

A PROJECT REPORT
on
STOCK MARKET PREDICTION

Submitted to
KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE AND ENGINEERING

BY

Siddharth Jena	1705851
Aakash Das	1705841
Meghna Verma	1705854
S. Subhashree Priyadarsini	1705861

UNDER THE GUIDANCE OF
Dr. Mainak Bandyopathy



SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAR, ODISHA - 751024
May 2020

A PROJECT REPORT
on
STOCK MARKET PREDICTION

Submitted to
KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE AND ENGINEERING

BY

Siddharth Jena	1705851
Aakash Das	1705841
Meghna Verma	1705854
S. Subhashree	1705861
Priyadarsini	

UNDER THE GUIDANCE OF
Dr. Mainak Bandyopathyay



SCHOOL OF COMPUTER ENGINEERING
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
BHUBANESWAE, ODISHA -751024
May 2020

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024



CERTIFICATE

This is certify that the project entitled

“NAME OF PROJECT“

submitted by

Siddharth Jena	1705851
Aakash Das	1705841
Meghna Verma	1705854
S. Subhashree Priyadarsini	1705861

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2019-2020, under our guidance.

Date: / /

(Dr. Mainak Bandyopathyay)
Project Guide

Acknowledgements

We are profoundly grateful to Dr. Mainak Bandyopathyay for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion.

Siddharth Jena
Aakash Das
Daibik Dasgupta
Meghna Verma
S. Subhashree Priyadarsini

ABSTRACT

Stock market prediction is the act of trying to determine the future value of a company's stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit. Even if you don't trade stocks for a living, or have any financial background whatsoever, being able to predict the future economy can be a huge benefit to your financial situation. Real estate can be sold at its highs, money can be drawn from mutual funds that are expected to rise indefinitely, and when things start to go bad, moving capital into tangible assets such as gold is a boon.

It is too simple to assume that with the steep decline of the market, it has already discounted epidemiologists' forecasts for Covid-19. By this logic, the stock market would fall further only if the virus turns out to be worse than forecast.

But the world has never seen an event quite like this before — a new pandemic that is being aggressively throttled by draconian shutdowns of whole industries, and by confining millions of people to their homes. The tools of statistical analysis and machine learning, powerful as they are, can't adequately assess what the world is experiencing. There isn't any stock market experience that is entirely analogous.

It's believed that the pandemic's effect on stock prices today is better understood as a series of emotional responses to unique events. People are trading stocks with their cellphones on their living room couches with the television news blaring about the pandemic. There is widespread foreboding, not just about the economy but about the possibility of grave illness or death in the weeks ahead.

Thus incorporating sentimental analysis along with prediction models(machine learning) may give enough data to the user to make educated guesses/predict the growth/ fall of the price.

Keywords: Stock Market Prediction, Emotional responses, Sentimental analysis, Machine learning

Contents

1	Introduction	1
2	Literature Survey	2
2.1	Long Short Term Memory	2
2.2	Natural Language Processing	2
3	Software Requirements Specification	3
3.1	Feasibility Study	3
3.2	Problem Statement	3
3.3	Objectives	4
3.4	System Overview	4
4	Requirement Analysis	5
4.1	Total Requirements	5
4.1.1	Function Requirements	5
4.1.2	Non-Function Requirement	5
4.2	External and User Requirement	5
5	System Design	6
5.1	System Overview	6
5.2	Design Constraints	6
5.3	System Architecture	6
6	System Testing	11
6.1	Test Cases and Test Results	11
6.2	Accuracy Comparison	11
7	Project Planning	12
7.1	Project Architecture	12
7.2	Dataset	12
7.3	Data Cleaning and Preprocessing	12
7.4	Training	12
7.5	Naive Bayes	13
7.6	LSTM	13
8	Implementation	14
8.1	Back-End Implementation and Automation	14
8.2	Linking of Front-End and Back-End using Firebase	15
8.3	Front-End development using Flutter	15
9	Screenshots of Project	16
9.1	Screenshots of Project (Android and iOS)	16
9.2	Screenshots of Project (Web)	20
9.3	Screenshot of APK Size	24
9.4	Screenshot of App Permission	24
10	Conclusion and Future Scope	25
10.1	Conclusion	25
10.2	Future Scope	25
11	References	26

List of Figures

1	Project Outline	6
2	SMP-20 App Outline	7
3	SMP-20 App Class Diagram	8
4	SMP-20 Use Case Diagram	9
5	SMP-20 Sequence Diagram	10

Chapter 1

Introduction

The information one has on the stock market is undeniably valuable. To be able to predict the future value of the stocks of a company can prove to be vital when making financial decisions, whether on an individual or an organizational level. Efficiently and accurately predicting these values has proven to be a challenge due to the very volatile conditions of the market.

Anomalous situations occur on a more or less frequent basis and affect the market status in various ways. So it is very challenging to continue to make correct predictions in these constantly changing conditions. The only alternative is to constantly update the model and means of prediction to maintain the accuracy of predictions. Looking thoroughly at this issue, we have found another means by which to predict the stock values without needing constant major updating in the method.

The Stock Market is very much known to be extremely unpredictable by any normal means. While one can rely on normal models and mathematical figures to obtain predictions, any unexpected situation or anomaly can result in a drastic and lasting change in the market, rendering the predictions to be inaccurate either temporarily or in the worst case, permanently. Therefore, any ordinary means of prediction may bring results initially but it would not be adaptable to the changes that are frequent in the Stock Market. Thus a viable alternative means are required by which to predict the stock values without needing constant updating.

Considering this, we have attempted a different approach to the prediction of Stock Market- one that does not rely on the regular methodology. Since the status of the market itself is so volatile, a viable alternative is to study the reaction of those investing in the market and study their opinion regarding the future. After collecting their sentiments regarding the future, we must thoroughly analyze that data and separate it into positive or negative feedback and accordingly decide whether or not the value of the related stocks are going to rise or fall.

This sentiment analysis would be very much adaptable to the volatile status of the market due to not relying on sheer mathematical values that can be influenced by unexpected factors, but instead relying on the sentiments of the investors and experts themselves to formulate a prediction about the value of the stock.

Chapter 2

Literature Survey

INTRODUCTION

Stock market prediction is an idea that is known to be inaccurate due to its random nature. But even the randomness must depend upon something that makes the stock prices move in a way that normal machine learning algorithms cannot predict. Thus this is a topic which was considered to be futile among most of the people in the domain of machine learning. But in the recent times, new methods of analysis of stocks have surfaced which help tackle the randomness of stock prices to an extent.

2.1 Long Short-Term Memory (LSTM)

As we tried to find out the best way to predict future prices(growth/downfall), LSTM was found out to be the way to process sequential data¹. LSTM thus is used in our project. The idea of using LSTM was strengthened by the fact that a similar usage of LSTM for prediction of stock returns in China increased the accuracy of the predictions by nearly 10% when compared to other machine learning algorithms. The efforts demonstrated the power of LSTM in stock market prediction in China, which is mechanical yet much more unpredictable². But stock market's unpredictability did not originate due to machine learning algorithms' inability to predict future prices.

2.2 Natural Language Processing (NLP)

One of the reasons stock prices are so random are due to the reaction of the people to the ongoing events in the real world which are related directly/indirectly to the stocks. NLP refers to a method of analysis of texts to find hidden correlations which can be used to find out sentiment of people towards specific stocks. One method includes extracting information from online review websites and then using the information extracted to predict the effect on the stocks³. We used this idea to instead extract information from social websites like Twitter so as to find out people's emotions towards certain stocks.

Chapter 3

Software Requirements Specification

3.1 Feasibility Study

Stock Market is very difficult to accurately predict. It is a highly complex problem that has far too many unpredictable variables. It is a central focal point of convergence for buyers and sellers where depending on which exceeds the other, the price increases or decreases accordingly. So there are factors which cause people to buy and sell, but it deals far more with emotion than logic. Due to the nearly unpredictable nature of emotion, the movements and status of the market itself is constantly changing. It is futile to attempt to predict this through normal logical means.

However in the proposed system, the predictions are likely to be superior to regular predictions due to changing the very basis for the prediction. Instead of relying on logical or mathematical factors, the model collects and analyzes the sentiments of the people purchasing and selling the stocks so as to summarize the sentiments into identifiable positives and negatives and then ascertain the values of the related stocks.

This system is a short term prediction and due to this it is also more feasible than long term predictions as it is constantly adapting to the changing status of the stock market. It is also taking in new information to provide output and as such is more reliable for these short term predictions.

3.2 Problem Statement

The Stock Market is very much known to be extremely unpredictable by any normal means. While one can rely on normal models and mathematical figures to obtain predictions, any unexpected situation or anomaly can result in a drastic and lasting change in the market, rendering the predictions to be inaccurate either temporarily or in the worst case, permanently. Therefore, any ordinary means of prediction may bring results initially but it would not be adaptable to the changes that are frequent in the Stock Market. Thus a viable alternative means are required by which to predict the stock values without needing constant updating.

3.3 Objectives

Considering this, we have attempted a different approach to the prediction of Stock Market- one that does not rely on the regular methodology. Since the status of the market itself is so volatile, a viable alternative is to study the reaction of those investing in the market and study their opinion regarding the future. After collecting their sentiments regarding the future, we must thoroughly analyze that data and separate it into positive or negative feedback and accordingly decide whether or not the value of the related stocks are going to rise or fall.

This sentiment analysis would be very much adaptable to the volatile status of the market due to not relying on sheer mathematical values that can be influenced by unexpected factors, but instead relying on the sentiments of the investors and experts themselves to formulate a prediction about the value of the stock.

3.4 System Overview

The system is called “Stock Market Prediction” and consists of a website as well as web application that can effectively predict the values of the stock market as well as provide a summary of the positive and negative sentiments of people regarding the related stocks.

Chapter 4

Requirement Analysis

4.1 Total Requirements

After analysis of the system, the requirements expected of the system have been categorized into two divisions-

4.1.1 Function Requirements

These features must be included in any system to satisfy the business needs and be accepted by the users:

- The system should be able to analyse the sentiments obtained from the source.
- The system should be able to summarize the information gleaned from the sentiments.
- The system should be able to provide a proper summary of data so as to allow predict the incoming change in the required stock price.

4.1.2 Non-Function Requirements

These requirements are a description of features, characteristics and attribute of the system as well as any constraints. They are as follows:

- The system should provide better accuracy.
- The system should have a proper and simple interface for users.
- The system should provide best possible performance in best possible time

4.2 External and User Requirements

These requirements are required for using the application in a proper way. They are as follows:

- Internet access.
- Understanding of pie charts and graphs.
- Understanding of how stock works.
- Android/IOS smartphone (Android 9 or above)

Chapter 5

System Design

5.1 System Overview

Our purpose of this project is to create an application which visualizes data related to stocks to give users an easy way to see the factors affecting stock prices at a glance and thus make educated decisions.

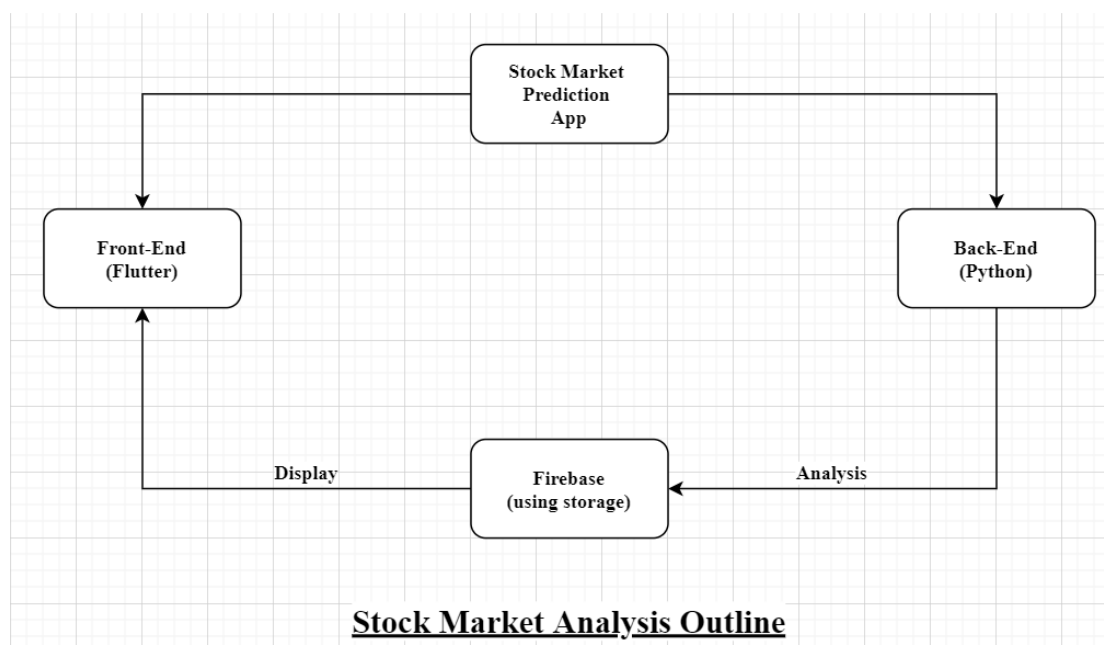
The application provides a graph showing the historical trend of the stock price as well as the price for the next day. Also it will provide tweets related to the stocks along with their sentiments and a pie chart showing different sentiments in total so as to compensate for the variable causing randomness of stock prices.

5.2 Design Constraints

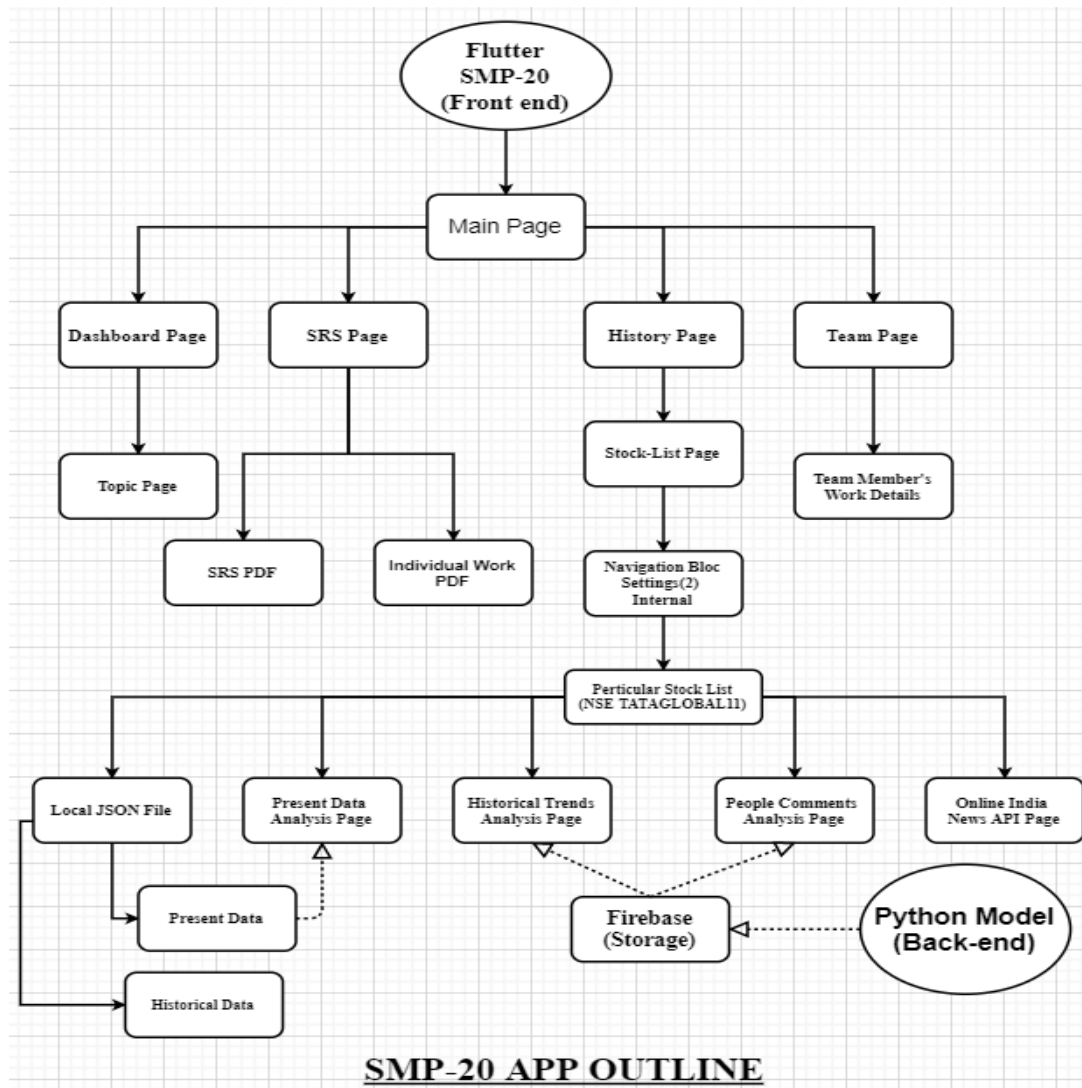
- User must understand how stocks work and how sentiments of people towards a stock affects its prices.
- Application will be available only in English.

5.3 System Architecture

1. Project Outline

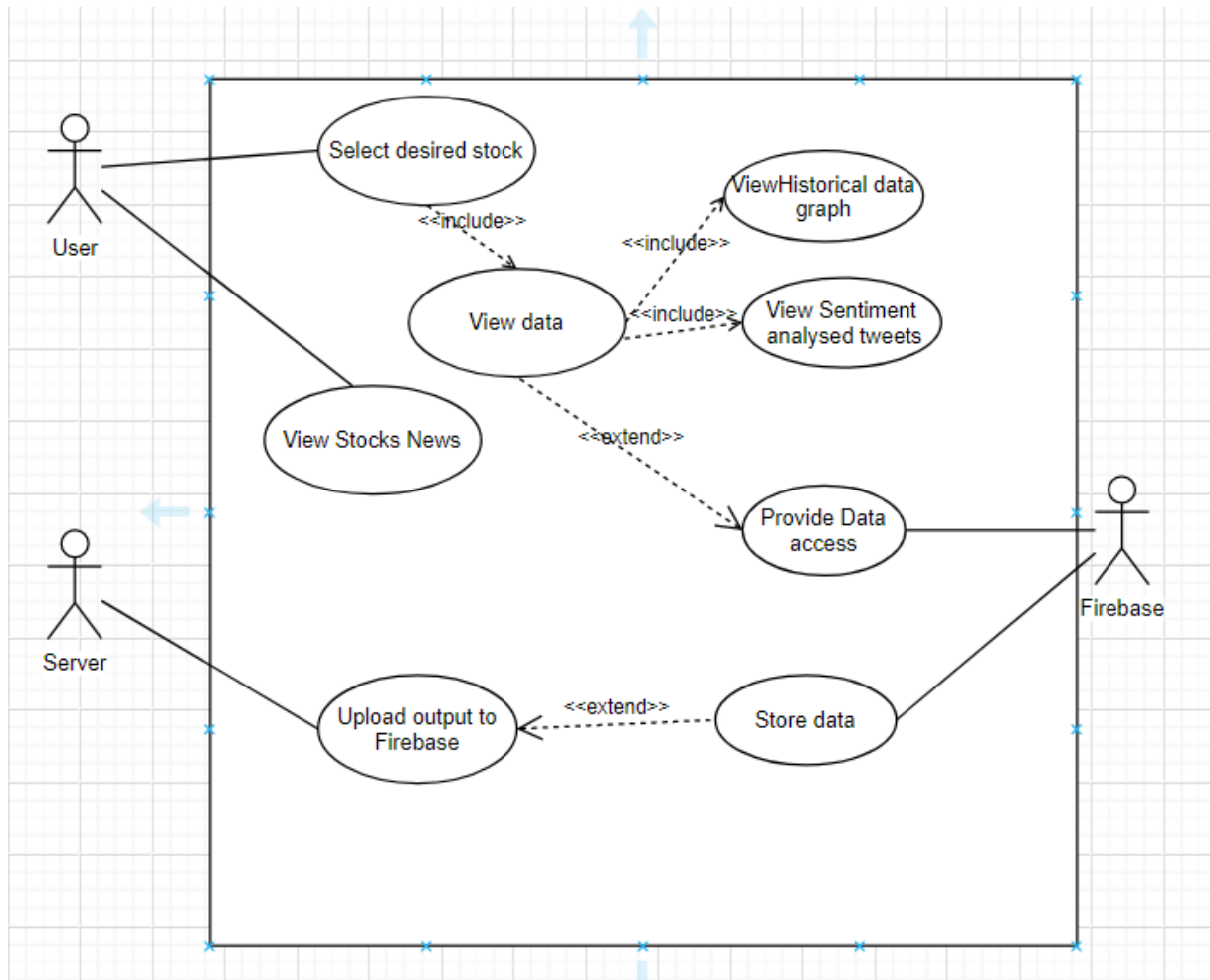


2. SMP-20 APP OUTLINE



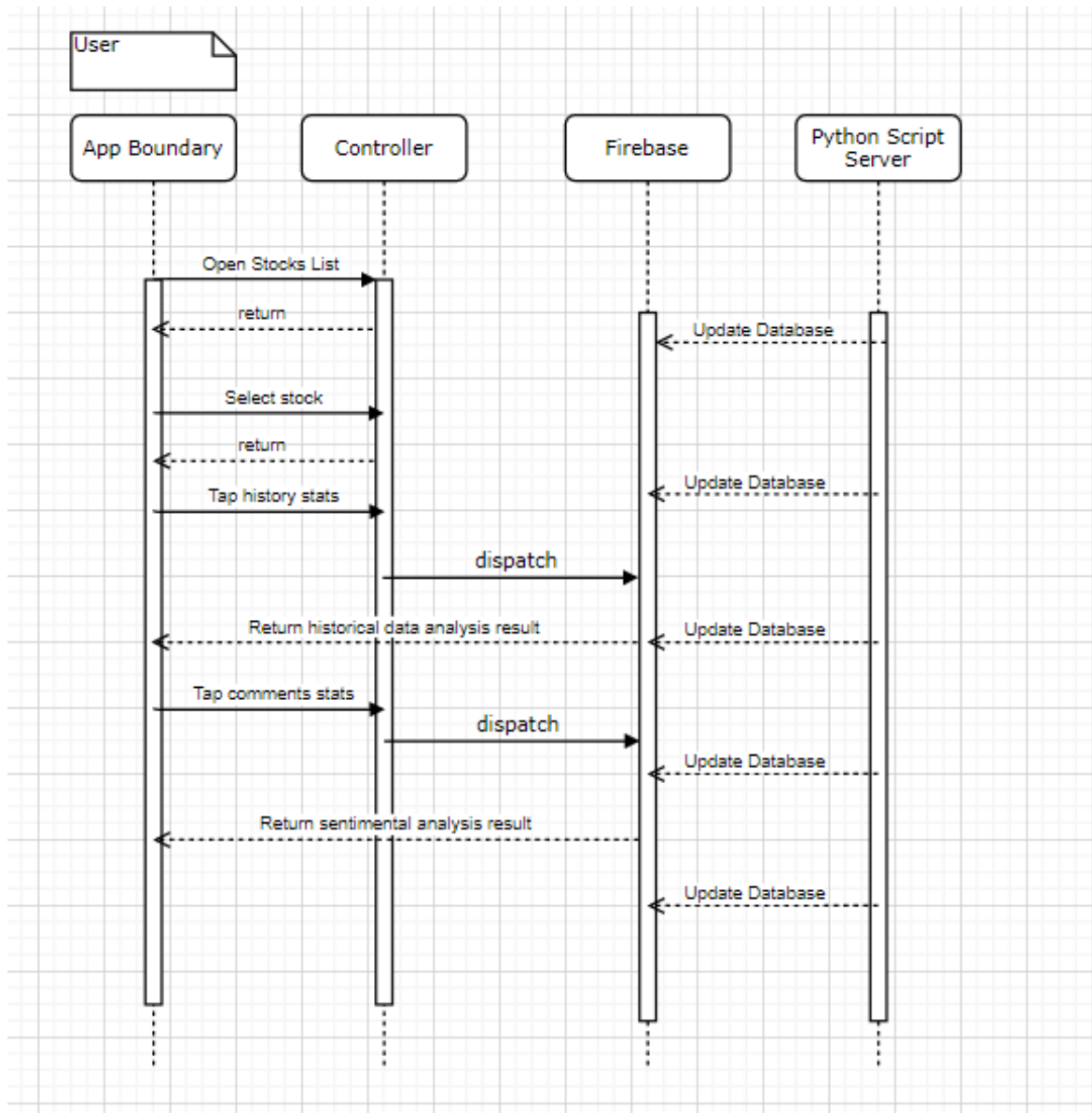
The app outline gives us a illustration of the flow of control in the app.

4. SMP-20 USE CASE DIAGRAM



A use case diagram is a graphical representation of the interaction of different users with the system.

5. SMP-20 SEQUENCE DIAGRAM



It shows the objects and classes involved in the scenario and the sequence of messages exchanged between the objects/classes needed to carry out the functional requirements.

Chapter 6

System Testing

As the application did not have any user input other than interacting with the user interface, the user interface leading to the desired data has been tested.

6.1 Test Cases and Test Results

Test ID	Test Case Title	Test Condition	System Behavior	Expected Result
T01	Historical Data	Tap on Historical stats	Shows historical trends graph with prediction	Should show historical trends graph with prediction
T02	Sentimental Analysis	Tap on Comments stats	Shows pie chart on different emotions of tweets	Should show pie chart on different emotions of tweets
T03	News	Tap on NEWS	Shows News	Should show News
T04	JSON	Tap on Present Data Files	Shows tweets with sentiments	Should show tweets with sentiments

6.2 Accuracy Comparison

Sentimental Analysis

	Pipeline components	Maximum accuracy achieved
1	CountVectorizer, Tfidf Transformer, Multinomial NB	72%
2	CountVectorizer, Multinomial NB	73%

Chapter 7

Project Planning

7.1 Project Architecture

The major components of this project's architecture include-

- Historical data analysis deep learning model
- Sentimental analysis machine learning model
- Firebase to store the output of the model
- A computer server running the models periodically
- User interface using flutter

7.2 Dataset

Dataset for historical data analysis was collected using the python module named yfinance⁴.

Dataset for sentimental analysis has been collected from multiple sources⁵. Multiple datasets were downloaded and were cleaned as mentioned in the following section.

7.3 Data Cleaning and Preprocessing

The major challenge for the datasets of sentimental analysis was that there were too many complex emotions/ not enough neutral emotions/ not a proper balance of positive, negative and neutral emotions. Thus multiple datasets were joined and their labelling was changed accordingly so as to counter all the challenges.

The obtained historical dataset had its date in string format which was changed to datetime format in python.

7.4 Training

The dataset for sentimental analysis was trained on using Naive-Bayes machine learning model and the dataset for historical data analysis was trained on using LSTM deep learning model.

7.5 Naive Bayes

The parameters of the model are-

```
pipeline(memory=None,
          steps=[('bow',
                  CountVectorizer(analyzer='word', binary=False,
                                  decode_error='strict',
                                  dtype=<class 'numpy.int64'>, encoding='utf-8',
                                  input='content', lowercase=True, max_df=1.0,
                                  max_features=None, min_df=1,
                                  ngram_range=(1, 1), preprocessor=None,
                                  stop_words=None, strip_accents=None,
                                  token_pattern='(?u)\\b\\w\\w+\\b',
                                  tokenizer=None, vocabulary=None)),
                  ('classifier',
                   MultinomialNB(alpha=1.0, class_prior=None, fit_prior=True))],
          verbose=False)
```

7.6 LSTM

The designed layers for LSTM are-

```
model = Sequential()
model.add(LSTM(units=60,
               return_sequences=True, input_shape=(x_train.shape[1],1)))
model.add(LSTM(units=60))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
```

Chapter 8

Implementation

Implementation refers to process of converting the system architecture into individual modules and then finally integrating them keeping in mind that all the functional and non functional requirements are met.

The implementation can be subdivided into-

- ✓ Back-end implementation and automation
- ✓ Linking of front-end and back-end using Firebase
- ✓ Front-end development using Flutter

8.1 Back-end Implementation and Automation

This included creation of one model based on Long Short-Term memory Deep learning for historical data analysis and future price prediction and one model based on Multinomial Naive Bayes for Sentimental Analysis. Both the models were created on Jupyter notebook using Python 3.7

For Sentimental Analysis

```
from sklearn.pipeline import Pipeline
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB
pipeline = Pipeline([
    ('bow', CountVectorizer()), # strings to token integer counts
    #('tfidf', TfidfTransformer()), # integer counts to weighted TF-IDF scores
    ('classifier', MultinomialNB()), # train on TF-IDF vectors w/ Naive Bayes classifier
])
```

For Historical Data Analysis

```
model = Sequential()
model.add(LSTM(units=60, return_sequences=True, input_shape=(x_train.shape[1],1)))
#model.add(Dropout(0.2))
model.add(LSTM(units=60))
#model.add(Dropout(0.2))
model.add(Dense(1))

model.compile(loss='mean_squared_error', optimizer='adam')
model.fit(x_train, y_train, epochs=1, batch_size=1, verbose=1)
```

For automation, i.e. for periodical running of the python models so as to ensure an updated output is displayed everytime, the python scripts are made to run every 15 minutes everyday on a computer using task scheduler.

8.2 Linking of front-end and back-end using Firebase

After the python program runs, it uploads the output to Firebase where it is stored. The Flutter application loads the output from Firebase for user's viewing.

8.3 Front-end development using Flutter

Flutter is an open-source UI software development kit created by Google. It is used to develop applications for Android, iOS, Windows, Mac, Linux, Google Fuchsia and the web. The first version of Flutter was known as codename "Sky" and ran on the Android operating system.

Points:

- ✓ Build a splash screen with app icon for android and iOS and change web icon also, this app fully runs in android and iOS and web (partially due to lack of package).
- ✓ Create a dashboard page with some swiped topics.
- ✓ Back-end development between Firebase and Python connected with Flutter.
- ✓ Secondly created a database screen where we attached our Project Individual Report and SRS Document (Local asset as PDF Format)
- ✓ Thirdly we design a particular stock list design which is also divided into 5 pages;
 - ✓ First one called JSON data where we attached present and historical JSON raw file according to positive, negative, and neutral statements analyzed by back-end.
 - ✓ Then, the next page is present data page where it analyzed locally by present JSON files.
 - ✓ Then the third and fourth page analyzed historical and people comments data from online by Firebase-storage and back-end process.
 - ✓ And the last one page; a online Indian news portal; which is updated by an outside API (collected from Internet randomly)
 - ✓ The 4th main page is the team page where the page displays our team members' picture and contribution in the project briefly.
 - ✓ The most important part of front end is 4 main part page was attached by curved bottom navigation and the stock list part was attached by navigation block.
 - ✓ The app-size also be minimized (i.e. 13.1 MB).

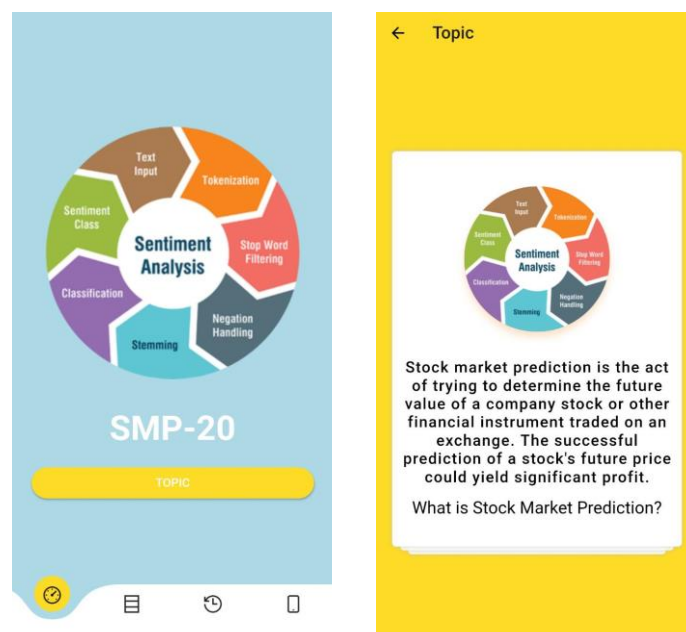
Chapter 9

9.1 Screen shots of Project(Android &iOS)

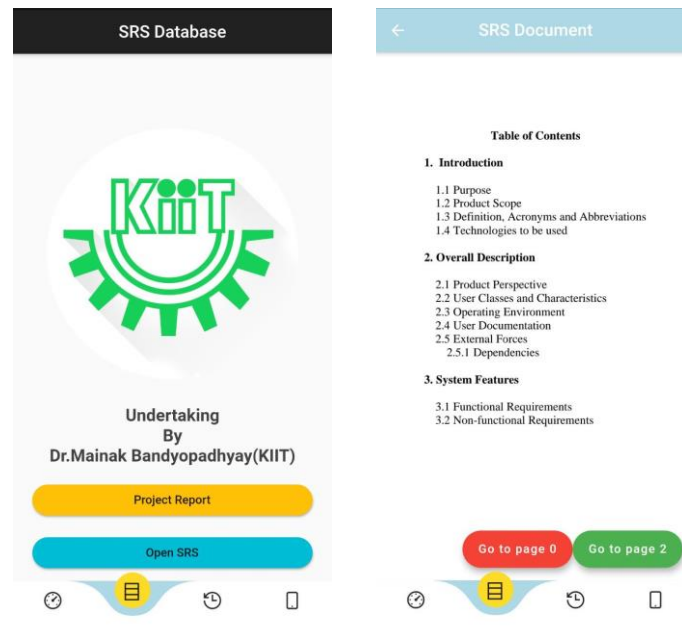
1.SMP-20 APP SPLASH SCREEN:



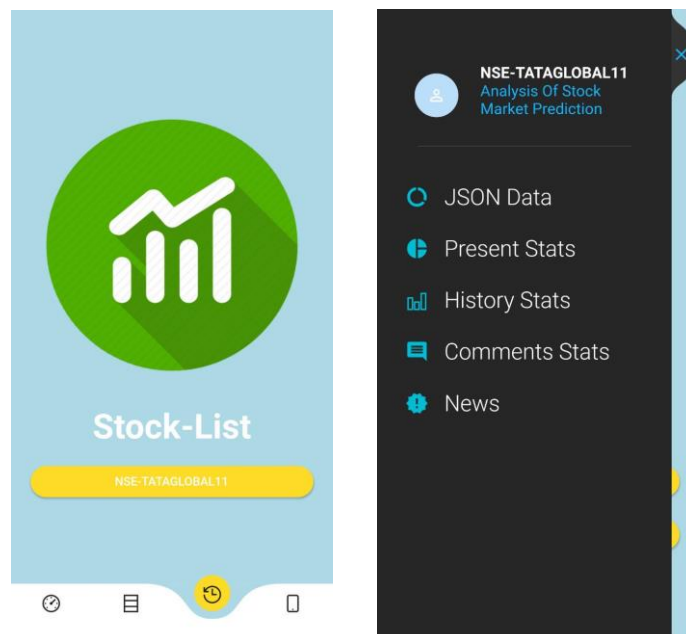
2.SMP-20 APP DASHBOARD PAGE:



3. SMP-20 APP DATABASE SCREEN:



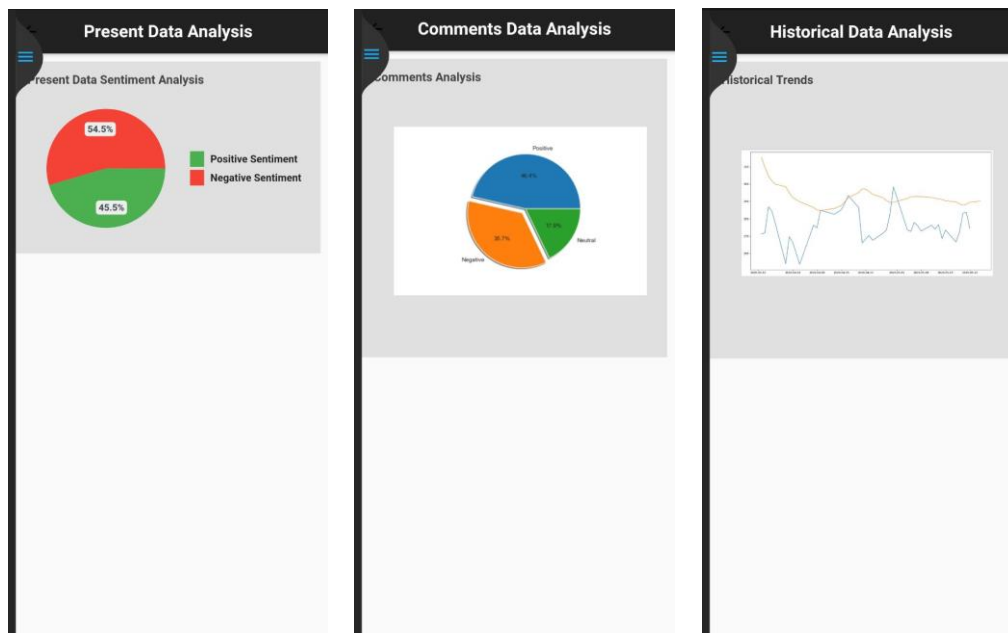
4. SMP-20 APP HISTORY & NAVIGATION SCREEN:



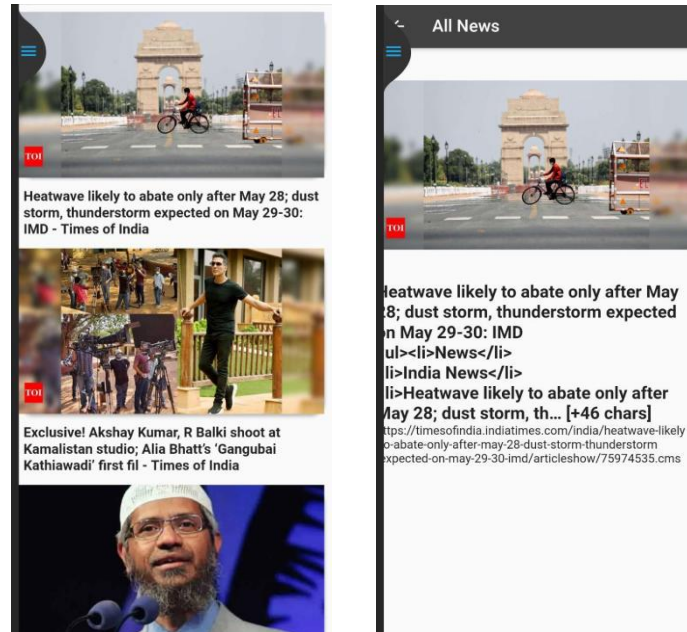
5. SMP-20 APP LOCAL JSON FILE SCREEN:



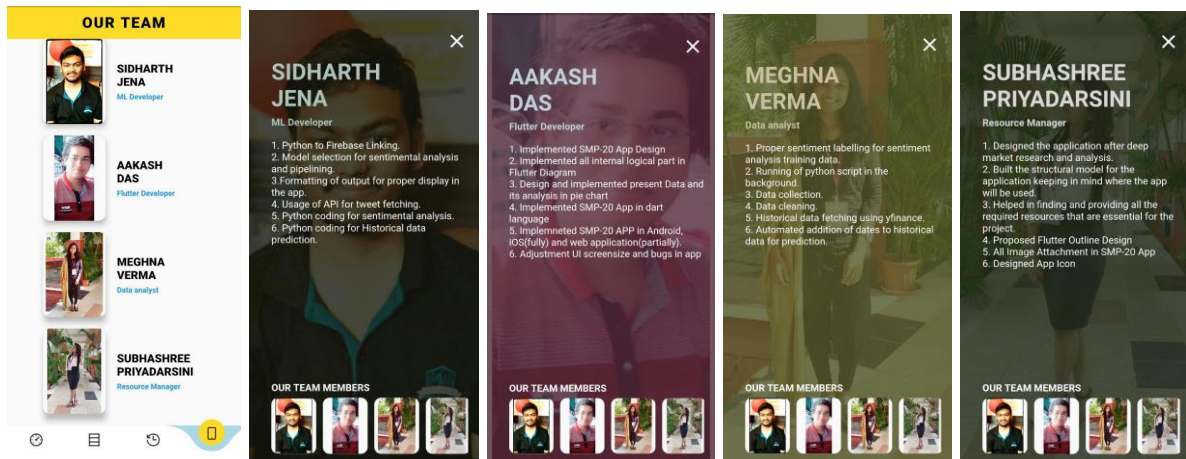
6. SMP-20 APP DATA ANALYSIS SCREEN:



7. SMP-20 APP ONLINE NEWS:

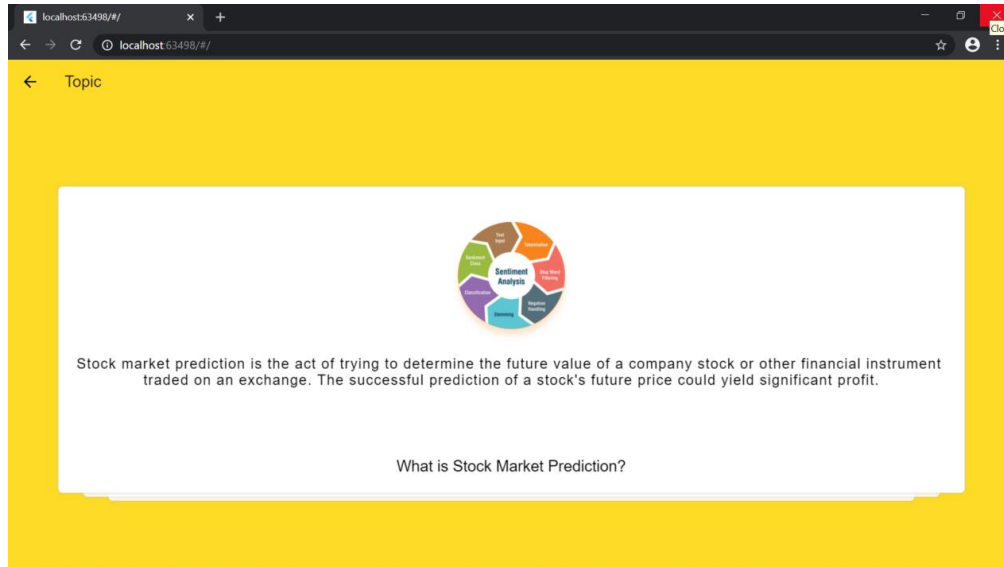


8. SMP-20 APP TEAM SCREEN:

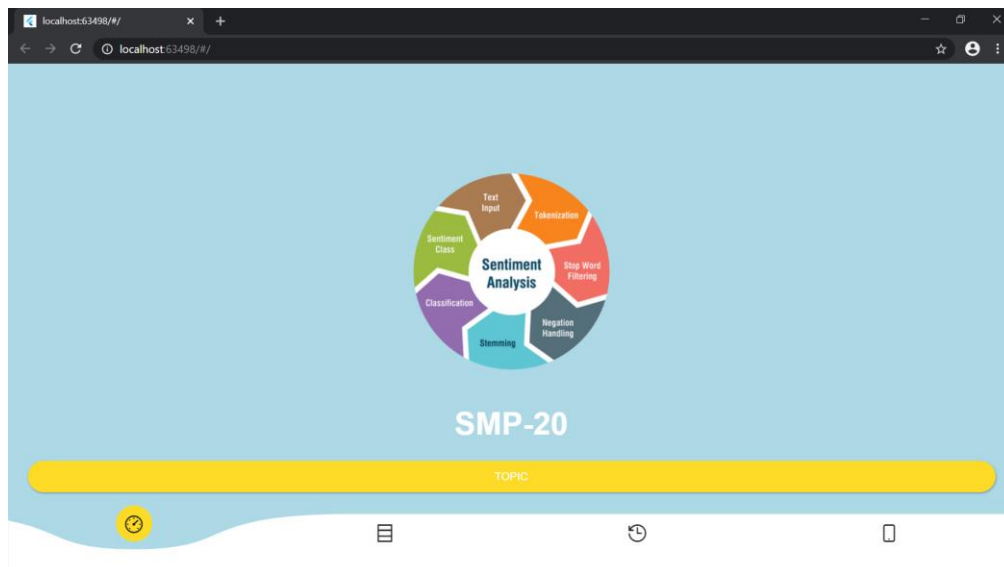


9.2 Screen shots of Project(Web)

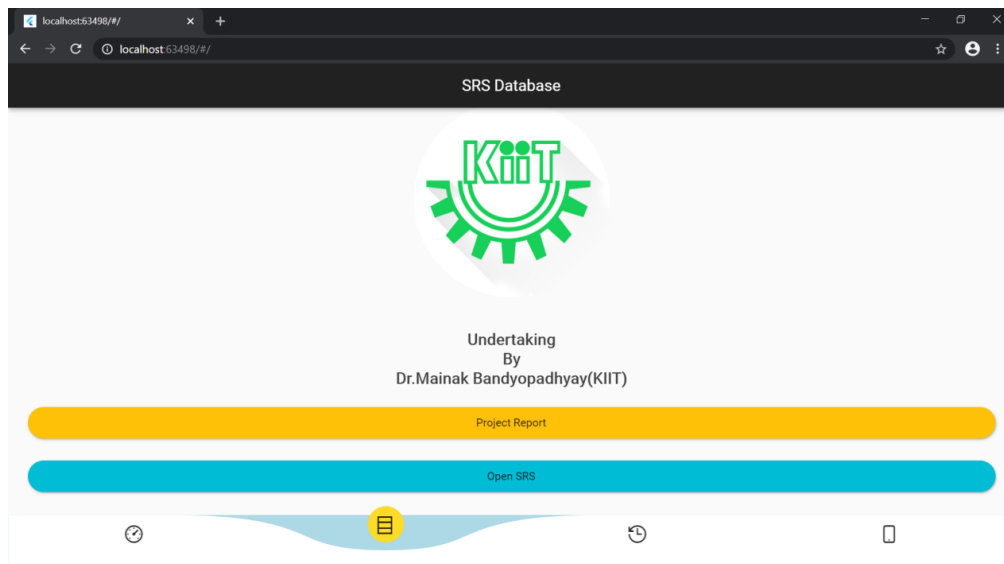
1.



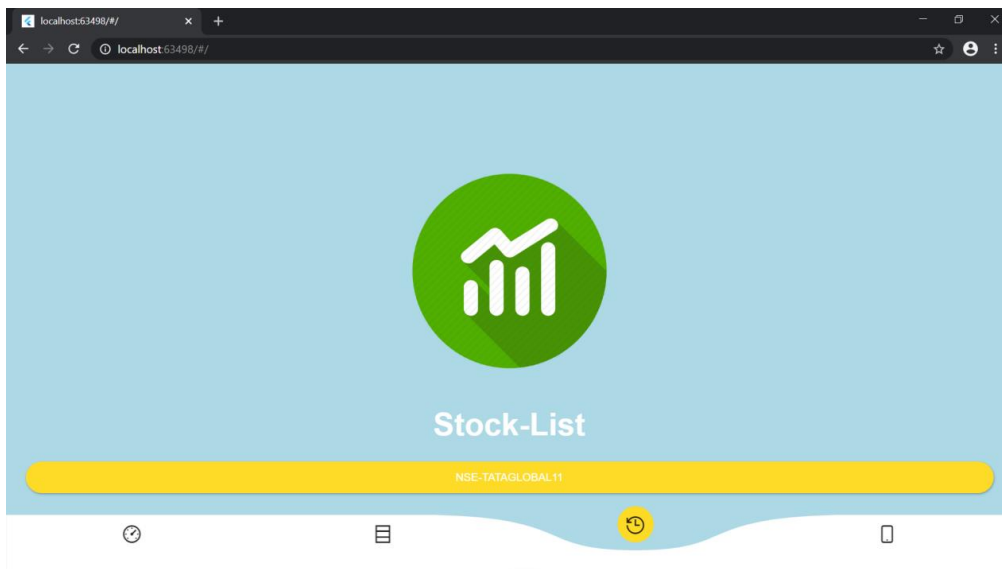
2.



3.



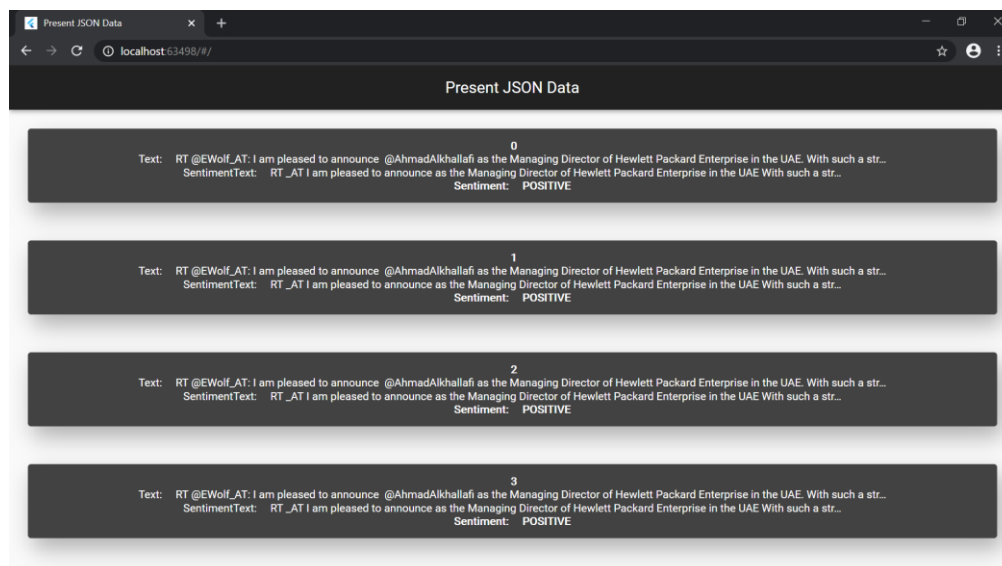
4.



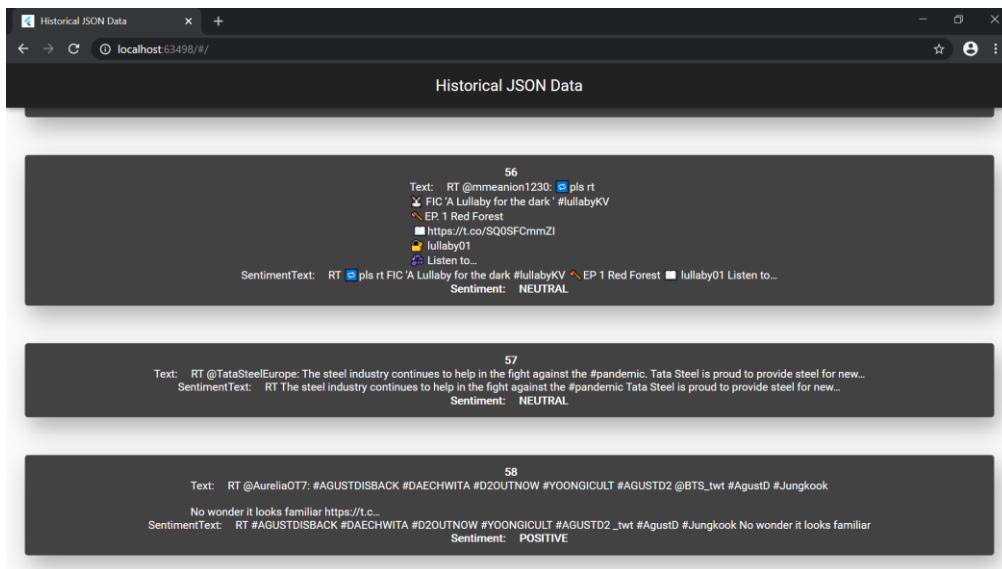
5.



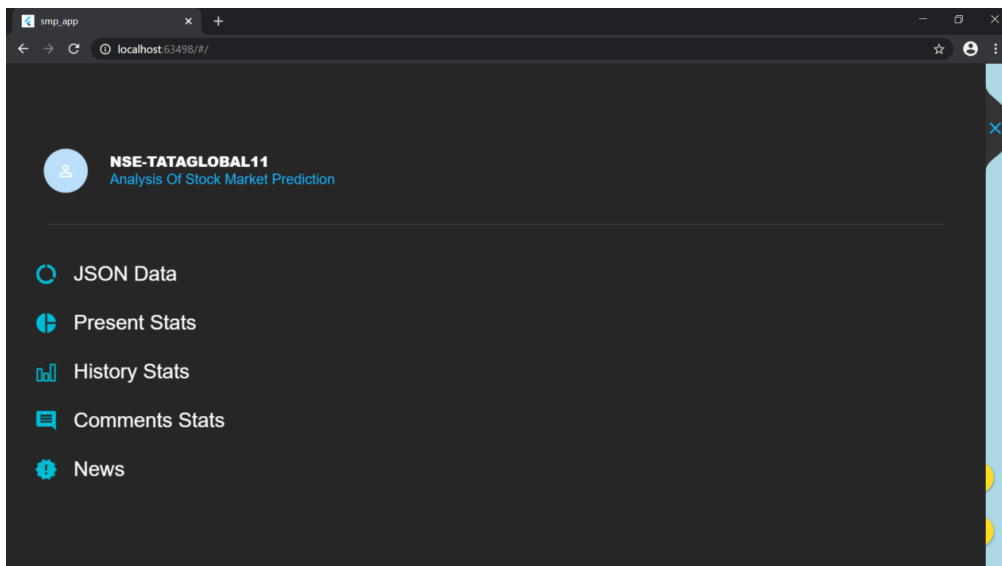
6.



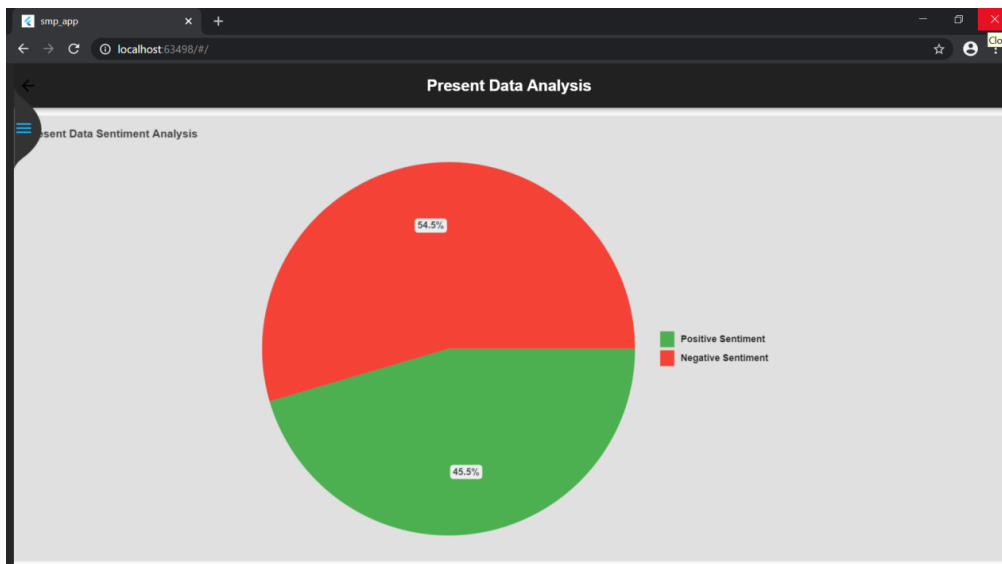
7.



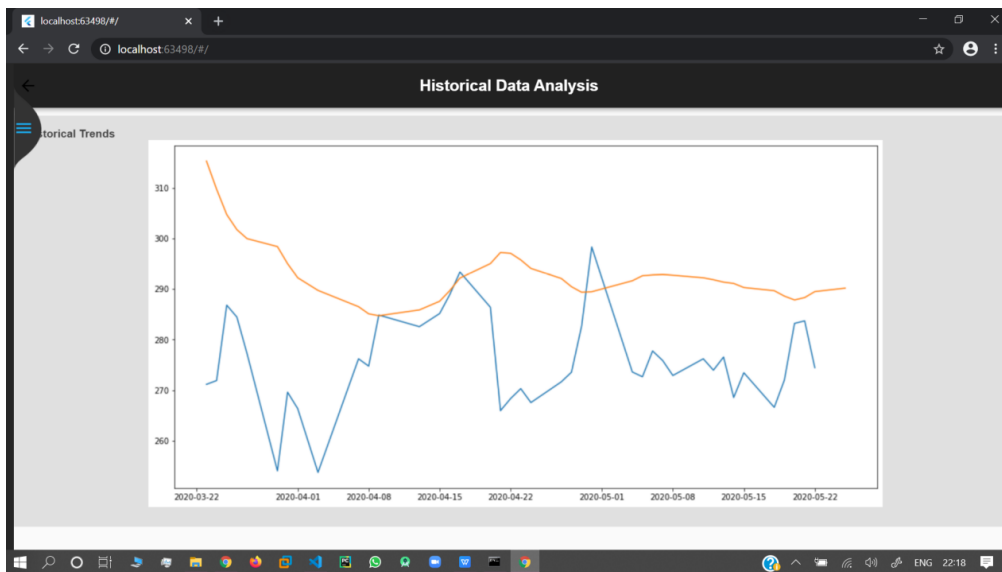
8.



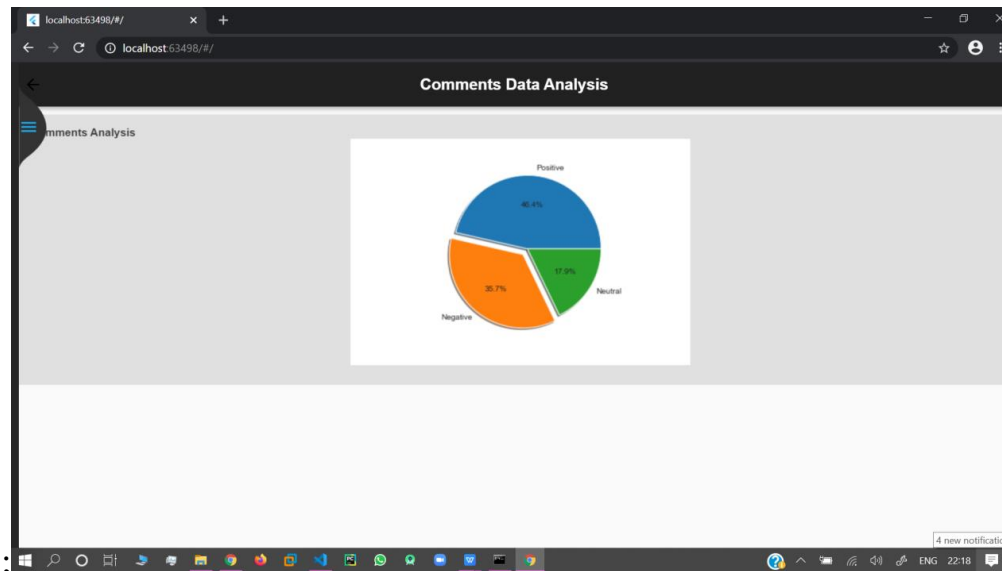
9.



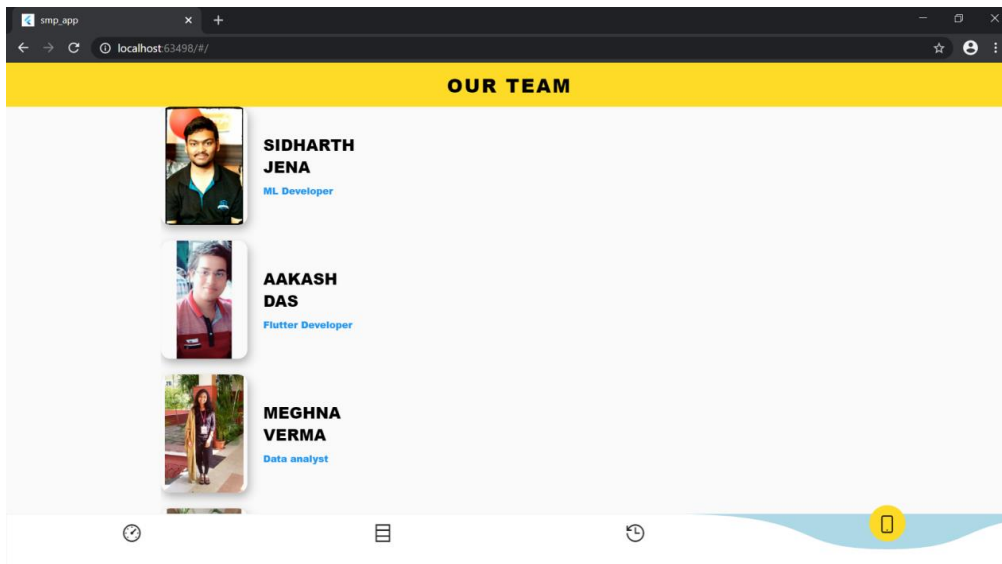
10.



11.



12.



9.3 Screenshot Of APK Size

```

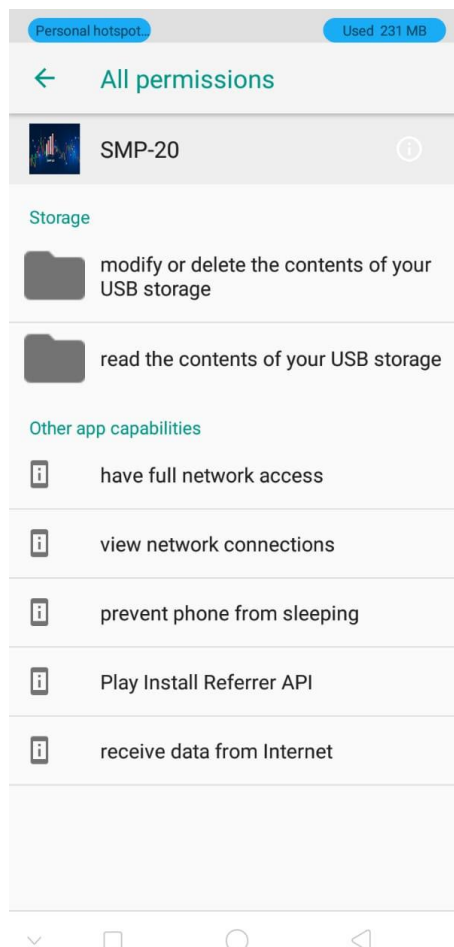
Command Prompt

C:\Users\KIIT\Desktop\college\smp_app>flutter build apk --no-shrink
You are building a fat APK that includes binaries for android-arm, android-arm64, android-x64.
If you are deploying the app to the Play Store, it's recommended to use app bundles or split the APK to reduce the APK size.
  To generate an app bundle, run:
    flutter build appbundle --target-platform android-arm,android-arm64,android-x64
    Learn more on: https://developer.android.com/guide/app-bundle
  To split the APKs per ABI, run:
    flutter build apk --target-platform android-arm,android-arm64,android-x64 --split-per-abi
    Learn more on: https://developer.android.com/studio/build/configure-apk-splits#configure-abi-split
Running Gradle task 'assembleRelease'...
Running Gradle task 'assembleRelease'... Done                    58.9s
✓ Built build\app\outputs\flutter-apk\app-release.apk (40.1MB).

C:\Users\KIIT\Desktop\college\smp_app>
C:\Users\KIIT\Desktop\college\smp_app>
C:\Users\KIIT\Desktop\college\smp_app>flutter build apk --release --split-per-abi --no-shrink
Running Gradle task 'assembleRelease'...
Running Gradle task 'assembleRelease'... Done                    19.2s
✓ Built build\app\outputs\flutter-apk\app-armeabi-v7a-release.apk (13.1MB).

```

9.4 App Permissions:



Chapter 10

Conclusion and Future Scope

10.1 Conclusion

The major issue with this project is the randomness of the stock prices. Many of the applications using machine learning did not take this into account and depended on the model to give them accurate predictions. Instead we used sentimental analysis to understand the randomness of stock prices in terms of people's sentiments towards the stock/company. Thus providing the historical analysis and sentimental analysis data to the user helped them to make a decision. The project does not give an absolute decision but instead leaves it up to the user to make the decision for himself/herself after showing them the stock's relevant data.

10.2 Future Scope

In the future, we expect to make the following changes-

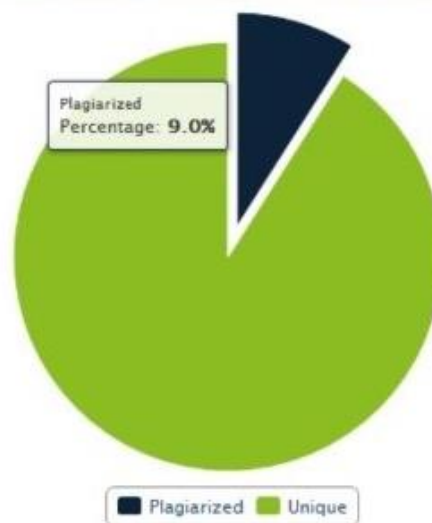
- Improving the accuracy of the LSTM model so as to get much more accurate predictions which are being presently limited due to the low computational power.
- Improving the sentimental analysis model by increasing the size of its dataset to make the analysis more accurate.
- Add more stocks to the list of stocks to provide users with more choices.

References

- [1] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink and J. Schmidhuber, "LSTM: A Search Space Odyssey," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222-2232, Oct. 2017, doi: 10.1109/TNNLS.2016.2582924.
- [2] K. Chen, Y. Zhou and F. Dai, "A LSTM-based method for stock returns prediction: A case study of China stock market," 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, 2015, pp. 2823-2824, doi: 10.1109/BigData.2015.7364089.
- [3] Frederick S.M. Herz, Lyle H. Ungar, M. Eisner and Walter Paul Labys, "Stock Market Prediction Using Natural Language Processing" Patent Id-US20130030981A1, dop- 2014.04.03
- [4] <https://aroussi.com/post/python-yahoo-finance>
- [5] github.com

PLAGIARISM REPORT

PlagiarismCheckerX Summary Report



Date	Saturday, May 30, 2020
Words	248 Plagiarized Words / Total 2844 Words
Sources	More than 26 Sources Identified.