STOCK PRICE PREDICTION AND ANALYSIS

A Project Work submitted in partial fulfillment of the

requirements for the degree of

Bachelor of Technology

in

Computer Science and Engineering

Submitted *By*

MD HAIDER ALI **16SCSE101818**

Under the Supervision of

Mr. Hradesh Kumar

Assistant Professor

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING GALGOTIAS UNIVERSITY, GREATER NOIDA – 201306 May 2020

DECLARATION

Project Title: **Stock Price Prediction and Analysis**

Degree for which the project work is submitted: **Bachelor of Technology in Computer Science and Engineering**

We declare that the presented project represents largely our own ideas and work in our own words. Where others ideas or words have been included, we have adequately cited and listed in the reference materials. The report has been prepared without resorting to plagiarism. We have adhered to all principles of academic honesty and integrity. No falsified or fabricated data have been presented in the report. We understand that any violation of the above will cause for disciplinary action by the Institute, including revoking the conferred degree, if conferred, and can also evoke penal action from the sources which have not been properly cited or from whom proper permission has not been taken.

Md Haider Ali

1613101389

Date: 17-05-2020

CERTIFICATE

It is certified that the work contained in this project entitled **"Stock Price Prediction & Analysis"** submitted by **Md Haider Ali (1613101389)** for the degree of Bachelor of Technology in Computer Science and Engineering is absolutely based on our own work carried out under my supervision and this project work has not been submitted elsewhere for any degree.

Mr. Hradesh Kumar

Assistant Professor

School of Computing Science and Engineering

Galgotias University

Greater Noida, UP, India

Date: 17-05-2020

ACKNOWLEDGEMENT

On the submission of our Synopsis report on "**Stock Price Prediction & Analysis**"

We would like to express our indebted gratitude and special thanks to our Assistant Professor Mr.Hradesh Kumar of Computing Science and Engineering who in spite of being extraordinarily busy, spare time for guidance and keep us on the correct path and allowing us to carry out our work in the previous one year. We truthfully appreciate and value his admired supervision and support from the start to the end of this project.

We were obliged to him for having helped us shape the trouble and providing insights towards the way out. We would like to offer our special thanks to for providing a solid backdrop for out studies and explore afterward. They have been great sources of motivation to me and I thank them from the core of my heart. Last but not the least I would like to thank each and every person who is involved directly or indirectly to make this project successful.

ABSTRACT

Nowadays, the stock market is attracting more and more people's notice with its high challenging risks and high return over. A stock exchange market depicts savings and investments that are advantageous to increase the effectiveness of the national economy. The future stock returns have some predictive relationships with the publicly available information of present and historical stock market indices. ARIMA is a statistical model which is known to be efficient for time series forecasting especially for short-term prediction. In this paper, we propose a model for forecasting the stock market trends based on the technical analysis using historical stock market data and ARIMA model. This model will automate the process of direction of future stock price indices and provides assistance for financial specialists to choose the better timing for purchasing and/or selling of stocks. The results are shown in terms of visualizations using python programming language. The obtained results reveal that the ARIMA model has a strong potential for short-term prediction of stock market trends.

Contents

1. Introduction

1.1Overall Description

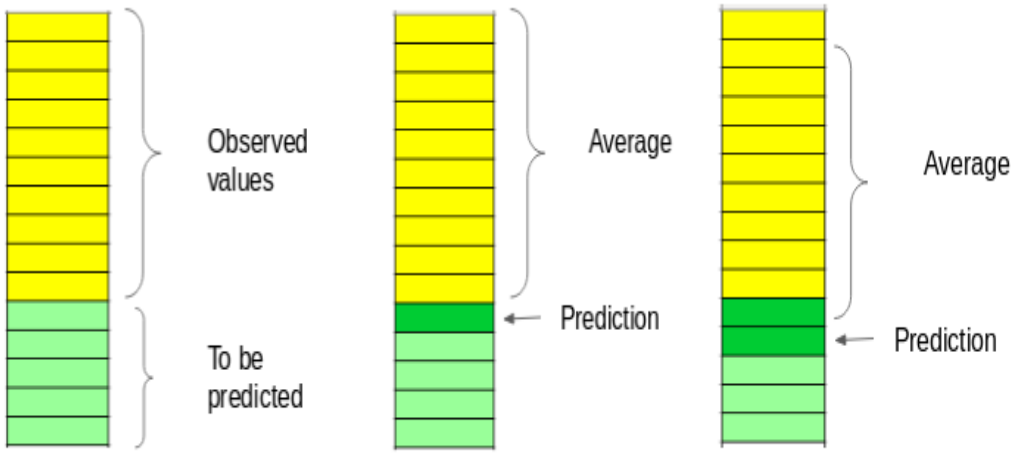
Stock Market prediction and analysis is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. Stock market is the important part of economy of the country and plays a vital role in the growth of the industry and commerce of the country that eventually affects the economy of the country. Both investors and industry are involved in stock market and wants to know whether some stock will rise or fall over certain period. The stock market is the primary source for any company to raise funds for business expansions. It is based on the concept of demand and supply. If the demand for a company's stock is higher, then the company share price increases and if the demand for company's stock is low then the company share price decrease. Another motivation for research in this field is that it possesses many theoretical and experimental challenges.

The Efficient Market Hypothesis (EMH), the hypothesis says that in an efficient market, stock market prices fully reflect available information about the market and its constituents and thus any opportunity of earning excess profit ceases to exist. One of the example of big exchange is New York Stock Exchange.Efficient Market Hypothesis was developed by Burton G. Malkiel in 1991. In Burton's hypothesis, he indicates that predicting or forecasting the financial market is unrealistic, because price changes in real world are unpredictable. All the changes in prices of financial market are based on immediate economic events or news. Investors are profit oriented, their buying or selling decisions are made according to most recent events regardless past analysis or plans. The argument about this Efficient Market Hypothesis has never been ended. So far, there is no strong proof that can verify if the efficient market hypothesis is proper or not.

1.1.1 Moving Average

'Average' is easily one of the most common things we use in our day-to-day lives. For instance, calculating the average marks to determine overall performance, or finding the average temperature of the past few days to get an idea about today's temperature – these all are routine tasks we do on a regular basis. So this is a good starting point to use on our dataset for making predictions.

The predicted closing price for each day will be the average of a set of previously observed values. Instead of using the simple average, we will be using the moving average technique which uses the latest set of values for each prediction. In other words, for each subsequent step, the predicted values are taken into consideration while removing the oldest observed value from the set. Here is a simple figure that will help you understand this with more clarity.



We will implement this technique on our dataset. The first step is to create a data frame that contains only the *Date* and *Close* price columns, then split it into train and validation sets to verify our predictions.

1.1.2 Problem Statement

There are multiple variables in the dataset – date, open, high, low, last, close, total_trade_quantity, and turnover.

- The columns *Open* and *Close* represent the starting and final price at which the stock is traded on a particular day.
- *High*, *Low* and *Last* represent the maximum, minimum, and last price of the share for the day.

- *Total Trade Quantity* is the number of shares bought or sold in the day and *Turnover (Lacs)* is the turnover of the particular company on a given date.

Another important thing to note is that the market is closed on weekends and public holidays.

1.1.3 Linear Regression

The most basic machine learning algorithm that can be implemented on this data is linear regression. The linear regression model returns an equation that determines the relationship between the independent variables and the dependent variable.

The equation for linear regression can be written as:

$$Y = \theta_1 X_1 + \theta_2 X_2 + \ldots \theta_n X_n$$

Here, x_1, x_2,….x_n represent the independent variables while the coefficients θ_1, θ_2, …. θ_n represent the weights. You can refer to the following article to study linear regression in more detail:

For our problem statement, we do not have a set of independent variables. We have only the dates instead. Let us use the date column to extract features like – day, month, year, mon/fri etc. and then fit a linear regression model.

1.1.4 Stock Analysis

Definition of Stock Analysis

Stock analysis is the evaluation of a particular trading instrument, an investment sector, or the market as a whole. Stock analysts attempt to determine the future activity of an instrument, sector, or market.

1.1.5 Technical Analysis

The second method of stock analysis is technical analysis. Technical analysis focuses on the study of past market action to predict future price movement. Technical analysts analyze the financial market as a whole and are primarily concerned with price and volume, as well as the demand and supply factors that move the market. Charts are a key tool for technical analysts as they show a graphical illustration of a stock's trend within a stated time period. For example, using a chart, a technical analyst may mark certain areas as a support or resistance level. The support levels are marked by previous lows below the currenttrading price, and the resistance markers are placed at previous highs above the current marketprice of the stock. A break below the support level would indicate a bearish trend to the stock analyst, while a break above the resistance level would take on bullish outlook.

Technical stock analysis is effective only when the price trend analysed is influenced by supply and demand forces. When outside factors are involved in a price movement, analysing stocks using technical analysis may not be successful. Examples of factors, other than supply and demand, that can affect a stock price includes stock splits, mergers, dividend announcements, a class action lawsuit, death of a company's CEO, a terrorist attack, accounting scandals, change of management, monetary policy changes, etc.

Both fundamental and technical analysis can be done independently or together. Some analysts use both methods of analyses, while others stick to one. Either way, using stock analysis to vet stocks, sectors, and the market is an important method of creating the best investment strategy for one's portfolio.

1.2. Purpose

In this work, we propose a prediction model for the time series stock market data. This modelwill automate the process of changeof stockprice indices based on technical analysisand provides assistancefor financial specialists to choose the better timing for purchasing and selling stocks. Data mining techniques are used to develop the prediction model and python programming language is used for visualization of results.

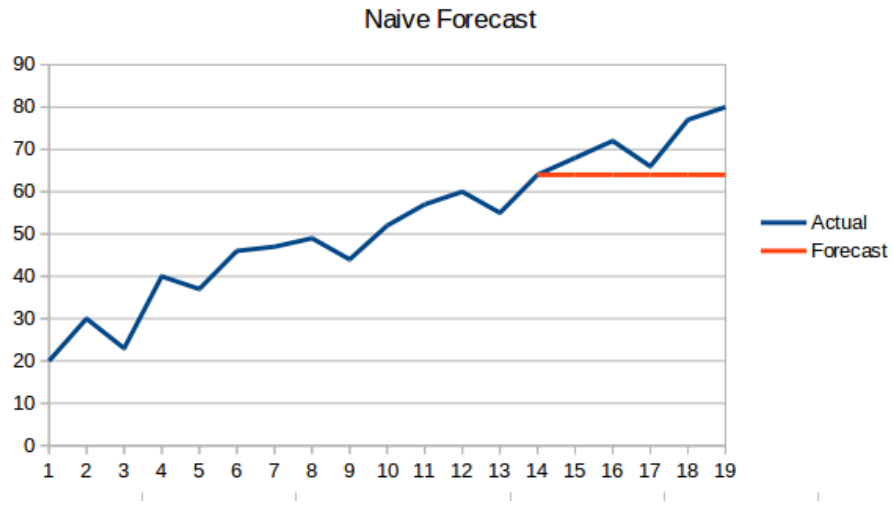
1.2.1 What is a time series?

Before we learn about the techniques to work on time series data, we must first understand what a time series actually is and how is it different from any other kind of data. Here is the formal definition of time series – It is a series of data points **measured at consistent time intervals**. This simply means that particular values are recorded at a constant interval which may be hourly, daily, weekly, every 10 days, and so on. **What makes time series different is that each data point in the series is dependent on the previous data points.**

1.2.2 Methods for time series forecasting

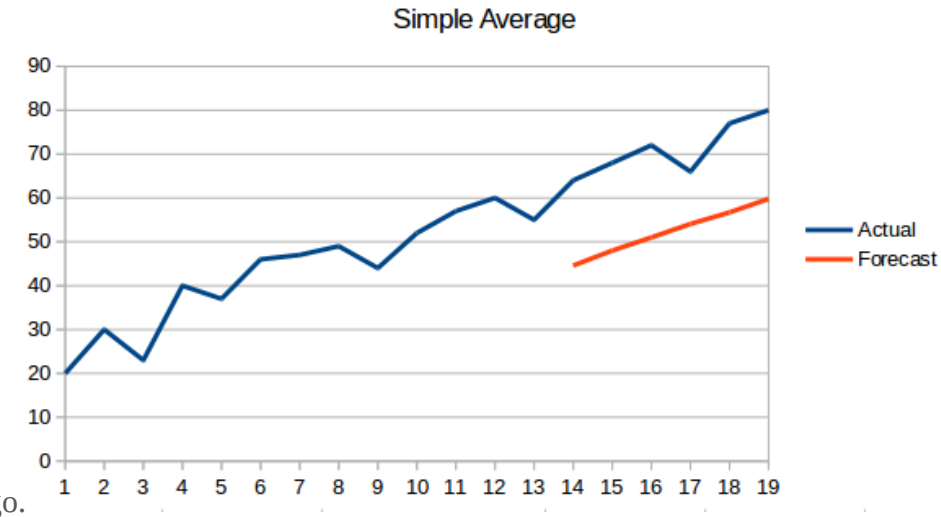
There are a number of methods for time series forecasting and we will briefly cover them in this section.

Naive Approach: In this forecasting technique, the value of the new data point is predicted to be equal to the previous data point. The result would be a flat line, since all new values take the previous values.

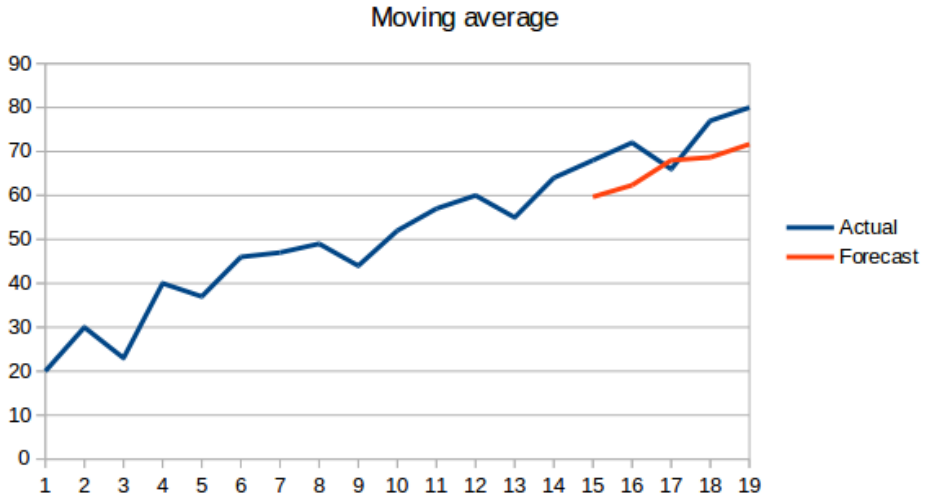


Simple Average: The next value is taken as the average of all the previous values. The predictions here are better than the 'Naive Approach' as it doesn't result in a flat line but here, all the past values are taken into consideration which might not always be useful. For instance, when asked to predict today's temperature, you

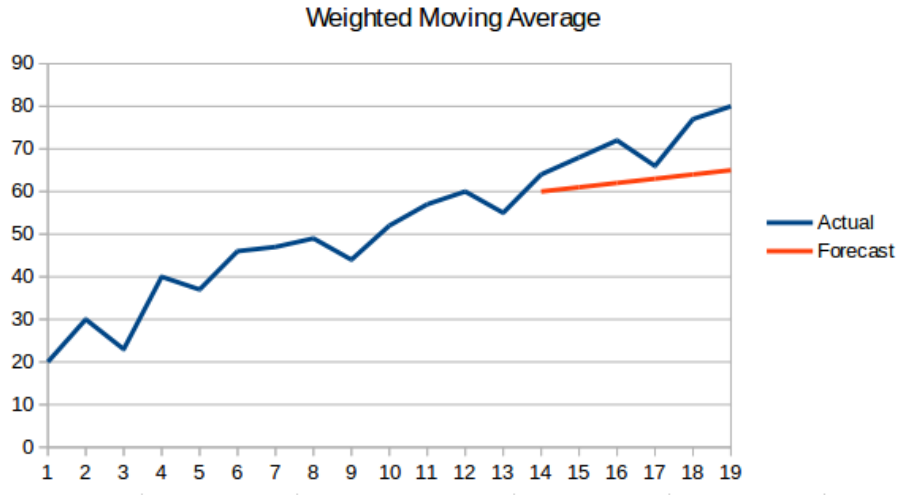
would consider the last 7 days' temperature rather than the temperature a month ago.



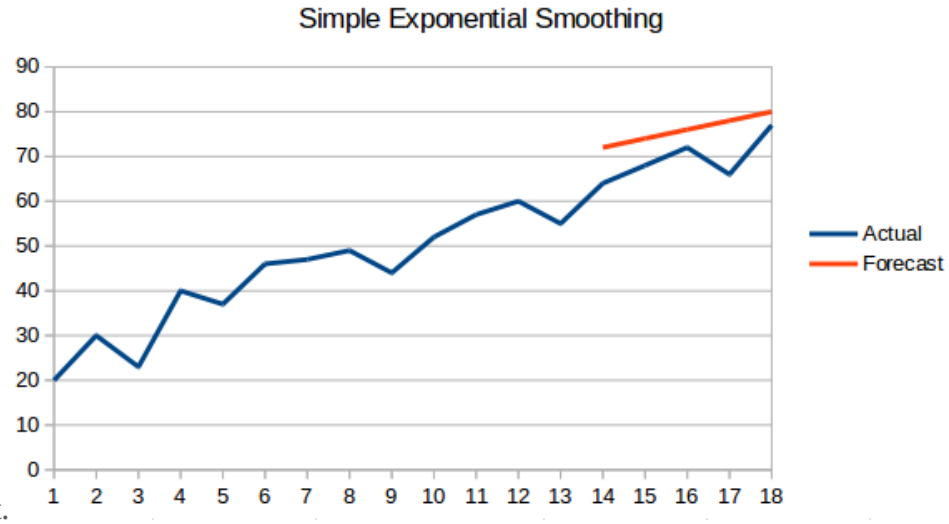
1. **Moving Average:** This is an improvement over the previous technique. Instead of taking the average of all the previous points, the average of 'n' previous points is taken to be the predicted value.



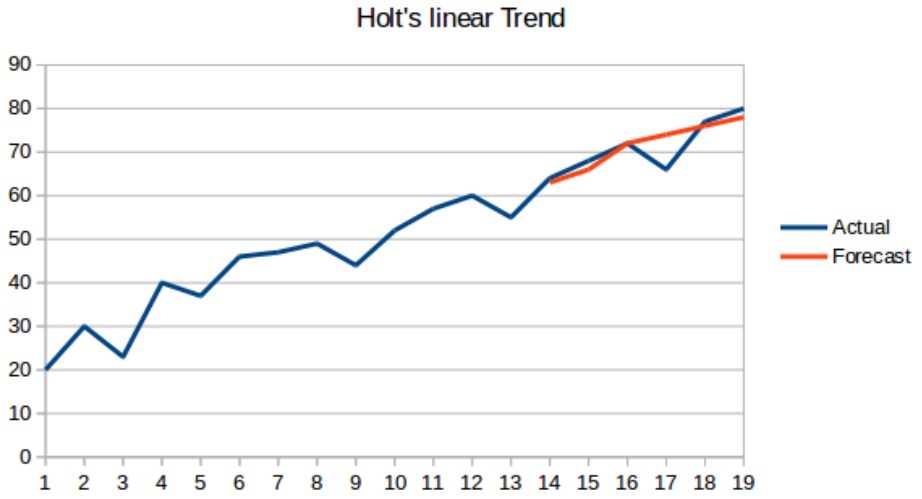
Weighted moving average: A weighted moving average is a moving average where the past 'n' values are given different weights.



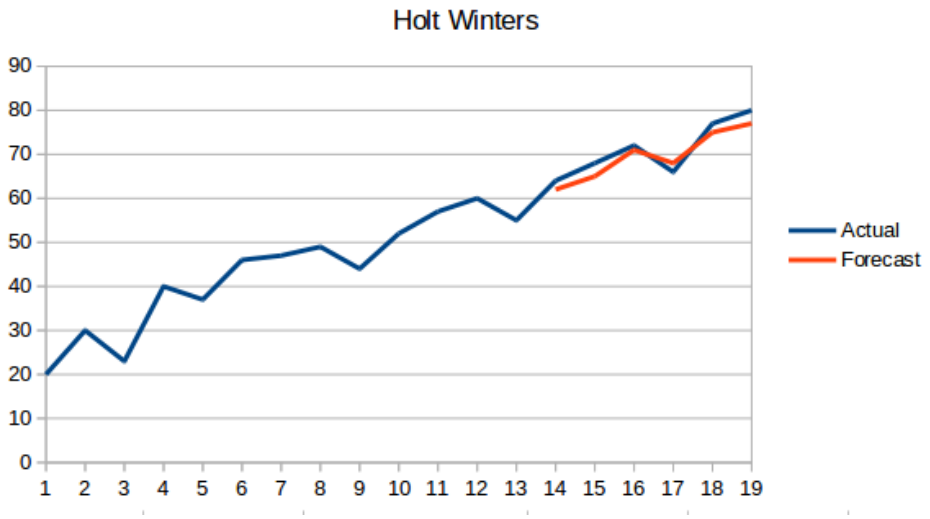
Simple Exponential Smoothing: In this technique, larger weights are assigned to more recent observations than to observations from the distant past.



Holt's Linear Trend Model: This method takes into account the trend of the dataset. By trend, we mean the increasing or decreasing nature of the series. Suppose the number of bookings in a hotel increases every year, then we can say that the number of bookings show an increasing trend. The forecast function in this method is a function of level and trend.



Holt Winters Method: This algorithm takes into account both the trend and the seasonality of the series. For instance – the number of bookings in a hotel is high on weekends & low on weekdays, and increases every year; there exists a weekly seasonality and an increasing trend.



ARIMA: ARIMA is a very popular technique for time series modelling. It describes the correlation between data points and takes into account the difference of the values. An improvement over ARIMA is SARIMA (or seasonal ARIMA). We will look at ARIMA in a bit more detail in the following section.

1.2.3 Introduction to ARIMA

In this section, we will do a quick introduction to ARIMA which will be helpful in understanding Auto Arima.

ARIMA is a very popular statistical method for time series forecasting. ARIMA stands for **Auto-Regressive Integrated Moving Averages**. ARIMA models work on the following assumptions –

- The data series is stationary, which means that the mean and variance should not vary with time. A series can be made stationary by using log transformation or differencing the series.
- The data provided as input must be a univariate series, since arima uses the past values to predict the future values.

ARIMA has three components – AR (autoregressive term), I (differencing term) and MA (moving average term). Let us understand each of these components –

- AR term refers to the past values used for forecasting the next value. The AR term is defined by the parameter ‘p’ in arima. The value of ‘p’ is determined using the PACF plot.
- MA term is used to define number of past forecast errors used to predict the future values. The parameter ‘q’ in arima represents the MA term. ACF plot is used to identify the correct ‘q’ value.

- Order of differencing specifies the number of times the differencing operation is performed on series to make it stationary. Test like ADF and KPSS can be used to determine whether the series is stationary and help in identifying the d value.

1.2.4 Steps for ARIMA implementation

The general steps to implement an ARIMA model are –

1. **Load the data:** The first step for model building is of course to load the dataset
2. **Pre-processing:** Depending on the dataset, the steps of pre-processing will be defined. This will include creating timestamps, converting the dtype of date/time column, making the series univariate, etc.
3. **Make series stationary:** In order to satisfy the assumption, it is necessary to make the series stationary. This would include checking the stationarity of the series and performing required transformations
4. **Determine d value:** For making the series stationary, the number of times the difference operation was performed will be taken as the d value
5. **Create ACF and PACF plots:** This is the most important step in ARIMA implementation. ACF PACF plots are used to determine the input parameters for our ARIMA model
6. **Determine the p and q values:** Read the values of p and q from the plots in the previous step
7. **Fit ARIMA model:** Using the processed data and parameter values we calculated from the previous steps, fit the ARIMA model
8. **Predict values on validation set:** Predict the future values
9. **Calculate RMSE:** To check the performance of the model, check the RMSE value using the predictions and actual values on the validation set

1.2.5 Why do we need Auto ARIMA?

Although ARIMA is a very powerful model for forecasting time series data, the data preparation and parameter tuning processes end up being really time consuming. Before implementing ARIMA, you need to make the series stationary, and determine the values of p and q using the plots we discussed above. Auto ARIMA makes this task really simple for us as it eliminates steps 3 to 6 we saw in the previous section. Below are the steps you should follow for implementing auto ARIMA:

1. Load the data: This step will be the same. Load the data into your notebook
2. Pre-processing data: The input should be univariate, hence drop the other columns
3. Fit Auto ARIMA: Fit the model on the univariate series
4. Predict values on validation set: Make predictions on the validation set
5. Calculate RMSE: Check the performance of the model using the predicted values against the actual values

We completely bypassed the selection of p and q feature as you can see. What a relief! In the next section, we will implement auto ARIMA using a toy dataset.

1.3 Motivations and Scope

1.3.1. Financial

Many people are interested in the financial market and need guidance and accurate predictions to invest wisely. Investors are always looking for the accurate future results.

The Firm Foundation theory defines the Stock Exchange as the reaction to a Company's Real Value by the Investors. Analysis of Present and Future prospects of a Company helps in deriving its Real Value. Whereas, the Castles in the Air theory defines the Stock Exchange as the reaction of investors to actions of other investors. Hence this theory pitches the only concern for an investor as to purchase a Stock at let's say INR 60/- today and sell the same at INR 70/- tomorrow without being concerned about the Real Value of the Company in which he/she invests. Study leads us to conclude that either the Stock Market is influenced by pure logic or it is purely influenced by psychology. A combination of both theories may prove out to be one of

the efficient ways to predict the behaviour of the market. This may focus on following factor that have a role in influencing the behaviour of the Stock Prices.

7. Global Indices
8. Indian Indices
9. Currency Prices
10. Sector Behaviour
11. Market Movers
12. General News
13. Company News
14. Company Financial Reports

In this project we propose a Stock Market prediction algorithm which will be combining both the absolute behaviour as well as the relative behaviour of a company stock with analytical focus on the past behaviour. In this project we will be implementing ways to identify the current situation of the Stock and map it up with a similar situation in the past.

1.3.2 Applications and News channels

There are many applications that try to predict the stocks but they do not give detailed information about the prediction. Thus with a successful model for stock prediction, we try to gain insight about market behavior over time, spotting trends that would otherwise not have been noticed.

Motivation behind using timeseriesprediction:

Time series analysis can be applied to real-valued, continuous data, discrete numeric data, or discrete symbolic data (i.e. sequences of characters, such as letters and words in the English language).

2.Literature survey

A share market could be a place of high interest to the investors because it presents them with a chance to learn financially by finance their resources on shares and derivatives of varied firms. It's a chaos system; that means the activity traits of share costs area unit unpredictable and unsure. To create some style of sense of this chaotic behavior, researchers were forced to search out a way, which may estimate the result of this uncertainty to the flow of share costs. From the analyses of varied applied math models, Artificial Neural Networks area unit analogous to non-parametric, nonlinear, regression models. So, Artificial Neural Networks (ANN) actually has the potential to tell apart unknown and hidden patterns in information which may be terribly effective for share market prediction. If successful, will this be useful for investors and finances which can completely contribute to the economy. There are unit totally different strategies that are applied so as to predict Share Market returns. The securities market reflects the fluctuation since itsinitial development. The securities market is characterized by bad, high-yield, thus investors are involved concerning the analysis of the securities market and making an attempt to forecast the trend of the securities market. However, securities market is wedged by the politics, economy and plenty of different factors, let alone the quality of its internal law, like value (stock index) changes within the non-linear, and shares knowledge with high noise characteristics, so the normal mathematical applied mathematics techniques to forecast the securities market has not yielded satisfactory results. Neural networks will approximate any advanced non-linear relations and hashardiness and fault-tolerant options. Therefore, it’s terribly appropriate for the analysis of stock knowledge. In dozens of neural network models that were suggests, researchers usually use the hop garden network. hop garden network is that the commonest feedback network model, it's one among the models that almost typically studied currently. The hop garden network is that the mono layer recognized by an equivalent vegetative cell, and is additionally a symmetrically connected associative network while not learning operates.

2.1 TimeSeriesPrediction

A time series is a sequence of data points made:

1) Over a continuous time interval.

2) Out of successive measurements across that interval.

3) Using equal spacing between every two consecutive measurements

4) With each time unit within the time interval having at most one data point.

Examples of time series are ocean tides, counts of sunspots, and the daily closing value of the Dow Jones Industrial Average.

Non-Examples: The height measurements of a group of people where each height is recorded over a period and each person has only one record in the data set.

Time series are very frequently plotted via line charts. Time series are used in statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, intelligent transport and trajectory forecasting, earthquake prediction, electroencephalography, control engineering, astronomy, communications engineering, and largely in any domain of applied science and engineering which involves temporal measurements.

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. While regression analysis is often employed in such a way as to test theories that the current values of one or more independent time series affect the current value of another time series, this type of analysis of time series is not called "time series analysis", which focuses on comparing values of a single time series or multiple dependent time series at different points in time.

Time series data have a natural temporal ordering. This makes time series analysis distinct from cross-sectional studies, in which there is no natural ordering of the observations (e.g. explaining people's wages by reference to their respective education levels, where the individuals' data could be entered in any order). Time

series analysis is also distinct from spatial data analysis where the observations typically relate to geographical locations (e.g. accounting for house prices by the location as well as the intrinsic characteristics of the houses).

A stochastic model for a time series will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values (see time reversibility).

2.2 PREDICTION AND FORECASTING

In statistics, predictionis apart ofstatistical inference. One particular approach to such inference is known as predictive inference, but the prediction can be undertaken within any of the several approaches to statistical inference. Indeed, one description of statistics is that it provides a means of transferring knowledge about a sample of a population to the whole population, and to other related populations, which is not necessarily the same as prediction over time. When information is transferred across time, often to specific points in time, the process is known as forecasting.

3. Proposed model

3.1 Programming language Requirement

Python is a widely used high-level programming language for general-purpose programming, created by Guido van Rossum and first released in 1991. An interpreted language, Python has a design philosophy which emphasizes code readability (notably using whitespace indentation to delimit code blocks rather than curly braces or keywords), and a syntax which allows programmers to express concepts in fewer lines of code than possible in languages such as C++ or Java. The language provides constructs intended to enable writing clear programs on both a small and large scale.

Python features a dynamic type system and automatic memory management and supports multiple programming paradigms, including object-oriented, imperative, functional programming, and procedural styles. It has a large and comprehensive standard library.

3.2 Library Requirements

3.2.1MATLAB

MATLAB (matrix laboratory) is a multi-paradigm numerical computing environment. A proprietary programming language developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, C#, Java, Fortran and Python.

Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the MuPAD symbolic engine, allowing access to symbolic computing

abilities. An additional package, Simulink, adds graphical multi-domain simulation and model-based design for dynamic and embedded systems.

As of 2017, MATLAB has roughly 1 million users across industry and academia.MATLAB users come from various backgrounds of engineering, science, and economics.

3.2.2NumPy

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

NumPy targets the CPython reference implementation of Python, which is a non-optimizing bytecode interpreter. Mathematical algorithms written for this version of Python often run much slower than compiled equivalents. NumPy addresses the slowness problem partly by providing multidimensional arrays and functions and operators that operate efficiently on arrays, requiring rewriting some code, mostly inner loops using NumPy.

Using NumPy in Python gives functionality comparable to MATLAB since they are both interpreted, and they both allow the user to write fast programs as long as most operations work on arrays or matrices instead of scalars. In comparison, MATLAB boasts a large number of additional toolboxes, notably Simulink, whereas NumPy is intrinsically integrated with Python, a more modern and complete programming language. Moreover, complementary Python packages are available; SciPy is a library that adds more MATLAB-like functionality and Matplotlib is a plotting package that provides MATLAB-like plotting functionality.

3.2.3 PANDAS

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data.

In 2008, developer Wes McKinney started developing pandas when in need of high performance, flexible tool for analysis of data.

Prior to Pandas, Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data — load, prepare, manipulate, model, and analyze.

Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

Various types of neural networks can be developed by the combination of different factors like network topology, training method etc. For this experiment, we have considered Time series analysis and ARIMA.
This section we will discuss the model of our system. Our system consists of several stages which are as follows:-

Stage 1: Raw Data

In this stage, the historical stock data is collected from https://in.finance.yahoo.comand this historical data is used for the prediction of future stock prices. We will use python built in library pandas function web-reader for web scrapping.

Stage 2: Data Preprocessing

The pre-processing stage involves

a) Data discretization: Part of data reduction but with particular importance, especially for numerical data

b) Data transformation: Normalization.

c) Data cleaning: Fill in missing values.

d) Data integration: Integration of data files.

After the dataset is transformed into a clean dataset, the dataset is divided into training and testing sets so as to evaluate. Here, the training values are taken as the more recent values. Testing data is kept as 5-10 percent of the total dataset.

Stage 3: Feature Extraction:

In this layer, only the features which are to be used in our analysis. We will choose the feature from Date, open, high, low, close, and volume.

Stage 4: Training Model: In this stage, we will use various statistical methods, investment formulas, Time series forecasting.

Stage 5: Output Generation:

In this steps output can be generated using the above method. We will usually show the output in visualization manner. We will also calculate error and based on that we will improve our model.

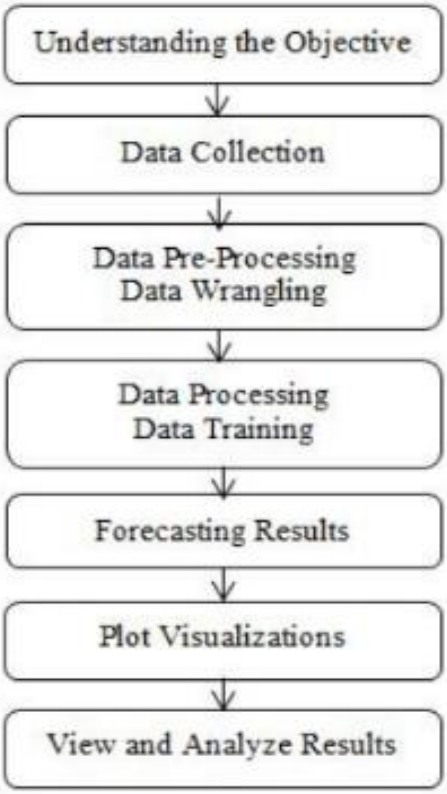
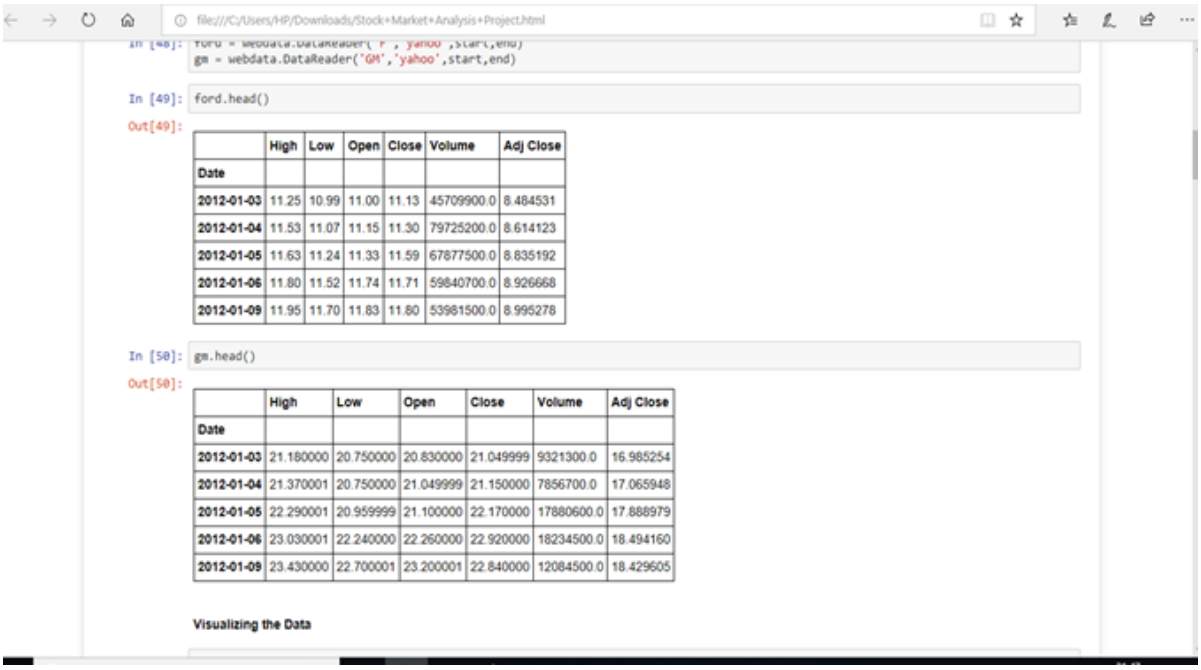


Fig. System Architecture

System architecture is a model that defines the behavior of a system in the conceptual model. The huge systems are decomposed into subordinate systems to provide similar set of services. The beginning layout strategy of perceiving these sub-systems and building up a structure for sub-systems control and cooperation is called architecture design.

4. Implementation



```
In [48]: ford = webdata.DataReader('F','yahoo',start,end)
         gm = webdata.DataReader('GM','yahoo',start,end)

In [49]: ford.head()
```

Out[49]:

	High	Low	Open	Close	Volume	Adj Close
Date						
2012-01-03	11.25	10.99	11.00	11.13	45709900.0	8.484531
2012-01-04	11.53	11.07	11.15	11.30	79725200.0	8.614123
2012-01-05	11.63	11.24	11.33	11.59	67877500.0	8.835192
2012-01-06	11.80	11.52	11.74	11.71	59840700.0	8.926668
2012-01-09	11.95	11.70	11.83	11.80	53981500.0	8.995278

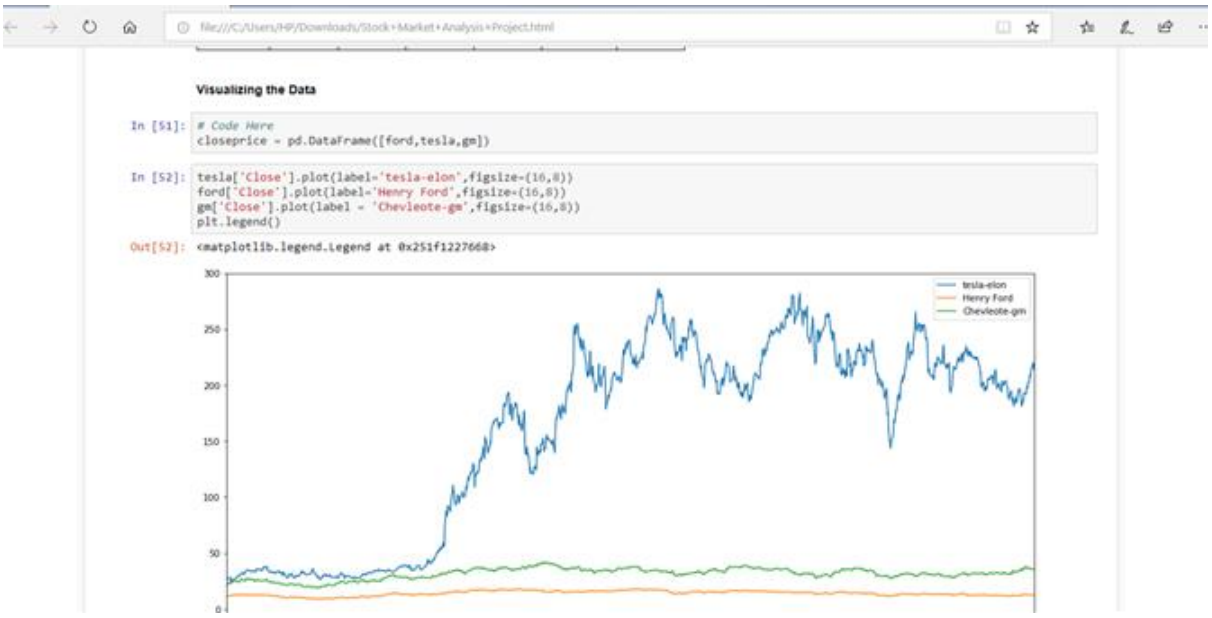
```
In [50]: gm.head()
```

Out[50]:

	High	Low	Open	Close	Volume	Adj Close
Date						
2012-01-03	21.180000	20.750000	20.830000	21.049999	9321300.0	16.985254
2012-01-04	21.370001	20.750000	21.049999	21.150000	7856700.0	17.065948
2012-01-05	22.290001	20.959999	21.100000	22.170000	17880600.0	17.888979
2012-01-06	23.030001	22.240000	22.260000	22.920000	18234500.0	18.494160
2012-01-09	23.430000	22.700001	23.200001	22.840000	12084500.0	18.429605

Visualizing the Data

Fig. Data Collection



Visualizing the Data

```
In [51]: # Code Here
         closeprice = pd.DataFrame([ford,tesla,gm])

In [52]: tesla['Close'].plot(label='tesla-elon',figsize=(16,8))
         ford['Close'].plot(label='Henry Ford',figsize=(16,8))
         gm['Close'].plot(label = 'Chevleote-gm',figsize=(16,8))
         plt.legend()
```

Out[52]: <matplotlib.legend.Legend at 0x251f1227668>

Fig. Comparison of Stock

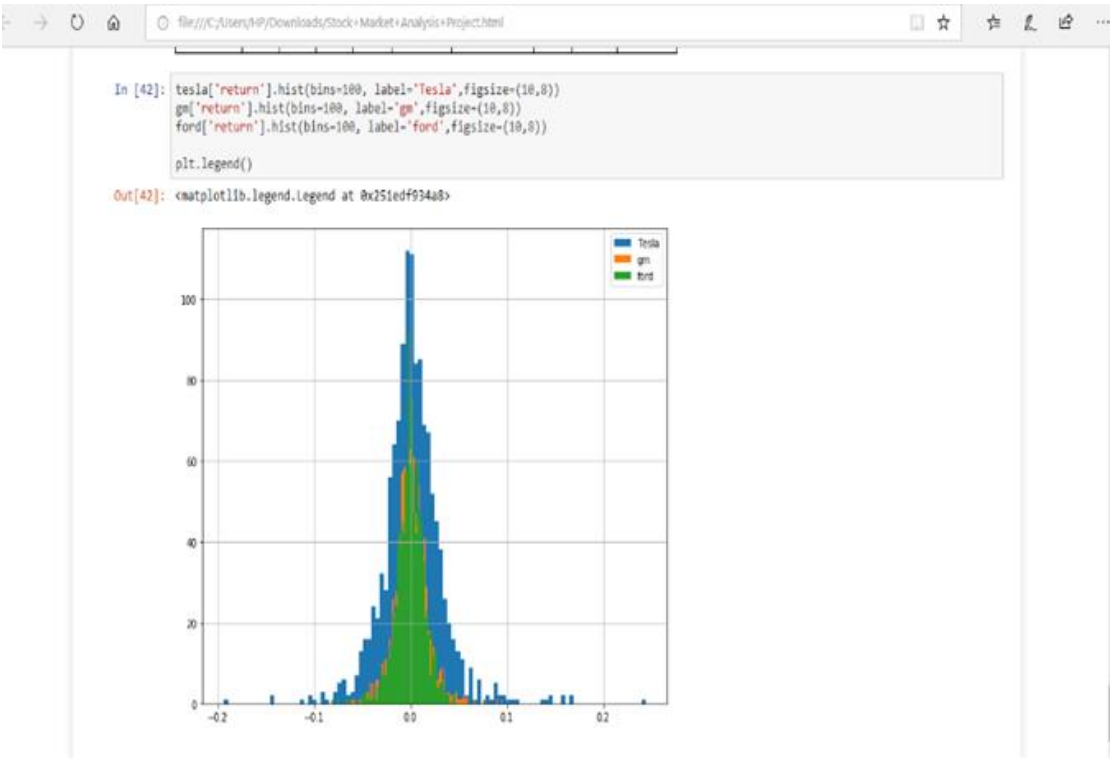
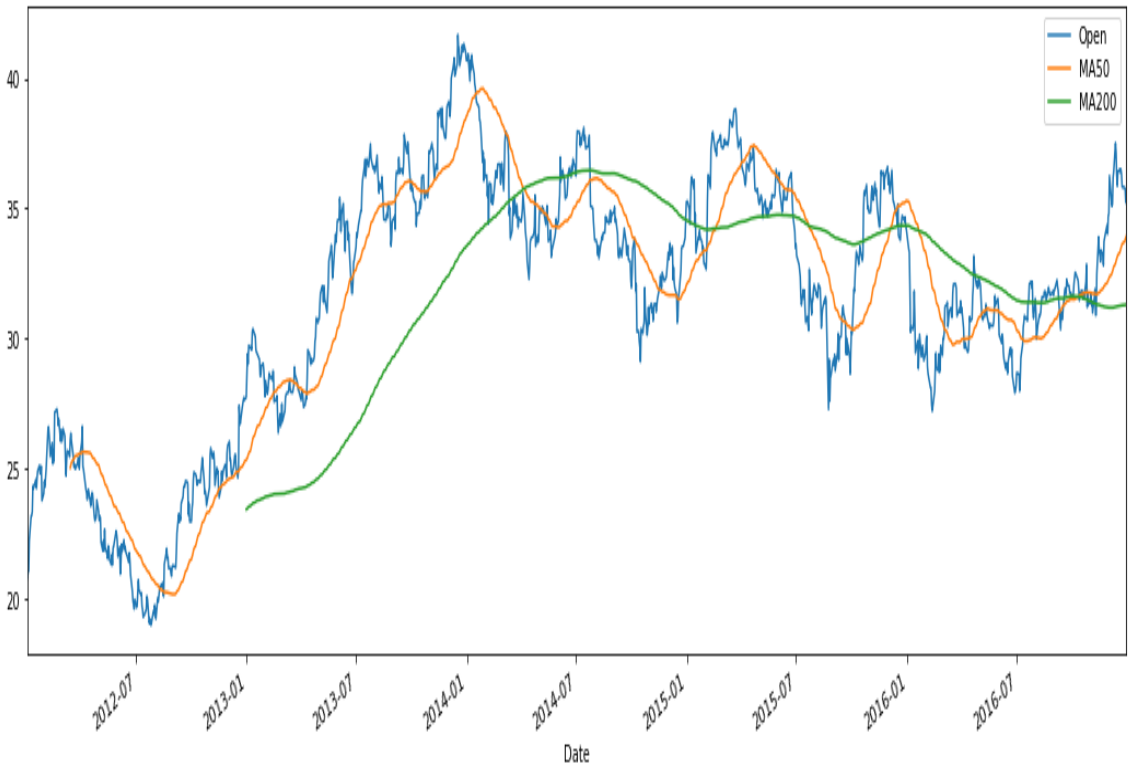


Fig. Density Diagram



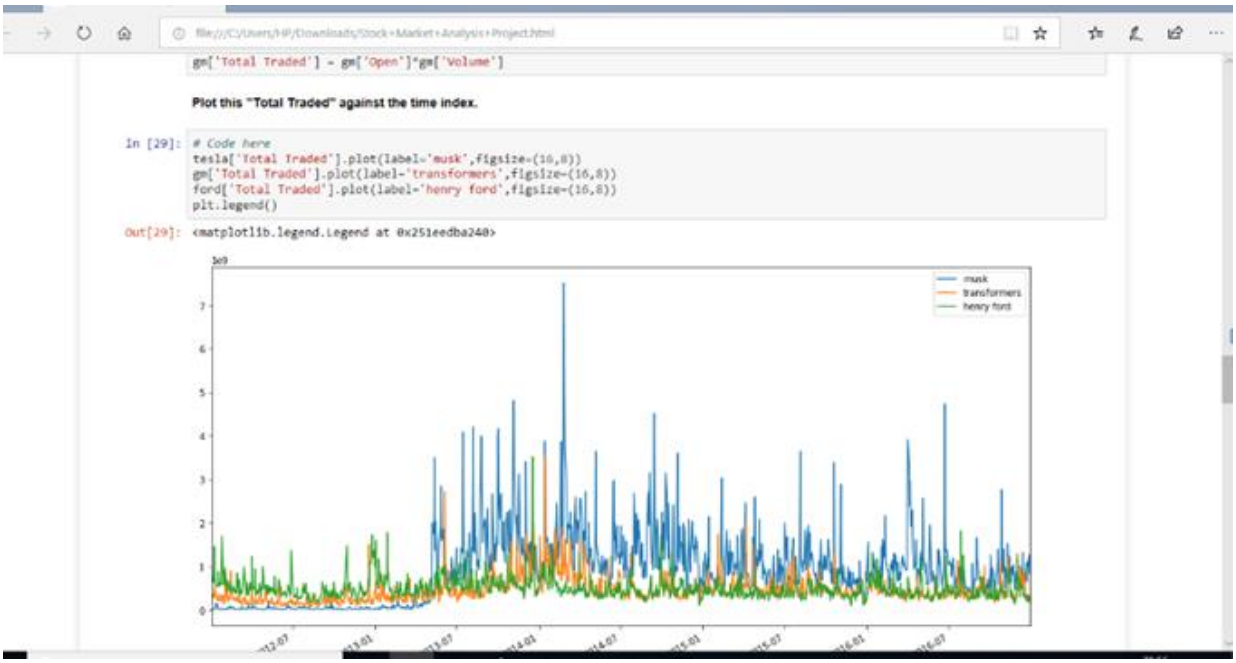
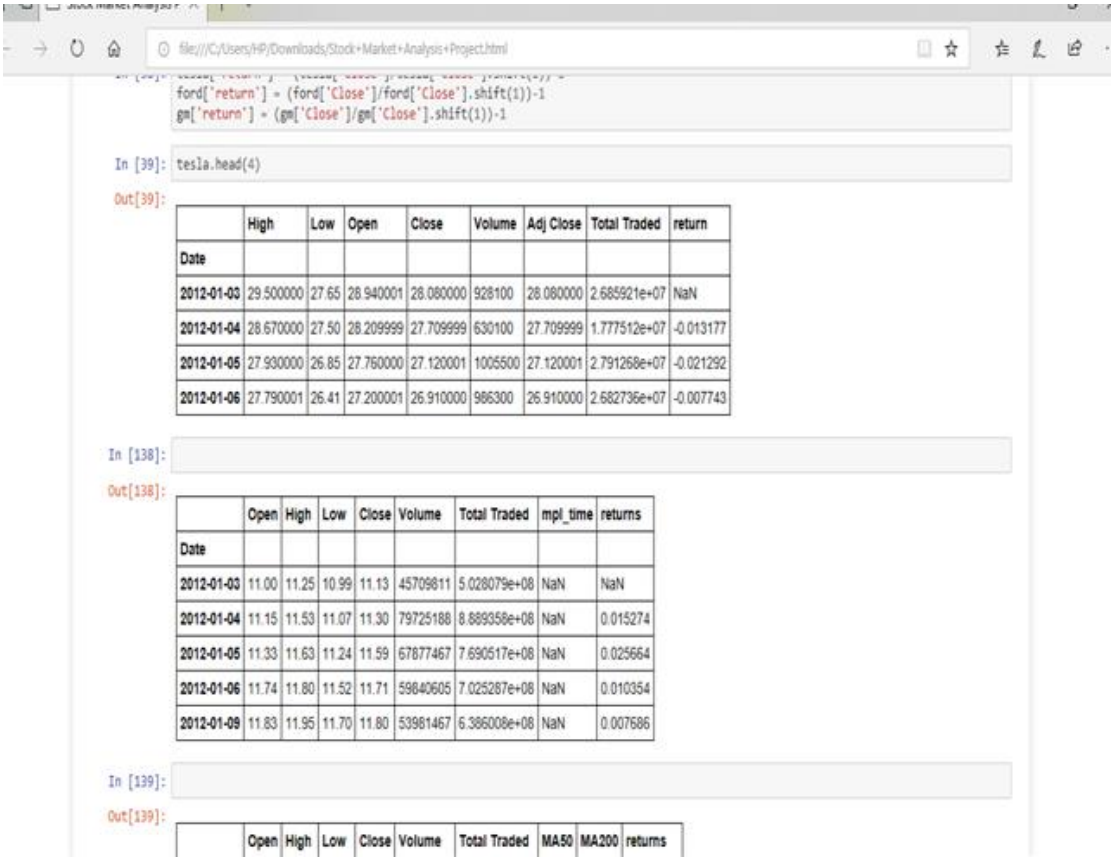


Fig. Time Series



```
ford['return'] = (ford['Close']/ford['Close'].shift(1))-1
gm['return'] = (gm['Close']/gm['Close'].shift(1))-1
```

In [39]:
```
tesla.head(4)
```

Out[39]:

	High	Low	Open	Close	Volume	Adj Close	Total Traded	return
Date								
2012-01-03	29.500000	27.65	28.940001	28.080000	928100	28.080000	2.685921e+07	NaN
2012-01-04	28.670000	27.50	28.209999	27.709999	630100	27.709999	1.777512e+07	-0.013177
2012-01-05	27.930000	26.85	27.760000	27.120001	1005500	27.120001	2.791268e+07	-0.021292
2012-01-06	27.790001	26.41	27.200001	26.910000	986300	26.910000	2.682736e+07	-0.007743

In [138]:

Out[138]:

	Open	High	Low	Close	Volume	Total Traded	mpl_time	returns
Date								
2012-01-03	11.00	11.25	10.99	11.13	45709811	5.028079e+08	NaN	NaN
2012-01-04	11.15	11.53	11.07	11.30	79725188	8.889358e+08	NaN	0.015274
2012-01-05	11.33	11.63	11.24	11.59	67877467	7.690517e+08	NaN	0.025664
2012-01-06	11.74	11.80	11.52	11.71	59640605	7.025287e+08	NaN	0.010354
2012-01-09	11.83	11.95	11.70	11.80	53981467	6.386008e+08	NaN	0.007686

In [139]:

Out[139]:

	Open	High	Low	Close	Volume	Total Traded	MA50	MA200	returns

Fig. 2012 Tesla Stock Data

5. Results and Discussions

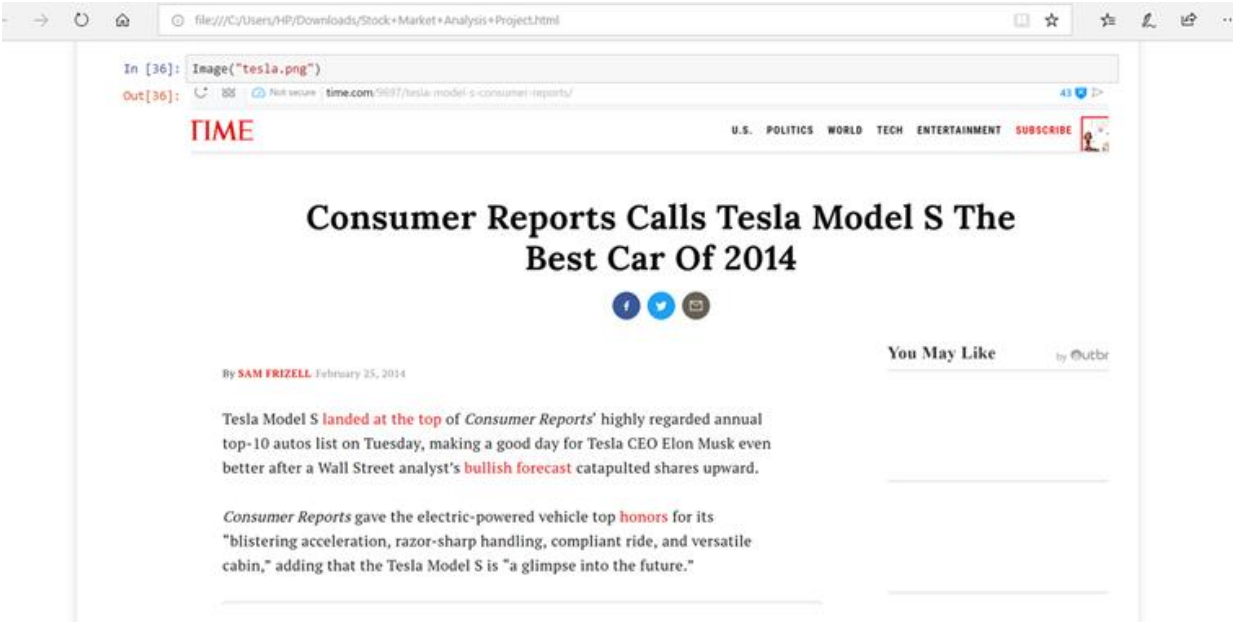
As we can see that we have collected dataset for Tesla, General Motors and Ford.We use time-series forecasting to predict future price of Stock. This can be done on volatile dataset as the formula can be performed over volatility. There is a sudden increasing movement in Tesla stock on 2014-01. This sudden rise in stock is caused because of “CONSUMER REPORTS CALLS TESLA MODEL S THE BEST CAR OF 2018” on 2014-01 which is the same day. So we can conclude that several factors are affecting stock prices like “announcement” in company, or certain volatility in stock.

5.1 Daily Percentage Change

Daily percentage change is defined by the following formula:

rt=pt/(pt-1)-1

This defines r_t (return at time t) as equal to the price at time t divided by the price at time t-1 (the previous day) minus 1. Basically, this just informs you of your percent gain (or loss) if you bought the stock on day and then sold it the next day. Its very helpful in analysing the volatility of the stock. If daily returns have a wide distribution, the stock is more volatile from one day to the next. Let's calculate the percent returns and then plot them with a histogram, and decide which stock is the most stable!



6. Conclusions and Future Works

7. In this paper, we made an attempt to develop a prediction model for forecasting the stock market trends based on the technical analysis using historical time series stock market data and data mining techniques. The experimental results obtained demonstrated the potential of ARIMA modelto predict the stock price indices on short-term basis. This could guide the investors in the stock market to make profitable investment decisions whether to buy/sell/hold a share. With the results obtained ARIMA modelcan compete reasonably well with emerging forecasting techniques in short-term prediction.

The implementation of this paper can be extended by integrating the technical analysisand fundamental analysis techniques. Through the evaluation of social media analysis particularly on public opinions using fundamental analysis techniques can be incorporated in order to obtain better results. In this way, we can provide the improved results for investors in the stock market to choose the better timing for profitable investment decisions.

7. References

[1] Banerjee, D., "Forecasting of Indian stock market using time-series ARIMA model", 2nd IEEE International Conference on Business and Information Management (ICBIM), January 2014, pp. 131-135.

[2] Li Bing, Chan, K. C. C., C. Ou, "Public sentiment analysis in Twitter data for prediction of a company's stock price movements", 11th IEEE International Conference on e-BusinessEngineering (ICEBE), November 2014, pp. 232-239.

[3] Tao Xing, Yuan Sun, Qian Wang, Guo Yu, "The analysis and prediction of stock prices", IEEE International Conference on Granular Computing (GrC), December 2013, pp. 368–373.

[4] Vishwanath R. Ha, Leena Sa, Srikantaiah K. Ca, K. Shreekrishna Kumar b., P. DeepaShenoya, Venugopal K. Ra, S. S. Iyengarc, L. M. Patnaik, "Forecasting stock time-series using data approximation and pattern sequence similarity", International Journal of Information Processing (IJIP), September 2013, pp. 90-100.

[5] Ayodele A. Adebiyi, Aderemi O. Adewumi, Charles K. Ayo, "Stock price prediction using the ARIMA model", 16th IEEE International Conference on Computer Modelling andSimulation (UKSim), March 2014, pp. 106 -112.

[6] Qasem A. Al-radaideh, Adel Abu Asaf, EmanAlnagi, "Predicting stock prices using data mining techniques", The International Arab Conference on Information Technology 2013.

[7] LiZhe; "Research on China's stock exchange markets: problems and improvements", International Conference on Education and Management Technology, 2010. pp 465-469.

[8] Hazem M. El-Bakry, and Wael A. Awad, “Fast forecasting of stock market prices by using new high speed time delay neural networks”, International Journal of Computer and Information
Engineering, February 2010, pp. 138-144.

[9] Han, J., Kamber, M., Jian P., “Data mining concepts and techniques”. San Francisco, CA: Morgan Kaufmann Publishers, 2011.

[10] Wang, Y.F., “Mining stock price using fuzzy rough set system”, Expert Systems with Applications, 2006, pp. 13-23.

[11] Enke, D., Thawornwong, S., “The use of data mining and neural networks for forecasting stock market returns”, Expert Systems with Applications, 2005, pp. 927-940.

[12] A. J. Conejo, M. A. Plazas, R. Espnola and B. Molina. “Day-ahead electricity price forecasting using the wavelet transform and ARIMA models”, IEEE Transactions on Power Systems, 2005, pp. 1035–1042.