

## **I. Abstract**

**Objective:** The project provides a simple tool for users to view the prediction of stock values of various companies listed on the stock market.

**Technology or Method:** Linear regression method is used to predict the stock market shares. Although the prediction can be further improved upon by using different classification algorithms.

**Results and Findings:** The tool currently displays the prediction details of the stock listed on the Bombay Stock Exchange(BSE) and the stocks listed on the National Stock Exchange(NSE).

**Conclusions:** In this software, we described the application of linear regression to the task of stock prediction. We described the theory behind linear regression and its salient features.

## **II. Introduction**

**STOCK PREDICTION TOOLKIT** helps to determine the future value of a company's stock traded on exchange.

The customer has to provide a valid username and password for signing into the toolkit predictor. If by any case, he/she is new to the predictor toolkit, he/she has to create a new account for signing in. Further changes will be reflected onto the database. The customer then has to choose between a set of stock symbols i.e. the stock in which the customer has invested in or is willing to invest. The time period for which the customer wants to hold on to the stock is then entered. An approximate future value is estimated by the predictor toolkit. Simultaneously, a graph is plotted for the ease of the customer in order to visualize the risks involved by investing in the respective company's share.

For the prediction of future value of a company's stock the concept of linear regression has been used in **STOCK PREDICTION TOOLKIT**. Linear regression is the most basic type of regression and commonly used predictive analysis. The overall idea of regression is to examine two things: (1) does a set of predictor variables do a good job in predicting an outcome variable? Is the model using the predictors accounting for the variability in the changes in the dependent variable? (2) Which variables in particular are significant predictors of the dependent variable?

For making the **STOCK PREDICTION TOOLKIT** user friendly, Graphical User Interface (GUI) has been designed so that the users can easily interact with the software.

## **III. Background**

A stock market, equity market or share market is the aggregation of buyers and sellers (a loose network of economic transactions, not a physical facility or discrete entity) of stocks (also called shares), which represent ownership claims on businesses; these may include securities listed on a public stock exchange as well as those only traded privately. Examples of the latter include shares of private companies which are sold to investors through equity crowd funding platforms. Stock exchanges list shares of common equity as well as other security types, e.g. corporate bonds and convertible bonds.

A stock exchange is a place where, or an organization through which, individuals and organizations can trade stocks. Many large companies have their stock listed on a stock exchange. This makes the stock more liquid and thus more attractive to many investors. It may also act as a guarantor of settlement.

Trade in stock markets means the transfer for money of a stock or security from a seller to a buyer. This requires these two parties to agree on a price. A potential buyer bids a specific price for a stock, and a potential seller asks a specific price for the same stock. Buying or selling at the market means you will accept any ask price or bid price for the stock. When the bid and ask prices match, a sale takes place, on a first-come, first-served basis if there are multiple bidders or askers at a given price.

The National Stock Exchange of India Limited (NSE) is the leading stock exchange of India, located in Mumbai. NSE was established in 1992 as the first demutualized electronic exchange in the country. NSE was the first exchange in the country to provide a modern, fully automated screen-based electronic trading system which offered easy trading facility to the investors spread across the length and breadth of the country. National Stock Exchange has a total market capitalization of more than US\$1.41 trillion, making it the world's 12th-largest stock exchange as of March 2016. NSE's flagship index, the NIFTY 50, the 51 stock index (50 companies with 51 securities inclusive of DVR), is used extensively by investors in India and around the world as a barometer of the Indian capital markets. However, only about 4% of the Indian economy / GDP is actually derived from the stock exchanges in India.

The Bombay Stock Exchange (BSE) is an Indian stock exchange located at Dalal Street, Kala Ghoda, Mumbai (formerly Bombay), Maharashtra, India. Established in 1875, the BSE is Asia's first stock exchange. It claims to be the world's fastest stock exchange, with a median trade speed of 6 microseconds. The BSE is the world's 11th largest stock exchange with an overall market capitalization of \$1.43 Trillion as of March, 2016. More than 5500 companies are publicly listed on the BSE. Unlike countries like the United States where 70% of the GDP is derived from larger companies and

the corporate sector, the corporate sector in India accounts for only 12-14% of the national GDP.

#### **IV. Objectives**

Not everyone can be a financial expert or have a financial advisor to help them. But this software provides a way for everyone to have their own financial advisor, at least about stock markets and investments in stock markets.

The stock market is one of the most important ways for companies to raise money, along with debt markets which are generally more imposing but do not trade publicly. This allows businesses to be publicly traded, and raise additional financial capital for expansion by selling shares of ownership of the company in a public market. The price of stocks and other assets is an important part of the dynamics of economic activity, and can influence or be an indicator of social mood. The purpose of a stock exchange is to facilitate the exchange of securities between buyers and sellers, thus providing marketplace. The exchanges provide real-time trading information on the listed securities, facilitating price discovery.

Thus, in STOCK PREDICTION TOOLKIT the future value of a company's stock is predicted. The prediction of future value is done using the concept of linear regression. The software helps the users to decide whether the investment made by them on the respective company's share is fruitful or not. The predictor also helps to visualize the risks involved in investing in the shares.

#### **V. Methodology**

We use the method of linear regression to forecast the stock market values.

Regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modelling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors').

Many techniques for carrying out regression analysis have been developed. Familiar methods such as linear regression and ordinary least squares regression are parametric, in that the regression function is defined in terms of a finite number of unknown parameters that are estimated from the data. Nonparametric regression refers to techniques that allow the regression function to lie in a specified set of functions, which may be infinite-dimensional.

In our approach to predict stock market values for the upcoming days, we use the method of linear regression. Linear regression is an approach for modelling the

relationship between a scalar dependent variable  $y$  and one or more explanatory variables (or independent variables) denoted  $X$ . The case of one explanatory variable is called simple linear regression. In simple linear regression, there is only one independent variable and only one dependent variable. So, there is only one dependency as opposed to multivariate linear regression. Linear regression is better as the points in the data set fit better. In  $n$  non-linear regression, the fitting of the points might be ambiguous. Linear regression can be basically used to predict the stock market values by analysing the behaviour of the market in the past, i.e., by looking at the past close values at each date.

We basically do the opposite of plotting a graph when an equation is given for a dependent variable ' $y$ ' and an independent variable ' $x$ '. While plotting a graph, calculate the values of  $y$  based on the equation given and plot a graph. In linear regression, we basically do the reverse. We try to derive an equation from a given set of values  $(x_i, y_i)$ .

- We plot the points in our data set  $(x_0, y_0)$ , and so on.
- We draw a line in the graph such that maximum number of plotted points are included in that line.
- The equation of the graph gives us the relation between the independent variable ' $x$ ' and dependent variable ' $y$ ', which essentially gives us the slope and intercept of the line.
- After finding the equation we can find values of  $y$  for a given value  $x$  which is not available in the original set of values which we used to derive the equation.
- After we get the value of  $y$ , we can use it to calculate the predicted percentage change in the value of  $y$  by comparing it other values.

For stock markets, the independent variable is the date and the dependent variable is the close value on that date. For a given stock market, the future value can be predicted by the above methodology.

Similarly, we can apply the same method to all stock markets to get their future close values and predicted percentage changes.

## VI. Implementation

In our code to forecast the stock market values, we used

- crawler to get historical stock market data of different stocks
- linear regression to predict the future values
- pyqt to design the GUI of the software

### A. *Getting the historical data of different stocks from the web*

First, we stored the stock symbols of the listed companies in a csv file. Using the stock symbols as the keyword, we downloaded the historical data of the stocks from the internet using urllib package. We begin by reading stock symbols from the “Tickr” file. The stock symbols encountered are taken up as arguments to be used by urllib module to find the past trading history of the stock. The same process was followed for other stocks and their data was downloaded from the internet, which gave us our required data set.

```
import urllib as ur
```

```
list_stock = {}
```

```
base_url =  
"http://ichart.finance.yahoo.com/table.csv?s="
```

```
def make_url(ticker_symbol):
```

```
    return base_url + ticker_symbol
```

```
output_path = "C:/Project"
```

```
def make_filename(ticker_symbol,  
directory="BSE"):
```

```
    return output_path + "/" + directory + "/" +  
    ticker_symbol + ".csv"
```

```
def pull_historical_data(ticker_symbol,  
directory="BSE"):
```

```
    try:
```

```
        ur.urlretrieve(make_url(ticker_symbol),  
make_filename(ticker_symbol, directory))
```

```
except ur.ContentTooShortError as e:
```

```
    outfile = open(make_filename(ticker_symbol,  
directory), "w")
```

```
    outfile.write(e.content)
```

```
    outfile.close()
```

```
def get_data1(file_name):
```

```
    data = pd.read_csv(file_name)
```

```
    for tickr in zip(data['Tickr']):
```

```
        string = str(tickr)
```

```
        pull_historical_data(string[2:(len(string)-3)])
```

```
get_data1('BSE_stock.csv')
```

### B. *Using Linear Regression to predict the future value of a company's stock*

Linear Regression was implemented using the sklearn package in python. Scikit-learn has built-in methods to implement linear regression on a data set and to derive the equation from the data set. From the derived equation, we can get the required values of the slope and the intercept. Our independent variable was the date and the dependent variable was the close value on that date. The date originally in ‘dd-mm-yy’ cannot be used directly as a variable. The date was then indexed as the oldest date in the record being 01 and the subsequent dates’ index were incremented by 1. The dependent variable was similarly taken to be the adjusted close value. The adjusted close value of a stock is its closing price on any given day of trading that has been amended to include any distributions and corporate actions that occurred at any time prior to the next day’s open. That gave us a data set with the indexed date value as the actual independent variable and the adjusted close value associated with the actual date as the actual dependent variable.

The date values and their corresponding close values were read from the .csv files which were downloaded earlier from the internet. Those values were then stored and used to draw the required line.

```
# Function for Fitting our data to Linear model
```

```
def
linear_model_main(X_parameters,Y_parameters
,predict_value):
```

```
    # Create linear regression object

    regr = linear_model.LinearRegression()

    regr.fit(X_parameters, Y_parameters)

    predict_outcome = regr.predict(predict_value)

    predictions = {}

    predictions['intercept'] = regr.intercept_

    predictions['coefficient'] = regr.coef_

    predictions['predicted_value'] =
predict_outcome

    return predict_outcome
```

### *C. Plotting the graph of a specific stock over a range of 30 days*

The graph was plotted and shown using a combination of scikit-learn package and matplotlib package. The scikit-learn gets the data set and finds the equation of the line while matplotlib plots the points in the data sets and draws the line by taking the points set by scikit-learn package's linear model variable. The matplotlib module then takes the calculated value after a few days and shows the user the stock chart of the selected stock.

#### **# Function to show the results of linear fit model**

```
def show_linear_line(X_parameters,
Y_parameters):
```

```
    # Create linear regression object

    regr = linear_model.LinearRegression()

    regr.fit(X_parameters, Y_parameters)

    plt.scatter(X_parameters, Y_parameters, color
= 'blue')

    plt.plot(X_parameters,
regr.predict(X_parameters),color='red',linewidth
= 4)

    plt.xticks(())

    plt.yticks(())
```

```
plt.show()
```

The GUI of the software was built using the pyqt package. The user login page contains the following information. To register, the user needs to input a user name, an email in proper format and a password, which he needs to re-enter again in another column for verification. If the user is already registered with our software, he can login and check the increase in stock prices during the last few days. The user can view a stock's predicted change after a few days individually through our software. From there, the user can get an abstract idea about the changes that might occur in the stock market in the upcoming few days.

The predicted values will have some level of inaccuracy. Further the date is from the current date, higher will be the inaccuracy of the predicted value. For the upcoming few days, it will remain up to 5%. After that, the inaccuracy or error can go higher.

## VII. Output & Results


### 1. Login Screen

The program provides a login screen at the beginning of the program for the user to login.

If the user is new and hasn't yet registered with the software, he/she needs to register with the software. The right side of the login screen will provide the necessary details required to register with the software. The user enters his/her username, his email id in proper format (e.g. kushanshekharr007@gmail.com) and then assigns the account with a password. The password needs to be re-entered for verification in another column of the correct password in case the user might have committed a typing

error during the assigning of the password. After all the above steps have been followed, the user will be registered with the software. A dialog box confirming the same will be displayed.

Now the user is already registered with the program. On the left side of the login window, the user has to enter his user id and password correctly to use the software. Once the correct user id and password is entered, the user is logged into the software when he clicks on the Log In button. The same is confirmed when a 'Successfully logged in' message is displayed in the text area on the left side.

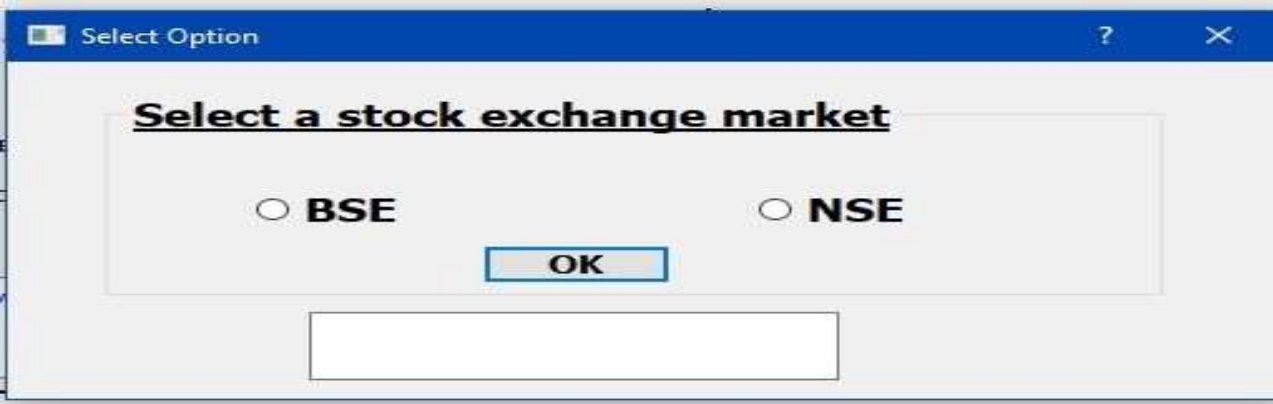


The screenshot shows a Windows application window titled "Stock Market Prediction Tool". The window is divided into two main sections by a vertical line. The left section is titled "Already a user? Login" and contains fields for "Username:" (with the value "imdebashishpatnaik@gmail.com") and "Password:" (with masked characters "\*\*\*\*\*"). Below these fields is a "Log In" button. A text area below the button displays "Successfully logged in.". The right section is titled "Create an account" and contains fields for "Full Name:", "E-mail:", "Password:", and "Re-Type Password:". Below these fields is a "Sign Up" button. At the bottom of the window, there is a promotional message: "Want to invest in stock market and don't know where to invest in. We're here to help you, join us and get the best stocks to invest in."

### 2. Select the stock exchange market

Then the software asks the user to select which stock market's information he would like. In the current program, the user is given two options:

BSE and NSE. The user selects one of them and the program displays the information of the stocks in the selected market.

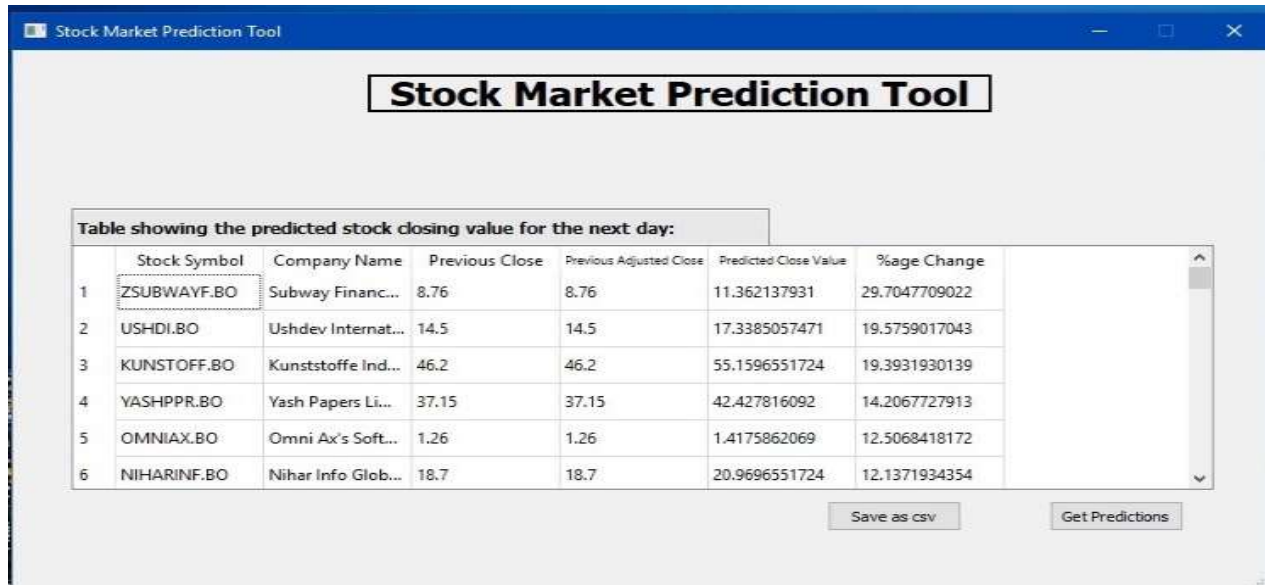


The screenshot shows a Windows dialog box titled "Select Option". The dialog box has a title bar with a question mark icon and a close button. The main content area is titled "Select a stock exchange market" and contains two radio button options: "BSE" and "NSE". Below these options is an "OK" button. At the bottom of the dialog box, there is an empty text input field.

### 3. Displaying Stock Market Data with Next Day Prediction

After the user selects a stock market and clicks on the OK button, the stocks' information will be displayed to the

user. The information is displayed in tabular format and is sorted in decreasing order of predicted percentage change. Only the stocks available in the selected market are displayed. The table contains the attributes stock symbol, the company name, the previous close value, the previous adjusted close value, the predicted close value for the next day and the predicted percentage change.



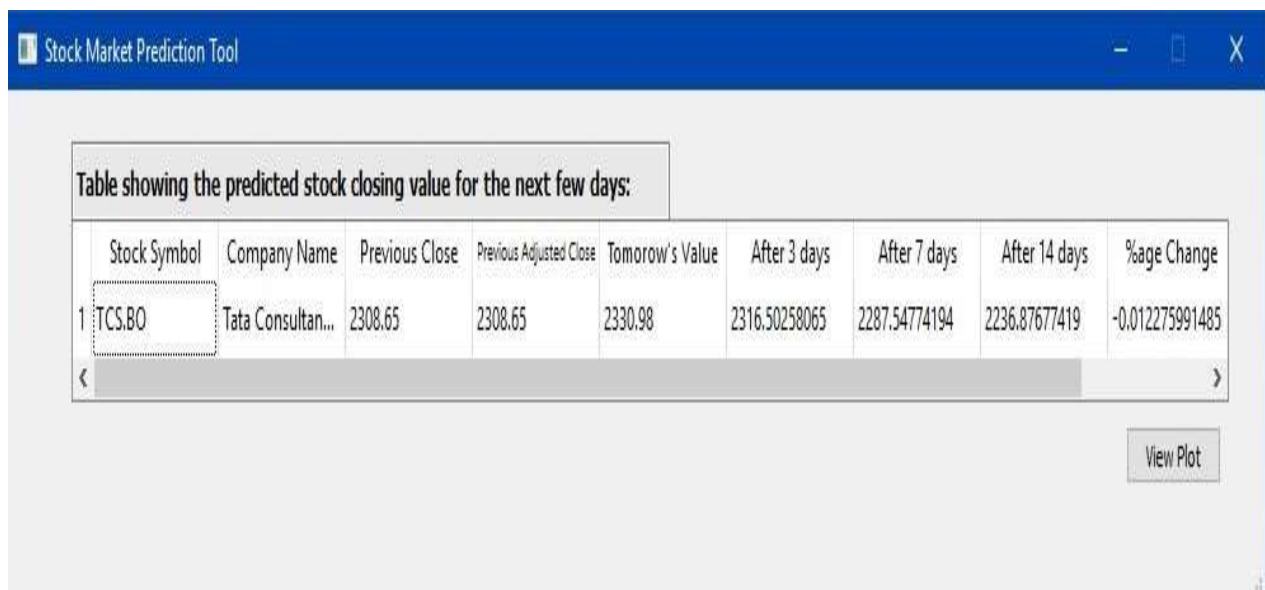
The screenshot shows a window titled "Stock Market Prediction Tool". Inside, there is a table with the caption "Table showing the predicted stock closing value for the next day:". The table has 7 columns: Stock Symbol, Company Name, Previous Close, Previous Adjusted Close, Predicted Close Value, and %age Change. There are 6 rows of data. Below the table are two buttons: "Save as csv" and "Get Predictions".

	Stock Symbol	Company Name	Previous Close	Previous Adjusted Close	Predicted Close Value	%age Change
1	ZSUBWAYF.BO	Subway Financ...	8.76	8.76	11.362137931	29.7047709022
2	USHDI.BO	Ushdev Internat...	14.5	14.5	17.3385057471	19.5759017043
3	KUNSTOFF.BO	Kunststoffe Ind...	46.2	46.2	55.1596551724	19.3931930139
4	YASHPPR.BO	Yash Papers Li...	37.15	37.15	42.427816092	14.2067727913
5	OMNIAX.BO	Omni Ax's Soft...	1.26	1.26	1.4175862069	12.5068418172
6	NIHARINF.BO	Nihar Info Glob...	18.7	18.7	20.9696551724	12.1371934354

### 4. Showing the stock prediction for several days of a specific company

Here, the user is provided with two options. He can either save the information depicted in tabular format or he can get predictions for a specific stock for the upcoming few days.

To get predicted values for the upcoming few days of a stock, the user has to select the row in which the stock's information is given. Then the user should click on 'Get Predictions' button. A new window will be opened which will show the user the predicted values of the stock in the upcoming few days.



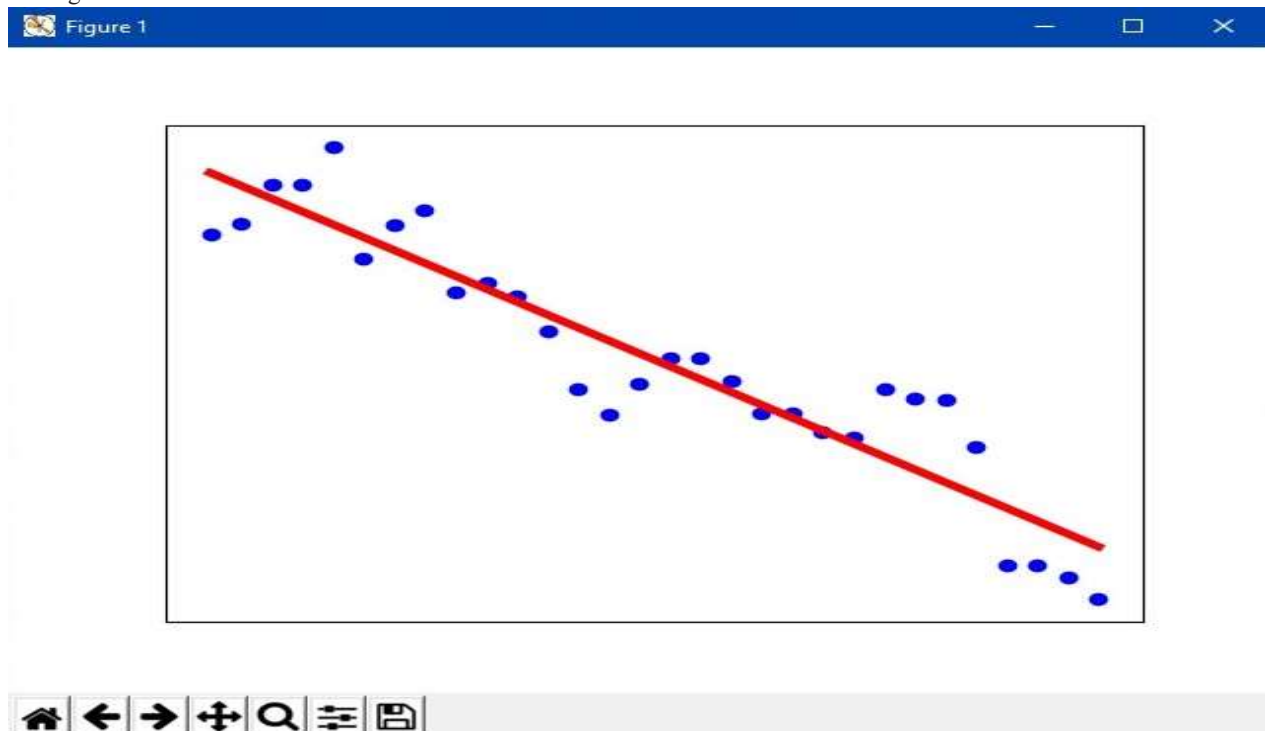
The screenshot shows a window titled "Stock Market Prediction Tool". Inside, there is a table with the caption "Table showing the predicted stock closing value for the next few days:". The table has 9 columns: Stock Symbol, Company Name, Previous Close, Previous Adjusted Close, Tomorrow's Value, After 3 days, After 7 days, After 14 days, and %age Change. There is 1 row of data. Below the table is a "View Plot" button.

	Stock Symbol	Company Name	Previous Close	Previous Adjusted Close	Tomorrow's Value	After 3 days	After 7 days	After 14 days	%age Change
1	TCS.BO	Tata Consultan...	2308.65	2308.65	2330.98	2316.50258065	2287.54774194	2236.87677419	-0.012275991485

##### 5. Viewing the plot of the linear regression line

The user can then select the button 'View Plot' to view the regressed line and how the stock has behaved in the

past and how it's going to behave in the future. The points in the graph depict the behaviour of the stock.



##### 6. Saving the output in csv format

The user can also save the information available about the stock market by clicking on the 'Save As CSV' button.

It will save the information in a tabular format in csv format which can be accessed using MS-Excel or Libre Editor or any software which supports csv format files.



The program correctly predicts the future stock values for the upcoming days fairly accurately within allowed limits.

So, if the market does not show bad behaviour, the user can follow and get some abstract idea about the markets he might want to invest in.

## **VIII. Conclusion**

In this software, we described the application of linear regression to the task of stock prediction. We described the theory behind linear regression and its salient features.

The predicted future value obtained by using the technique of linear regression was fairly accurate. As it is fairly evident from Figure1, the prediction is fairly accurate unless there is a huge and sudden variation in the actual data. On the other hand, this also proves the hypothesis that the stock markets are actually unpredictable.

Thus, we can say that Linear regression is an effective tool for stock market prediction and can be used on real world datasets.

## **IX. Implications & Recommendations**

### **Implications:**

We can predict the stock prices for the upcoming few days with an average absolute error of 5%. Although we correctly forecast the stocks within an allowed limit, we assumed various things like no abnormal behaviour in the market, no unforeseen circumstances in the future etc.

### **Recommendations:**

- The project can be further modified by including the user's financial status. The program will suggest the user the best stock in which he can invest according to his financial status and needs.
- The program does not take into account any misbehaviour in the market in the future due to unforeseen circumstances like natural disaster, death, terrorist attacks, war, etc.
- The program can forecast stocks better if methods other than linear regression such as random forest binary classification, K nearest neighbours binary classification, SVM binary classification, etc. are also used and the best method is chosen from them.
- The program assumes that no unforeseen circumstances have occurred from the time where historical data has been considered for the forecasting.
- The program was a demo version. It can serve a much larger purpose by updating the stock every day or every time the user wants to access information about the stock.
- The program currently only includes Indian stock markets namely BSE and NSE. It can be further developed by adding more stock markets including markets centred outside India.



- The program can be made into a website for easy access of the software to users worldwide.
- In future implementations, we can use some kind of database like MySQL or SQLite to store and retrieve the data instead of depending upon csv files.
- Storing the data in a database will make the accessing and updating of data a much faster and easier process.
- By integrating our software with **A-LOGIN** which was developed by Group 9 of CSE-B(I) 4<sup>th</sup> Semester, we can automate the whole process of stock prediction.

## **X. References**

1. <https://stackoverflow.com>
2. <https://docs.python.org>
3. <https://wiki.python.org>
4. <https://www.w3schools.com>
5. <https://in.finance.yahoo.com>
6. <https://www.youtube.com>
7. <https://forum.qt.io>
8. <http://www.qtcentre.org>
9. <https://www.riverbankcomputing.com>
10. <http://scikit-learn.org>
11. <http://dataconomy.com>
12. <https://www.google.com/finance>
13. <https://www.investopedia.com>