School of Computer Science and Engineering

Review 1:slot:A2+TA2 **Course code: CSE 4015**

HUMAN COMPUTER INTERACTION

TITLE: " Stock market advisor/predictor using statistical analysis technique ,qualitative finance and news analysis"

TEAM MEMBERS:

18BCE0779	**SAPAN KUMARSHRIVASTAVA**
18BCE0776	**ANURAGKASHYAP**
18BCE0789	**BHAVESHSAHU**
18BCE0261	**SAYAK MAJI**

..

TABLE OF CONTENTS

TOPIC	Page number
Abstract	3
Introduction	3-5
Related Works	5-7
Real life Applicability	7
Individual Contribution	8
Tools and Technology Used	8-9
Proposed system process flow	10-14
Working methodology	14-15
Implementation Result and UI	16-23
Neilson 10 point heuristics	23-28
Comparative analysis with other existing technologies.	29-30
Conclusion and Future scope	30-31
References	31
Appendix	32

ABSTRACT:

It is rightly said that money makes money. But it's rate of multiplication depends completely on an individual's investments. Everyone wants their money to grow, but due to lack of knowledge many fail to do so. We want to help people know which companies are suitable for buying stocks and which is suitable for selling. This can be achieved via proper prediction of stock prices of companies upto a certain period of time. We will try to use data from twitter and other news headlines for sentiment analysis and use user's reviews about a company to predict the stock prices. Apart from that we will also use metrics like moving average, fibonacci retracement method improvise our stock price predicting ability. The predicted prices will be visually represented to the user via graphs, charts, etc. For past data about companies we will use standard datasets from their websites to study a trend of rise and fall of stock prices. We will also use semantic analysis on latest news that we will extract from various sites about a particular company. Stock prediction and analysis has been a hot topic under research since many years. Even now very few companies are able to do it with considerable accuracy. Our attempt to this involves sentiment analysis of users from various social networks as well as news headlines and standard datasets from official company websites.

INTRODUCTION:

A share price can be defined as the cost a single share has out of a number of saleable stocks of some particular company, derivative or other financial asset. In simple words, the price of a stock is the highest amount someone will voluntarily pay to buy it, or the lowest amount for which it can be bought. Supply and demand greatly affects the appeal and thus the value and price of a stock.

Although it might appear that there are other factors at play, such as the health of the economy and company earnings, these are merely just drivers of supply and demand. In other words, even when a stock is undervalued, its upto the market what the real price of stock is. Thus there are dynamic fluctuations of prices of stocks and this refers to market risks.

In case of limited supply, if more buyers enter the market, the price of shares go up. When supply and demands are just about equal, the prices of stocks move about a narrow range of fluctuations. Thus, there is a need to know the proper time to buy as well as sell stocks to get maximum profit. It has been observed that many companies show a gradual increase in their stock prices despite local

rise and fall of stocks. Many of these data can be obtained from nifty index of companies.

There are graphs we are going to use for analysis:-

line graph

We plan on making a dynamic line graph to display a company's predicted stock prices over a period.

candlestick Chart

This indicator indicates that the market is open, high, low, and close to the day. The candlelight has a wider part, called the "real body."

This real body represents the price range between the opening and closing of a trade for that day. When the real body is full inside or dark, it means the closure was lower than the opening. If the real body was empty, it meant that the closure was higher than the opening.

Following are indicators we are going to use for our prediction:-

a. Moving Average

It is a technical analysis indicator that constantly provide us average. It is calculated over a specific period of time, for example ten days, twenty minutes, 32 weeks or any time period the trader chooses.

a. On balance volume

OBV refers to a simple indicator using volume and price to measure the intensity of buying and selling. Purchase pressure is seen when the positive volume exceeds the negative volume and the OBV line rises. Sales pressure exists when a negative volume exceeds the appropriate volume and the OBV line falls.

b. Boiler Band

It is a technical indicator developed by John Bollinger, used to measure market volatility and to identify "extended" or "additional" conditions. Basically, this little tool tells us whether the market is quiet or the market is a highly active.

c. Fibonacci Retracement Levels

These are horizontal lines indicating areas where support and resistance can or are likely to occur. They are based on Fibonacci numbers. Each level is associated with a percentage. Percentage is how much previous movement price will be retraced.

Live Feed News

The most critical aspect of intraday trading is short-term trading and the immediate effects of the market on security price movements.

In order to take advantage of this opportunity, intraday marketers must be aware of troubling issues and obtain these real stories so that they can predict or calculate the outcome of the news at security prices and take action or position appropriately.

A trader who receives and holds good news at the right time has a high hand. Stories can be used directly, or they can be used to understand the insights that cause stories and can also affect other bonds and movements as well.

Website

Used HTML5, CSS3, Javascript and various API's to get data and it's representation.

Target Audience:

Common people who wants to grow their money with stable growth and low risk. Also many people who want to choose stock market as their career, but are reluctant to do so due to the lack of knowledge and platform.

RELATED WORKS:

1. Machine learning to predict stock market indicators on the basis of historical data and Twitter sentiment analysis data by Alexander Porshnev, Ilya Redkin, Alexey Shevchenko:-

Main concept behind this project is analysis of reviews and comments about certain company and predicting their stock. It includes following steps:-

Sentiment analysis:-

In this paper the author applies various supervised learning models (logistic regression, Naive Bayes algorithm, SVM, etc.) on the data obtained from news headlines and find a relation between stock price movements and sentiments in tweets. It was revealed that the highest accuracies were provided by Naive Bayes and SVM algorithms each around 69.4%.

Prediction of the growth of stock market:-

To test the main hypothesis two machine learning algorithms are used which allow us to do classification by appearance of events and use created model for prediction. They are Neural Networks and Support Vector Machine.
Source:- https://ieeexplore.ieee.org/document/6753954

2. Machine LearningTechniques and Usage of Event Information for Stock Market Prediction: A Survey and Evaluation by P.D. Yoo , M.H. Kim , T. Jan.

i. In this paper, various ML techniques were implemented for stock market prediction.
ii. Comparison of linear regression, multivariate regression, neural networks, SVM and Case Based Reasoning models showed that Neural Networks offer ability to predict market directions more accurately compared to others.
iii. It was also found that incorporating event information with prediction model plays a very important role for more accurate prediction.

Source:-https://ieeexplore.ieee.org/document/1631572

3. Stock Price Prediction Using Financial News Articles by M.I Yasef Kaya and M. ElefKarshgil

i. In this paper, analysis of correlation between contents of financial news articles and the stock prices is done.
ii. Articles from news were labeled positive and negative depending on their effect on stock market.
iii. Here, the authors defined features to b analysed as woed couples rather than words. A word couple comprised of combination of noun and verb. Analysis of articles was done by SVM classifier.
iv. Results showed that evaluation of word couples showed higher precision and recall compared to evaluation of individual words.

Source:-https://www.semanticscholar.org/paper/Stock-price-prediction-using-financial-news-Kaya-Karsligil/043821300b4800a87f684b27fc7038eaed968df9

4. Forecasting of Stock Market Indices Using Artificial Neural Network by Dr. Jay Joshi, Nisarg A Joshi

i. In this paper, the authors have devised an approach to predict the S&P CNX Nifty 50 Index.

ii. Artificial Neural Network(ANN) based model has been used to predict the direction of moving pattern of closing value of the index by prediction of the moving average value change after seven days.

iii. Performance of neural network was compared with performances of random walk and linear autoregressive models. It was found that Neural Network outperforms linear autoregressive and random walk models by all metrics. Average accuracy of the model was found to be 62%.

Source:-
https://www.researchgate.net/publication/254965043_Forecasting_of_Stock_Market_Indices_Using_Artificial_Neural_Network

REAL LIFE APPLICABILITY:

- Many Indians are reluctant in investing in stock market because of the lack of knowledge and uncertainty in the trends of the market. There is lack of tools that helps in understanding the market.
- Presently only 2% of Indians spend in stock stock market. But the numbers are increasing enormously.
- This gave us the motivation to develop a web app which can guide common people who doesn't have any idea of investing in share market and help them grow their investments profitably by suggesting the good stocks . In this way people can easily invest and be profitable.

INDIVIDUAL CONTRIBUTION:

BHAVESH SAHU: Write Machine Learning code and prepare algorithm for LTSM technique. Work on front end specifically on charts and getting data from websites using API's.

ANURAG KASHYAP: I will work on most of the charts, moving average, live news feed for analyzing market sentiment and on the machine learning models for predicting recommendations. Inthe UI/UX part, I worked on wireframing the UI and creating Graphical models of data for the users.

SAYAK MAJI: Research of various pre-existing models of stock prediction. I will work on the overall UI/UX of our application. Also I will work on scrapping of various headlines and user reviews of stocks of various companies for the dataset of our model.

SAPAN SHRIVASTAVA: Most of the front- end part. Working on the wire frame of the UI, involvement in Contributing live news for market analysis. Collecting user reviews regarding user interface and making changes in app, for business and marketing aspects.

TOOLS AND TECHNOLOGIES USED:

We have used the following technologies to help the users in assessing the stocks and for suggesting the stocks:

LSTM

Since the stock prices mainly depend on the previous stock prices Long Short Term Memory Networks (LSTM) can be used to find long term dependencies in the stock prices. It also remove unnecessary values. Inspired by YoshuaBengio deep learning architectures [5] we have included their one model layer in our code.

Bollinger Band

The Bollinger Band technology display has bands that usually set the standard deviation off a simple moving average. In general, going to a higher level suggests that the goods have been overused, while proximity to the lower band suggests that the goods are being repossessed.[6] Foundations and trends Since the general deviation is used as a statistical measure of volatility, this indicator also adapts to market conditions.

Moving average

The moving average, is calculated by taking the arithmetic definition of a given set of values. In other words, a set of numbers , prices in case of financial instruments are added together and divided by the price values in the set. The formula to calculate is:

$$SMA = \frac{A_1 + A_2 + \ldots + A_n}{n}$$

where:

$A =$ average in period n

$n =$ number of time periods

Standard deviation

Standard deviation is measurement of financial performance that, when applied at the annual rate of return on investment, highlights the historical instability of that investment. The greater the general deviation from the security, the greater the variance between each price and the description, which indicates the range of the maximum values. For example, a volatile stock has a normal deviation, and the deviation of a stable blue stock is usually very small.

The Formula for Standard Deviation

$$\text{Standard Deviation} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$

where:

$x_i =$ Value of the i^{th} point in the data set

$\overline{x} =$ The mean value of the data set

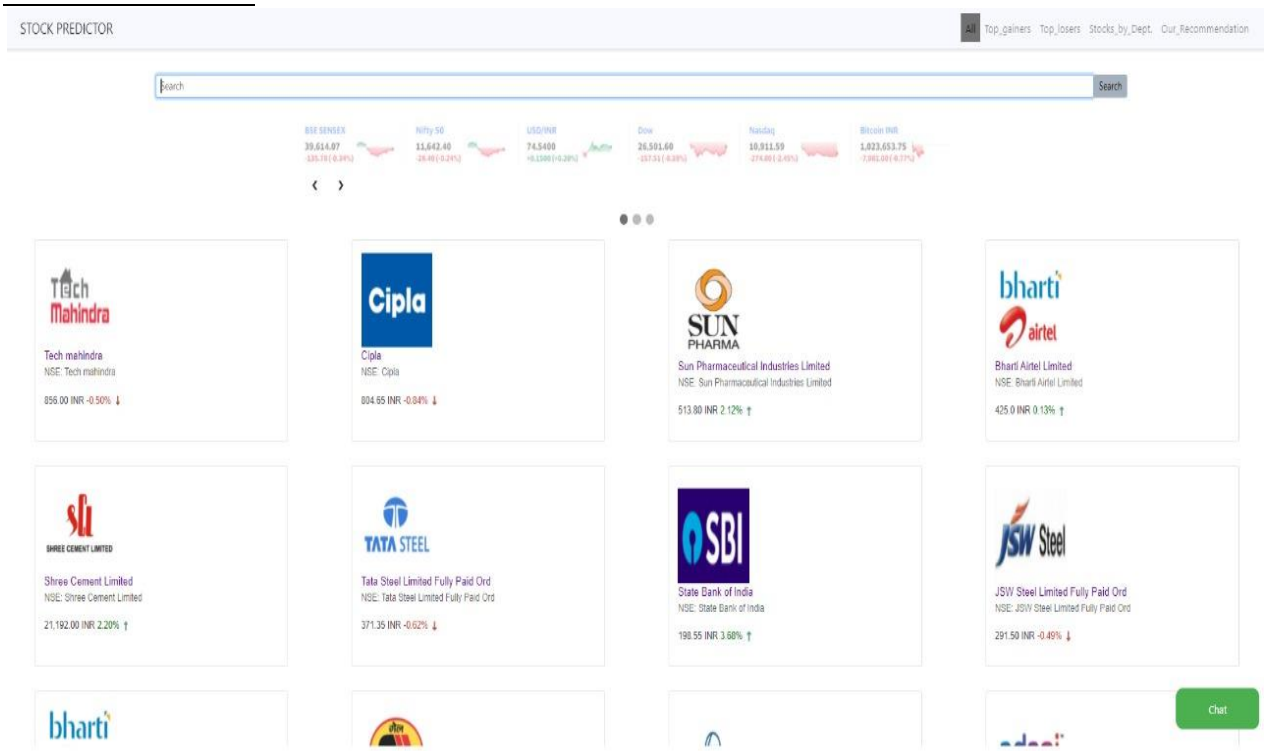
$n =$ The number of data points in the data set

News sentiment analysis

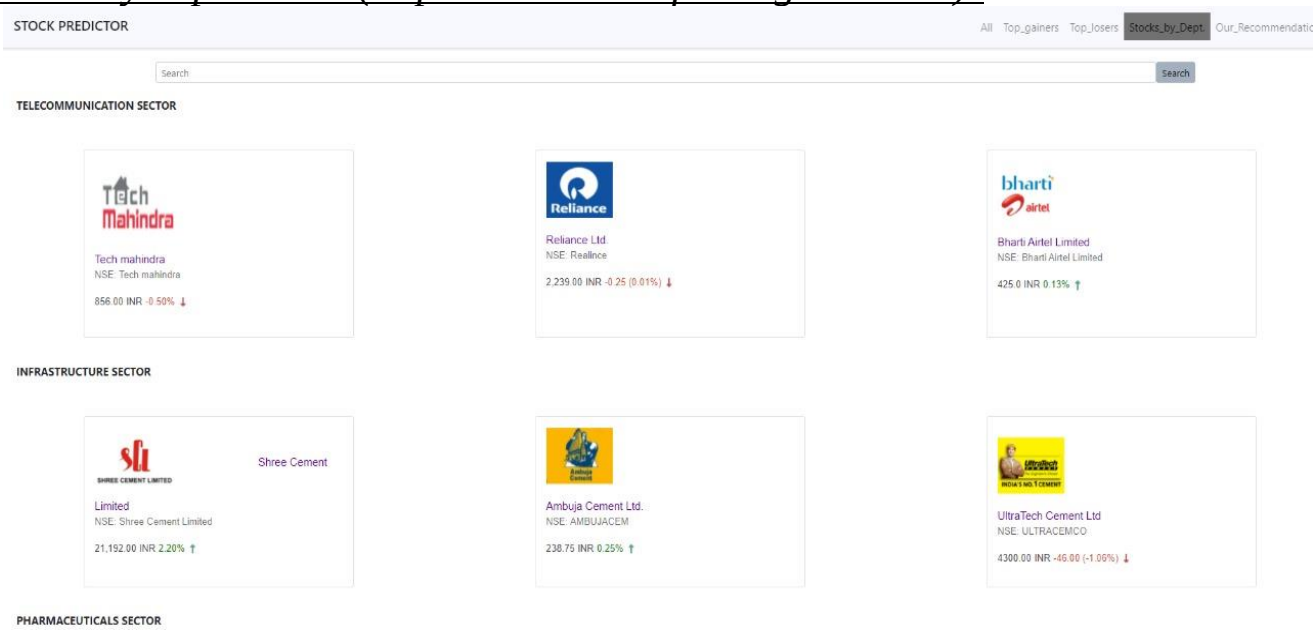
By doing sentiment analysis we know what market is thinking about a particular company. We will do it on some finance site like moneycontrol or yahoo finance. We have decided to give 21 percent weightage to semantic analysis outcome.

PROPOSED SYSTEM PROCESS FLOW:

HOME PAGE:

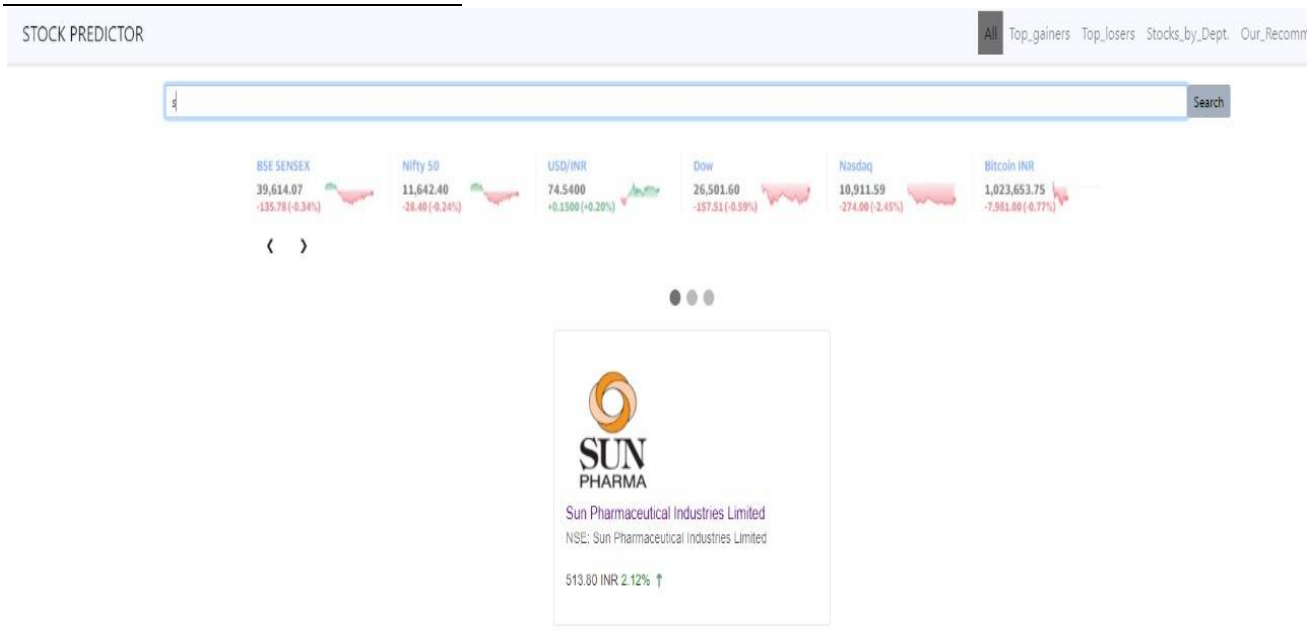


Stock by department (Implementation of Navigation bar):



..

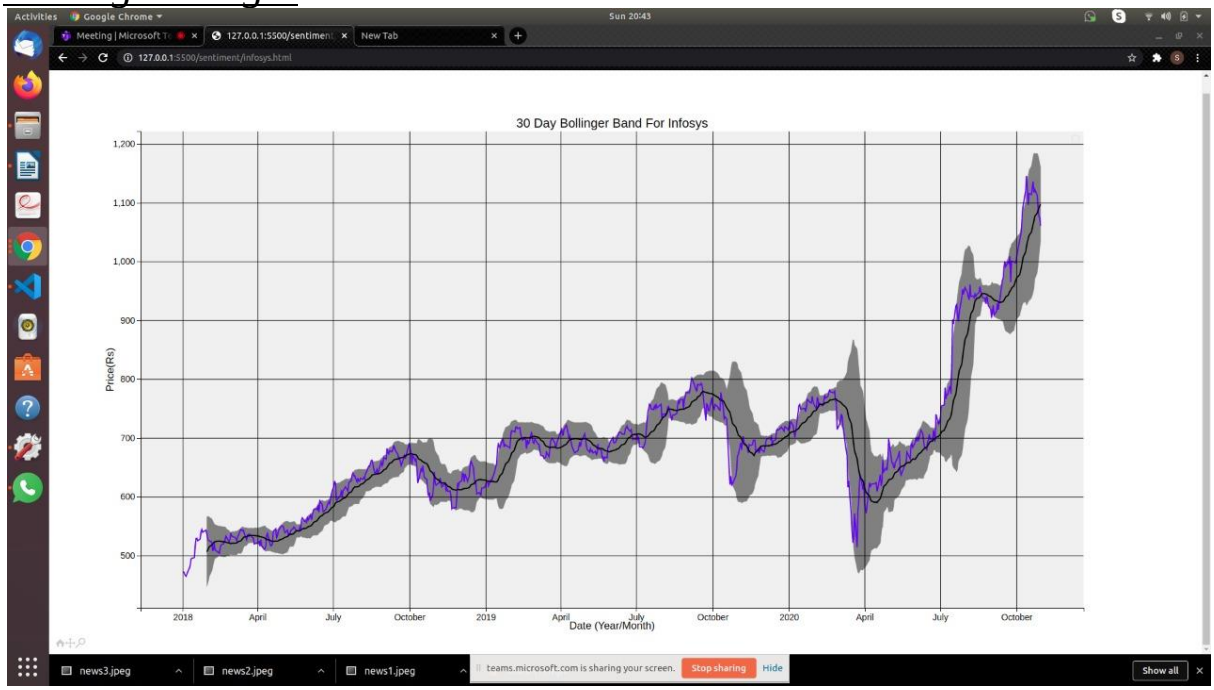
Search bar demonstration:



Opening a company card (ex : CIPLA) :

Moving through the 4 buttons:

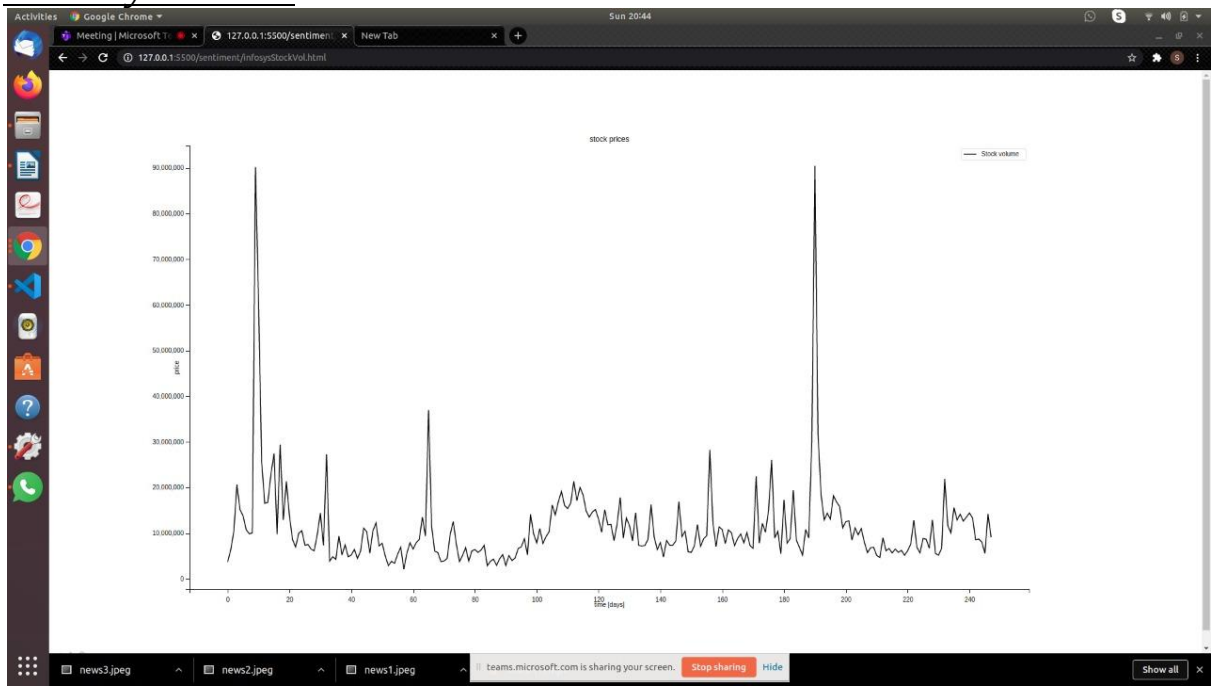
Moving Average:



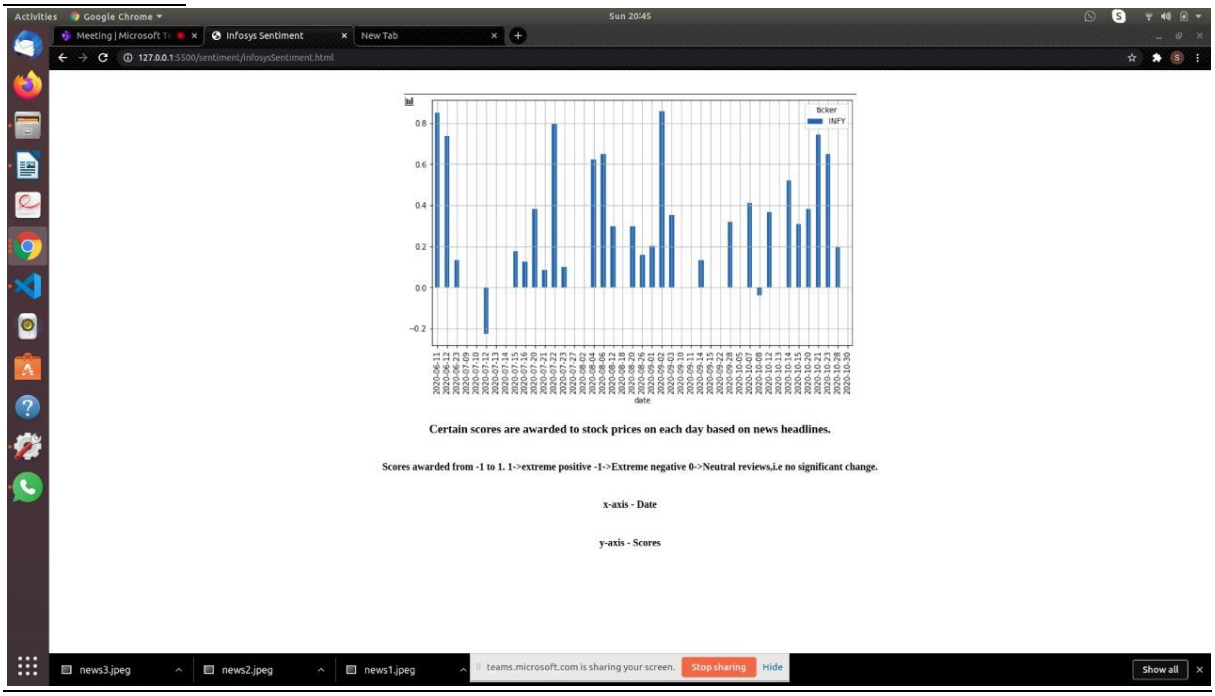
LSTM Model:

..

Stock by volume:



Sentiment:



Certain scores are awarded to stock prices on each day based on news headlines.

Scores awarded from -1 to 1. 1->extreme positive -1->Extreme negative 0->Neutral reviews,i.e no significant change.

x-axis - Date

y-axis - Scores

WORKING METHODOLOGY:

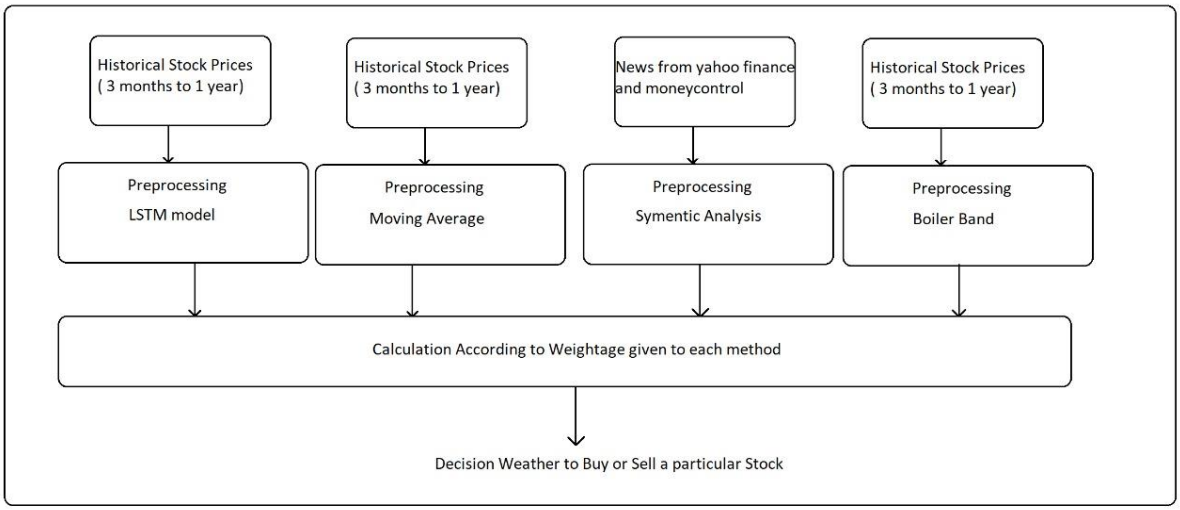


Fig 1.0: Architecture of project

The methodology of our project is simple yet efficient. Our web will not only enable the user to assess a particular stock but it will also suggest them. As mentioned in the tools and technologies section, we are going to use the features of LSTM, Moving average, Bollinger Band and news analysis to provide suggestion to the users. Out of 100 percent weightage we will give 33 percent weightage to LSTM model, 23 percent weightage to moving average, 23 percent weightage boiler band, 21 percent weightage to semantic analysis.
By doing this we will we able to take decision which is calculated based on various dimension of stock market, thus predicting greater number of right decisions.

IMPLEMENTATION RESULT AND UI:

Frontend

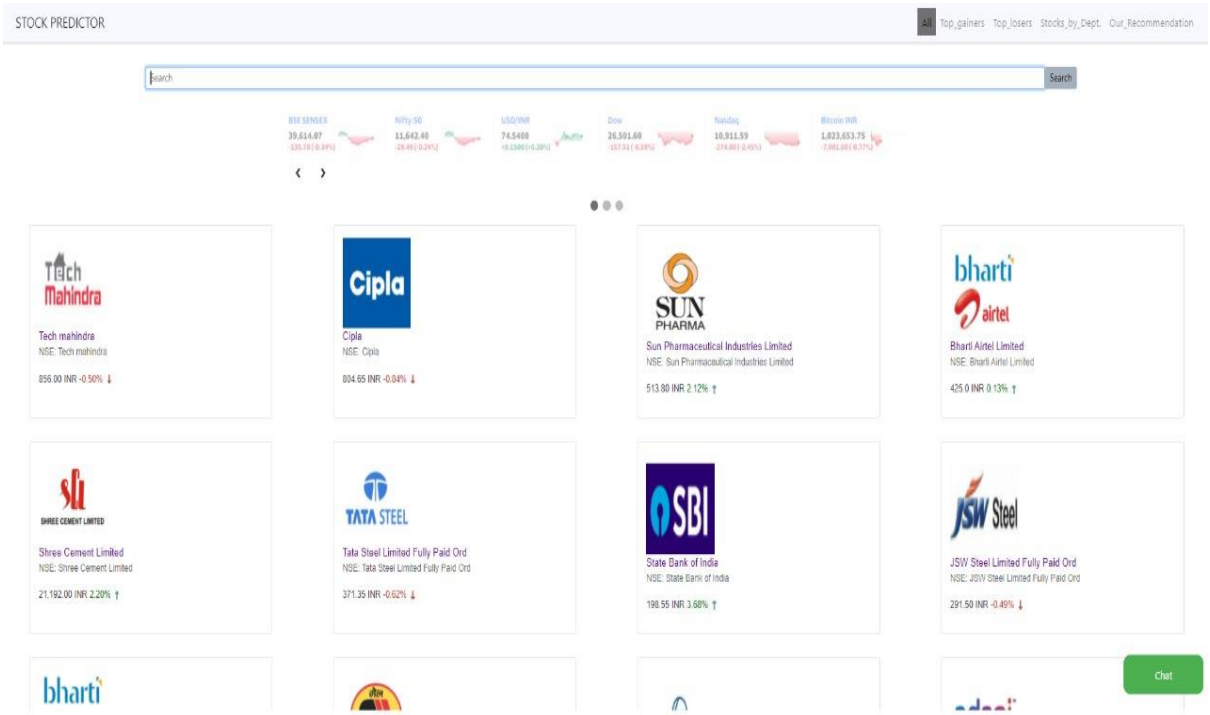


Fig 3.1 *Represents Front page of our stock advisory website. It contains many stocks that one can study and make a decision to buy or sell*

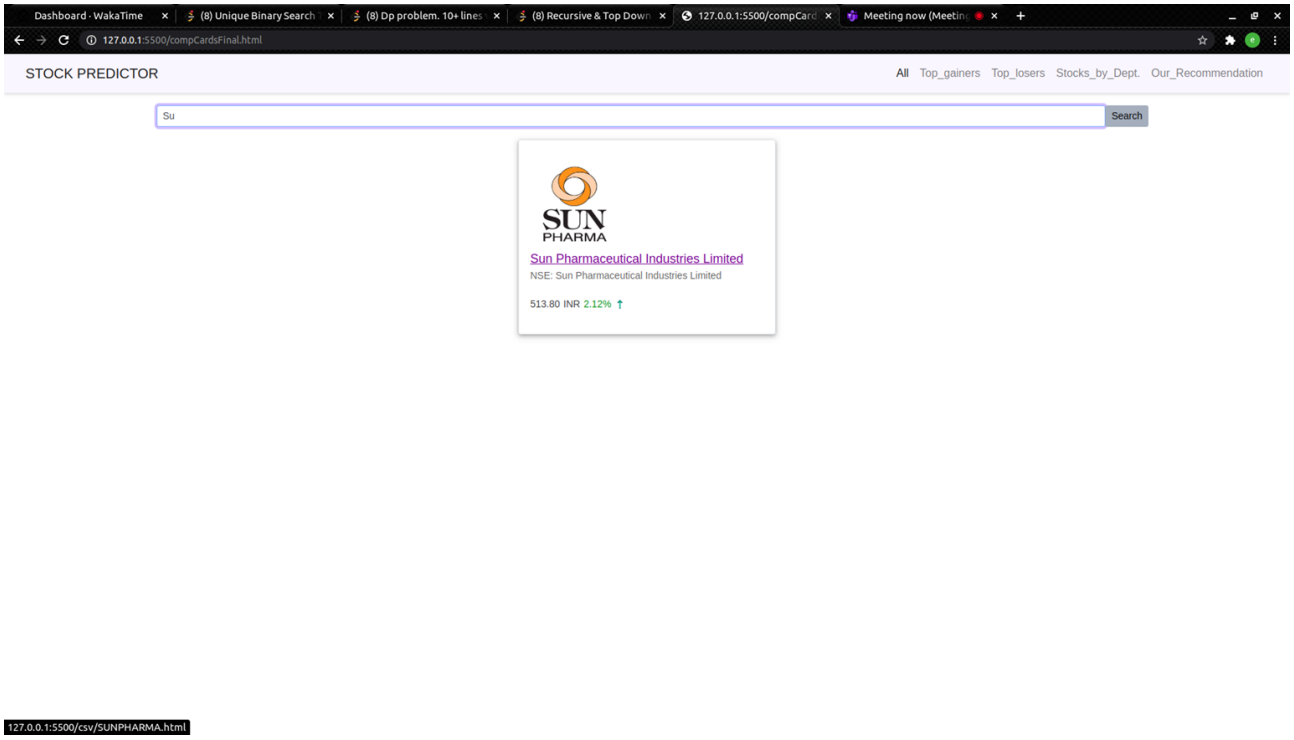


Fig 3.2 Represents search option of our website that can be used to search companies

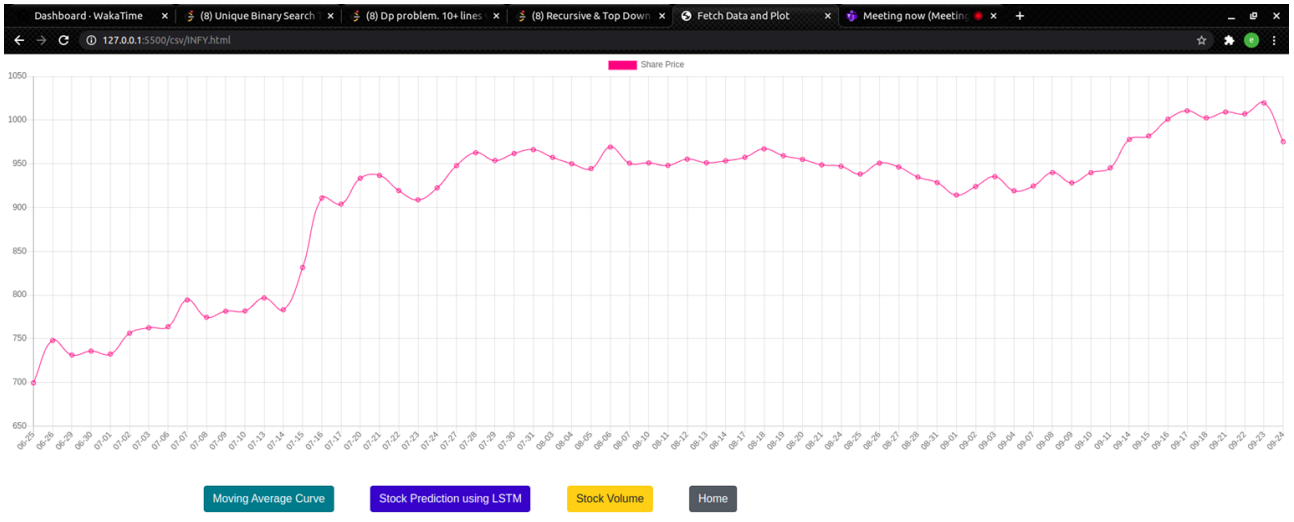


Fig 3.2 Price vs Days

Fig 3.2 In the above graph x-axis represents number of days y-axis represents stock price of that day

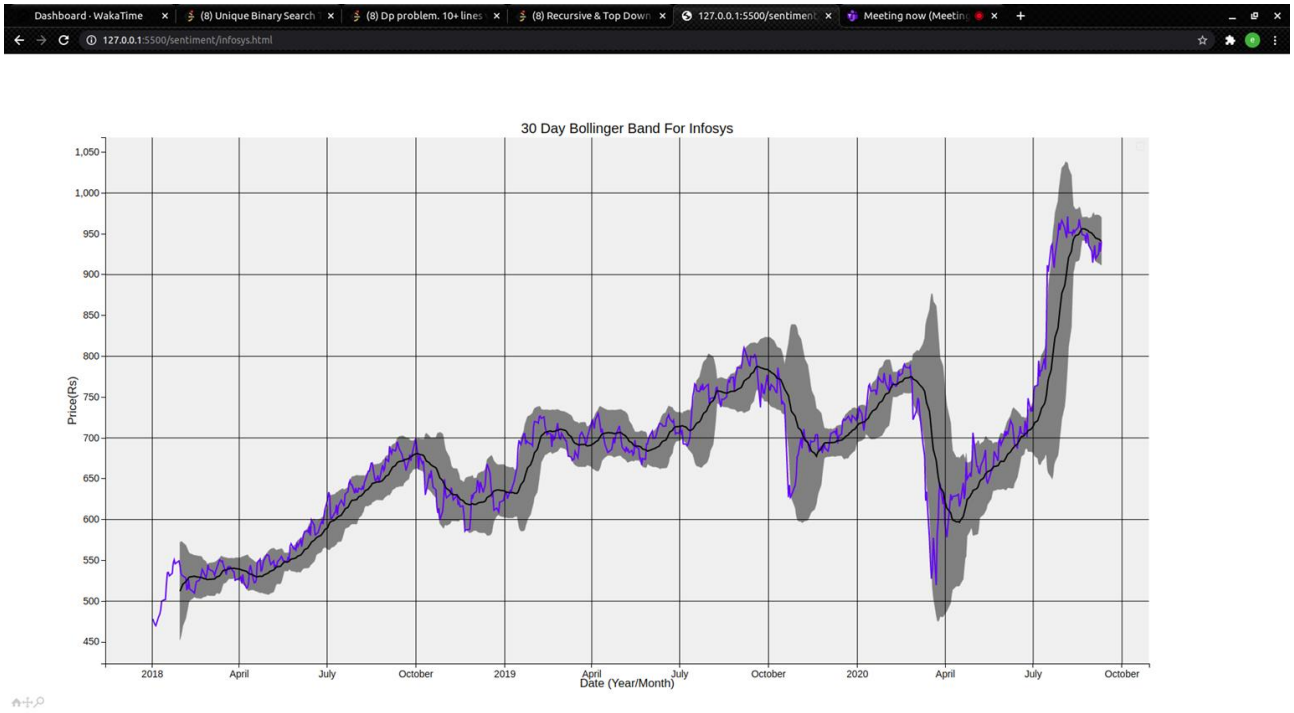


Fig 3.3 Moving average and Bollinger Band

In the above graph x-axis represents months y-axis represents stock price of that month. The shaded portion represents the standard deviation from mean stock price If blue line goes above shade, time to sell, if blue line goes below shade, time to buy.

Fig 3.4 LSTM algorithm

..

In the above graph x-axis represents days starting from 3 months earlier y-axis represents stock price of that day. Blue line represents predicted price according to LSTM model. Red line represents original stock price

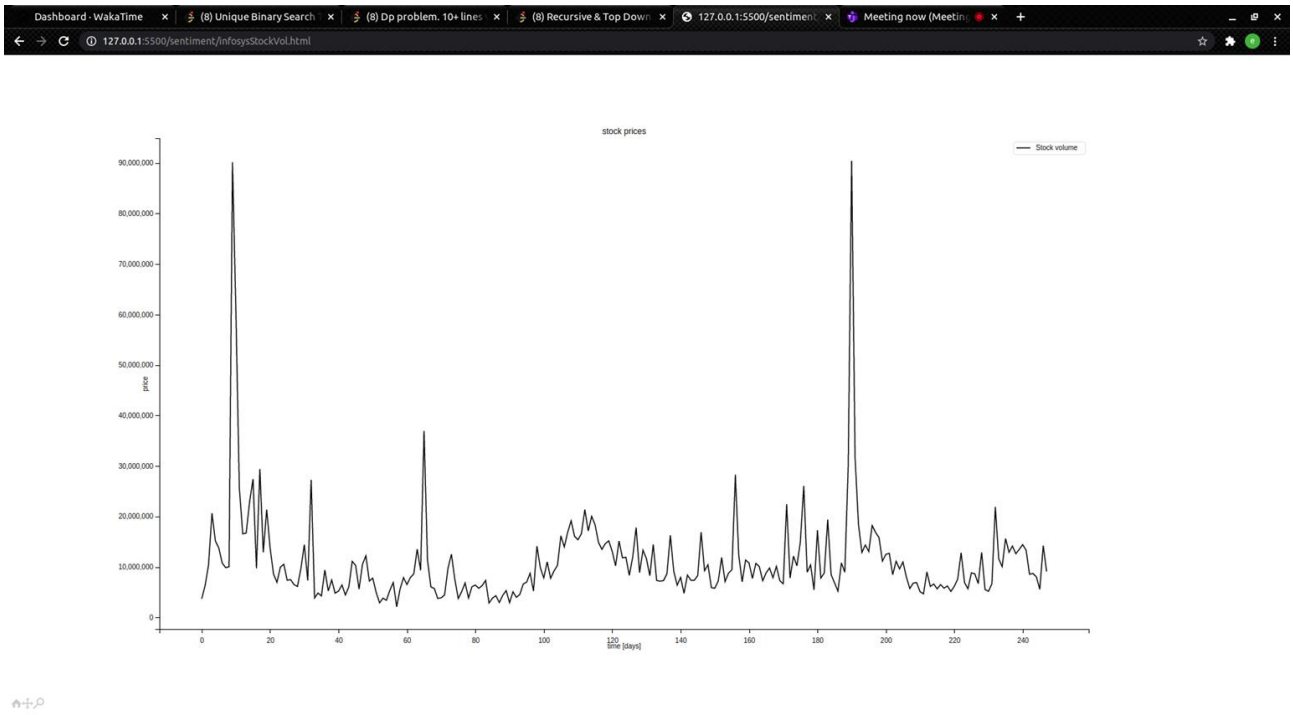


Fig 3.5 Stock Volume

In the above graph x- represents days starting from 3 months earlier y-axis represents stock volume (number of buyers and sellers on that day) of that day

Code:-

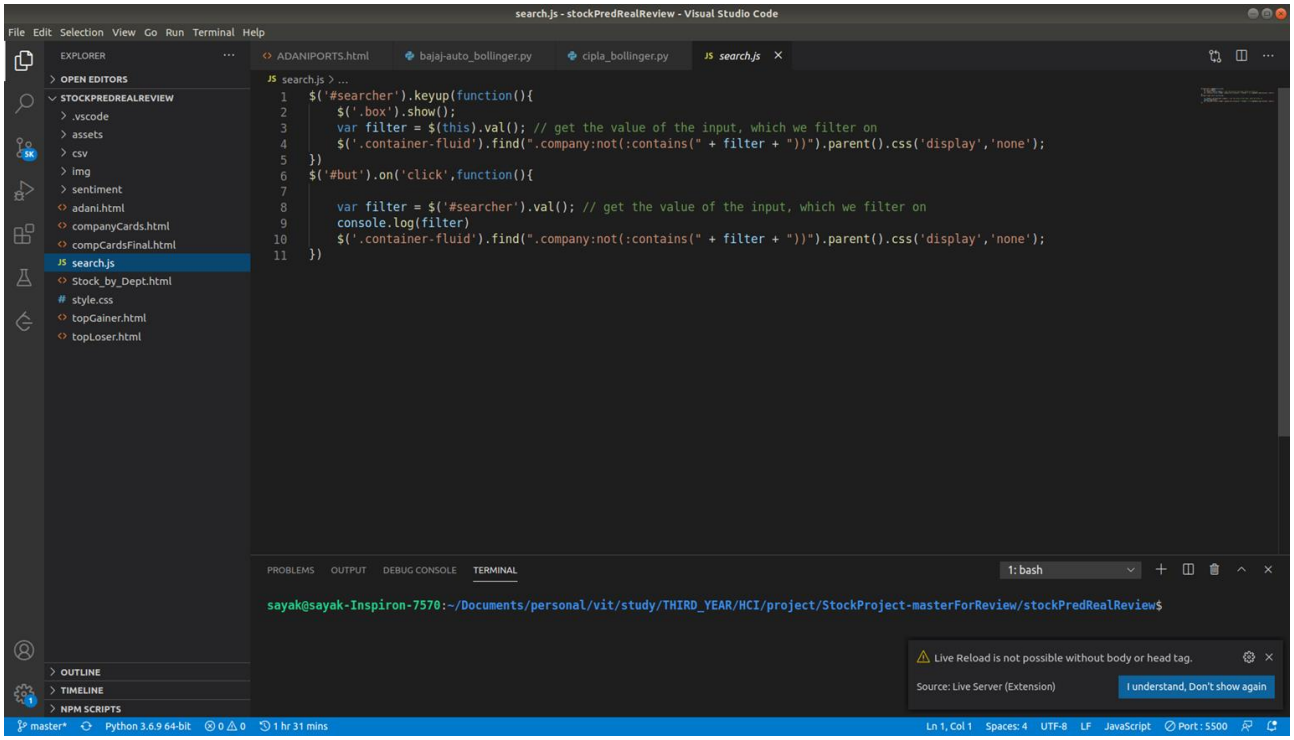


Fig 4.1 Search feature

The above code is the implementation of search option of stocks

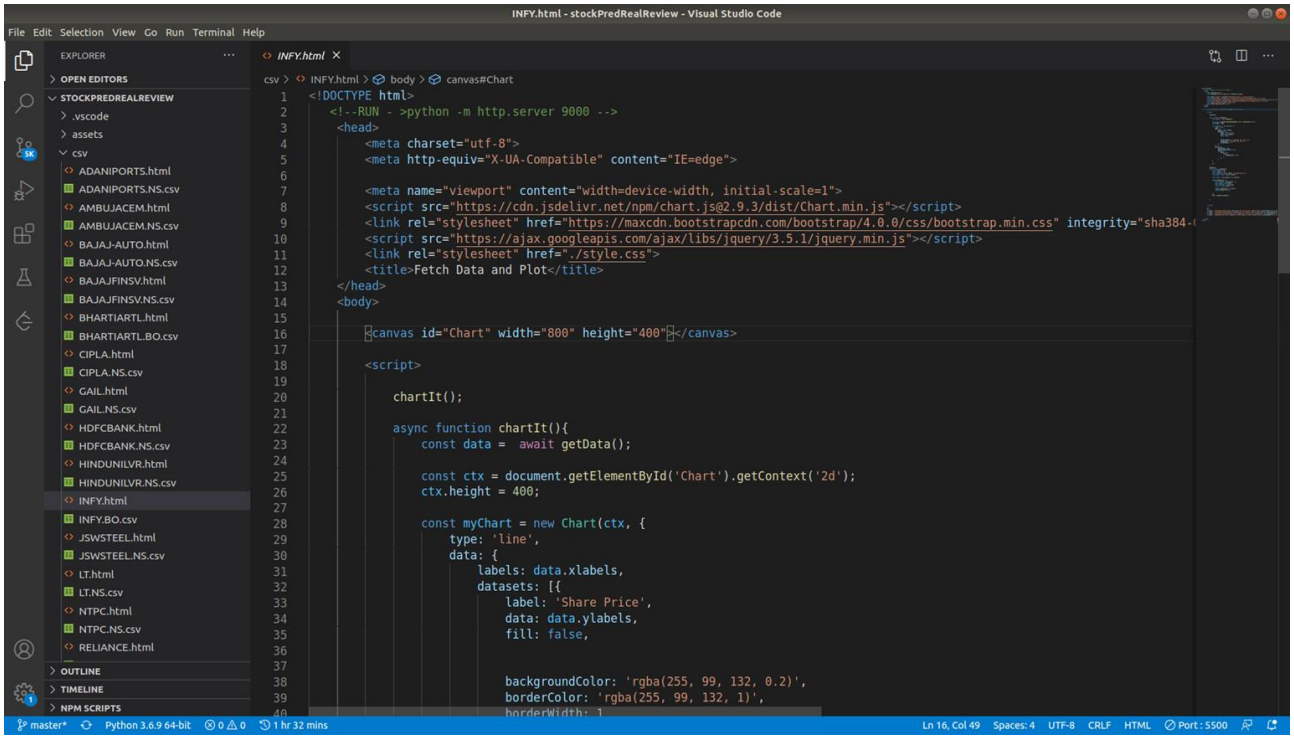


Fig 4.2 *represents the code for chart function which charts stock price vs days*

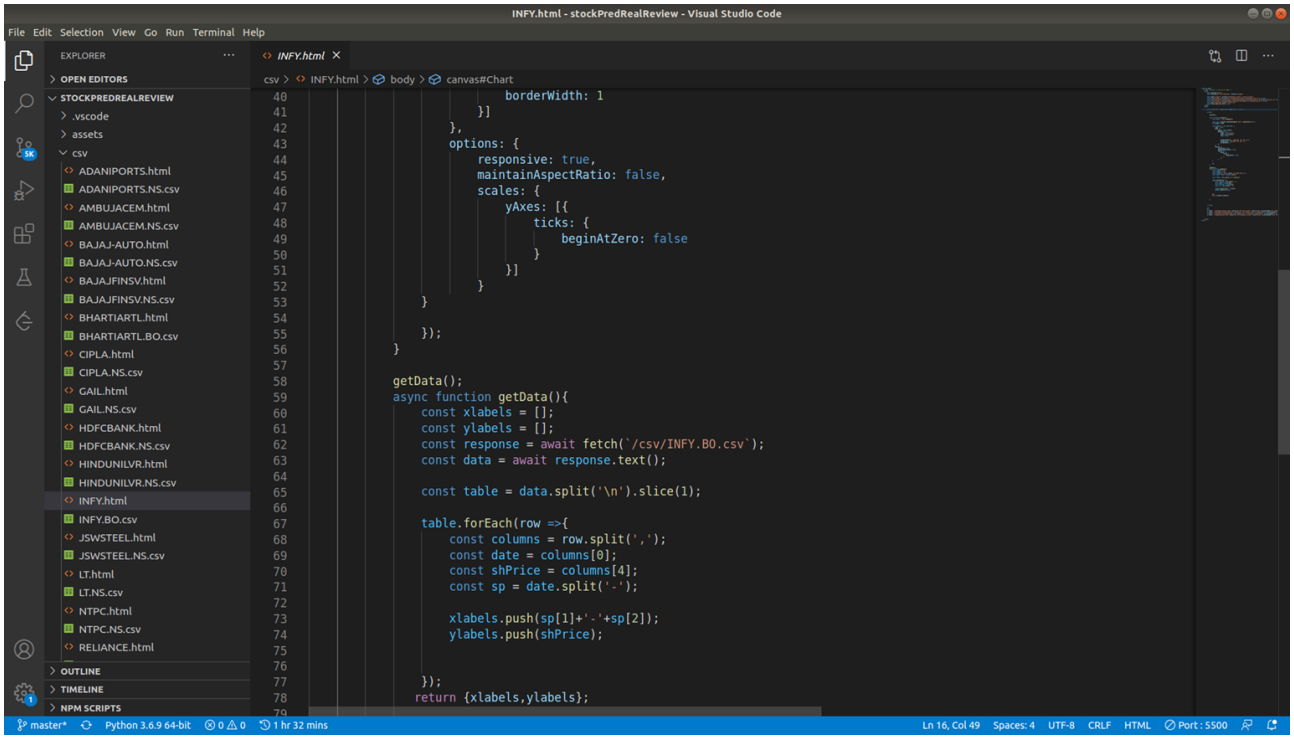


Fig 4.3 *represents the code for chart function which charts stock price vs days*

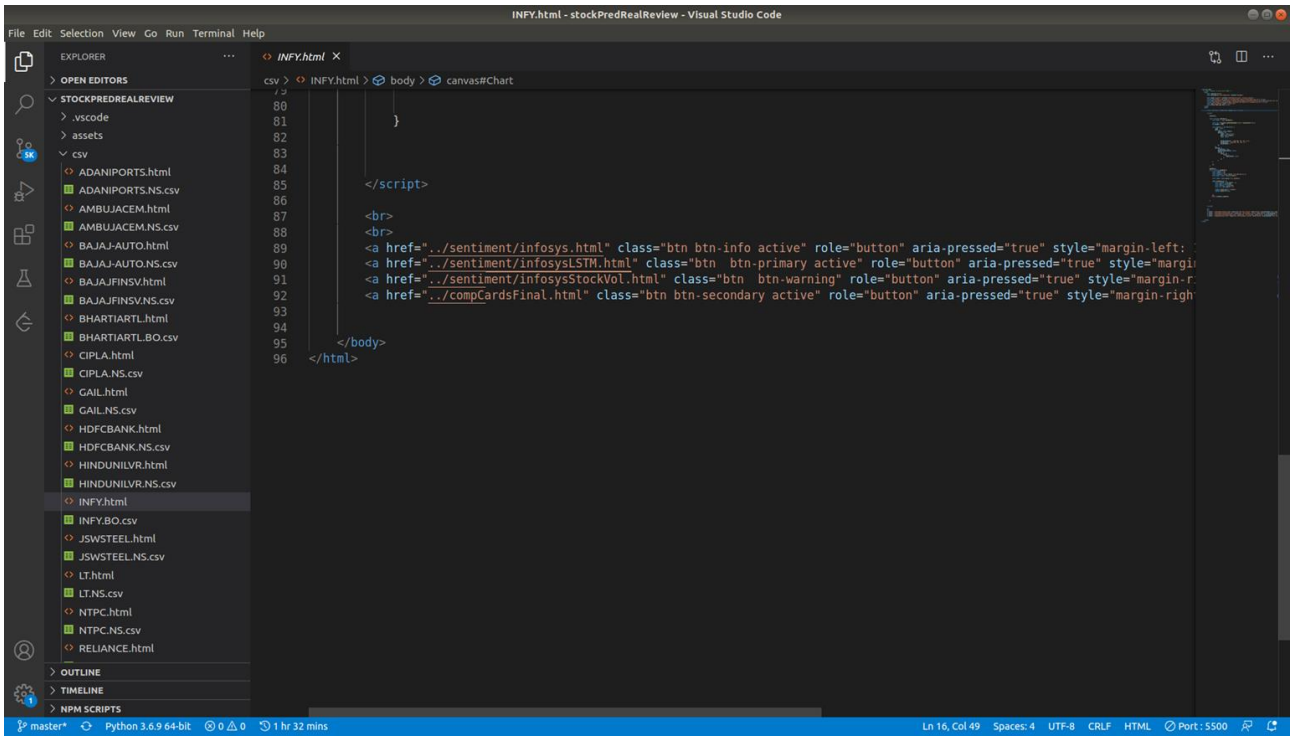


Fig 4.2 *represents the code for chart function which charts stock price vs days*

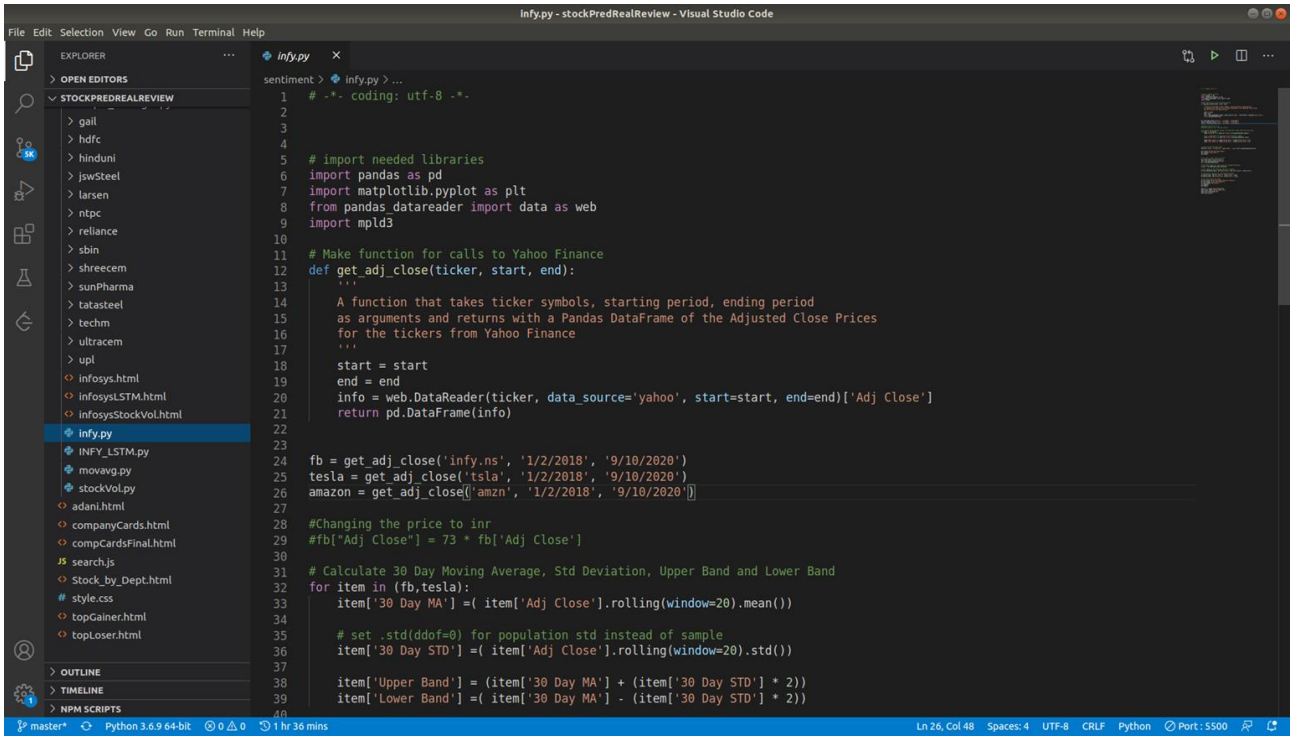


Fig 4.2 *The code is for Moving average and bollinger bands*

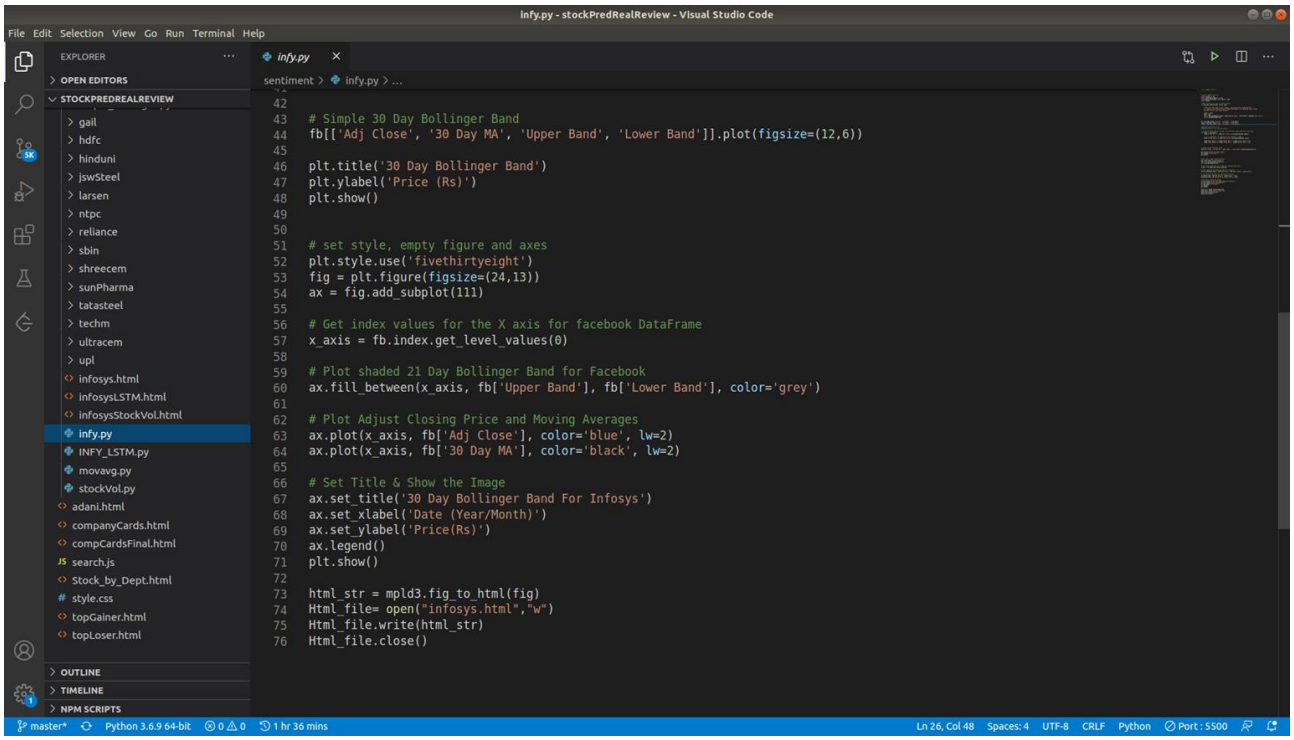


Fig 4.2 *represents the code for chart function which charts stock price vs days*

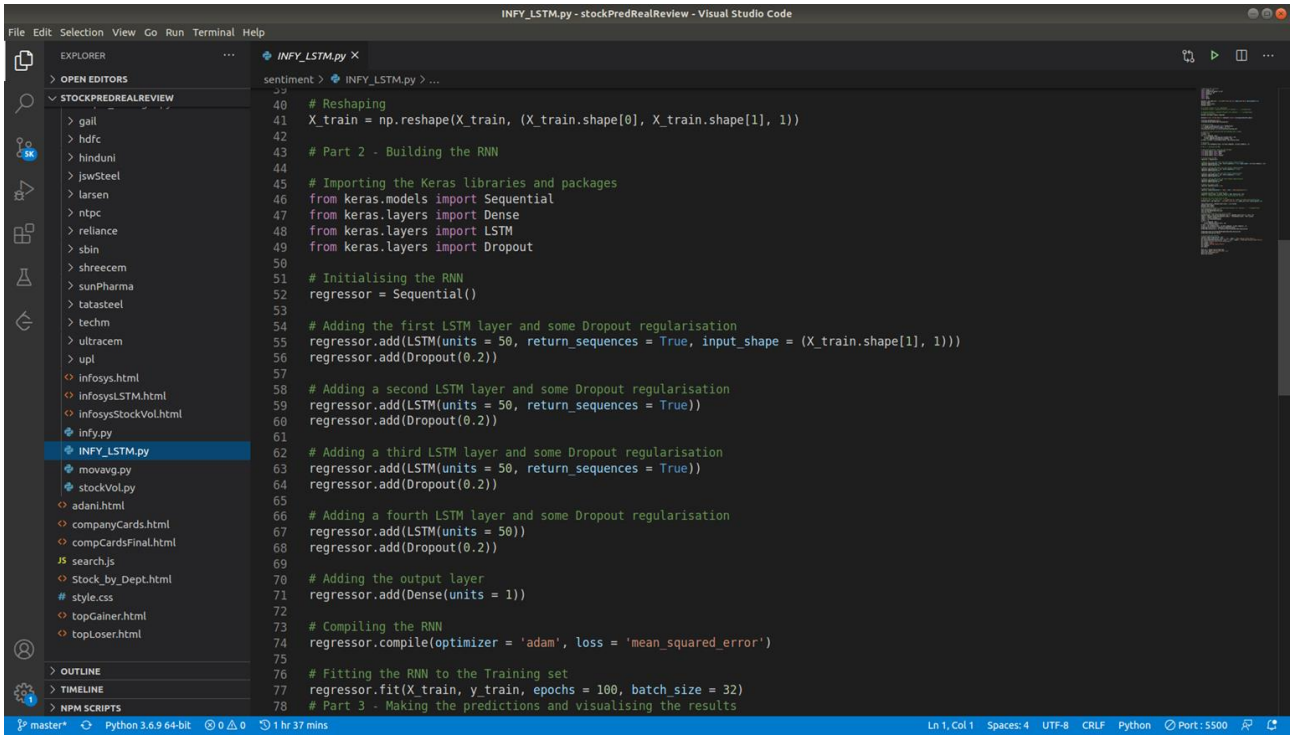
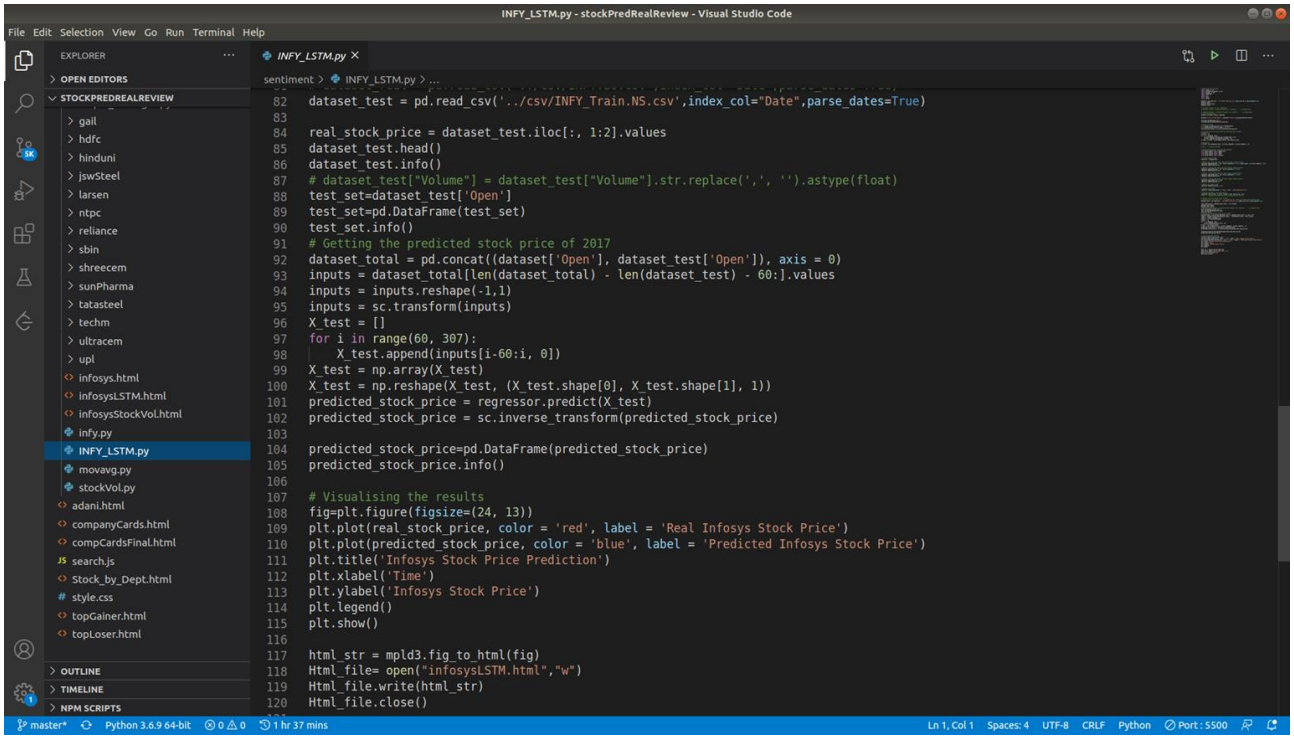


Fig 5.1 *Represents the code for LSTM Model*



..

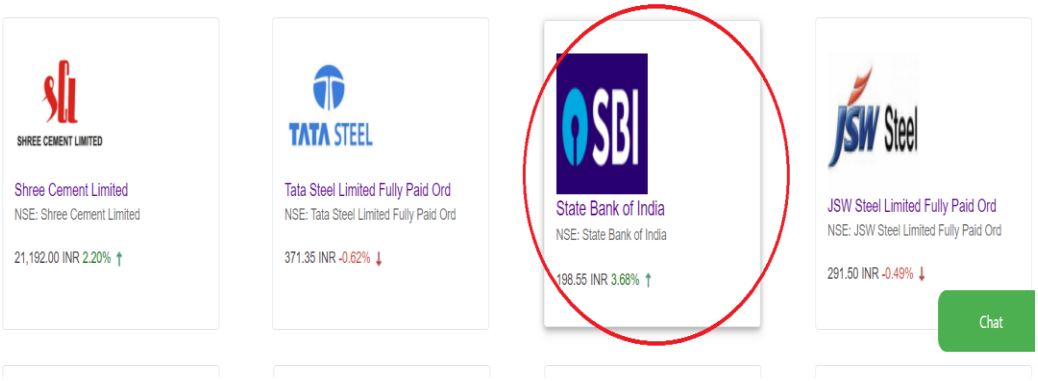
Fig 5.2 LSTM Algorithm for Stock price detection

Fig 5.2 *Represents the code for LSTM Model*

INTERFACES VALIDATION WITH NIELSEN'S 10 POINT HEURISTICS:

1) **Visibility of system status.**

a) The website features the glow and expand effect when , the cursor points on the individual boxes.



b)

Our website also features different types of dynamic graph(Real time movement of graph) for the prediction, of either buying or selling the graph.

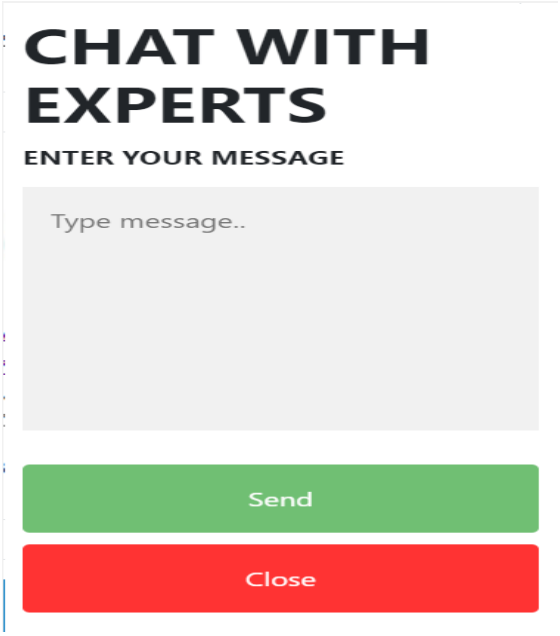
2)Match between the system and real world.

a) The stocks data of different company, uses real time data from websites such as yahoo finance , to show user the exact and current value of stock prices to the user .

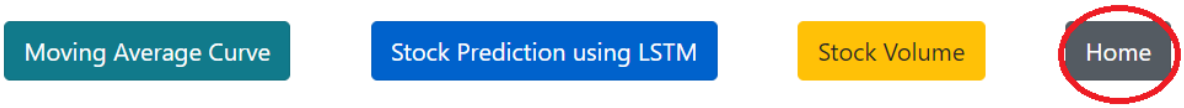
b) The stock graph page provides user a moving average button , through which user can have a general idea about stock trend from a certain period of time.

3) <u>User control and Freedom</u>

a) for the new user , which is almost new to stocks chat with experts option is a boon. This will make user to have a sense of freedom and control.

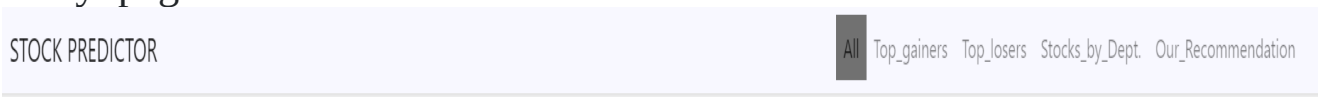


b) The home button in the page of graph for stock status.(like an emergency exit)



4)<u>Consistency and standards.</u>

a) The navigation bar provides a sense of consistency to the user , on every page of the website.



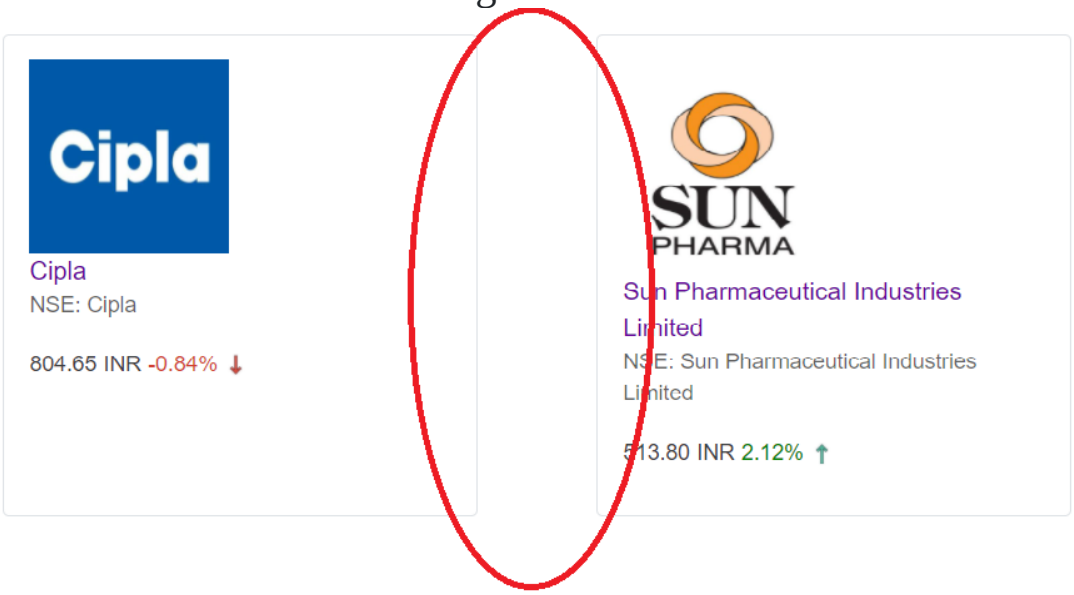
b) In addition to that the website contains the different company containers in same fashion , on different tabs.

5) <u>Error prevention</u>

a) For minimizing error in selecting the company by the users , we have added the option of glow and expand , so that the user can ensure his desired company is selected.

..

b) the same effect have been applied to the switch tab buttons and navigation bar button , to avoid errors in selecting them.
c) we have provided decent amount of space between the company's container to avoid clustering.



6) RECOGNITION RATHER THAN RECALL

a) We have put the icons for each and every container in the boxes , so that many users can directly jump to the desired company.
b) Also we have added the stocks_by_dept in the navigation bar so that that the user can search the companies on the basis on the sectors. So that he can chose to buy or sell the stocks sector wise.

7) Flexibility and efficiency of use.

The **demo chat box** ,**search box** and the features such as **search_by_dept** increases the flexibility of the website , and helps the immature users to understand the functions of website in a better fashion.

8) <u>Aesthetic and minimalistic design.</u>

a) In Our project , to achieve the minimalistic design and to beautify the frontend we have tried to keep the pages of website cluster free and have maintained proper spaces.

b) Also the color of the website have a contrast feature, in order to give a good and irritation free view.

9) <u>Help users recognize, diagnose and recoverfrom errors.</u>

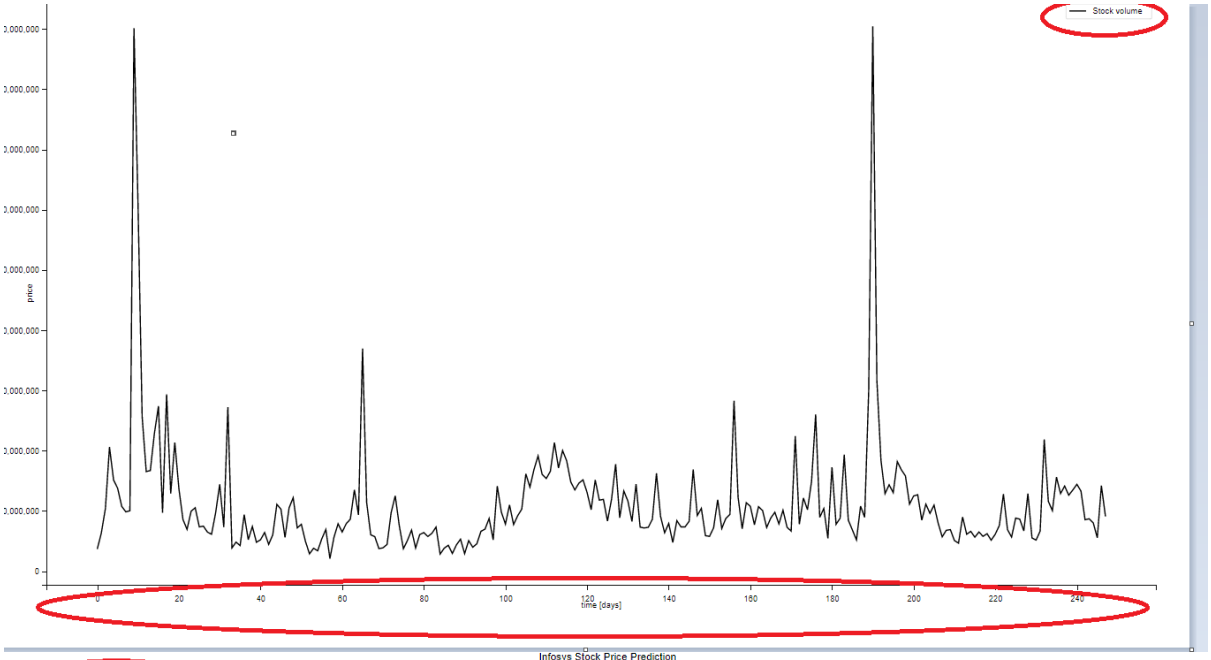
a) for the user to recognize we have put the option of icons as well as a fast search bar and a separate tab for stocks by dept , so that the user can easily jump to the desired stock company.

b) for the diagnose part we have created a notification under the graph of the company, which is suggesting the user to buy, sell or hold for the particular stock .

c) To keep our website from clustering free, we have chosen a proper layout in order to put the boxes , navigation bar, search bar etc.

10) <u>HELP AND DOCUMENTATION.</u>

For the help and documentation system, we have made a proper message under the graphs defining its axis , its use and the current situation of the stock for now.


Stock volume
price
time (days)


Infosys Stock Price Prediction
Real Infosys Stock Price
Predicted Infosys Stock Price
Time

..

COMPARATIVE ANALYSIS WITH OTHER EXISTING TECHNOLOGIES:

1. In the first paper, sentiment analysis is done with the use of dictionary approach. Datasets of tweets about a company is collected over a span of time and a brief Mood Introspection Scale with eight scales and two adjectives have been used to represent each mood state for starting point to create dictionaries.

Results of applying stock prediction algorithm on the dataset formed:-

Day	Basic	Our Model	Number of experiments
1	56.21%	63.41%	23
2	45.41%	56.87%	45
3	53.02%	51.33%	34
4	34.03%	54.67%	53
5	41.6%	59.60%	12
6	34.73%	63.78%	45
7	31.61%	73.77%	24

Fig 2.1 Comparison

Fig 2.1 *Represents the percentage of accuracy of other model compared to ours*

The SVM model is applied on thre datasets Basic, Basic&WHF and Basic&8EMO and results are above.

2. In this paper we see that taking into account certain events on web like some political or economic crisis often leads to sudden changes in prices. The authors here developed an automated system(KBNMiner) that acquires event-knowledge from the internet in order to predict interst rates. An increase in accuracy is claimed here but no numerical data has been provided.

3. In this stock exchange program financial articles are used. The system automatically analyzes and analyzes news articles and builds recommendations for investors. We found 61% accuracy in our study. This level of accuracy is greater than random predictions, with 50% accuracy. The results potray that a strong correlation between financial issues and stock price movements is present. The success rate of this system can be increased through the use of relevant news articles.

Our Results:- In this model data cleaning is required often. If you want correct result then data cleaning is very important. Where is in our LSTM model it automatically removes invalid / useless data making work of engineer easier and error free.

4. In the above tested networks they are able to beat the purchase and capture index benefits. Couples of networks have surpassed the market with an important margin. The return of 70.42% compared to the 47.41% return of purchases and holdings in the 20-20-1 structure with zero log sigmoid activity is truly astonishing. The model predicted the direction of the simple seven-day walking average after seven days.

Our Results:- If there happens market crash then above model will not be able to react to it as much as quickly as our model because we have used semantic analysis and this will help to take decision very fast if there is a market crash then it will immediately tell to sell

Our Work:-

i. Graph is obtained after application of autoregressive model on past datasets of the company.
ii. A graph comparing actual and predicted stock price after application of LSTM on datasets obtained is displayed.
iii. In the end we will display confusion matrix of our result.

iv. LSTM is theoretically more accurate than traditional Naive Bayes algorithm and SVM algorithm and should give more accuracy. Along with it, we would use moving average to improve our results.

CONCLUSION & FUTURE SCOPE:

We built a system which performs following tasks:-

1. Taking a dataset with closing prices of a company's stock, we use first 60-70% data for training our LSTM model. Then we predict the remaining closing prices. We have shown a curve comparing our predicted results with actual results.

2. We scrape live data of a company from news sites like yahoo finance, money control and plot a moving average curve and bollinger bands. The bollingerbands(upper and lower) give the user information about whether to buy or sell the stock. If the actual stock price goes above the upper

band we predict its time to sell the stock. If the actual price goes below the lower band we predict that its time to buy that particular stock.

Along with news analysis, these algorithms are expected to give great results with a high expectancy of providing atleast 12-15% returns on investment for the naïve users.

Future Work:

This area is at its nascent stage. A lot of parameters can be added to get more reliable predictions.

- A better neural network and ML algorithm can be implemented.
- Discussions forums can be created.
- On call, advisory service can be added
- We can add sentiment analysis on user's reviews from discussion forum. This would definitely improve the accuracy of prediction. We hope to work on that and merge it to our existing algorithm.

REFERENCES:-

[1] Machine learning to predict stock market indicators on the basis of historical data and Twitter sentiment analysis data by Alexander Porshnev, Ilya Redkin, Alexey Shevchenko
https://ieeexplore.ieee.org/document/6753954

[2] Machine Learning Techniques using Events Information for prediction of Stock Market: A Survey and Evaluation by P.D. Yoo , M.H. Kim , T. Jan.
https://ieeexplore.ieee.org/document/1631572

[3] Stock Price Prediction using Financial News Articles by M.I Yasef Kaya and M. ElefKarshgil

[4] Forecasting of Stock Market Indices with use of Artificial Neural Network by Dr. Jay Joshi, Nisarg A Joshi

[5]YoshuaBengio. Learning deep architectures for AI. Foundations and trends in Machine Learning

[6] Antoine Bordes, Jason Weston, Ronan Collobert, YoshuaBengio, et al. Learning structured embeddings of knowledge bases

[7] Paul C Tetlock. Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance

[8] Yue Zhang and Stephen Clark. Syntaactic processing by using the generalized perceptron and beam search. Computational Linguistics

APPENDIX:

Link to PPT:

https://drive.google.com/file/d/1mvWGFfN0D0Wy8dlK7AnVU50hLrubNlr9/view?usp=sharing

Link to Pre-recorded Demonstration video:

https://drive.google.com/file/d/1cTCQQpq5xqyDUSlpXzQjR-irWxjGP-30/view?usp=sharing

Link to access source files and file containing steps to execute the project:

https://github.com/GeekyGeek3371/Stock_Prediction_using_LSTM_sentimentAnalysis_Moving_Average