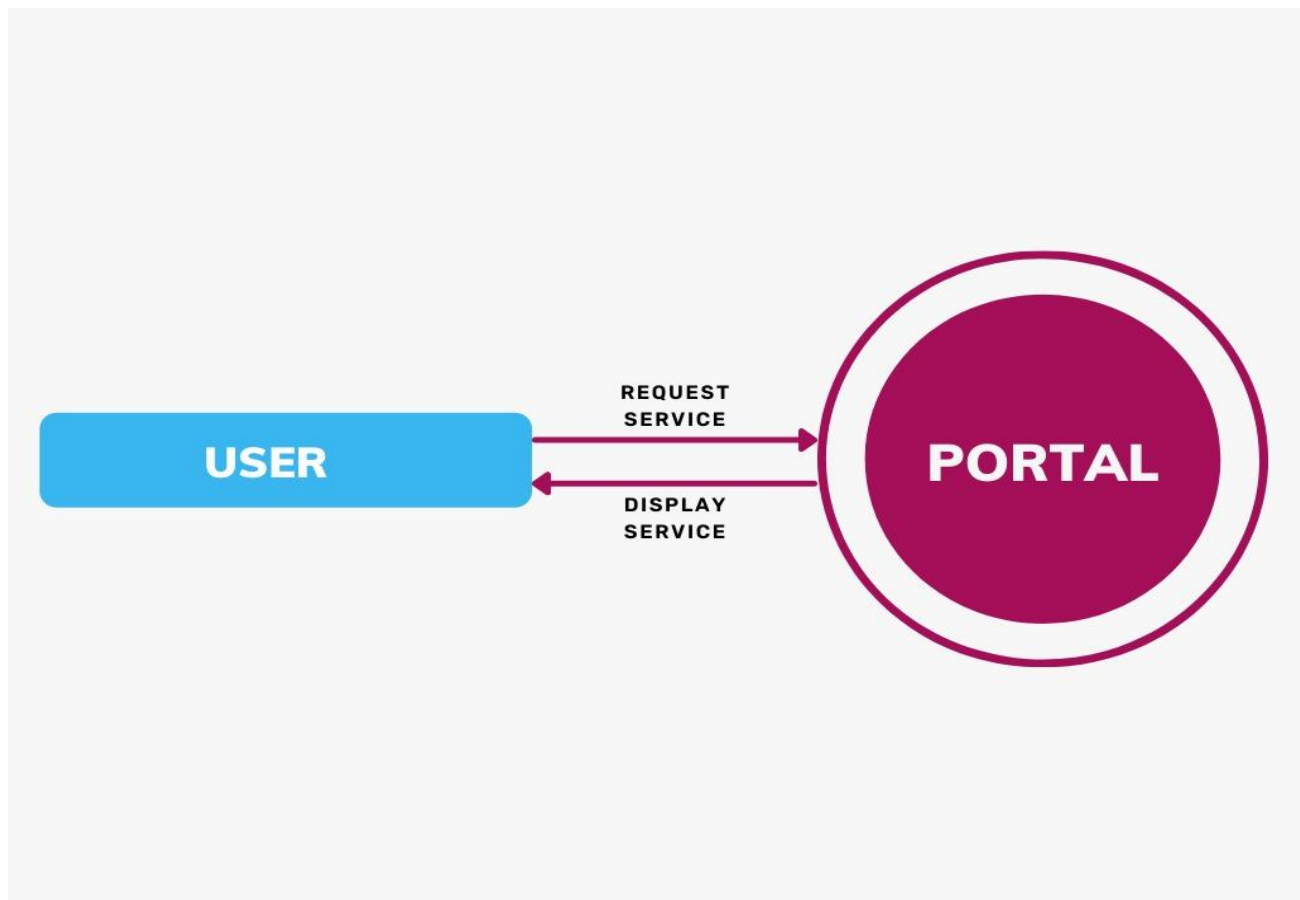


Fig 6.2.1: Data Flow Diagram Level 0

## 6.2.2 Data Flow Diagram Level 1

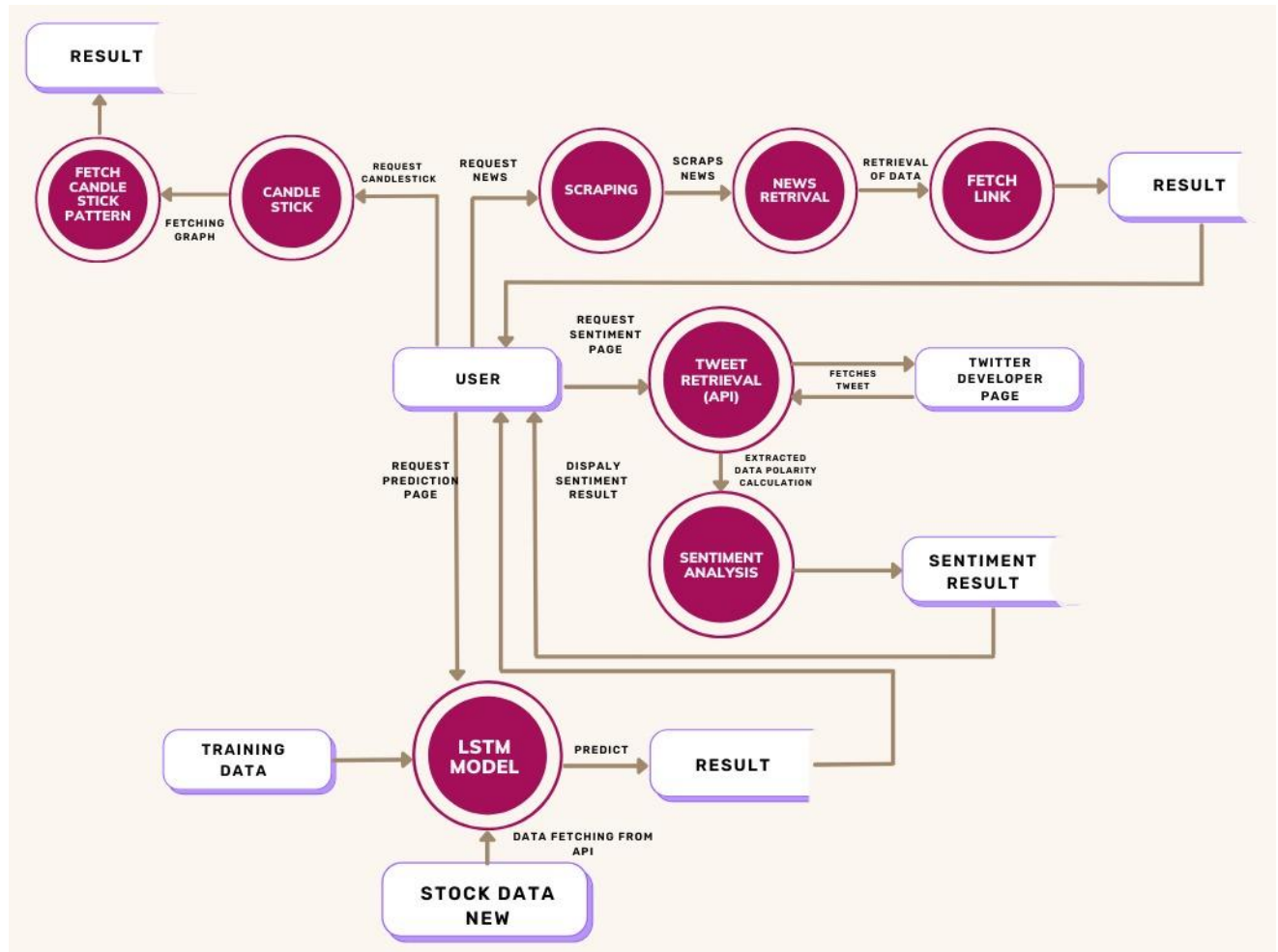


Fig 6.2.2: Data Flow Diagram Level 1

## 6.3 Sequence Diagram

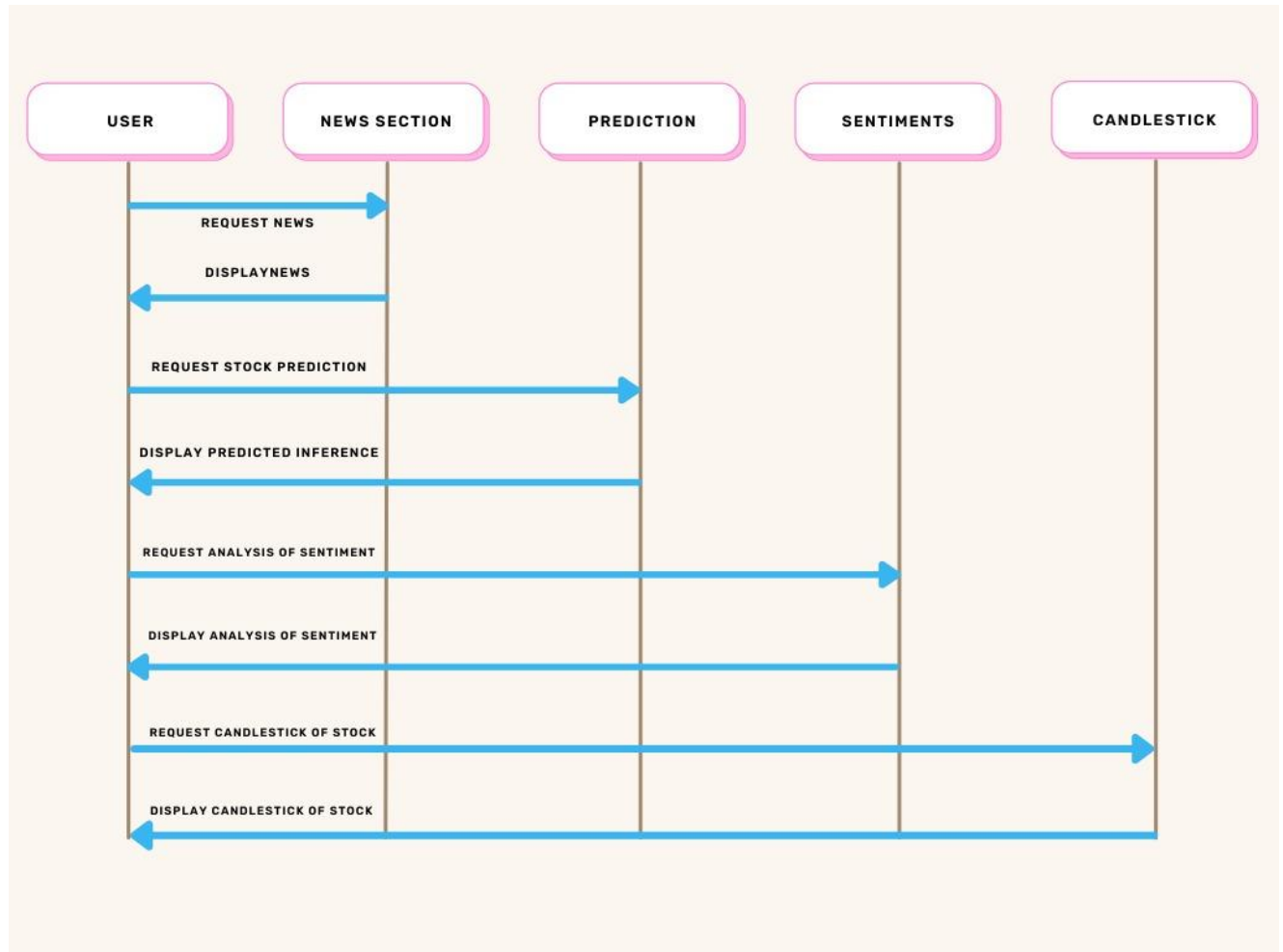


Fig 6.3: Sequence Diagram

## 7. Requirement Analysis & Methodology

Operating Systems such as Windows, Linux or MAC can be used, A minimum of 4 to 8 GB RAM is required to process the heavy models. At least 2.8Ghz CPU speed would be required. Visual Studio Code is a code editor redefined and optimized for building and debugging modern web and cloud applications. For running a project in data science or data visualization, one could use Jupyter notebooks inside VS Code. Python is an interpreted, high-level and general-purpose programming language. The entire GUI and processing in this project will be done in python. In this project, we are going to use python's web-based framework, Flask. Flask's primary goal is to ease the creation of complex websites. CSS is used to style an HTML document. CSS describes how HTML elements should be displayed. Bootstrap is a framework which is used to create user interfaces in web applications. It provides css, js and other tools that help to create required interfaces. In Flask, we can use bootstrap to create more user-friendly applications.

Web scraping is the process of extracting raw data from any website in an efficient way where large amounts of Data can be scraped within a fraction of time. Scraped data can be later used for analysis in different ways. Not every website has an Application Programming Interface(API) nor they offer data sets, so with the help of web scraping, we can get the data and save it in different formats.

BeautifulSoup is a Python library that is used for web scraping purposes to pull the data out of HTML and XML files. It creates a parse tree from page source code that can be used to extract data in a hierarchical and more readable manner.

Candlestick charts are displayed using the Boken library, Bokeh is a Python library which is used to plot interactive graphs.

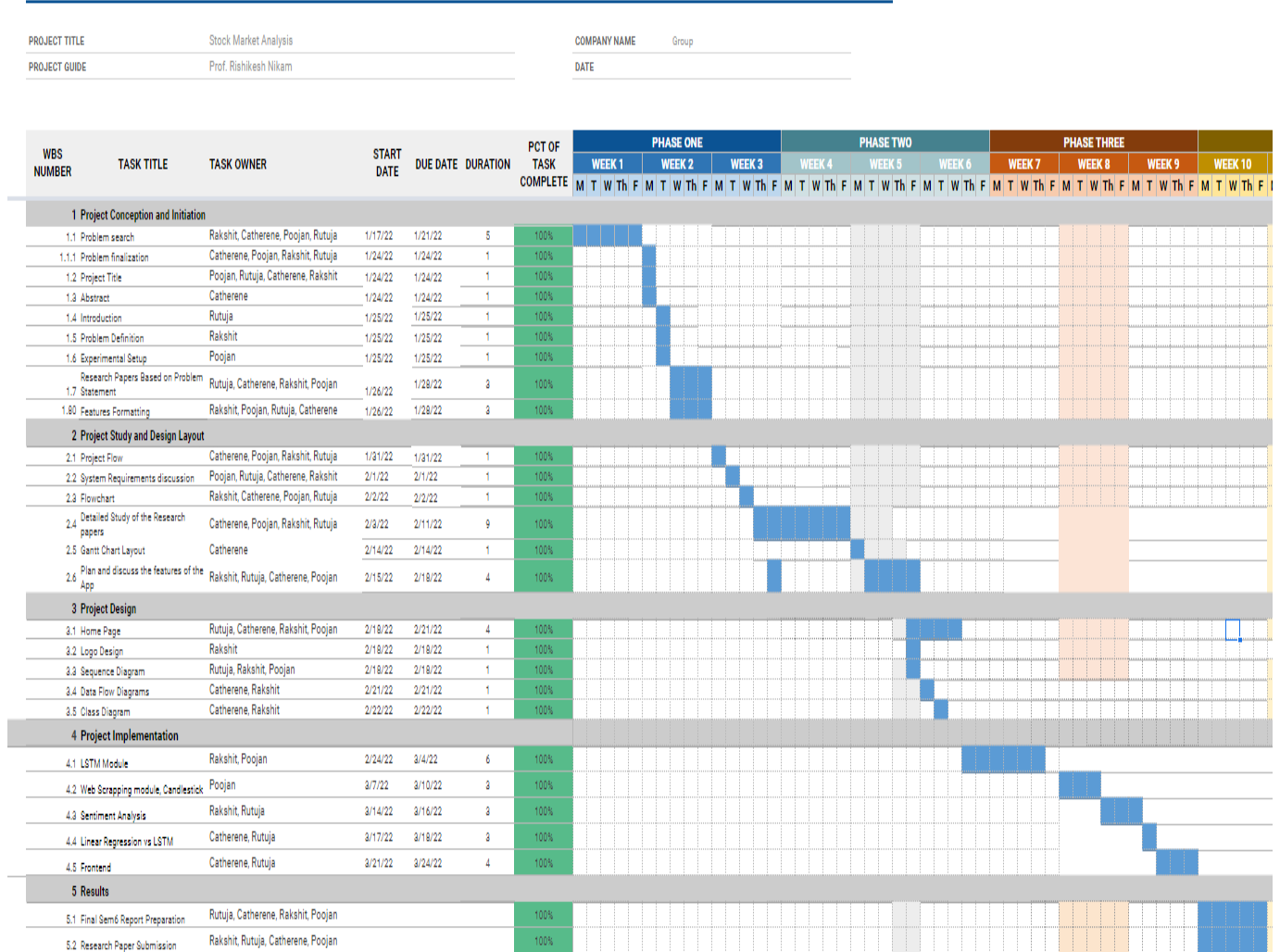
Python Libraries Required for web scraping - BeautifulSoup, Request Library.

To scrape data from a website, we need to extract the content of the webpage. Once the request is made to a website, the entire content of the webpage is available in the form of plain text.

The tweepy library is used for accessing the Twitter API. Tweepy uses API key of the users Twitter developer account and on the basis of request response mechanism, it fetches recent tweets for related hashtags and then sends them to the textblob field. Using textblob, a python library, the polarity of a text is calculated. If the polarity of text is 0 then consolidation value is incremented by 1. If the polarity of text is greater than 0 then bullish value is incremented by 1. If the polarity of text is less than 0 then bearish value is incremented by 1. Percentages of all- consolidation, bullish and bearish values are calculated. Using the python library plotly, bullish, bearish and consolidation percentages are plotted in separate images. These graphs are then shown on the website using Flask whenever the user requests for the sentiment analysis page.

# 8. Project Planning and Scheduling

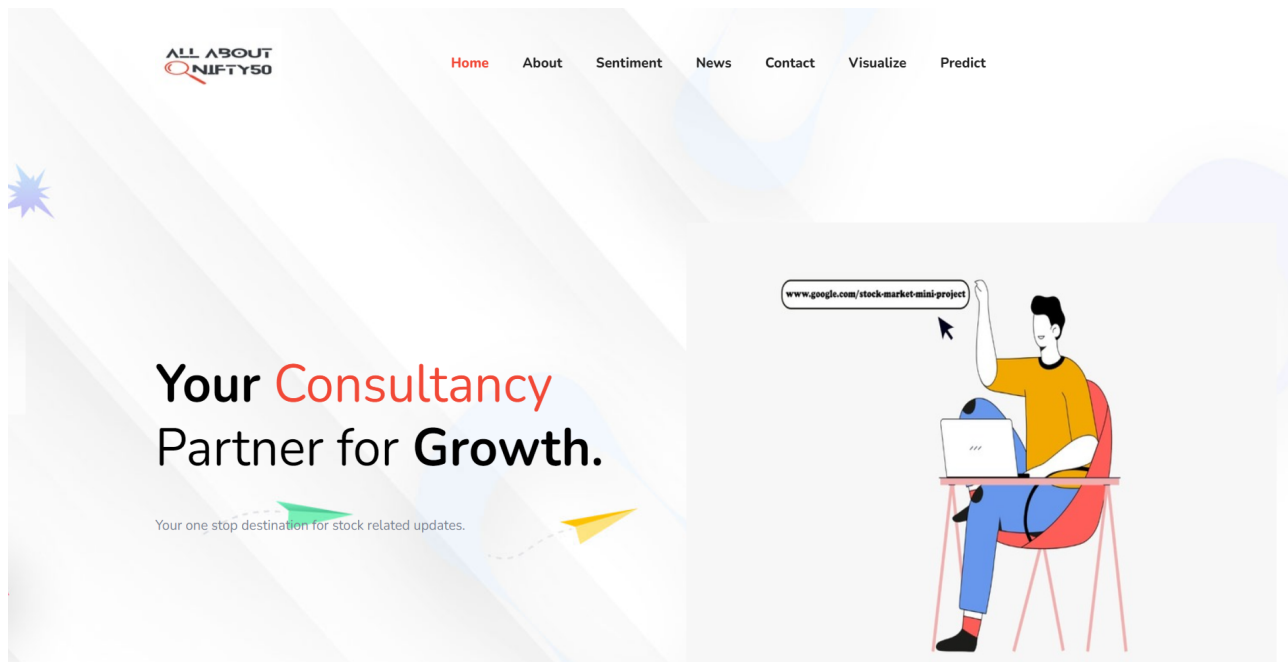
## GANTT CHART

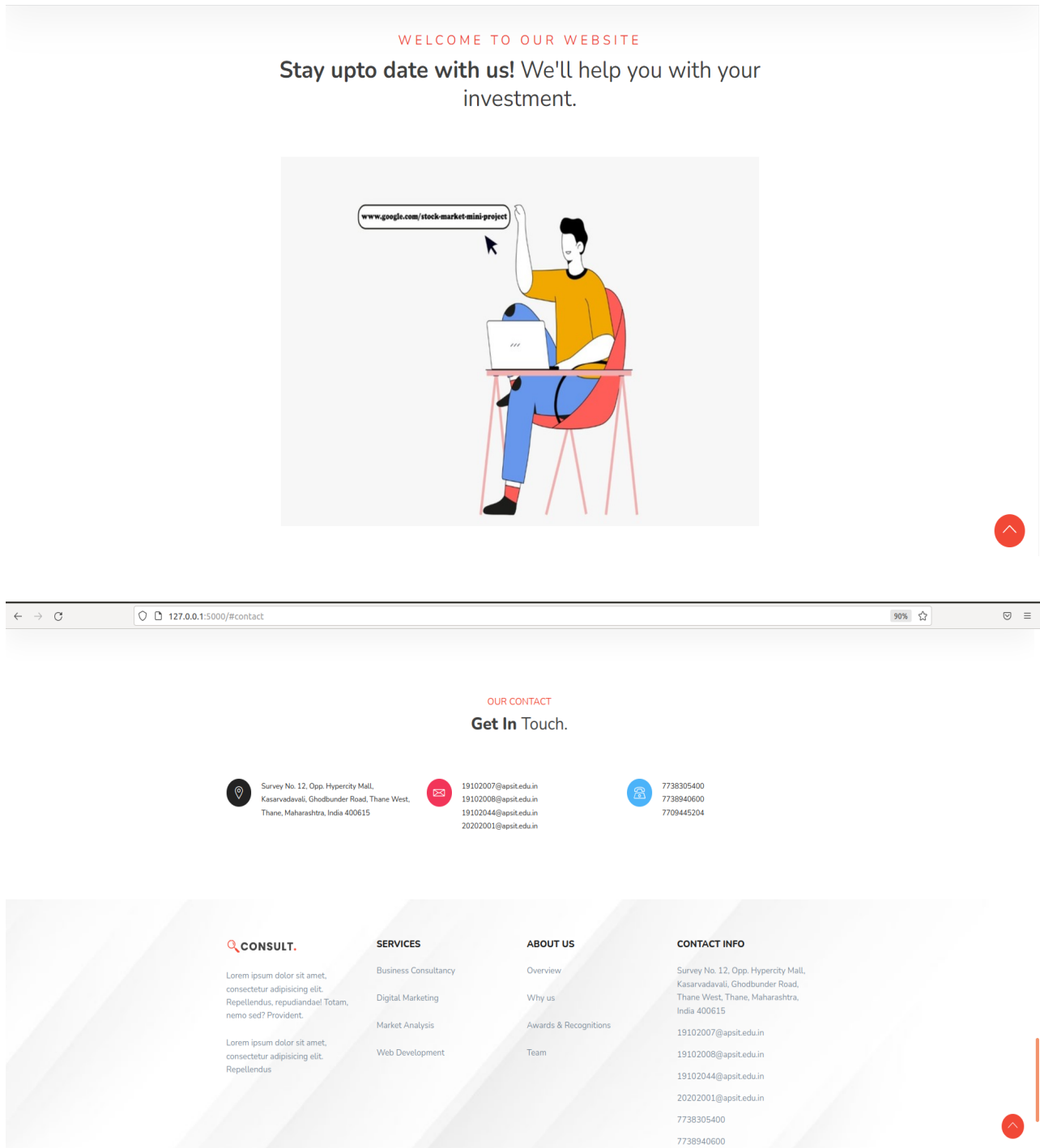




# 9. System Design and Implementation

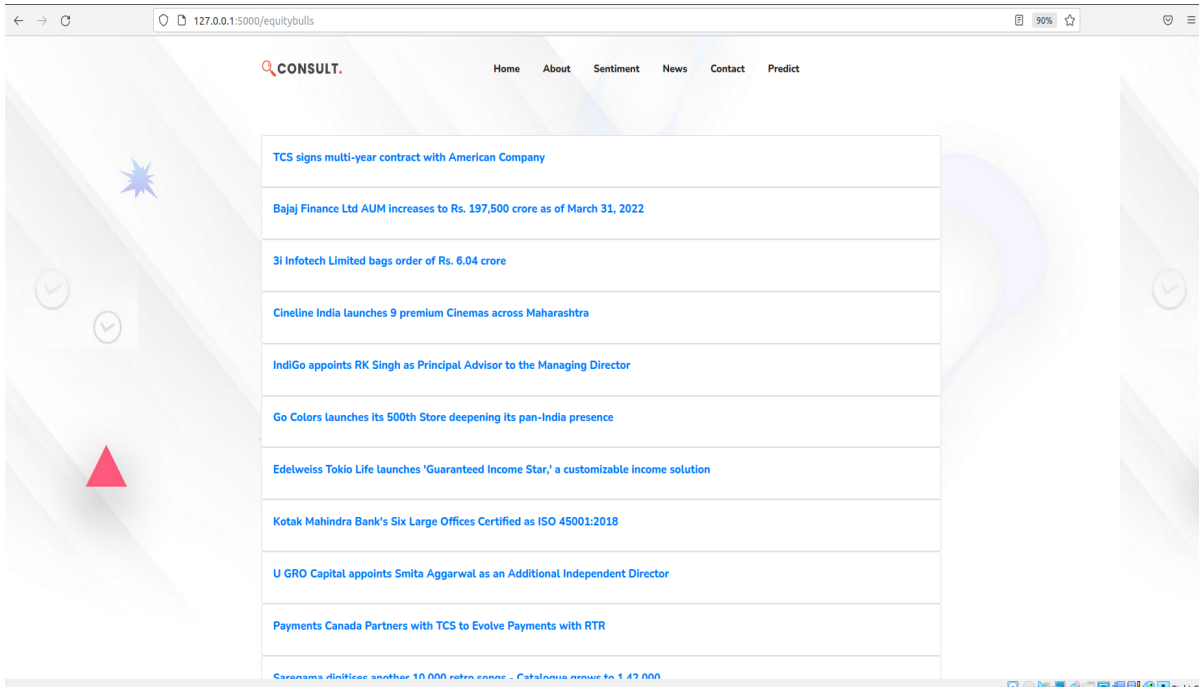
## 9.1 Home Page





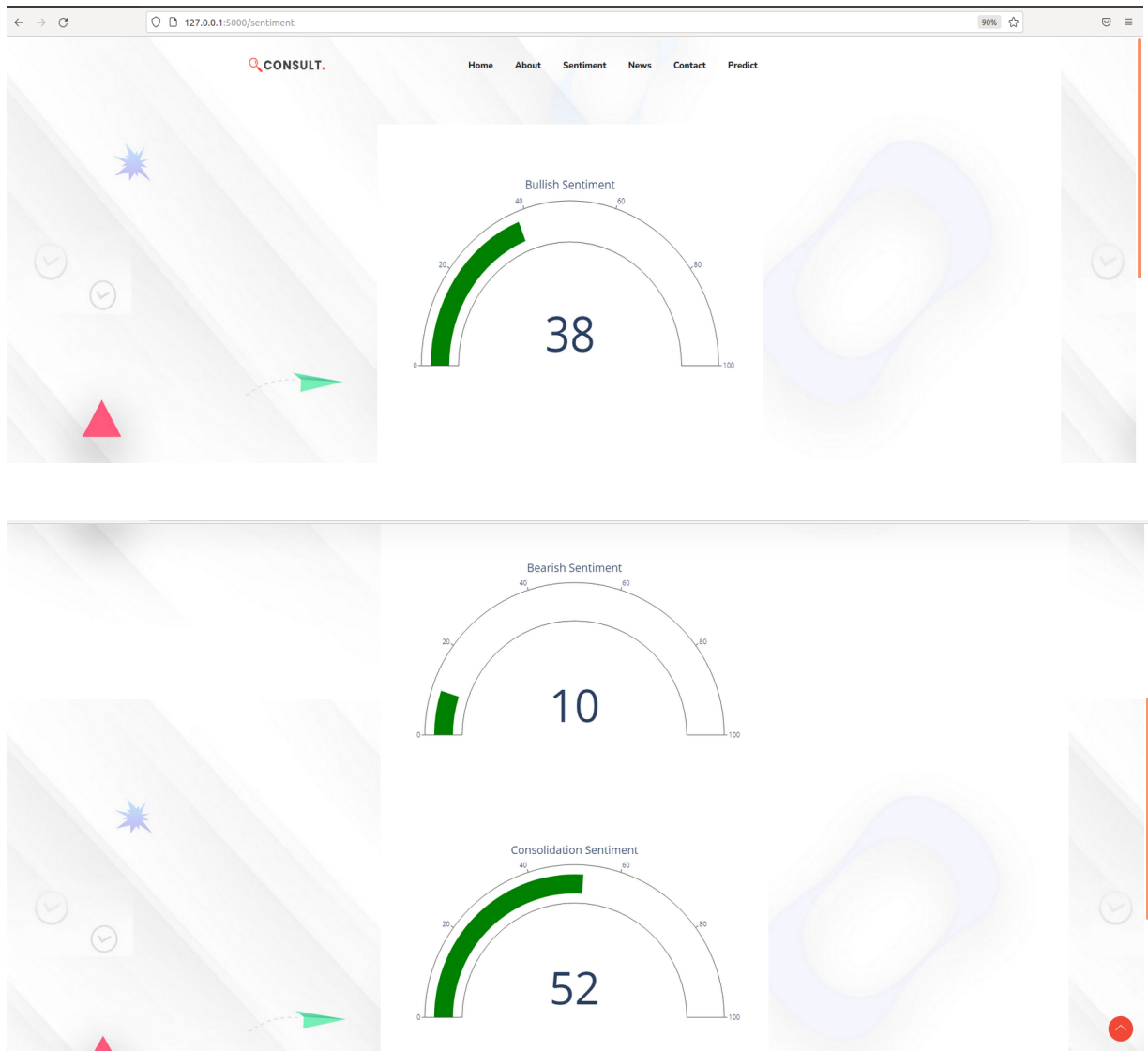
On navigating to the main page, the user would reach the home page. The home page gives an overview of the services provided to the customer.

## 9.2 News Section



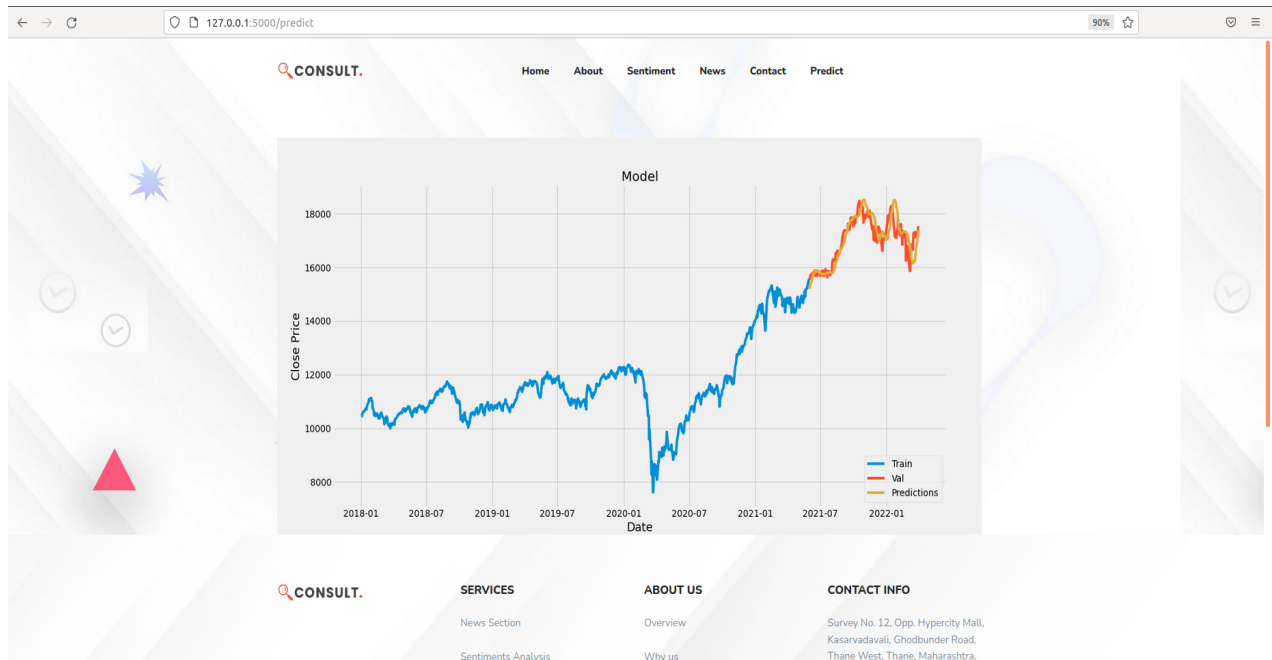
On navigating to the News section, the user would be able to browse through the latest news scraped from popular and reliable websites.

## 9.3. Sentiment Analysis



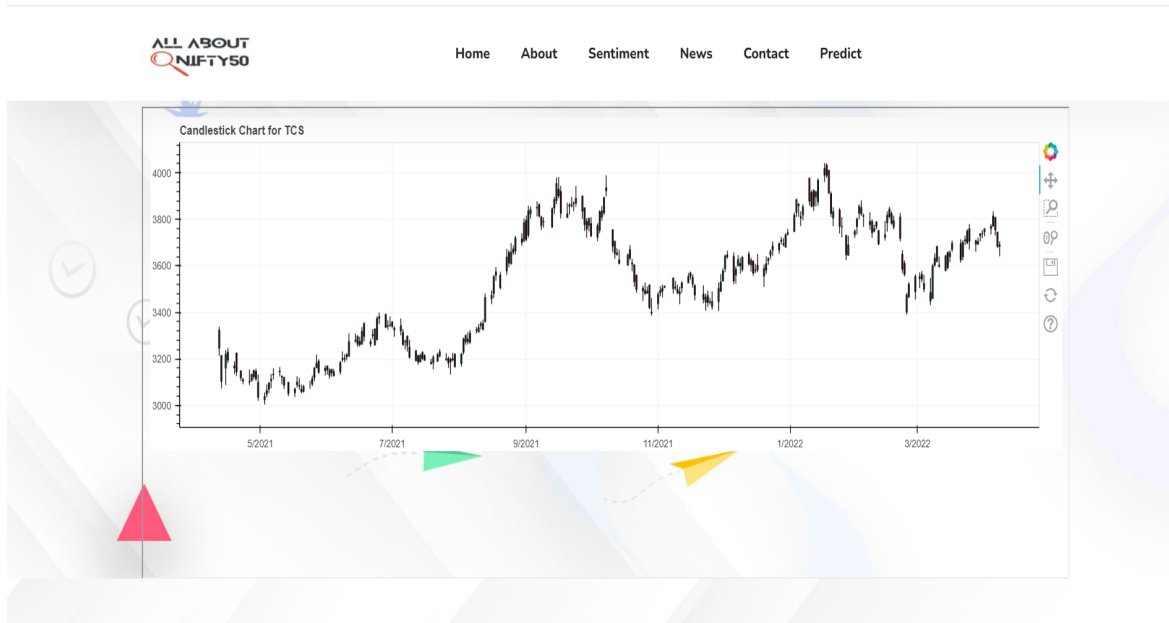
The sentiment analysis section displays the sentiments of the stock in terms of Bullish, Bearish, Consolidation Sentiments.

## 9.4 Prediction Section



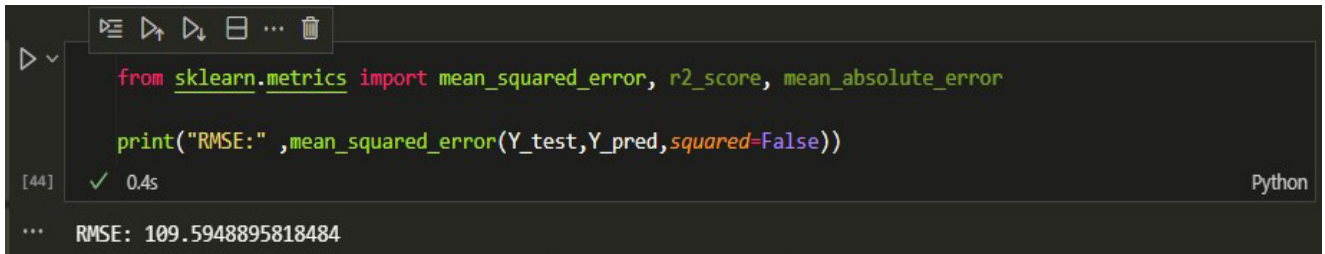
The prediction section displays the prediction chart on the basis of the LSTM Model used on the live dataset.

## 9.5 Visualize



The visualization section displays the candlestick chart on the basis of the past trends observed on the live dataset.

## 10. Results & Discussion

A screenshot of a Jupyter Notebook cell. The code imports metrics from sklearn and prints the RMSE for a Linear Regression model. The output shows an RMSE of 109.5948895818484.

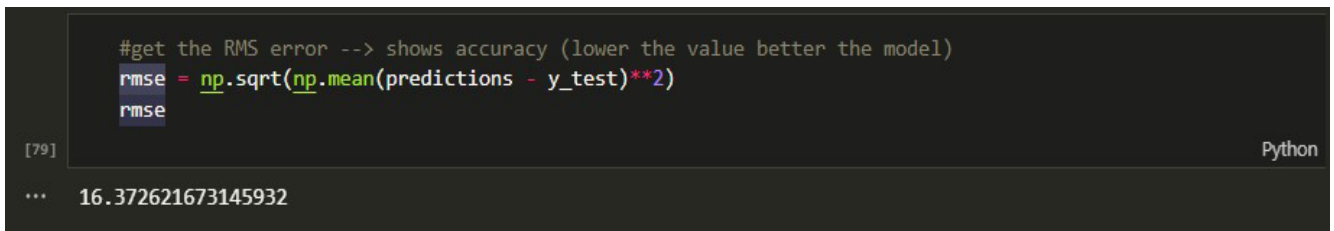
```
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error

print("RMSE:" ,mean_squared_error(Y_test,Y_pred,squared=False))
```

[44] ✓ 0.4s Python

... RMSE: 109.5948895818484

Fig. 10.1. RMSE value of Linear Regression.

A screenshot of a Jupyter Notebook cell. The code calculates the RMSE for an LSTM model using a custom formula. The output shows an RMSE of 16.372621673145932.

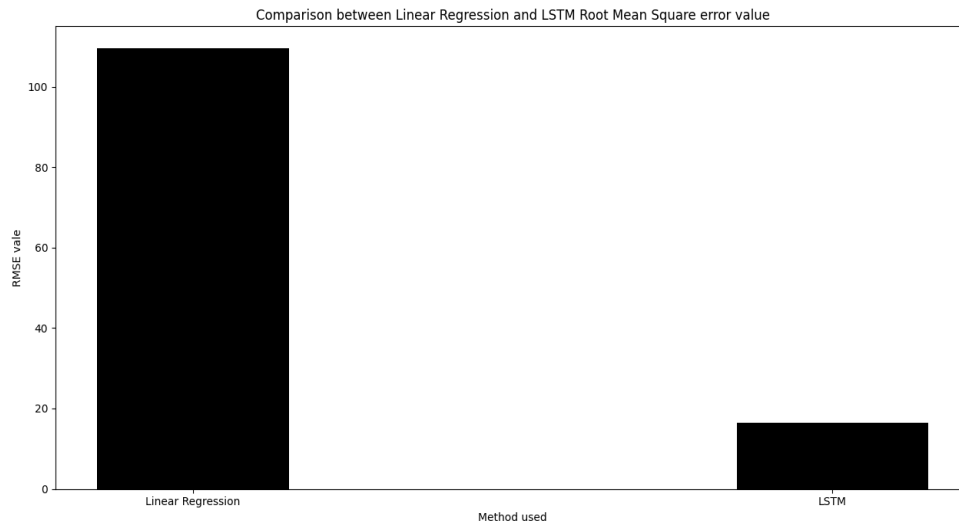
```
#get the RMS error --> shows accuracy (lower the value better the model)
rmse = np.sqrt(np.mean(predictions - y_test)**2)
rmse
```

[79] Python

... 16.372621673145932

Fig 10.2: RMSE value of LSTM Model.

By performing the following method and seeing the result and checking the RMSE of both Linear Regression and Long short-term memory (LSTM) and the difference between them we can conclude that Long short-term memory (LSTM) model is best fit for Time Series Dataframe and can give a better accurate result. And hence we also conclude that Linear Regression is more suited for non Time series data and should be used for that.



Graph 10.1: Bar Graph Depicting the Difference Between Both the RMSE

The data when given with respect to time can be unpredictable hence a Recurrent Neural Network such as Long short-term memory (LSTM) model who has feedback mechanism available shall be used for better accuracy and predictions.



## 11. Conclusion

In the future, we are planning to add a user database so that the users can subscribe to the latest news customized to their search history with the help of Machine Learning. The user would be led to a payment gateway. As Cloud is suitable for machine learning workloads, this project could be integrated with Cloud. Their inbuilt services could be used such as Cloud Natural Language API. As compared to the current stock related websites, our web application makes this whole process smooth and efficient for the investor. One can learn how well the desired stock is doing with just a few clicks on the screen.

# References

- [1] D.M.Thomas, S. Mathur, “Data Analysis by Web Scraping using Python”, in the Third International Conference on Electronics Communication and Aerospace Technology [ICECA 2019]
- [2] Z. Zhu. “Web Scraping Yahoo Finance News”, in *Towards Data Science* [Nov 6, 2020]
- [3] A. Biswal, “An Easy Guide to Stock Price Prediction Using Machine Learning”,, 2021
- [4] F. Ahemad, “Stock Market Analysis in Python”, 2022
- [5] Polamuri S.R, Shrinivas K, A. Krishna M. “A Survey on Stock Market Prediction Using Machine Learning Techniques, 2020
- [6] Y. Bengio. R. Pascanu, T. Mikolov “Learning Long-Term Dependencies with Gradient Descent”, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 5, NO. 2, MARCH 1994.
- [7] V. R. Vyshnavi, A. Malik “Efficient Way of Web Development Using Python and Flask”, 2019
- [8] P. Ashiwal , S.R.Tandan , P Tripathi , R Miri, “Web Information Retrieval Using Python and BeautifulSoup “ in International Journal for Research in Applied Science & Engineering Technology (IJRASET),2016
- [9] A. Sharma, S. Modak, Eashwaran Sridhar, “Data Visualization and Stock Market and Prediction” in International Research Journal of Engineering and Technology (IRJET), 2019
- [10] V. S. Pagolu, K. N. R. Challa, G. Panda, B. Majhi, “Sentiment Analysis of Twitter Data for Predicting Stock Market Movements” in International conference on Signal Processing, Communication, Power and Embedded System (SCOPES), 2016

- [11] B. Lindemann, T. Müller, H. Vietz, N. Jazdi, M. Weyrich, “*A Survey on Long Short-Term Memory Networks for Time Series Prediction*” 2016

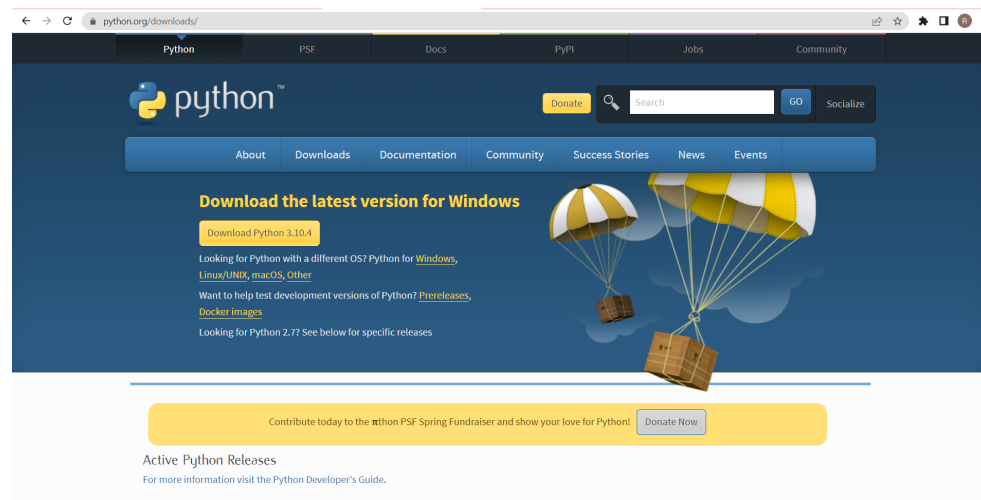
# Appendix

## Appendix A

### Python Download and Installation

The Python download requires about 25 Mb of disk space; keep it on your machine, in case you need to re-install Python. When installed, Python requires about an additional 90 Mb of disk space.

1. Go to the official website of Python ie. <https://www.python.org>.
2. Click on the Download button which will redirect you to <https://www.python.org/downloads/>.
3. Click the download button.(Python 3.10 64 bit)
4. Provide the location for the python file on your device
5. Open the downloaded python.exe file and proceed for installation.
6. Double click on the .exe file and select ADD PYTHON TO PATH checkbox to add it automatically to the windows environment variable.
7. Click on Install now.
8. A popup window will appear showing the installation progress.
9. Once the setup is completed we will see a message saying “Click on the close button to finish the installation of python”
10. Once python is installed, go to the Windows search bar and search for Python and we will see a desktop app named Python 3.10(64 bit).
11. Click on that and a command prompt will appear.

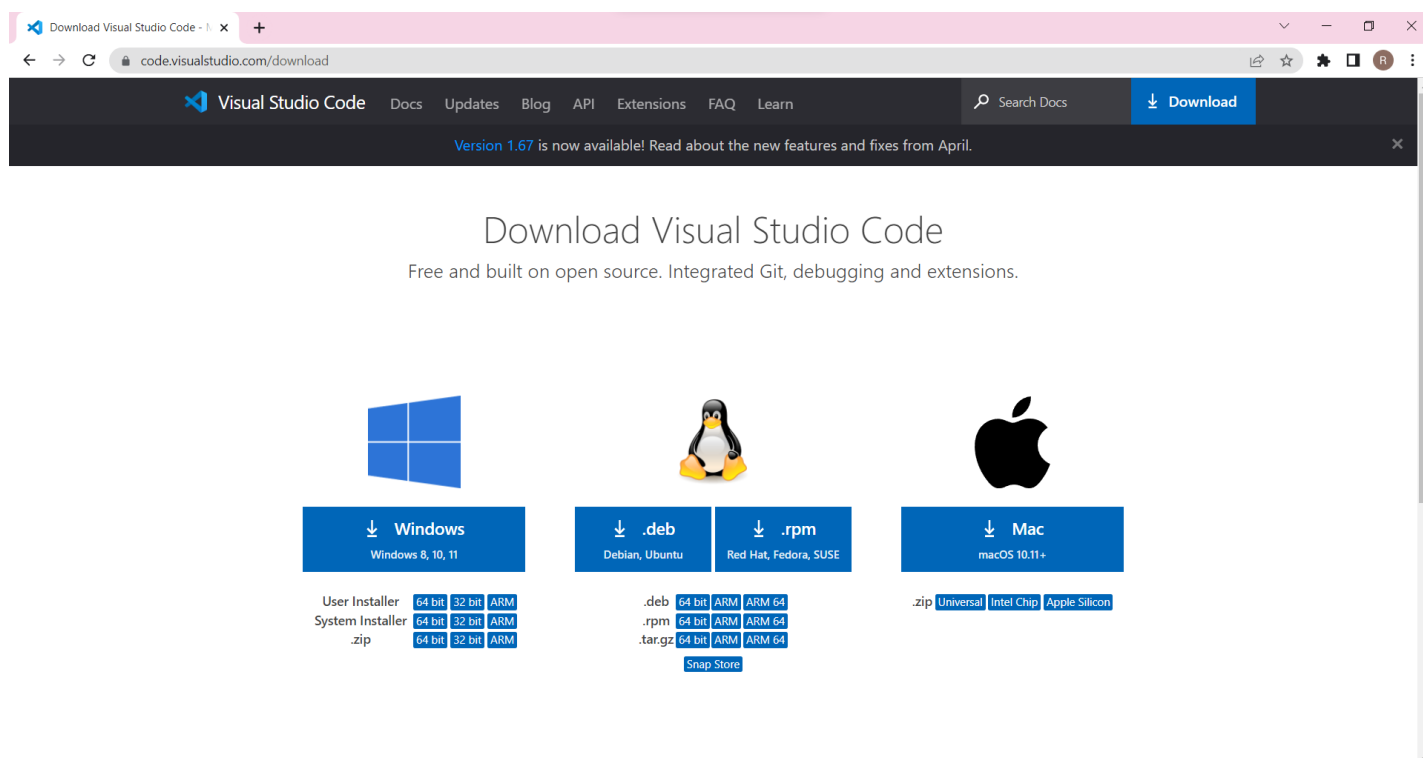


# Appendix

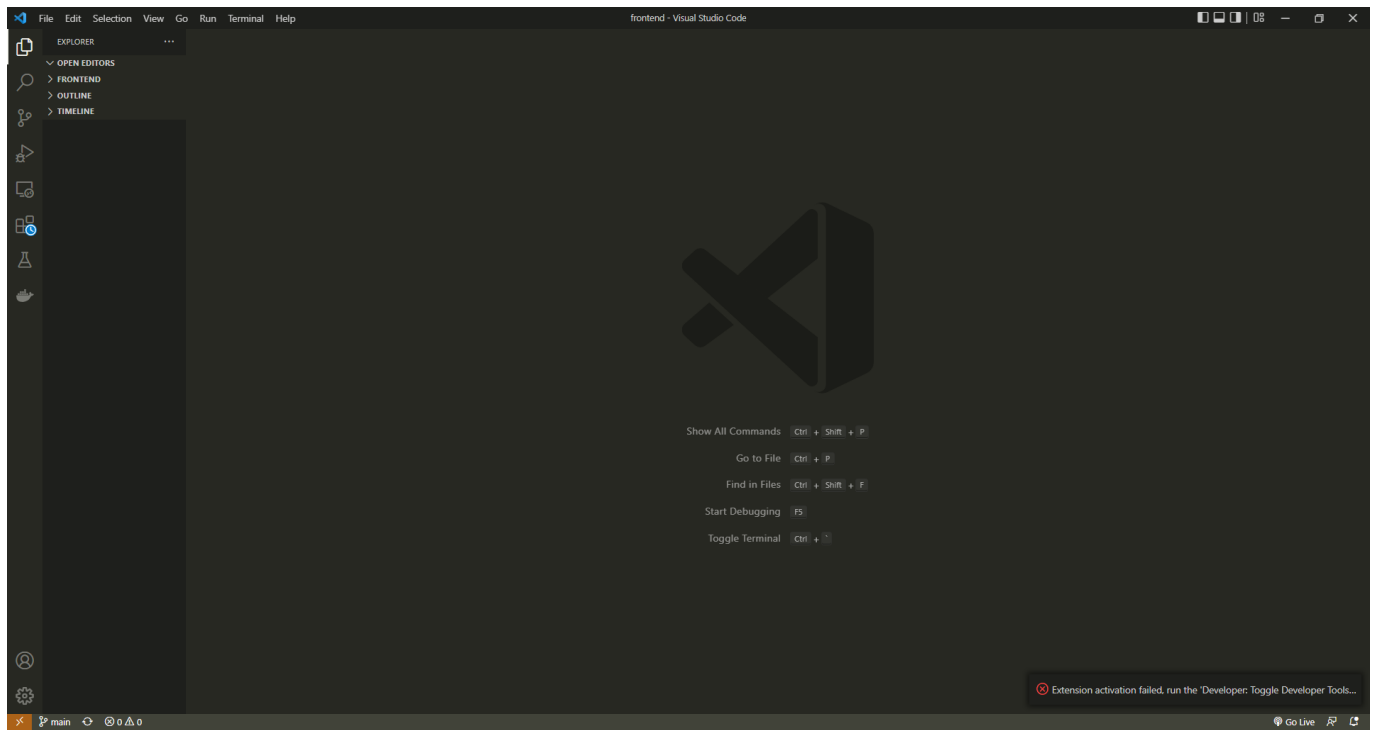
## Appendix B

### VS Code Download and Installation

1. Visit the official website of VS code <https://code.visualstudio.com/>
2. Go to the downloads section on this url <https://code.visualstudio.com/download/>



3. Download the Visual Studio Code installer for windows.
4. Post downloading double click on the file to start the installation.
5. Run the installer VSCodeUserSetup-{version}.exe.
6. Accept and agree to the terms and click Next.
7. After accepting all the requirements press the finish button.
8. By default VS code will get installed under  
"C:\users{username}\AppData\Local\Programs\Microsoft VS Code."
9. Once the installation is done you will see the following.



## **Acknowledgement**

We have great pleasure in presenting the mini project report on “**Stock Market Analysis**”. We take this opportunity to express our sincere thanks towards our guide **Prof. Rushikesh Nikam**, Department of Computer Engineering, APSIT Thane for providing the technical guidelines and suggestions regarding line of work. We would like to express our gratitude towards his constant encouragement, support and guidance through the development of the project.

We thank **Prof. Sachin Malave, Head of Department**, Computer Engineering, APSIT for his encouragement during the progress meeting and providing guidelines to write this report. We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues of APSIT for their encouragement.

**Student Name1: Rakshit Shah**

**Student ID1: 19102008**

**Student Name2: Catherene Joshi**

**Student ID2: 19102007**

**Student Name3: Rutuja Jain**

**Student ID3: 19102044**

**Student Name4: Poojan Shah**

**Student ID4: 20202001**

