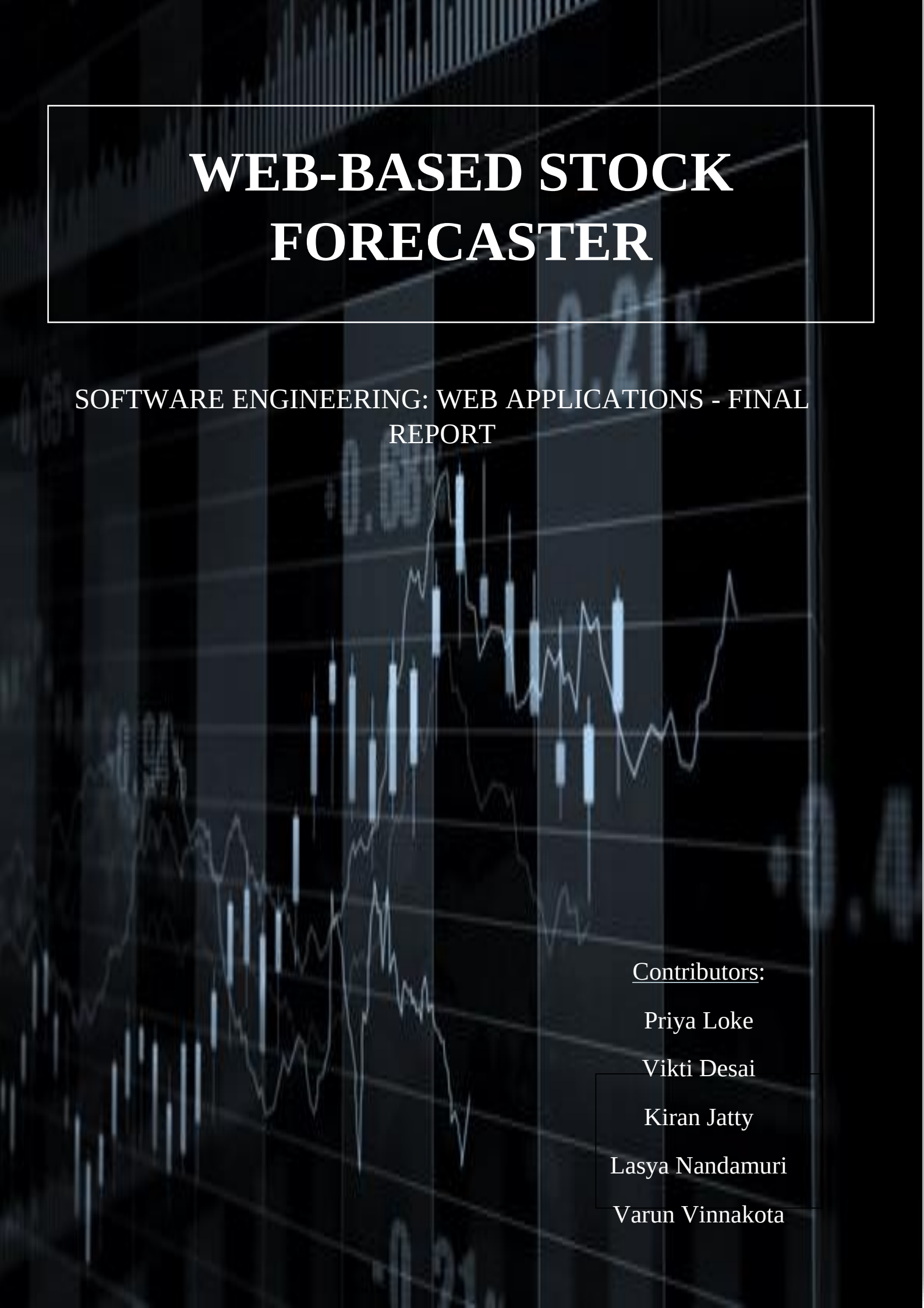
WEB-BASED STOCK FORECASTER

SOFTWARE ENGINEERING: WEB APPLICATIONS - FINAL REPORT



Contributors:

Priya Loke

Vikti Desai

Kiran Jatty

Lasya Nandamuri

Varun Vinnakota

Table of Contents

1. BACKGROUND:

In today's time, more than 50% people invest money in stocks, either directly, or indirectly. They experience financial security through mutual funds, retirement accounts or other managed assets. Few human activities have been so exhaustively studied during the past century, from so many angles and by so many different sorts of people, as has the buying and selling of corporate securities.

Since the glorious days of the East India Company or in today's times of New York Stock Exchange, capital investment in stocks has been a risky affair. Often, it is addressed as "gambling" since the company's fate, in which a person would invest money in hopes of multiplying returns, is always uncertain. In today's difficult times of high taxes and low wages the investing in the stock market seems like an escape to quick money. Financial analysts have been making an effort for a sure and gainful method to predict nature of the stock market thereby helping investors to head in the right direction. However, some argue that stock market is impossible to predict and is of random nature. The rewards which the stock market holds out to those who read it right are enormous; the penalties it exacts from careless, dozing, or "unlucky" investors are calamitous.

There are two categories of market predicting models – Fundamental Analysis and Technical Analysis.

Fundamental Analysis is a method of evaluating a security that entails attempting to measure its intrinsic value by examining related economic, financial and other qualitative and quantitative factors. It includes economic analysis, industry analysis and company analysis. On the basis of these, the intrinsic value of the shares are determined. This is considered as the true value of the share. Fundamental analysts attempt to study everything that can affect the security's value, including macroeconomic factors (like the overall economy and industry conditions) and company-specific factors (like financial condition and management). The end goal of performing fundamental analysis is to produce a value that an investor can compare with the security's current price, with the aim of figuring out what sort of position to take with that security (underpriced=buy, overpriced=sell or short).This method of security analysis is considered to be subjective, the opposite of technical analysis.

Technical Analysis is a method of evaluating securities by analysing statistics generated by market activity, such as past prices and volume. Technical analysts do not attempt to measure a security's intrinsic value, but instead use charts and other tools to identify patterns that can suggest future activity.

Technical analysts believe that the historical performance of stocks and markets are indications of future performance. Technical Analysis utilizes objective mathematical indicators to reflect all relevant information of the market.

In a shopping mall, a fundamental analyst would go to each store, study the product that was sold, and then decide whether to buy it or not. By contrast, a technical analyst would sit on a bench in the mall and watch people go into the stores. Disregarding the intrinsic value of the products in the store, the technical analyst's decision would be based on the patterns or activity of people going into each store.

Technical analysis is widely used among traders, financial professionals and is very often used by active day traders. Technical analysis analyzes the price, volume and other market information, whereas fundamental analysis looks at the facts of the company, market, currency or commodity.

2. INTRODUCTION:

Technical analysis of stocks comprise of determining the future price of a stock solely based on the trends of the past price (a form of time series analysis. Artificial neural networks and generic algorithms can also be deployed to analyze the stock price.

Predicting stock market with 100% precision is impossible, but people still manage to develop some analysis techniques to help us get much more comprehensive understanding of the market and make better decisions in our investments.

The **goal** of our project is to build a web-based prediction system that provides investors with advice regarding investment in the stock market. It predicts the future values of stock using the historic (past) price to help the investors to invest smartly.

An application runs in the background extracting the current and past values of stocks from a web source and stores the data into a database. We used the cross-platform software stack XAMPP which provides us with all the tools needed for hosting a web server, a database server etc. The stock values are then analyzed to implement the prediction model. Following are more details regarding the main modules of the project.

We aim to collect and provide information about around 10 stocks. We provide the current stock prices and make long term and short term predictions for these stocks. We use indicators to provide graphical representation of the stock prices.

We use Yahoo Finance’s restful API and Kibot API to collect data of the stocks.

Machine learning techniques: For the purpose of prediction, we use Bayesian Curve Fitting, Support Vector Machine (SVM) and Artificial Neural Networks (ANN).

3. REQUIREMENTS:

We create a user friendly web interface for users to get predictions of stock prices, view graphs and get valuable suggestions regarding different stocks. The system requirements to develop such an application are as follows:

A. Requirements -

1. Connection with a web-server
2. A database to store the data acquired
3. Knowledge of a scripting language to communicate with the database
4. A web source for acquiring stock information
5. Queries for accessing the database

B. Resources –

XAMPP – A cross-platform solution stack which contains all the resources required to connect to the web server viz. server application (Apache), database (MariaDB), scripting language (PHP or Perl).

- *Server Application*: Apache - A widely used HTML web-server
- *Database*: MariaDB - A community developed for the MySQL RDBMS.
- *Scripting Language*: PHP - It is a server-side scripting language designed for web-development
- *Web Source*: Kibot API - For acquiring the values of stocks
- *Queries*: SQL was used to create the queries

4. SOFTWARE ENVIRONMENT:

I. XAMPP

XAMPP is a free and open source cross-platform web server solution stack package developed by Apache Friends, consisting mainly of the Apache HTTP Server, MariaDB database, and interpreters for scripts written in the PHP and Perl programming languages. XAMPP stands for Cross-Platform (X), Apache (A), MariaDB (M), PHP (P) and Perl (P). It is a simple, lightweight Apache distribution that makes it extremely easy for developers to create a local web server for testing purposes. Everything needed to set up a web server – server application (Apache), database (MariaDB), and scripting language (PHP) – is included in an extractable file.

A. APACHE

The Apache HTTP Server, colloquially called Apache is the world's most used web server software. Originally based on the NCSA HTTPd server, the development of Apache began in early 1995 after work on the NCSA code stalled. Apache played a key role in the initial growth of the World Wide Web, quickly overtaking NCSA HTTPd as the dominant HTTP server, and has remained most popular since April 1996. In 2009, it became the first web server software to serve more than 100 million websites.

Key Features:

- Loadable Dynamic Modules
- Multiple Request Processing modes (MPMs) including Event-based/Async, Threaded and Pre-fork.
- Highly scalable (easily handle more than 10,000 simultaneous connections
- Handling of static files, index files, auto-indexing and content negotiation
- .htaccess support
- Reverse proxy with caching

B. PHP

PHP is a server-side scripting language designed for web development but can also be used as a general- purpose programming language. PHP code is very flexible and can be simply mixed with HTML code, or it can be used in combination with various web frameworks. PHP code is processed by a PHP interpreter, which is usually implemented as a web server's native module or a Common Gateway Interface (CGI) executable. After the PHP code is interpreted and executed, the web server sends resulting output to its client in form of generated web page; for example, PHP code can generate a web page's HTML code, an image, or some other data.

PHP has a direct module interface called Server Application Programming Interface (SAPI), which is supported by many web servers including Apache HTTP Server, Microsoft IIS.

In terms of keywords and language syntax, PHP is similar to most high level languages that follow the C style syntax. IF conditions, FOR and WHILE loops, and function returns are similar in syntax to languages such as C, C++, C#, Java and Perl.

PHP 5 included new features like support for object-oriented programming, the PHP Data Objects PDO extension which defines a lightweight and consistent interface for accessing databases.

We used PHP as our scripting language to implement various features of our website.

C. MariaDB

MariaDB is a community-developed fork of the MySQL relational database management system intended to remain free under the GNU GPL. It is notable for being led by the original developers of MySQL, who forked it due to concerns over its acquisition by Oracle. Contributors are required to share their copyright with the MariaDB Foundation.

MariaDB intends to maintain high compatibility with MySQL, ensuring a "drop-in" replacement capability with library binary equivalency and exact matching with MySQL APIs and commands. It includes the XtraDB storage engine for replacing InnoDB, as well as a new storage engine, Aria, that intends to be both a transactional and non-transactional engine perhaps even included in future versions of MySQL.

D. SQL

SQL- Structured Query Language is a special-purpose programming language designed for managing data held in a relational database management system (RDBMS), or for stream processing in a relational data stream management system (RDSMS).

We primarily used SQL to create and access tables in our database.

II. PYTHON:

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

Often, programmers fall in love with Python because of the increased productivity it provides. Since there is no compilation step, the edit-test-debug cycle is incredibly fast. Debugging Python programs is easy: a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it raises an exception. When the program doesn't catch the exception, the interpreter prints a stack trace. A source level debugger allows inspection of local and global variables, evaluation of arbitrary expressions, setting breakpoints, stepping through the code a line at a time, and so on. The debugger is written in Python itself, testifying to Python's introspective power. On the other hand, often the quickest way to debug a program is to add a few print statements to the source: the fast edit-test-debug cycle makes this simple approach very effective.

Python supports lot of libraries for scientific and numeric computing. It provides simple and efficient tools for data mining and data analysis. This project uses ‘scikit’ machine learning library for prediction strategies. It is based on modules numpy, scipy and matplotlib which help in computing and visualization of the data.

III. HTML:

Hyper Text Markup Language, commonly referred to as HTML, is the standard markup language used to create web pages. It is written in the form of HTML elements consisting of tags enclosed in angle brackets (like <html>). HTML tags most commonly come in pairs like <h1> and </h1>, although some represent empty elements and so are unpaired, for example <img>. The first tag in such a pair is the start tag, and the second is the end tag (they are also called opening tags and closing tags).

Web browsers can read HTML files and render them into visible or audible web pages. Browsers do not display the HTML tags and scripts, but use them to interpret the content of the page. HTML describes the structure of a website semantically along with cues for presentation, making it a markup language, rather than a programming language.

HTML elements form the building blocks of all websites. HTML allows images and objects to be embedded and can be used to create interactive forms. It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items. It can embed scripts written in languages such as JavaScript which affect the behavior of HTML web pages.

IV. CSS:

Cascading Style Sheets (CSS) is a style sheet language used for describing the look and formatting of a document written in a markup language. While most often used to change the style of web pages and user interfaces written in HTML and XHTML, the language can be applied to any kind of XML document, including plain XML, SVG and XUL. Along with HTML and JavaScript, CSS is a cornerstone technology used by most websites to create visually engaging webpages, user interfaces for web applications, and user interfaces for many mobile applications[1].

CSS is designed primarily to enable the separation of document content from document presentation, including elements such as the layout, colors, and fonts.[2] This separation can improve content accessibility, provide more flexibility and control in the specification of presentation characteristics, enable multiple HTML pages to share formatting by specifying the relevant CSS in a separate .css file, and reduce complexity and repetition in the structural

content, such as semantically insignificant tables that were widely used to format pages before consistent CSS rendering was available in all major browsers. CSS makes it possible to separate presentation instructions from the HTML content in a separate file or style section of the HTML file. For each matching HTML element, it provides a list of formatting instructions. For example, a CSS rule might specify that "all heading 1 elements should be bold," leaving pure semantic HTML markup that asserts "this text is a level 1 heading" without formatting code such as a <bold> tag indicating how such text should be displayed.

The entire project was developed on a Windows platform, hence the installation process will guide through the steps required on Windows platform.

- First install XAMPP software bundle on your laptop.
- Then install Python in order to run the scripts
- In addition to python, the following dependencies for Python also need to be installed: numpy, scipy, matplotlib, sklearn

5. IMPLEMENTATION:

XAMPP is used to make a connection to the "localhost/phpmyadmin/". A database with the required name was created.

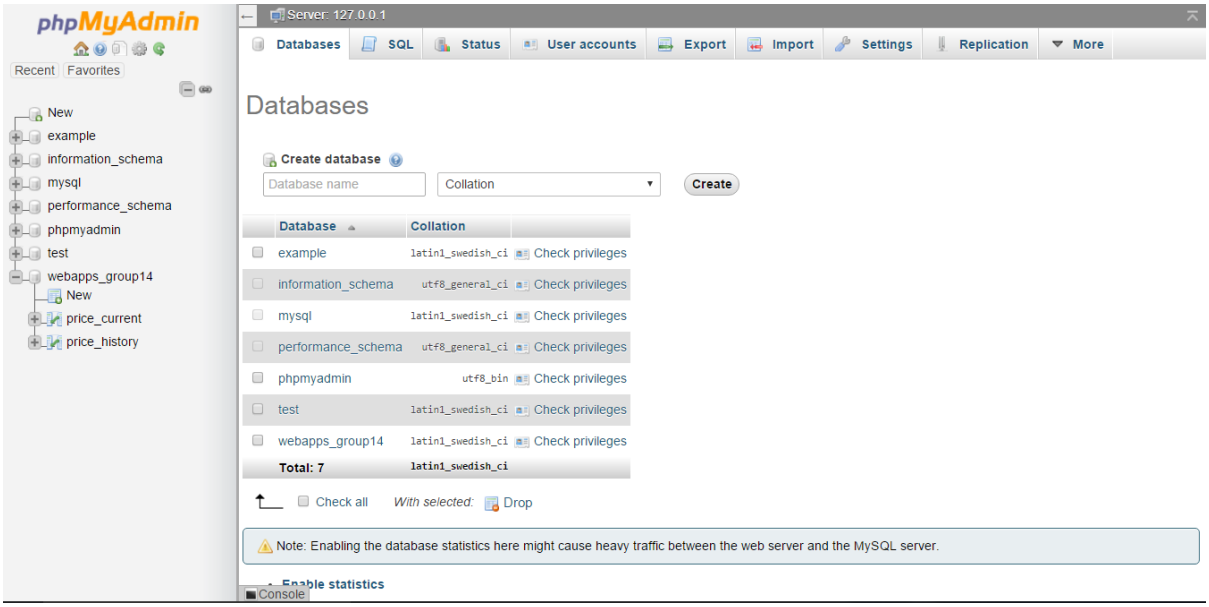


Figure 1: Creating a new database in /localhost/phpmyadmin/

Two tables were created in the database using SQL queries in php.

Database Connection successful.
Table created successfully

Figure 2: Output of the php file on creation of the two database tables “price_current” and “price_history”

The table price_history has the stock values for 10 stocks over a period of 1 year.

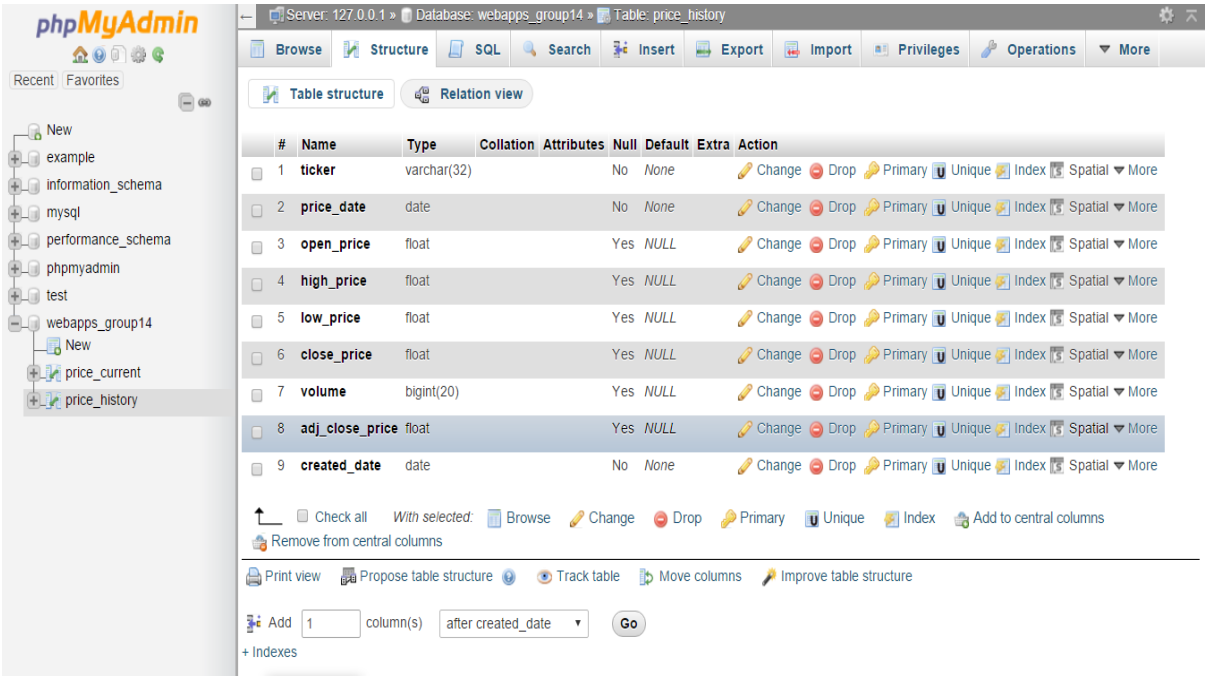


Figure 3: Structure of the table “price_history”

The table price_current is populated by running a php script that collects stock values of 10 stocks every 1 minute i.e. 60 seconds.

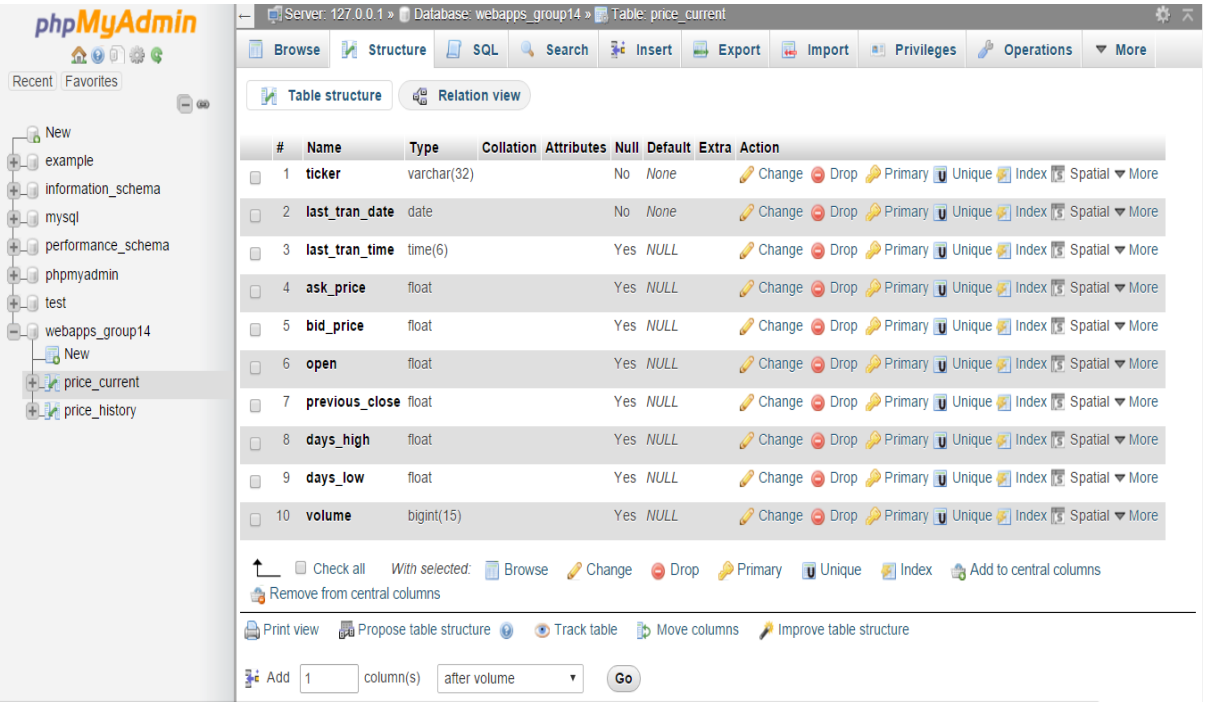


Figure 4: Structure of the table "price_current"

DATA COLLECTION: The data required is collected from Yahoo finances' API. Yahoo finance provides the URLs to download the required .csv files.

A. HISTORICAL DATA:

Data for the past one year is collected by writing a php program. The algorithm used for this purpose is as follows:

- Connect to the database "webapps_group14" in phpmyadmin
- Current date (today's date) and previous date(which can be changed accordingly) are defined
- This url used to access the data from history "http://ichart.finance.yahoo.com/table .csv?s=".$id."&d=".$crnt_date[0]."&e=".$crnt_date[1]."&f=".$crnt_date[2]."&g=d&a=".$pr_date[0]."&b=".$pr_date[1]."&c=".$pr_date[2].'&ignore=.csv'
- The value of the id can be changed in the program to obtain the results for different companies
- Inserts the data from .csv into corresponding "price_history" table in the database in proper format

Connection successful
Current date is:03/03/16
Previous date is:02/28/15
Inserted into Table price_history

Figure 5: Output of the php file on collection of historical data

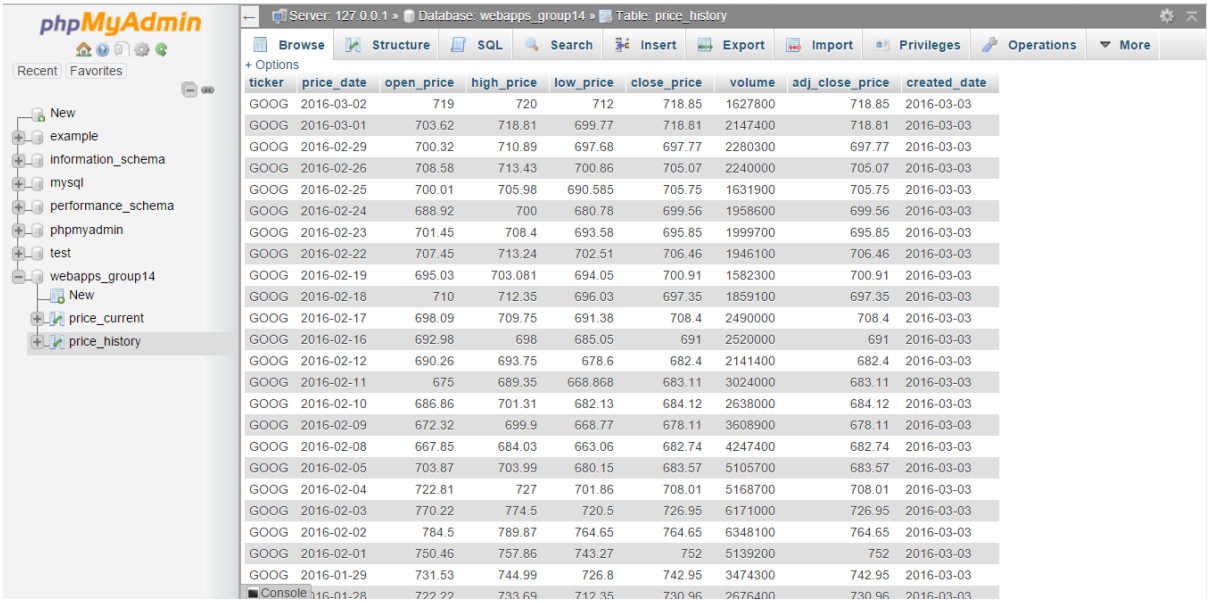


Figure 6: Collection of historical data for GOOG, reflected in the database

B. CURRENT VALUES:

The stock data for every 60 seconds for all the 10 companies is collected. The algorithm used in the currentprice.php program is as follows.

- Connect to the database "webapps_group14" in phpmyadmin
- The localhost/xampp/currentprice.php page refreshes

- This url used to access the data from history ("http://finance.yahoo.com/d/quotes .csv?s=".$id."&f=sd1t1abophgv")
- The variable id is an array and is put in loop to run for ten different companies.
- Inserts the data from .csv into corresponding "price_current" table in the database in proper format

Symbols	**Symbol Names**
ERIC	Ericsson
CSCO	Cisco Systems, Inc.
MSFT	Microsoft Corporation
INTC	Intel Corporation
FB	Facebook
AAPL	Apple
BAC	Bank of America
GOOG	Google Inc.
YHOO	Yahoo Inc.
C	Citi

Database Connection successful.

Fetching data for GOOG
Data fetched is: "GOOG","3/3/2016","11:31am",711.61,711.13,717.54,718.85,719.45,706.02,931274
Data Inserted For GOOG

Fetching data for YHOO
Data fetched is: "YHOO","3/3/2016","11:31am",32.46,32.45,32.71,32.91,33.21,32.42,4637453
Data Inserted For YHOO

Fetching data for MSFT
Data fetched is: "MSFT","3/3/2016","11:31am",51.79,51.78,52.97,52.95,52.97,51.78,9229193
Data Inserted For MSFT

Fetching data for INTC
Data fetched is: "INTC","3/3/2016","11:31am",30.62,30.61,31.12,30.54,31.35,30.60,10897526
Data Inserted For INTC

Fetching data for CSCO
Data fetched is: "CSCO","3/3/2016","11:31am",26.63,26.62,26.95,26.90,26.97,26.59,7073260
Data Inserted For CSCO

Insertion Completed in table price_current

Figure 7: Output of the PHP file on collection of current data

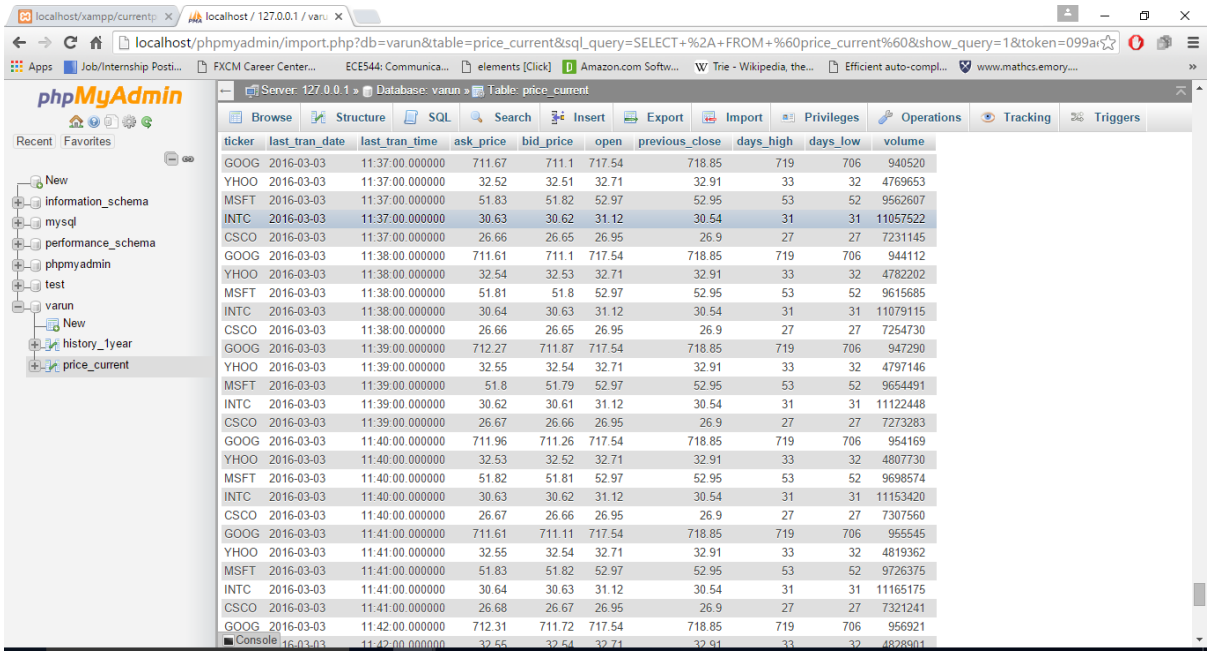


Figure 8: Collection of current data for GOOG, YHOO, MSFT, INTC, CSCO reflected in the database

6. DATABASE SCHEMA

The database schema of our database is as follows:

It has 2 tables as mentioned above, the price_history table that stores the historical prices and the price_current tables that stores the current values of stocks.

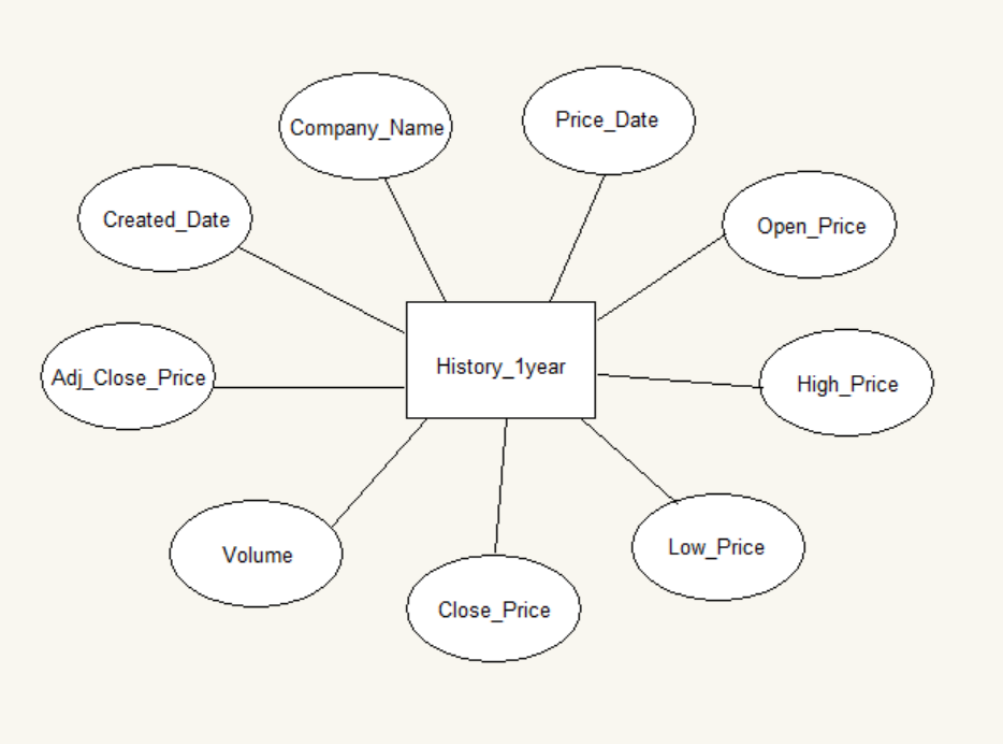


Figure 9: Database schema for historical data

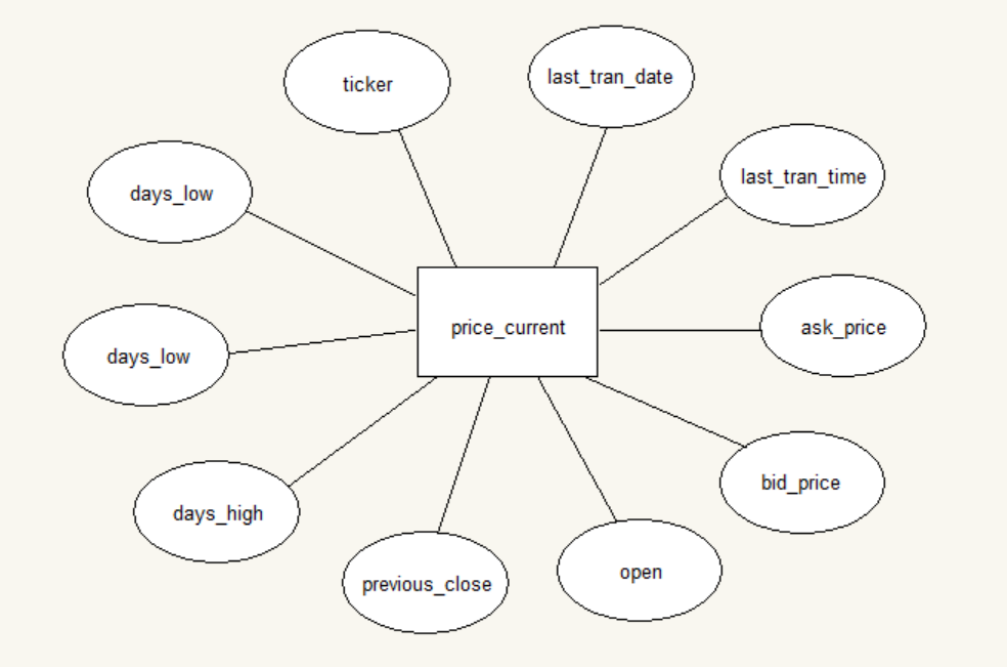


Figure 10: Database schema for current quotes

A. ATTRIBUTES OF DATABASE TABLES:

TICKER: It is the symbol, used to look up the stocks of any company

PRICE DATE: It shows the date on which the stock data is recorded

OPEN PRICE: It is the price at which security first trades upon the opening of an exchange on a given trading day

CLOSE PRICE: It refers to the last price at which a stock trades during a regular trading session

VOLUME: It is the number of shares that change hands during a given day

ADJUSTED CLOSING PRICE: It is a stock's closing price on any given day of trading that has been amended to include any distributions and corporate actions that occur at any time prior to the next day's open.

TODAY'S HIGH: Today's high is the highest price at which a stock traded during the course of the day. Today's high is typically higher than the closing or opening price. More often than not this is higher than the closing price.

TODAY'S LOW: Today's low is the lowest price at which a stock trades over the course of a trading day. Today's low is typically lower than the opening or closing price.

CREATED DATE: It specifies the date when the table in the database is filled.

ASK PRICE: The Ask price represents the minimum price that a seller is willing to receive for security.

BID PRICE: The Bid price represents the maximum price that a buyer is willing to pay for a security.

LAST TRANSACTION DATE: The last transaction date represents the date on which ownership is transferred for the most recent transaction.

LAST TRANSACTION TIME: The last transaction time represents the date on which ownership is transferred for the most recent transaction.

PREVIOUS CLOSE: Previous close can refer to the prior day's value of a stock, bond, commodity or any other security.

Once the data has been collected, the next step was to use this data for the purpose of prediction.

7. PREDICTION SCHEMES:

The financial domain is affected by known and unknown factors that can cause the stock prices to vary quickly and hence the stock prices are considered dynamic in nature. In relation to this, a lot of research goes into applying machine learning algorithms to predict stock prices, analyze patterns and index changes, based on the information stored in the database. This helps to identify the fluctuations in the market before it actually happens thus helping the trader to gain profit at the right time or avoid a loss before incurring it.

In this project, we are using short term and long term prediction strategies to help the trader make decision for day trading purposes and for daily and weekly monitoring and trading. We are using Bayesian Prediction methodology for short term prediction. For long term prediction, we use Artificial Neural Networks to train our algorithm to predict the close prices for the next 5 days. We then use SVM (Support Vector Machine) to indicate whether the trend of the price is rising or falling, so as to suggest whether a user should BUY or SELL the stock in order to reap benefits.

A. Long term prediction:

Artificial Neural Networks (ANN) are a family of models inspired by biological neural networks and are used to estimate or approximate functions that can depend on a large number of inputs and are generally unknown. The word network in the term 'artificial neural network'

refers to the inter-connections between the neurons in the different layers of each system. An example system has three layers. The first layer has input neurons which send data via synapses to the second layer of neurons, and then via more synapses to the third layer of output neurons. The synapses store parameters called "weights" that manipulate the data in the calculations.

An ANN is typically defined by three types of parameters:

1. The interconnection pattern between the different layers of neurons
2. The learning process for updating the weights of the interconnections
3. The activation function that converts a neuron's weighted input to its output activation.

What has attracted the most interest in neural networks is the possibility of *learning*. Most of the algorithms used in training artificial neural networks employ some form of gradient descent, using back propagation to compute the actual gradients. Back propagation, an abbreviation for "backward propagation of errors", is a common method of training artificial neural networks used in conjunction with an optimization method such as gradient descent. The method calculates the gradient of a loss function with respect to all the weights in the network. The gradient is fed to the optimization method which in turn uses it to update the weights, in an attempt to minimize the loss function. Back propagation requires a known, desired output for each input value in order to calculate the loss function gradient. It is therefore usually considered to be a supervised learning method.

In our implementation of ANN, we trained the algorithm using back propagation with 1 year of data for each stock symbol separately. The neural network developed has 3 input neurons or nodes, for the average, minimum and maximum values of 5 consecutive days ending on the current day. The network has 3 hidden neurons and 1 output neuron, which is the estimated closing price for the next day. Prediction for multiple days is provided by making it a closed loop system and using the predicted value in calculating the average, minimum and maximum for 5 days. Thus our implementation predicts the closing price of the stock for the next 5 days. In this way, the algorithm can also be modified to predict the values for as many days as required. But it has to be kept in mind that a slight error in a predicted value carried forward has the chance of deviating from the actual value if used for a really long period.

Support vector machine (SVM) is a very specific type of learning algorithms characterized by the capacity control of the decision function, the use of the kernel functions and the sparsity of the solution. Established on the unique theory of the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization error, SVM is shown to be very resistant to the over-fitting problem, eventually achieving a high generalization performance. Another key property of SVM is that training SVM is equivalent to solving a linearly constrained quadratic programming problem so that the solution of SVM is always unique and globally optimal, unlike neural

networks training which requires nonlinear optimization with the danger of getting stuck at local minima.

SVM is proven to be good for single feature classification and hence was only used to provide the direction of trend of stocks.

Our implementation of the SVM algorithm uses one year of historical stock high price per stock as the training dataset. The predicted stock price using ANN is used as a test dataset. SVM helps determine whether the predicted stock price follows a rising or falling trend. This can be used to suggest the user to either sell the stocks or buy more stocks.

B. Short term prediction:

Bayesian estimator is based upon the simple Bayes' theorem. Given a set of observed data, it is straight forward to get the probability of choosing the model. It is called a prior probability. But, given a model, what is the probability of choosing the data set becomes tricky. This is called the posterior probability. Bayes' theorem helps to identify the posterior probability with the help of a likelihood function.

The input to the Bayesian estimator are the training data set stock prices(x) at time points (t) and the future time point (T) for which the stock price (X) has to be predicted. The mean and variance are determined from the given data set. The future stock price is given by the probability function

p(X | T, t,x) = N(X | m(x), s2(x))

Where the mean and variance are given by

m(x) = (t)T S n=1N (tn) xn
s2(x) = -1 + (t)T S (t)

Here the S matrix is given by

S-1 = I + n=1N (tn) (t)T

Where I is the unit matrix and the vector (t) is defined by the elements

n(t) = tn for n = 0,1... M
where M is the order of the polynomial. The values are chosen to be 5 10-3 and 11.1 respectively.

The input to the Bayesian predictor is the price collected every minute during the market hours over a period of a day. The output is the prediction value for the next minute.

8. INDICATORS:

Indicators are calculations based on the price and the volume of a security that measure things like money flow, trends, volatility and momentum. Indicators are used in two main ways:

- Leading: It moves before the price movements, indicating a predictive quality. It is thought to be the strongest during periods of sideways or non-trending trading ranges.
- Lagging: It follows the movements of prices.

A. Simple Moving Average (SMA):

A simple moving average (SMA) is a simple, or arithmetic moving average that is calculated by adding the closing price of the security for a number of time periods and then by dividing this total by the number of time periods. Short-term averages respond quickly to changes in the price.

B. Exponential Moving Average (EMA):

This type of moving average reacts faster to recent price changes than a simple moving average. The 12- and 26-day EMAs are the most popular short-term averages, and they are used to create indicators like the moving average convergence divergence (MACD) and the percentage price oscillator (PPO).

Exponential moving averages have less lag and are therefore more sensitive to recent prices and recent price changes. Exponential moving averages will turn before simple moving averages. Simple moving averages, on the other hand, represent a true average of prices for the entire time period. As such, simple moving averages may be better suited to identify support or resistance levels.

The direction of the moving average conveys important information about prices. A rising moving average shows that prices are generally increasing. A falling moving average indicates that prices, on average, are falling. A rising long-term moving average reflects a long-term uptrend. A falling long-term moving average reflects a long-term downtrend.

In summary, an exponential moving average (EMA) can be considered a type of moving average that is similar to a simple moving average, except that more weight is given to the latest data. The exponential moving average is also known as exponentially weighted

moving average. In general, the 50 and 200-day EMAs are used as signals of long-term trends.

9. USE CASES:

The use cases that have been implemented and are supported by our application are as follows:

Case	Action
Register	To create a new account by registering
Login	To login to the account using user name and password which they have entered while registering
Logout	To sign-out from the website
Predict	To get short term and long term predictions for the required stock
Suggest Me	Suggests whether to buy or sell stocks based on the predicted value
About	Brief description of prediction algorithms used
Show Graph	To be able to view the charts for the selected stock symbol and number of days specified by user
Queries	To be able to view the five queries of specified stock
Contact us	To post their questions and suggestions

Register:

The user is required to create a new account and activate it using their email address, username and password. The details of the user are stored into the database in order to keep track of the users. Users can enter into webpage and access its features only through this account. After registering the user is directed to login page again.

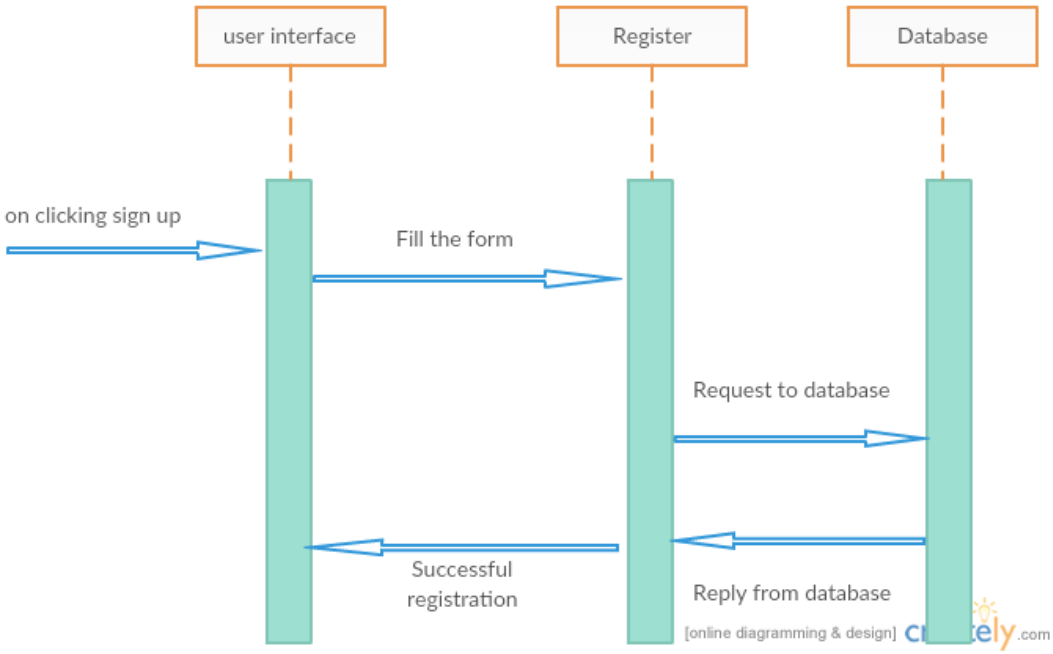


Figure 11: User registration

A more detailed version of the interaction diagram is shown below:

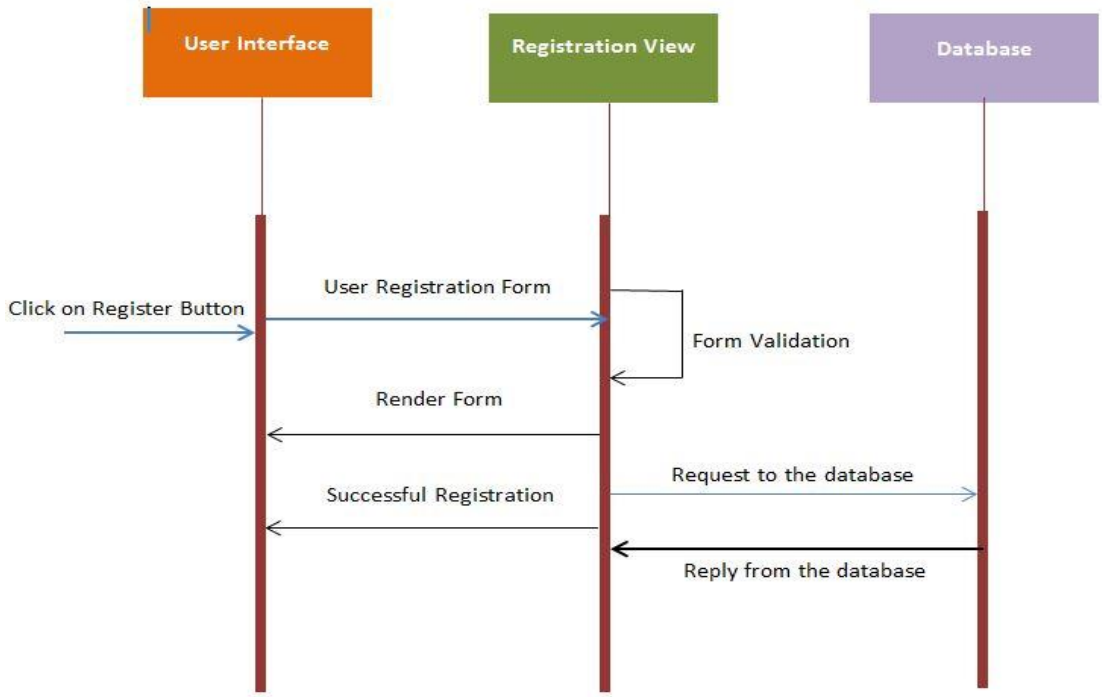


Figure 12: User registration

Login:

The user can login using the username and password specified in the registration form. This login is validated by comparing the details with the data that is already present in the database. On logging in the user will be guided to the home page which shows various operations offered by the webpage.

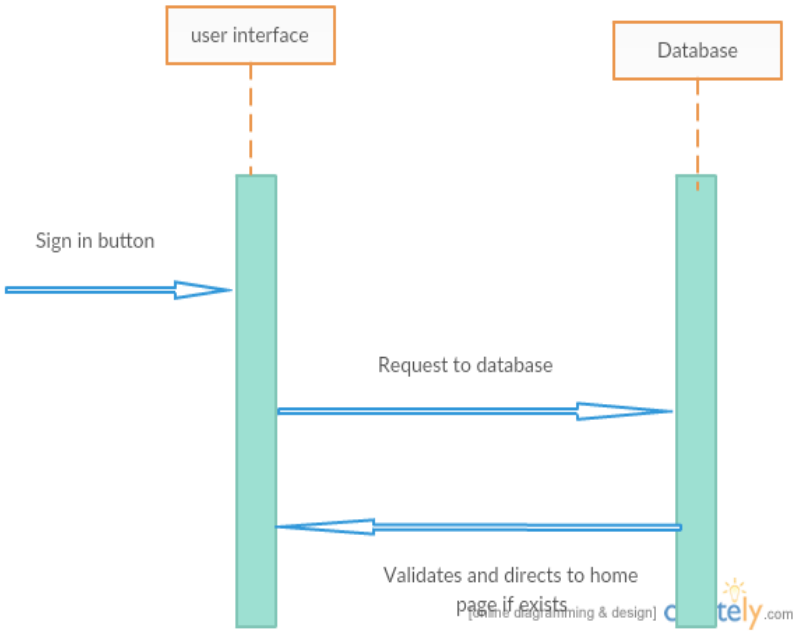
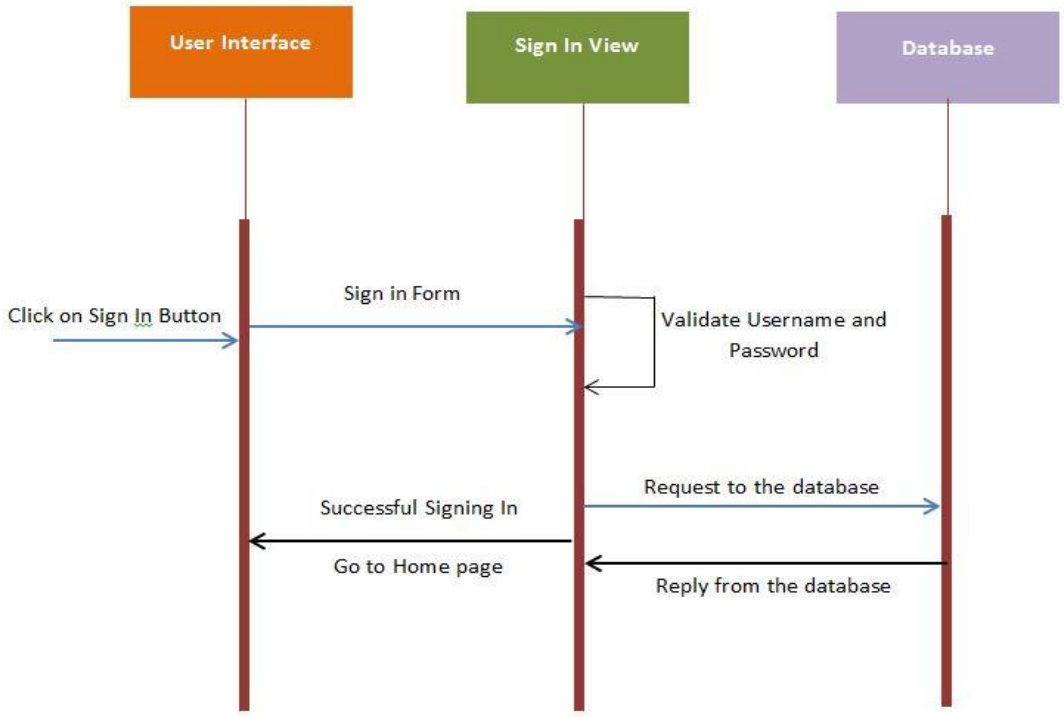


Figure 13, 14: User login

A more detailed version of the interaction diagram is shown below:



Logout:

To sign-out from the webpage, on clicking it directs the user to the login page.

Predict:

It provides two options-current prediction and 5-day prediction. When the user selects current prediction, the PHP script is executed which runs the Bayesian algorithm. This in turn accesses the real time data present in the database and based on these values predicts the next minute value.

When 5-day prediction is selected, a request for the calculating long term prediction is made. After the request is made, analyzer.php is executed (which contains the Python code for implementing Artificial Neural Networks) which in turn uses neuralNetwork.php to train the data and predict the open price for the next 5 days.

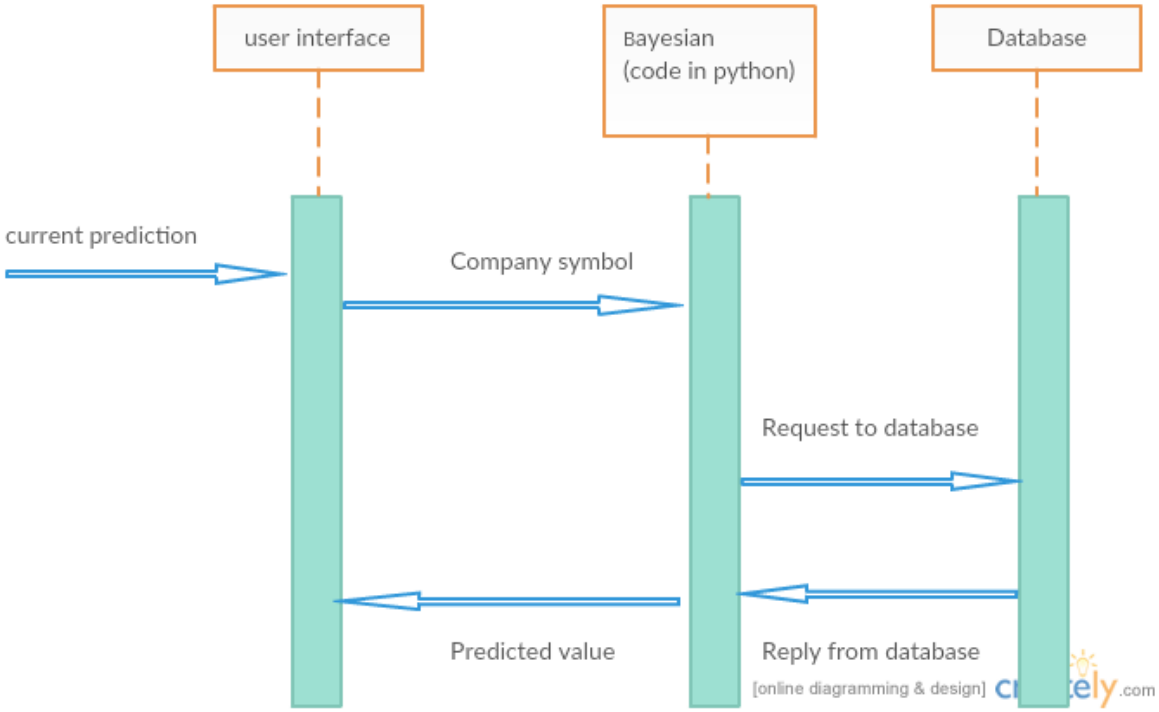


Figure 15: Short term prediction

A more detailed version of the interaction diagram is shown below:

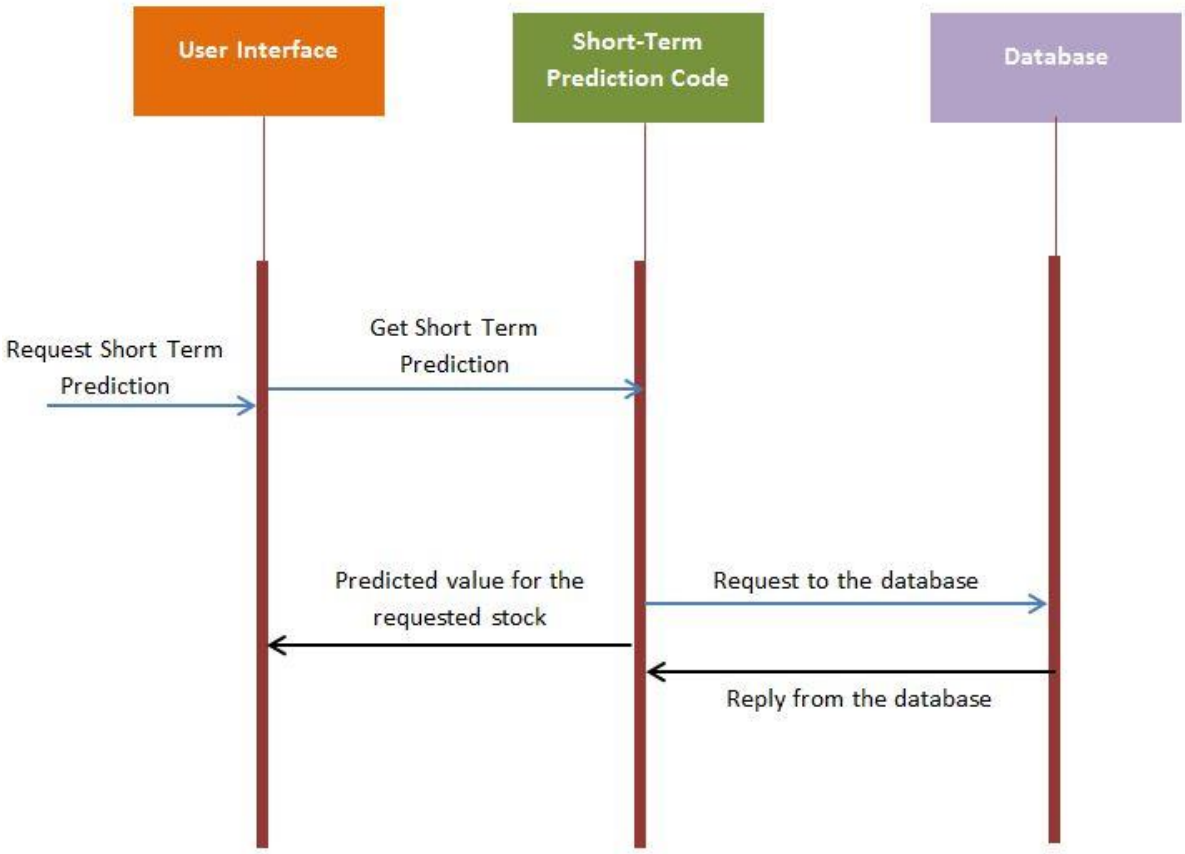


Figure 16: Short term prediction

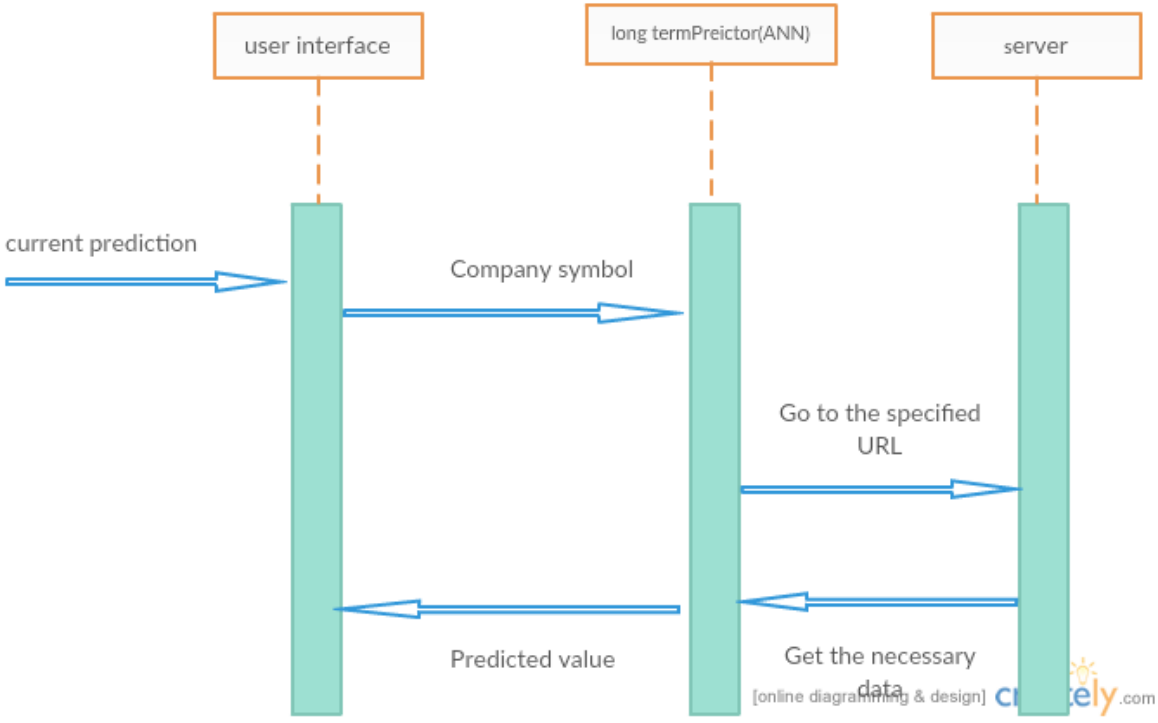


Figure 17: Long term prediction

A more detailed version of the interaction diagram is shown below:

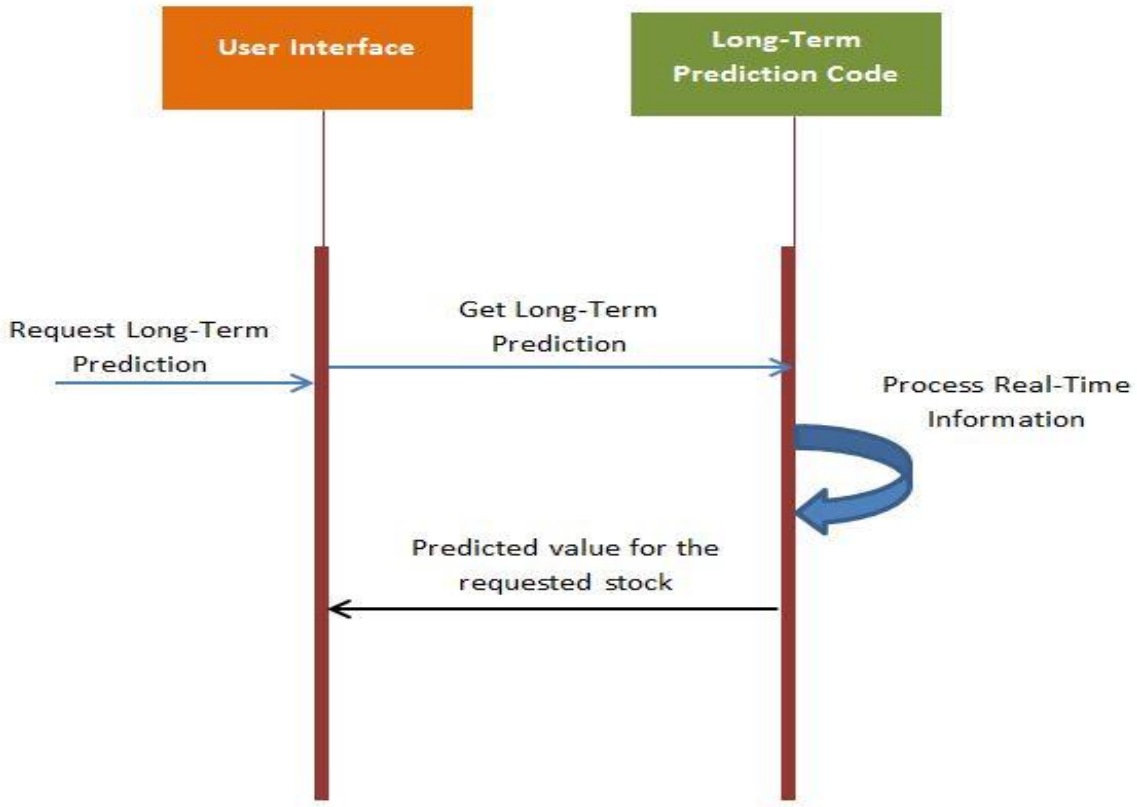


Figure 18: Long term prediction

SuggestMe:

When this icon is clicked, a value is predicted using ANN. This predicted value will be fed to SVM predictor as the test data set where the SVM validates if the predicted value follows the trend of the historical high and low stock price (which is the training data). Here this also provides a confidence level whether the prediction is strong or weak. This predicted stock price is now compared with the computed MACD and suggests the user whether to buy or sell the stock depending on whether the stock is following a rising or a falling trend.

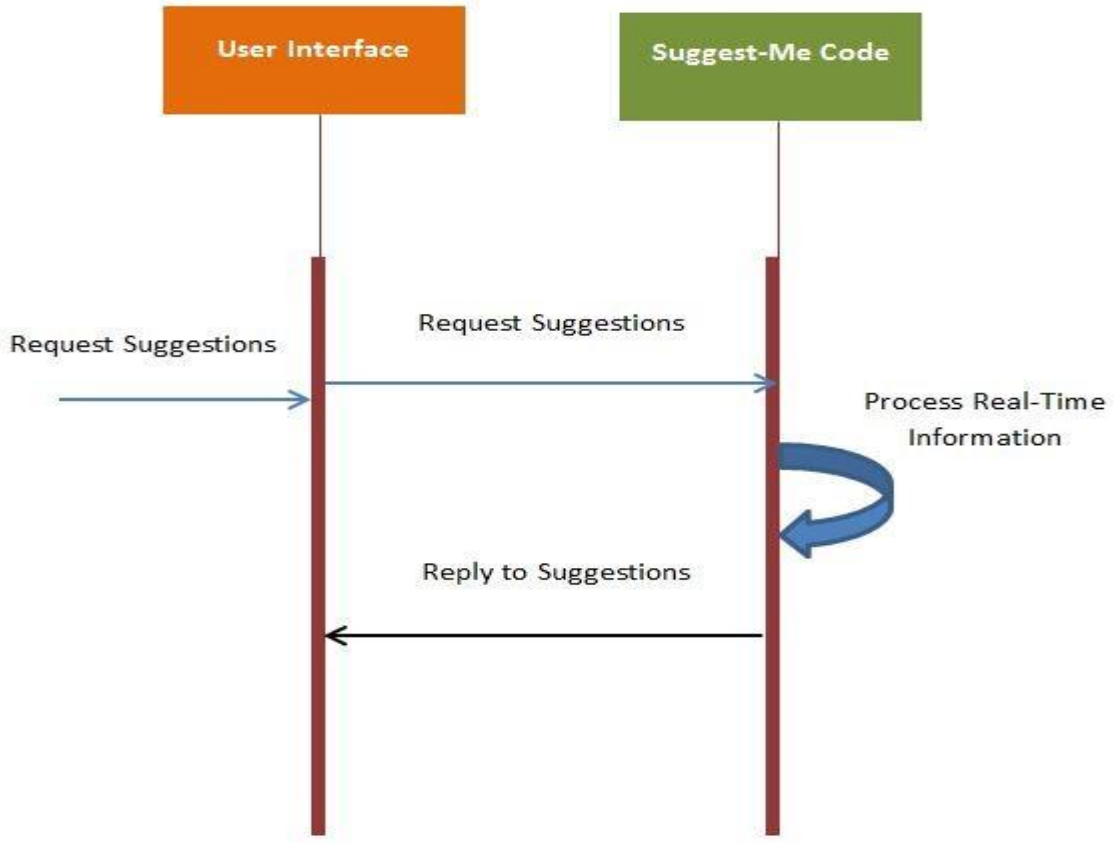


Figure 19: Suggest Me

About:

This page gives a brief description about the long term and short term prediction algorithms used.

Show Graph:

The user selects the company name and enters the number of days, this links to python code which access the URL and collects the value for the specified number of days. The indicators of EMA and SMA are used to set the trend lines.

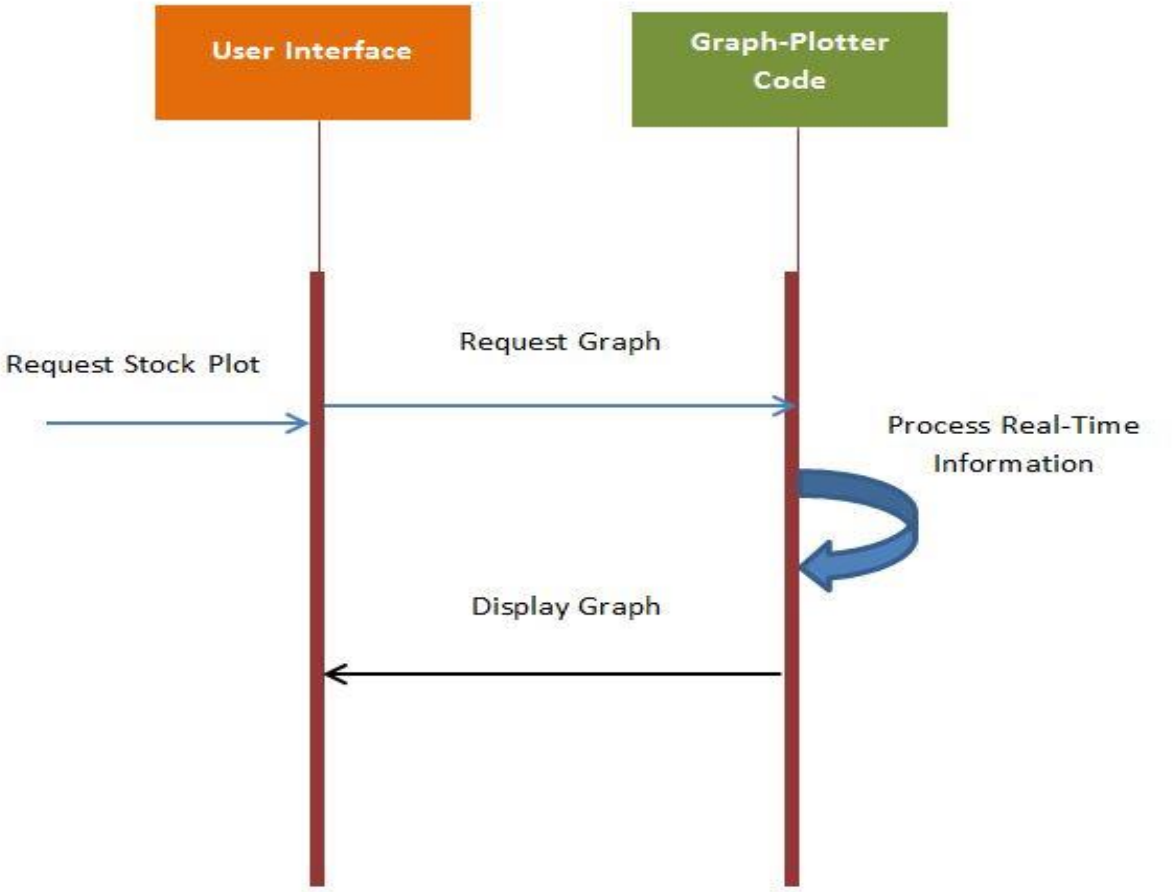


Figure 20: Display graph

Queries:

The five queries provided are:

- Shows the list of all companies in the database along with their latest stock price.
- The highest stock price of Google in the last ten days
- Average stock price of Microsoft in the latest one year
- Lowest stock price for each company in the latest one year
- List of ids of companies along with their name who have the average stock price lesser than the lowest of Google in the latest one year.

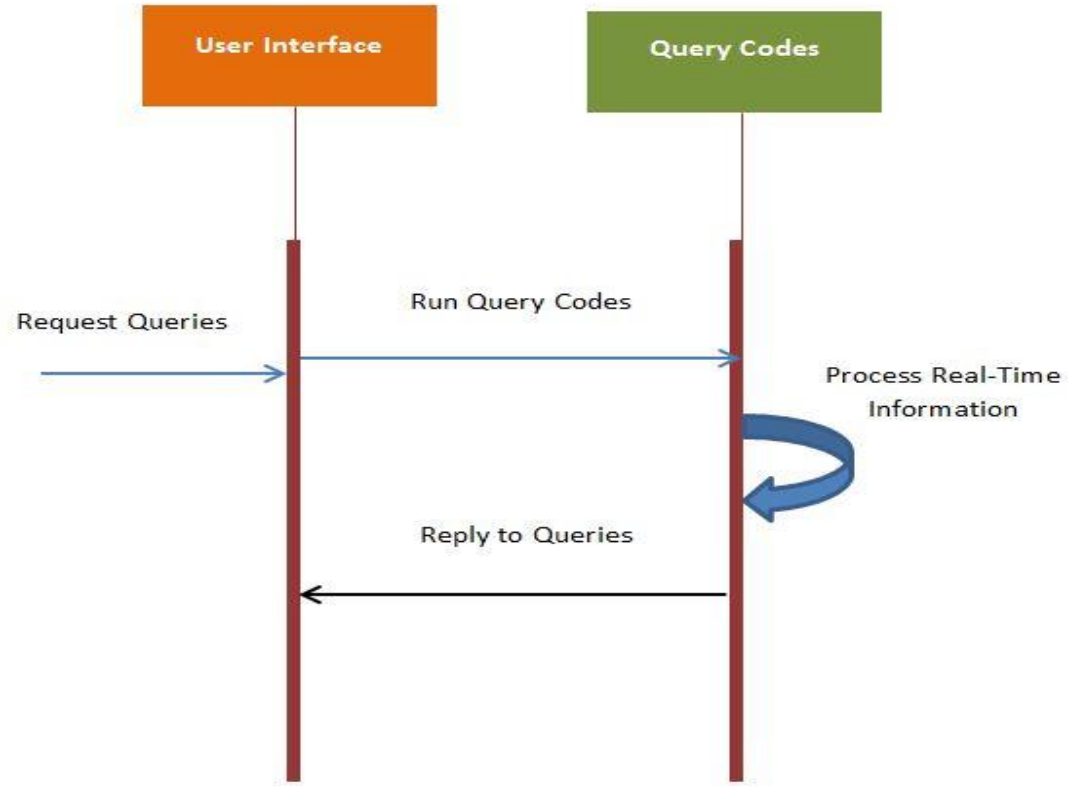


Figure 21: Queries

Contact us:

This allows a user to contact us. The user needs to enter his/her email id and name. They can post any questions or add suggestions. These go to the email id created for the website.

All the use cases mentioned above can be summarized as shown in the diagram below:

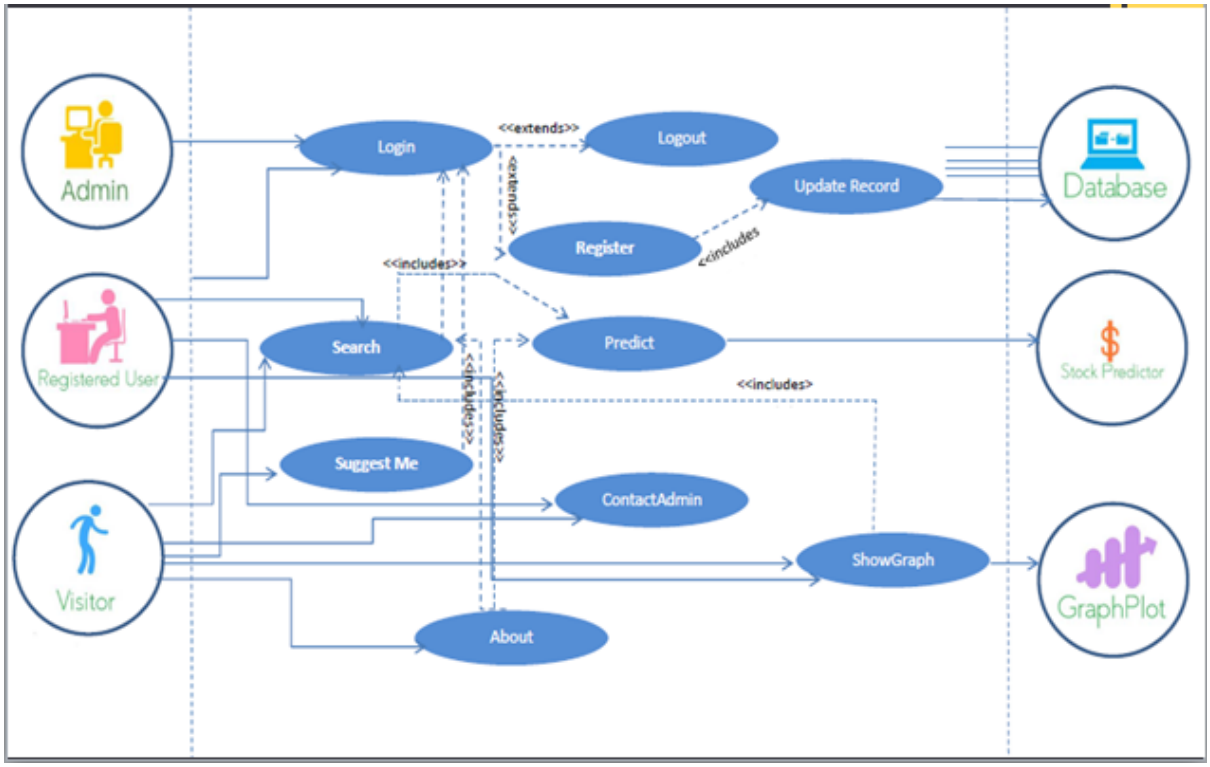


Figure: Use Case Diagram

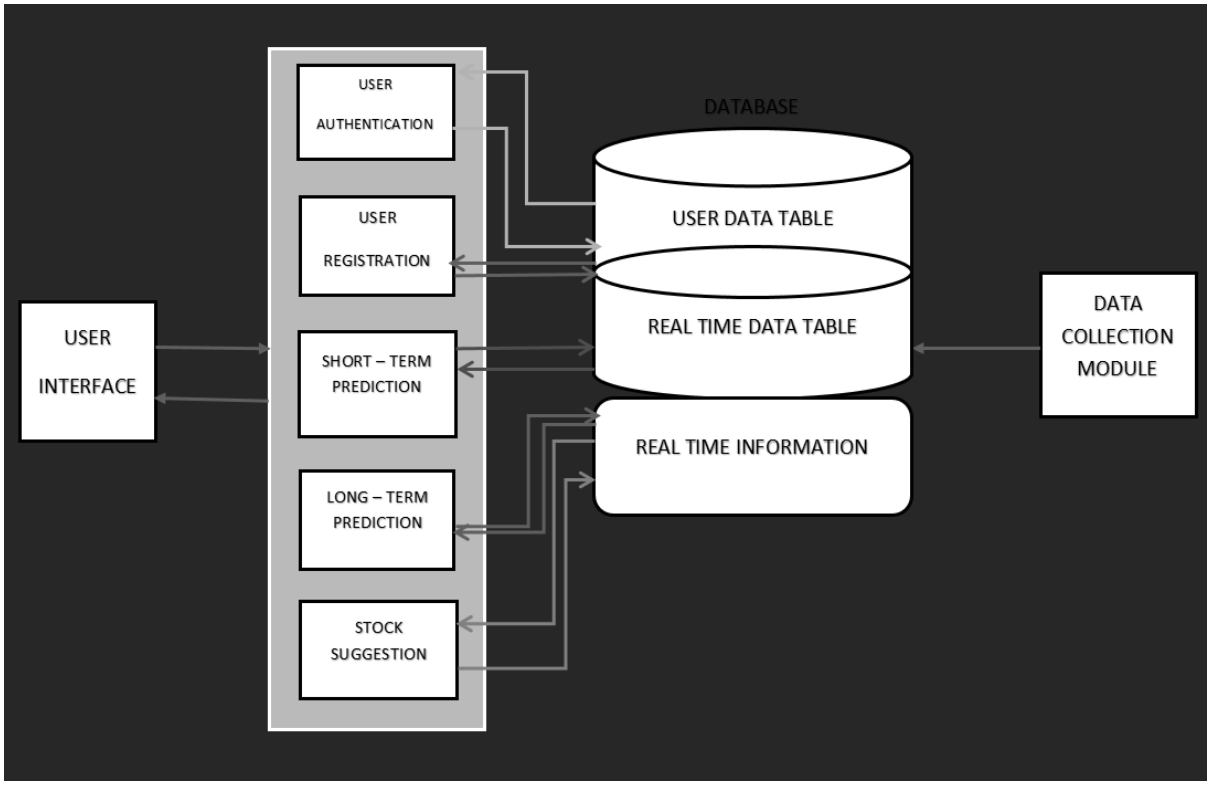
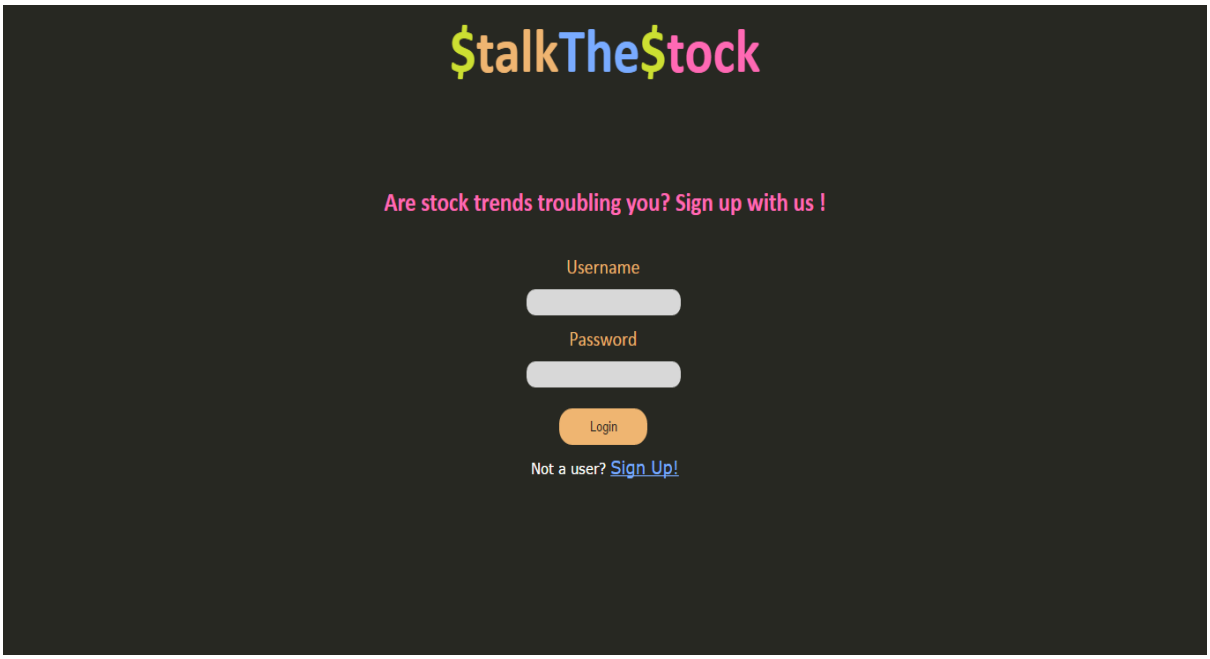


Figure 22: Global architecture

10. USER INTERFACE:

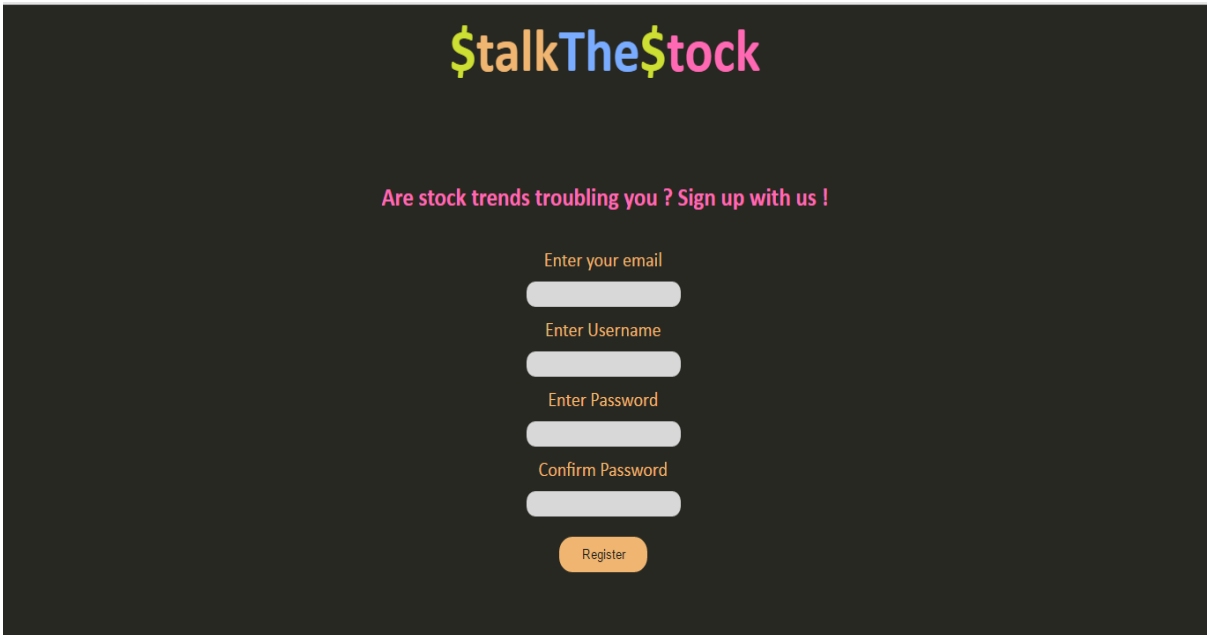
The UI of our web application has the following webpages:

Login Page: The user can have the access by logging in with User name and Password. If the user is not already registered he need to Sign up. If the user name is not already available in the database or the password is incorrect an alert dialog box is displayed.



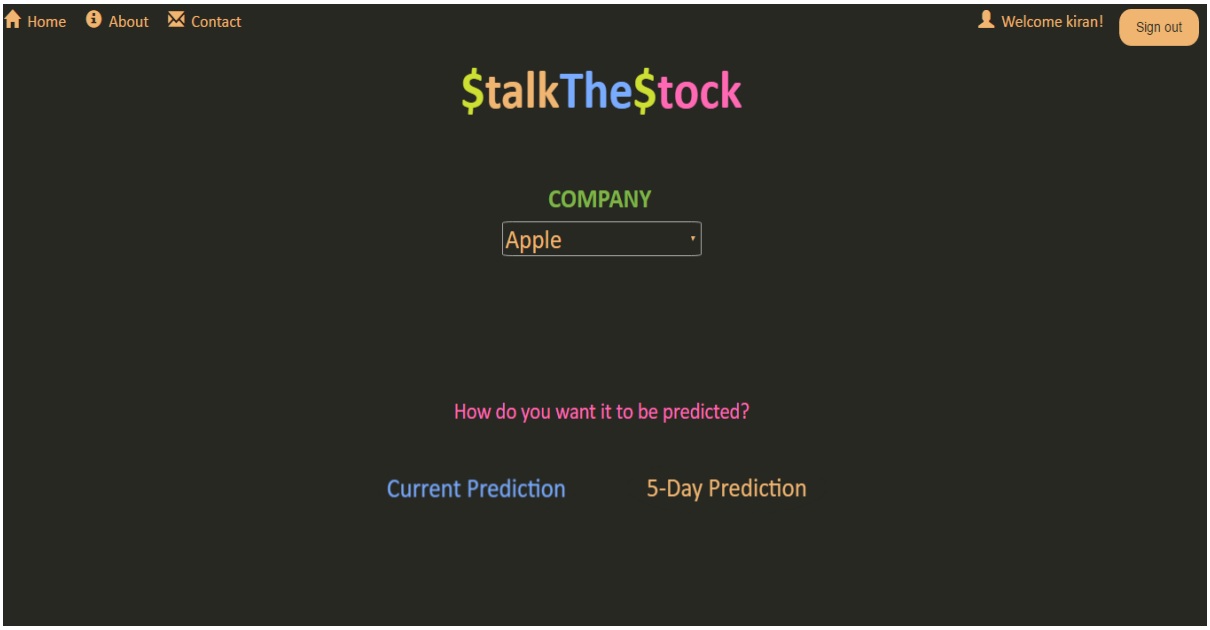
Register Page:

The user need to give his mail id, user name and password and should reconfirm the password to get registered. If both the passwords do not match then it shows an alert box. Once the user gets registered then it directs the user to login with his user name.



Home Page:

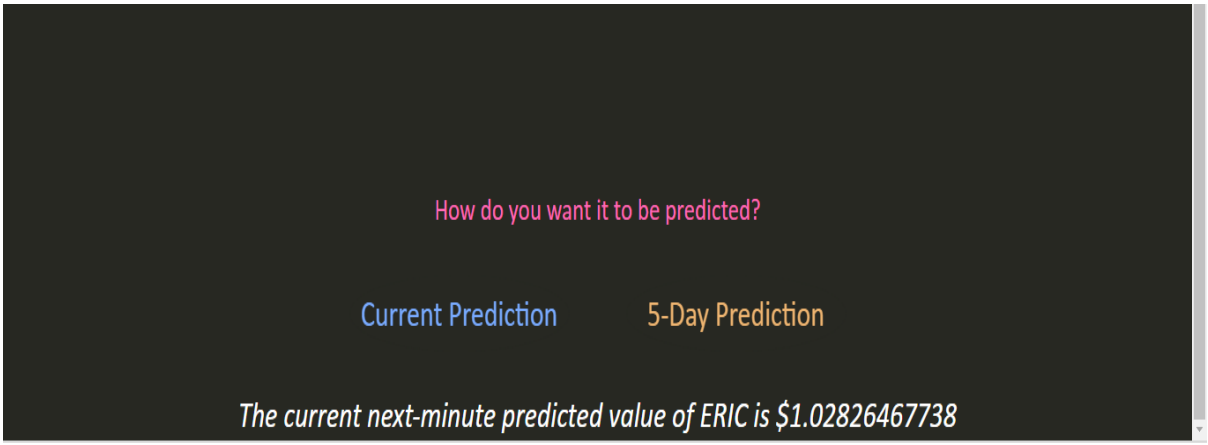
The user is directed to the home page from the login page. It has Predict, Suggest, Show graph and Query buttons.



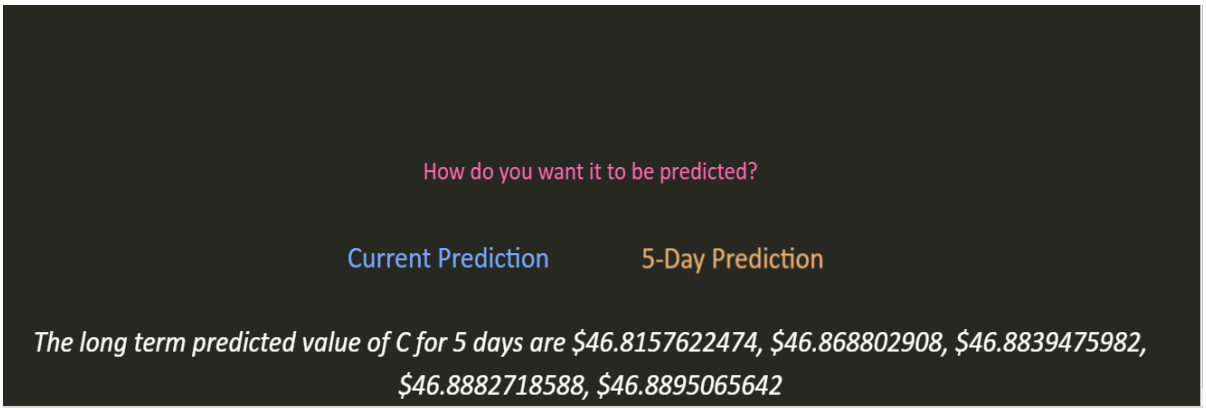
Predict:

The user is directed to the page which asks the user to select the name of the company and choose the type of prediction i.e, current prediction and 5-day prediction.

1-Day Prediction: It predicts the values for the next minute.



5-Day Prediction: It predicts the values for next 5 days

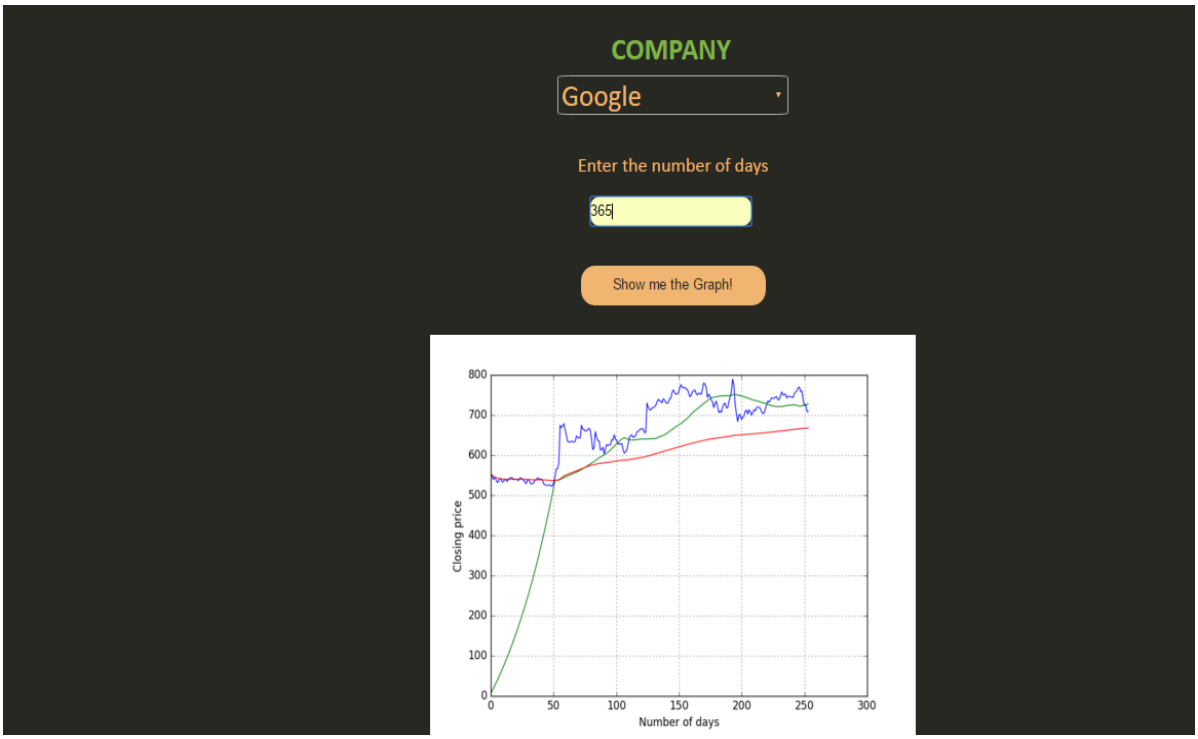


Suggest:

The user is directed to a page which suggests the user to either buy or sell the stock based on the trend of the predicted values for all the stocks.

Show Graph:

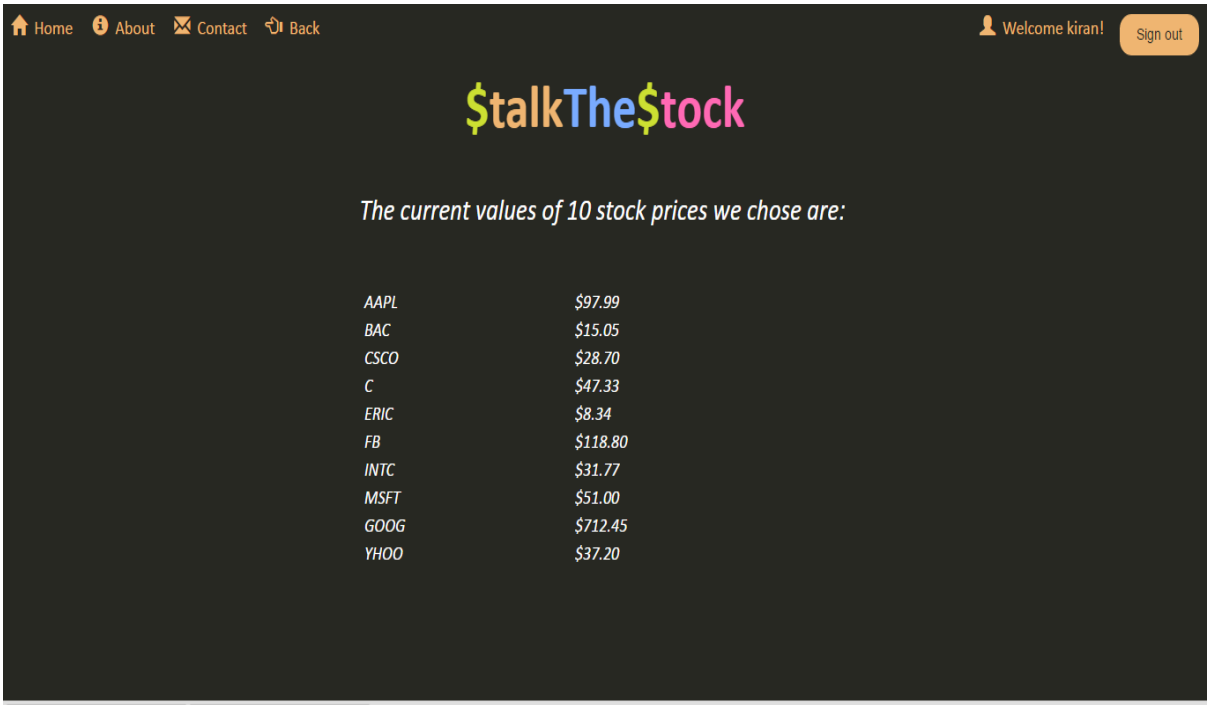
It asks the user to select the required ticker and the number of days he need the graph. It displays the graph showing trend line, SMA (Simple Moving Average) and EMA (Exponential Moving Average)



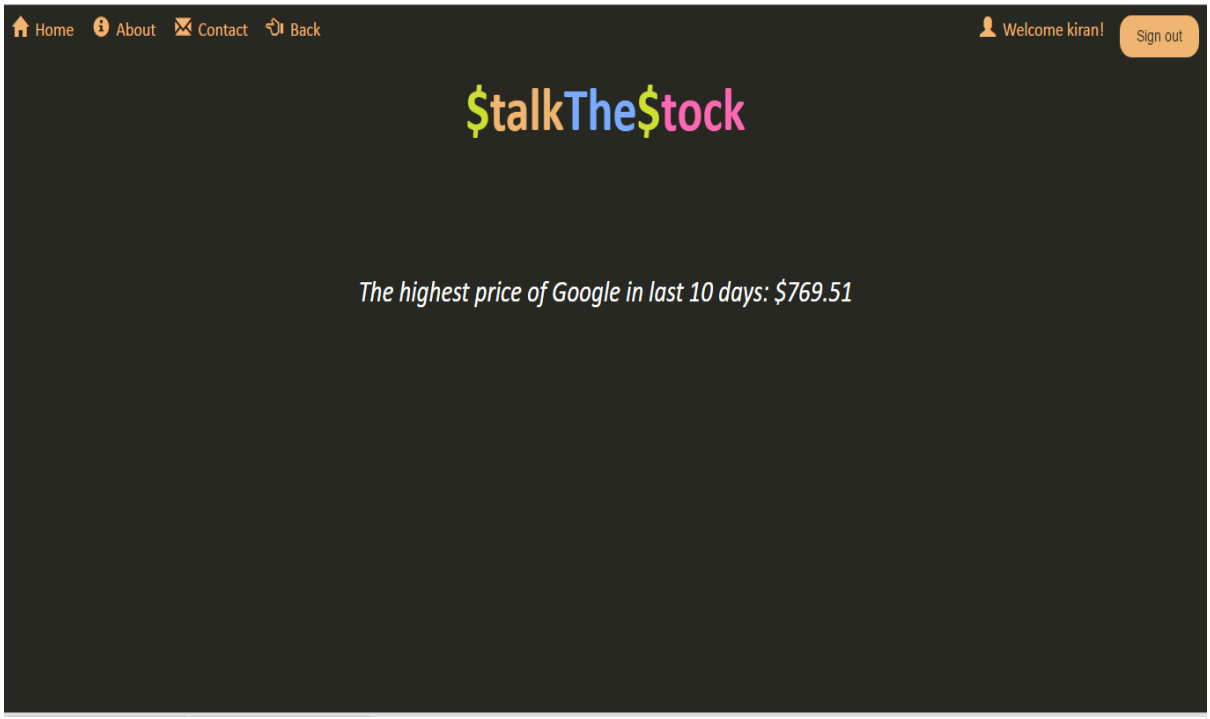
Queries:

The user is directed to a page which asks him to select the query and redirects to the result page.

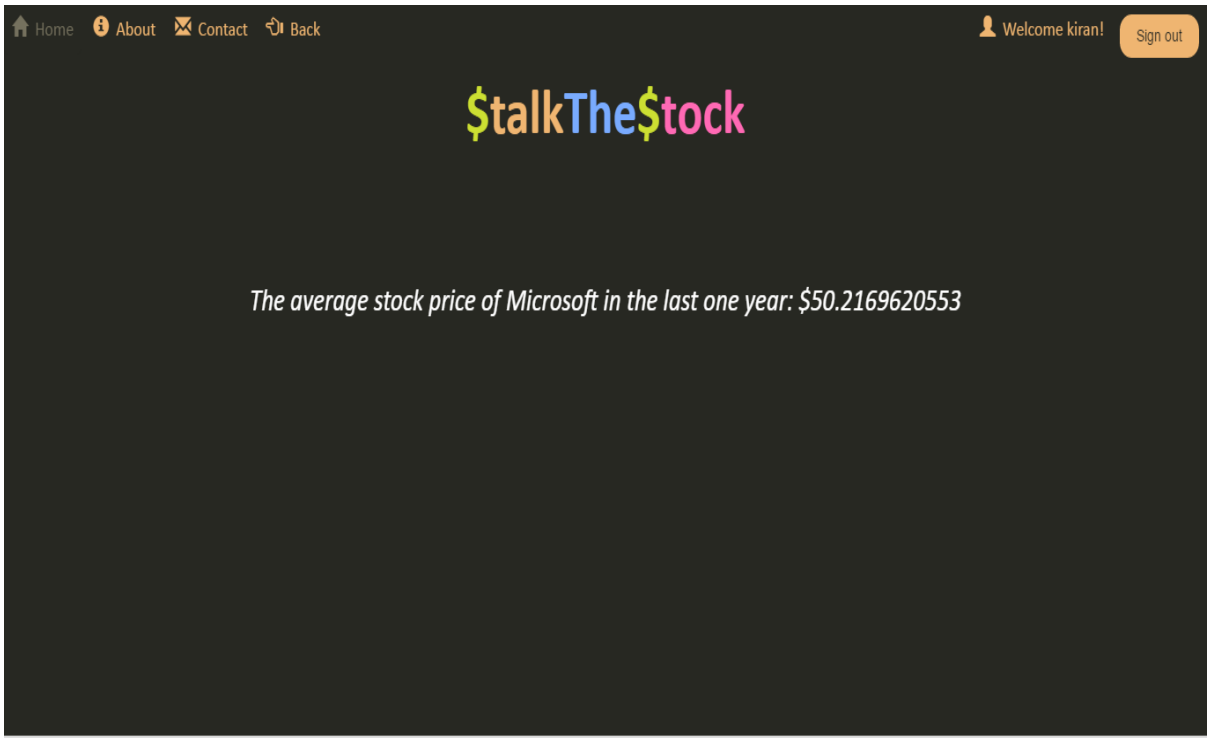
Query 1: Shows the list of all companies in the database along with their latest stock price.



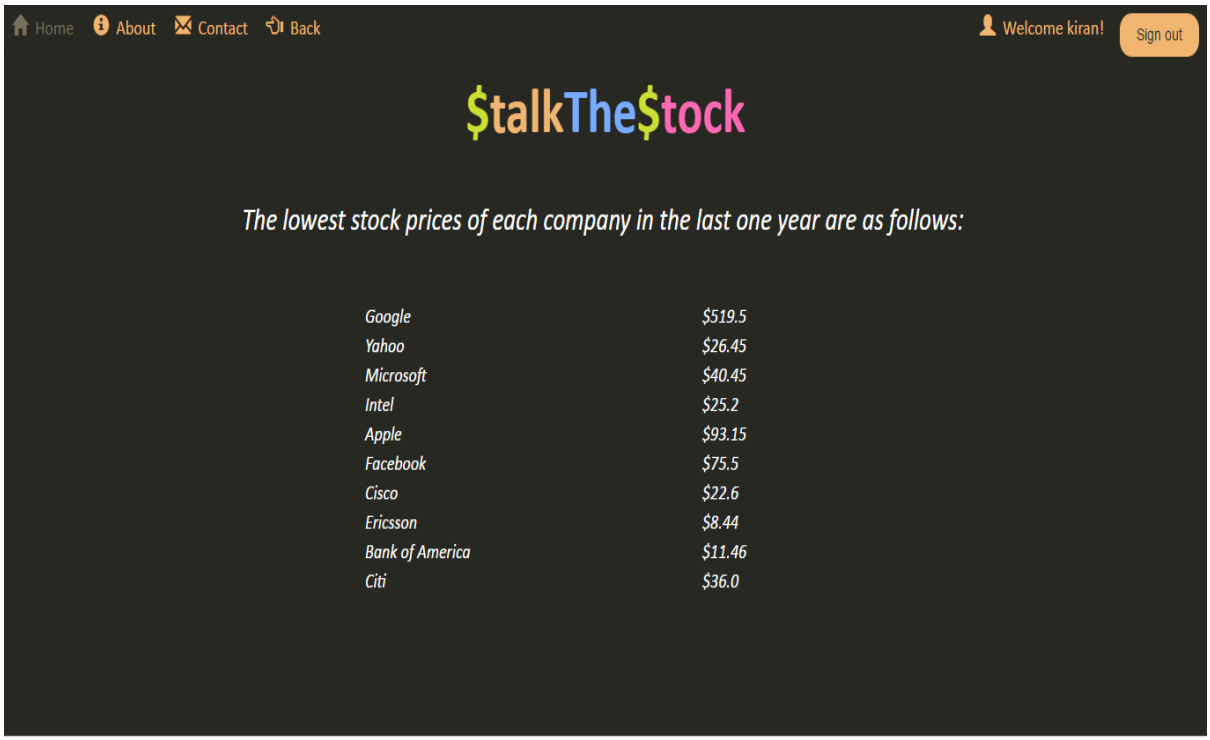
Query 2: The highest stock price of Google in the last ten days



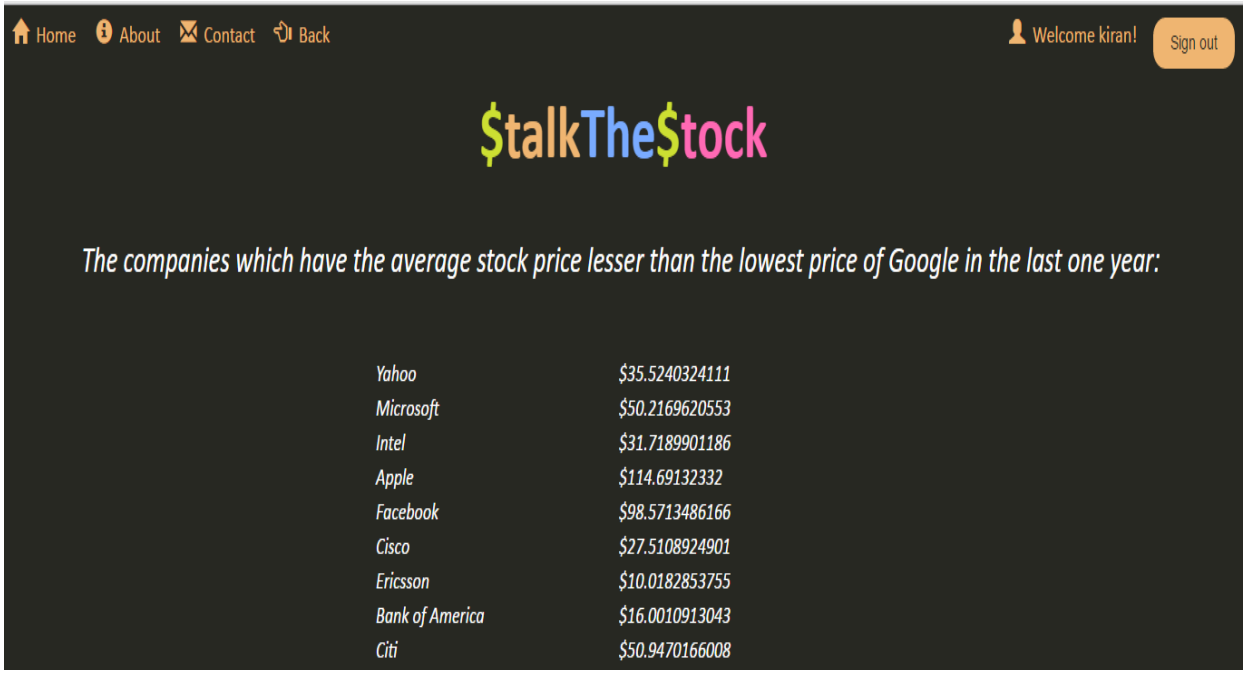
Query 3: Average stock price of Microsoft in the latest one year



Query 4: Lowest stock price for each company in the latest one year



Query 5: List of ids of companies along with their name who have the average stock price lesser than the lowest of Google in the latest one year.



Contact:

This page enables the user to contact us by submitting any his query. We access the query and responds to him through email.

About:

This page gives the information about the basic prediction algorithms we used to predict the stock prices.

11. CONCLUSION:

In this project, we are developing a web application to predict stock prices and their movement using Bayesian predictor for short term stock prediction and ANN (Artificial Neural Networks) and SVM (Support Vector Machine) predictor for long term stock prediction and predict the stock values in advance. The additional features of our website is to allow the user to contact us if they have any queries. They were sent an email notification with the predicted values each time they choose to show the predictions. We implemented a feature **Suggest** which would suggest the user by displaying the trend (Buy/Sell) for all the companies. This would greatly help the users to make decisions by looking at the predicted trend of the stock.

In the beginning we faced many technical challenges with regression models. It involved gathering and processing of large amount of data which demanded utmost precision and accuracy. However, we managed to implement the model successfully with a good understanding of the model. The step by step procedure involved in software engineering development style has helped us to build these use cases from a conceptual point to a realized module.

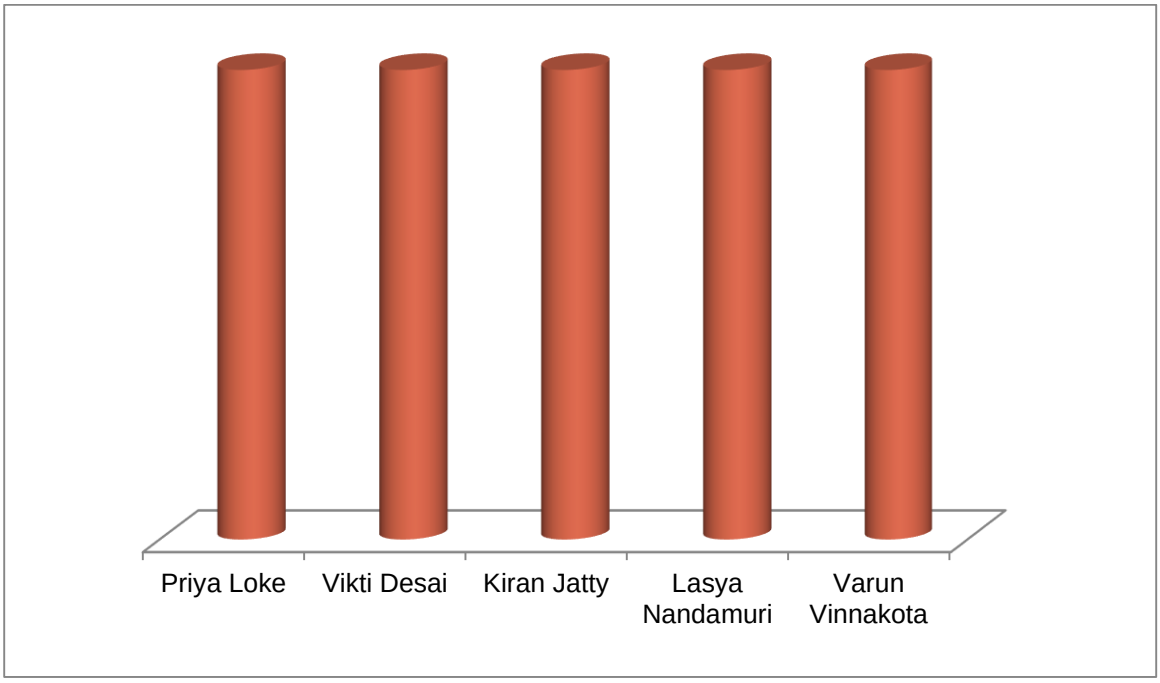
Though we found the integration task to be hard in the beginning, we were able to overcome it by hard work, team work and perseverance. The course Software Engineering has been of great help to us in terms of developing good knowledge of languages like Python, HTML, CSS and PHP.

12. FUTURE WORK:

- The system can be integrated with price alert feature.
- With this feature whenever the stock price value falls in a particular range or reaches a particular value a notification e-mail or text message will be sent to the user.
- The user can be prompted with the two companies best to Buy and best to Sell by looking and the trends of various stocks and choosing the one that best fits the above two queries.
- Having a blog on the latest trends of the Stocks would be helpful to the users

13. CONTRIBUTION:

All the group members contributed equally in the completion of the project. Each member of the team contributed equally in each module of the project including the GUI, prediction algorithms and integration.



14. REFERENCES:

1. http://finance.yahoo.com/
2. http://php.net/manual/en/function.explode.php
3. http://php.net/manual/en/function.implode.php
4. http://php.net/manual/en/reserved.variables.get.php
5. http://www.w3schools.com/php/default.asp
6. https://en.wikipedia.org/
7. http://www.investopedia.com/