A PROJECT ON

“STOCK MARKET PREDICTION ”

SUBMITTED IN
PARTIAL FULFILLMENT OF THE REQUIREMENT
FOR THE COURSE OF
DIPLOMA IN
BIG DATA ANALYTICS FROM CDAC

SUNBEAM INSTITUTE OF INFORMATION TECHNOLOGY
‘Plot no R/2’, Market yard road,
Behind hotel Fulera,Gultekdi
Pune – 411037.
MH-INDIA

SUBMITTED BY:

Naveen Kumar Singh

UNDER THE GUIDENCE OF:
MS. Sarika pol
Faculty Member
Sunbeam Institute of Information Technology, PUNE.

CERTIFICATE

This is to certify that the project work under the title ‘Stock Market Prediction’ is done by Naveen Kumar Singh in partial fulfillment of the requirement for award of Diploma in Big Data Analytics Course.

MS. Sarika pol
Project Guide

Date: 30/July/2018

ACKNOWLEDGEMENT

A project usually falls short of its expectation unless aided and guided by the right persons at the right time. We avail this opportunity to express our deep sense of gratitude
towards Mr. Nitin Kudhale (Center Coordinator, SIIT, Pune) and Mr. Nilesh Ghule and Project Guide MS. Sarika pol.
We are deeply indebted and grateful to them for their guidance, encouragement and deep concern for our project. Without their critical evaluation and suggestions at every stage of the project, this project could never have reached its present form.
Last but not the least we thank the entire faculty and the staff members of Sunbeam Institute of Information Technology, Pune for their support.

Naveen Kumar Singh
DBDA February 2018 Batch,
SIIT Pune

TABLE OF CONTENTS

5. FINDINGS AND CONCLUSIONS

1. Introduction

Predicting the Stock Market has been the bane and goal of investors since its existence. Everyday billions of dollars are traded on the exchange, and behind each dollar is an investor hoping to profit in one way or another. Entire companies rise and fall daily based on the behavior of the market. Should an investor be able to accurately predict market movements, it offers a tantalizing promises of wealth and influence. It is no wonder then that the Stock Market and its associated challenges find their way into the public imagination every time it misbehaves. The 2008 financial crisis was no different, as evidenced by the flood of films and documentaries based on the crash. If there was a common theme among those productions, it was that few people knew how the market worked or reacted. Perhaps a better understanding of stock market prediction might help in the case of similar events in the future.

1.1. Statement of the problem

Stock market is very vast and difficult to understand. It is considered too uncertain predictable due to huge fluctuation of the market. Stock market prediction task is interesting as well as divides researchers and academics into two groups, those who believe that we can devise mechanisms to predict the market and those who believe that the market is efficient and whenever new information comes up the market absorbs it by correcting itself, thus there is no space for prediction.
Investing in a good stock but at a bad time can have disastrous result, while investing in a stock at the right time can bear profits. Financial investors of today are facing this problem of trading as they do not properly understand as to which stocks to buy or which stocks to sell in order to get optimum result. So, the purposed project will reduce the problem with suitable accuracy faced in such real time scenario.

1.2 Relevant current/open problems

1. Data-are-humongous, nowadays we are seeing a rapid-explosion of numerical-stock quotes and textual-data. They are provided from all different-sources.
2. Demand forecasts are important since the basic op management process, going from the vendor raw-materials to finished goods in the customers'

hands, takes some time. Most firms cannot-wait for demand to elevate and then give a reaction. Instead, they make-up their mind and plan according to future demand so-that they can react spontaneously to customer's order as they arrive.

3. Generally, demand forecasts-lead to good-ops-and great-levels of customer satisfaction, while bad forecast will definitely-lead to costly ops and worst-levels of customer satisfaction.

4. A confusion for the forecast is the horizon, which is, how distant in the future will the forecast project? As a simple rule, the away into the future we see, the more blurry our vision will become -- distant forecasts will be inaccurate that short-range forecasts

1.3 Technical analysis

In contrast to fundamental analysis, technical analysis does not try to gain deep insight into a company's business. It assumes the available public information does not offer a competitive trading advantage. Instead, it focuses on studying a company's historical share price and on identifying patterns in the chart. The intention is to recognize trends in advance and to capitalize on them.

1.4 Goal

The goal was to build a system capable of the following tasks:

1. Collecting fundamental and technical data from the internet

The system should be able to crawl specific websites to extract fundamental data like news articles and analyst recommendations. Furthermore, it should be able to collect technical data in the form of historical share prices.

2. Simulating trading strategies

The system should offer ways to specify and simulate fundamental and technical trading strategies. Additionally, combining the two approaches should be possible.

3. Evaluating and visualizing trading strategies

The system should evaluate and visualize the financial performance of the simulated strategies. This allows a comparison to be made between technical, fundamental and the combined approaches.

1.5 Project Goals and Scope

Despite its prevalence, Stock Market prediction remains a secretive and empirical art. Few people, if any, are willing to share what successful strategies they have. A chief goal of this project is to add to the academic understanding of stock market prediction. The hope is that with a greater understanding of how the market moves, investors will be better equipped to prevent another financial crisis. The project will evaluate some existing strategies from a rigorous scientific perspective and provide a quantitative evaluation of new strategies. It is important here to define the scope of the project. Although vital to any investor operating in the real world, no attempt is made in this project at portfolio management. Portfolio management is largely an extra step done after an investor has made a prediction on which direction any particular stock will move. The investor may choose to allocate funds across a range of stocks in such a way to minimize his or her risk. For instance, the investor may choose not to invest all of their funds into a single company lest that company takes unexpected turn. A more common approach would be for an investor to 3 invest across a broad range of stocks based on some criteria he has decided on before . This project will focus exclusively on predicting the daily trend (price movement) of individual stocks. The project will make no attempt to deciding how much money to allocate to each prediction. More so, the project will analyses the accuracies of these predictions.

1.6 Overview of proposed solution approach

1. Basically the main objective of this project is to collect the stock information for some previous years and then accordingly predict the results for the predicting what would happen next. So for we are going to use of two well-known techniques Machine Learning and data mining for stock market prediction. Extract useful information from a huge amount of data set and data mining is also able to predict future trends and behaviors through neural

network. Therefore, combining both these techniques could make the prediction more suitable and much more reliable

2. As far as the solutions for the above problems, the answer depends on which way the forecast is used for. So the procedures that we will be using have proven to be very applicable to the task of forecasting product demand in a logistics system. Many techniques, which can prove useful for forecasting-problems, have shown to be inadequate to the task of demand forecasting in logistics systems.

1.7 Novelty/Benefits:

The rich variety of on-line information and news make it an attractive resource from which one can get data. Stock market predictions can be aided by data mining and analysis of such financial information. Numerical stock quotes collected from morningstar India finance are available in organized manner but we have to apply some techniques to parse textual news information about stock market is collected from websites released daily

2. Product Overview and Summary

2.1 Purpose

The aims of this project are as follows:

1. To identify factors affecting share market
2. To generate the pattern from large set of data of stock market for prediction of stock market movements
3. To predict weather the stock price is going to increase or decrease or steady in next seven days
4. To provide analysis for users through Desktop application

The project will be useful for investors to invest in stock market based on the various factors. The project target is to create web application that analyses previous stock data of companies and implement these values in data mining algorithm to determine the value that particular stock will have in near future with suitable accuracy. These predicted and analyzed data can be observed by individual to know the financial status of companies and their comparisons. Company and industry can use it to breakdown their limitation and enhance their stock value. It can be very useful to even researchers, stock brokers, market makers, government and general people.

2.2 Scope

Stock market includes daily activities like sen-sex calculation, exchange of shares. The exchange provides an efficient and transparent market for trading in equity, debt instruments and derivatives.

The stock values of company depend on many factors, some of them are:

1. Demand and Supply of shares of a company is a major reason price change in stocks. When Demand Increase and Supply is less, price rises. and viceversa.

2. Main Strength in hands of share buyer. Popularity of a company can effect on buyers. Like if any good news of a company, may result in rise of stock

price.

The stock value depends on other factors as well, but we are taking into consideration only these main factors.

3. Requirements and Feasibility study

3.1. Feasibility Study

Simply put, stock market cannot be accurately predicted. The future, like any complex problem, has far too many variables to be predicted. The stock market is a place where buyers and sellers converge. When there are more buyers than sellers, the price increases. When there are more sellers than buyers, the price decreases. So, there is a factor which causes people to buy and sell. It has mo re to do with emotion than logic. Because emotion is unpredictable, stock market movements will be unpredictable. It's futile to try to predict where markets are going. They are designed to be unpredictable.

There are some fundamental financial indicators by which a company's stock value can be estimated. Some of the indicators and factors are: Price-to-Earning (P/E) Ratio,Price-to-Earning Growth (PEG) Ratio, Price-to-Sales (P/S) Ratio, Price/Cash Flow (P/CF) Ratio, Price-to-Book Value (P/BV) Ratio and Debt-to-Equity Ratio. Some of the parameters are available and accessible on the web but all of them aren't. So we are confined to use the variables that are available to us.

The proposed system will not always produce accurate results since it does not account for the human behaviors. Factors like change in company's leadership, internal matters, strikes, protests, natural disasters, change in the authority cannot be taken into account for relating it to the change in Stock market by the machine.

The objective of the system is to give a approximate idea of where the stock market might be headed. It does not give a long term forecasting of a stock value. There are way too many reasons to acknowledge for the long term output of a current stock.
Many things and parameters may affect it on the way due to which long term forecasting is just not feasible.

3.2. Requirement Analysis

After the extensive analysis of the problems in the system, we are familiarized with the requirement that the current system needs. The requirement that the system needs is categorized into the functional and non-functional requirements.
These requirements are listed below:

1. Functional Requirements
2. Non-Functional Requirements

3.2.1 Functional Requirements

Functional requirement are the functions or features that must be included in any system to satisfy the business needs and be acceptable to the users. Based on this, the functional requirements that the system must require are as follows:

1. The system should be able to generate an approximate share price.
2. The system should collect accurate data from the stock market website in consistent manner.
3. The prediction shall abide by the following functional requirements:
4. Prior to application of stock recommendations, the database is updated by the latest values.
5. The charts and comparison of the companies would be done only on the latest data stock market data.
6. The user can look previous data Information which was collected.
7. The user can also be recommended on the basis of the trending stocks which would require the data regarding the stocks.

3.2.2 Non-Functional Requirements

Non-functional requirement is a description of features, characteristics and attribute of the system as well as any constraints that may limit the boundaries of the proposed system. The non- functional requirements are essentially based on the performance, information, economy, control and security efficiency and services. Based on these the non-functional requirements are as follows:

1. The system should provide better accuracy.
2. The system should have simple interface for users to use.
3. To perform efficiently in short amount of time.

1. Reliability:

The reliability of the product will be dependent on the accuracy of the dataset of purchase, how much stock was purchased, high and low value range as well as opening and closing figures. Also the stock data used in the training would determine the reliability of the software.

2. Security:

The user will only be able to access the website using his login details and will not be able to access the computations happening at the back end.

3. Maintainability:

The maintenance of the product would require training of the software by recent data so that there commendations are up to date. The database has to be updated with recent values.

4. Portability:

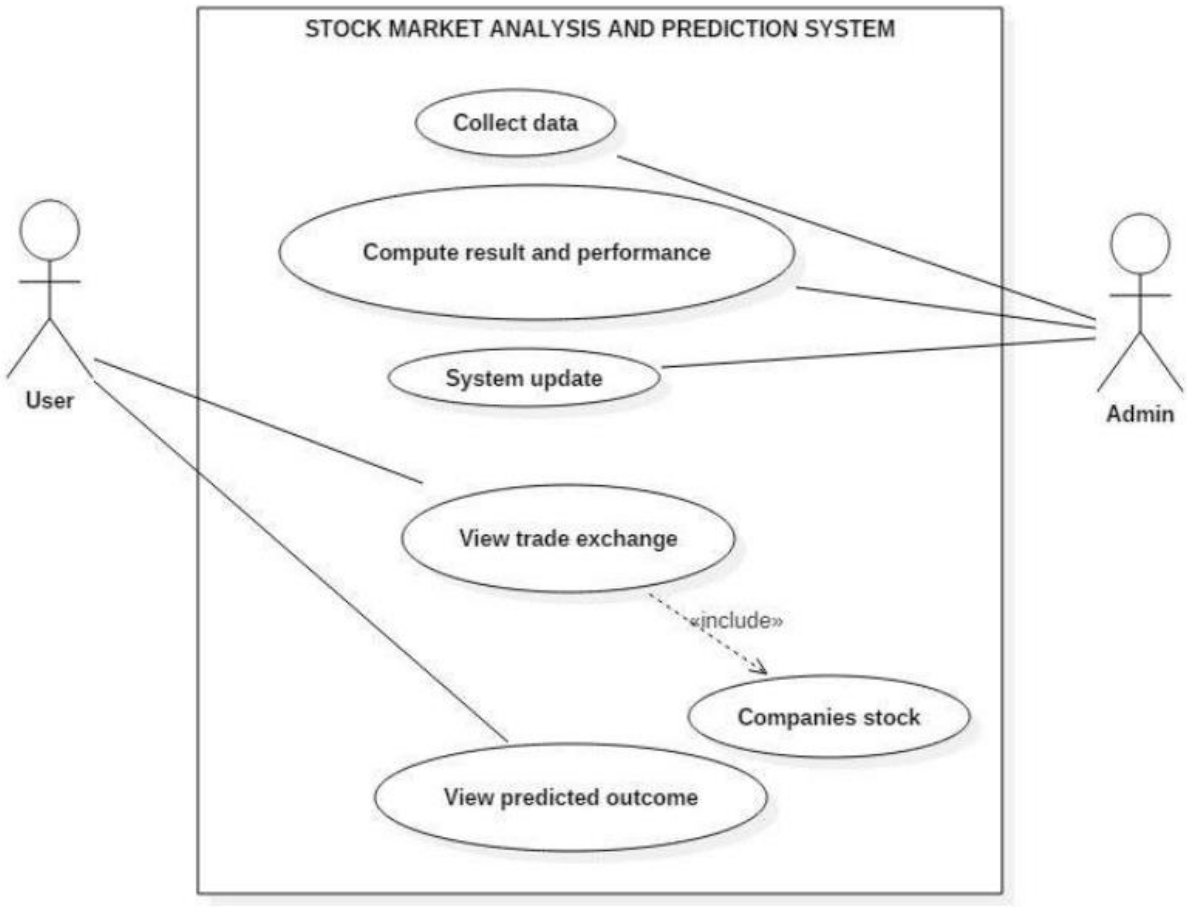
The website is completely portable and the recommendations completely trustworthy as the data is dynamically updated.

5. Interoperability:

The interoperability of the website is very high because it synchronize all the database with the wamp server.

4. System Design and Architecture

4.1 Use Case Diagram



1-Use case index

Use case ID	Use case name	Primary actor	scope	complexity	priority
1	Collect data	admin	in	high	1
2	Compute result and prepare	admin	in	high	1
3	System update	admin	in	high	1
4	View trade exchange	user	in	medium	2
5	Company stock	user	in	medium	2
6	View predicted outcome	user	in	high	1

2-Use case description:

Use case ID:1
Use case name: Collect data
Description: Every required data will be available in stock exchange. System will be able to collect the data for system.

Use case ID:2
Use case name: Compute result and performance
Description: Prediction result will be handled and generated by System. The system will be built, through which the result of prediction and system performance will be analyzed.

Use case ID: 3
Use case name: System update
Description: With the change of market and technology regular update of system is required. Beside there the predict result of stock exchange and their actual price will be updated by system automatically in regular basis.

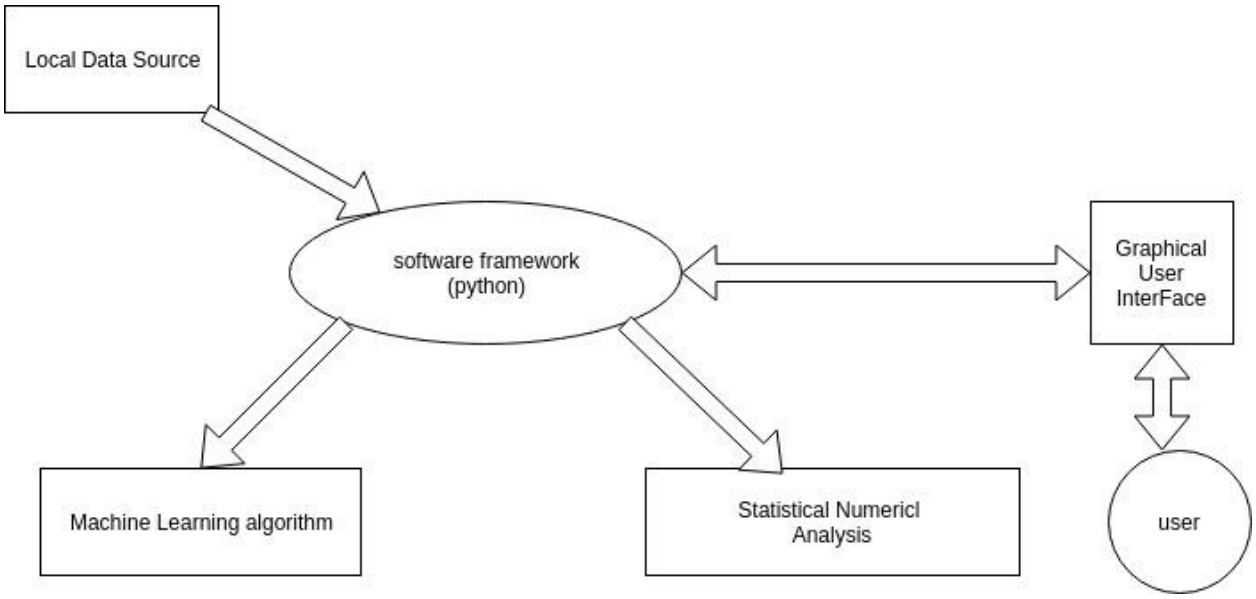
Use case ID: 4
Use case name: View traded exchange

Description: Company trading which is held at Stock exchange can be viewed by user.

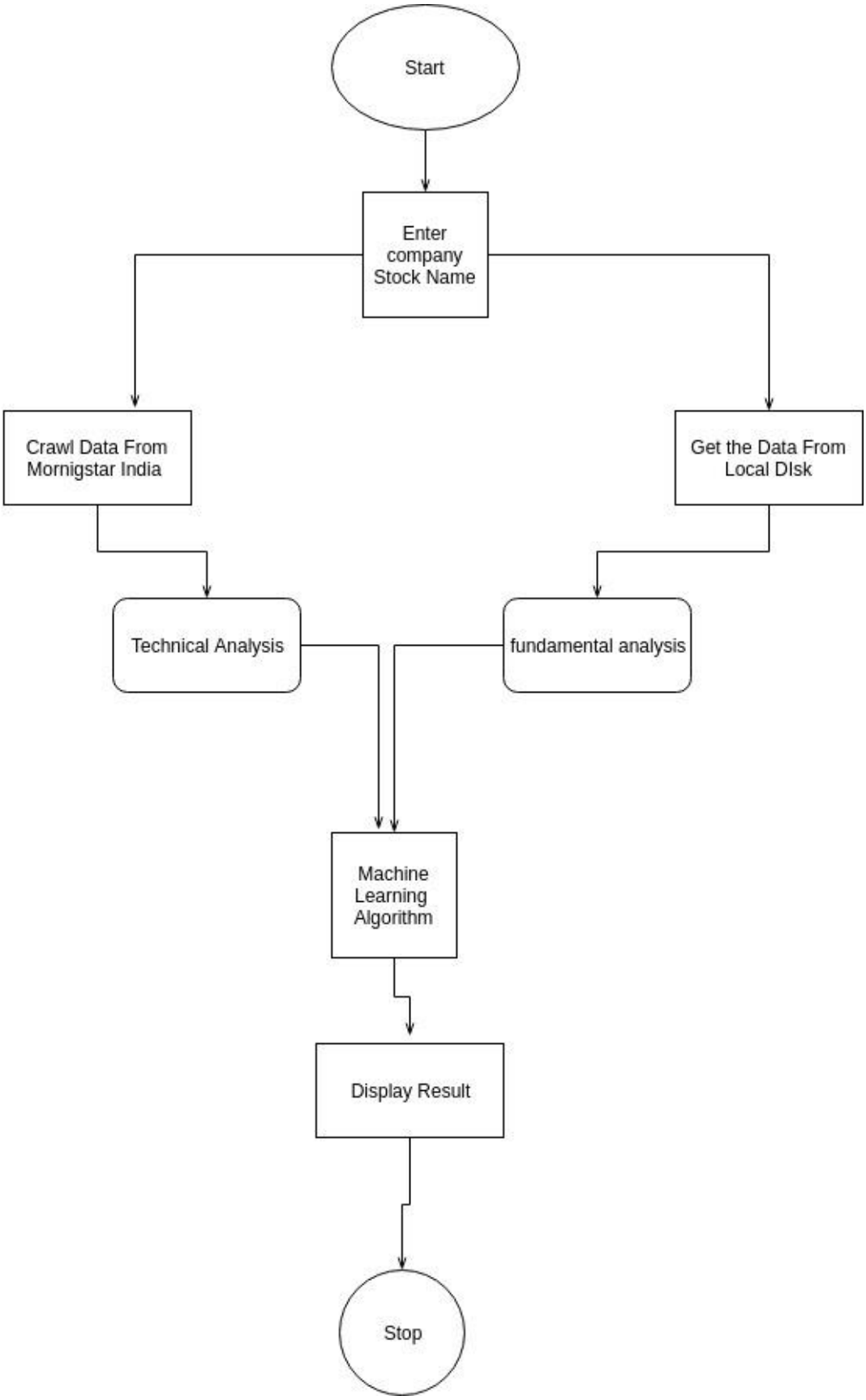
Use Case ID: 5
Use Case Name: Company Stock
Description: It is extended feature of view traded exchange. This includes the stock value of particular company.

Use Case ID: 6
Use Case Name: View predicted outcome
Description: This use case is must important in whole project. The key feature of this project is to predict the stock value of hydro power companies. Thus, this will be available in user interface and viewer can observe them.

4.2. System Flow diagram



4.3 CONTROL FLOW DIAGRAM


Start
Enter company Stock Name
Crawl Data From Mornigstar India
Get the Data From Local DIsk
Technical Analysis
fundamental analysis
Machine Learning Algorithm
Display Result
Stop

4.4 Proposed Algorithm

1. Support Vector Machines The support vector machines (SVMs) were proposed by Vapnik in 1999. There are two main categories for support vector machines: support vector classification (SVC) and support vector regression (SVR). SVM is a learning system using a high dimensional feature space. Khemchandani and Chandra stated that in SVM, points are classified by means of assigning them to one of two disjoint half spaces, either in the pattern space or in a higher-dimensional feature space. Khemchandani et al (2009). The main objective of support vector machine is to identify maximum margin hyper plane. The idea is that the margin of separation between positive and negative

2. Random forest Decision tree learning is one of the most popular techniques for classification. Its classification accuracy is comparable with other classification methods, and it is very efficient. ID3 presented by Quinlan (1986), C4.5 presented by Quinlan (1993) and CART presented by Breiman et al (1984) are decision tree learning algorithms. Details can be found in article of Han et al (2006). Random forest belongs to the category of ensemble learning algorithms. It uses decision tree as the base learner of the ensemble. The idea of ensemble learning is that a single classifier is not sufficient for determining class of test data. Reason being, based on sample data, classifier is not able to distinguish between noise and pattern. So it performs sampling with replacement such that given n trees to be learnt are based on these data set samples. Also in our experiments, each tree is learnt using 3 features selected randomly. After creation of n trees, when testing data is used, the decision which majority of trees comes up with is considered as the final output. This also avoids problem of over-fitting

3. K-nearest neighbor The K-nearest neighbor (KNN) classification approach is an instant-based learning algorithm that uses the nearest distance in determining the category of new vector in the training data set. During the training stage, the feature space is divided into multiple regions and the training data points are mapped into these regions according to the similarity of their contents. The unlabeled input data points are categorized to a particular category by finding the closet or distance from input data point

and that particular category. The KNN approach needs only a small number of training data points and this has contributed to the simplicity of the KNN which makes it outperforms other classification approaches, The most commonly and widely used distance function for the KNN classifier is the Euclidean distance formula and it is used to calculate the distance between the new unlabeled data point and the training data points. The main step in the classification stage of the KNN is to measure the distance in order to identify the nearest neighbors of the new input data point.

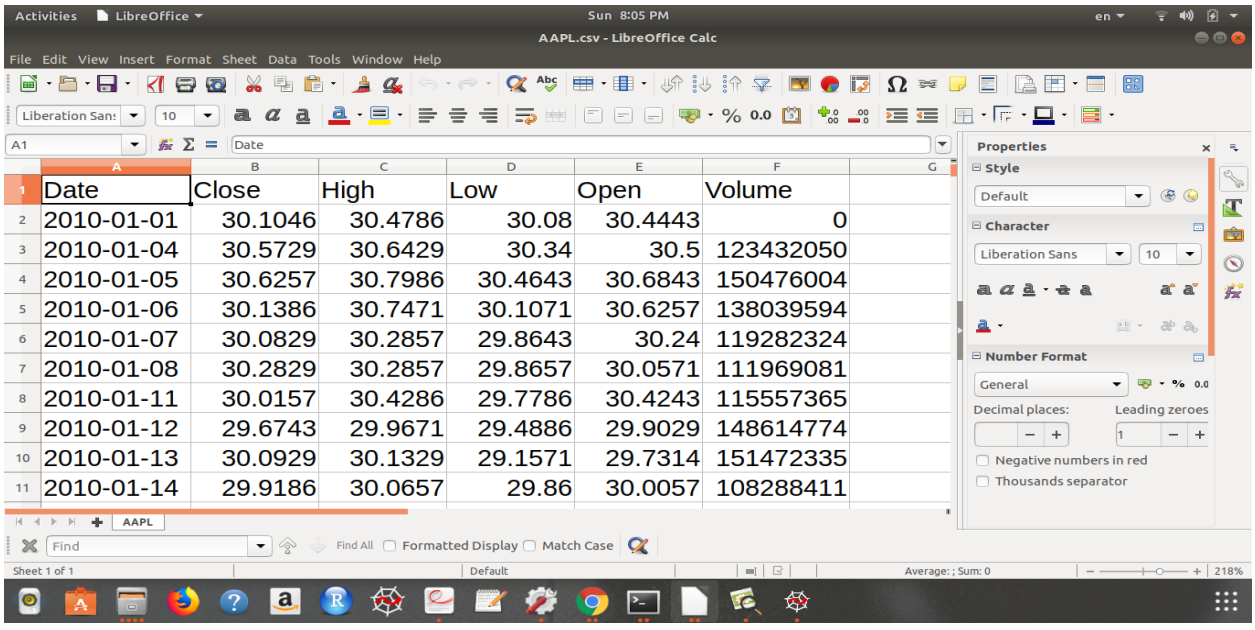
5.FINDINGS AND CONCLUSIONS

The system evaluation on the stocks from Stock Exchange is carried out. For given day's open index, day's high, day's low, volume and adjacent values along with the stock news textual data, our forecaster will forecast the closing index value for particular trading day.

Our predictive model is evaluated on stock market on the financial historical stock data over the training period of June 2000to july 2018. The news data is collected from the financial web sites **https://www.morningstar.in .** The news data is collected once in day. The stock quotes corresponding to each trading day were downloaded from **https://www.morningstar.in** .

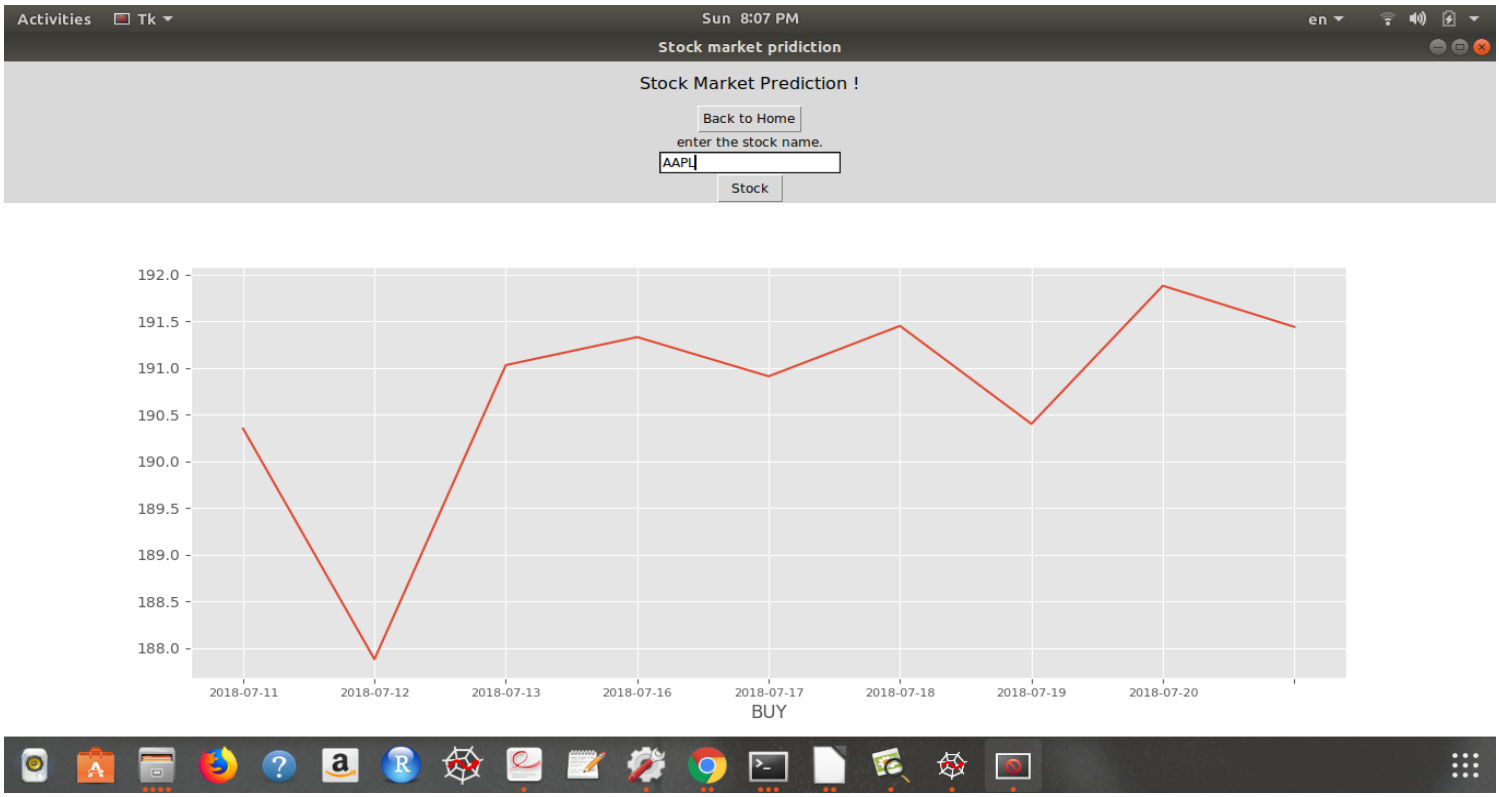
The accuracy of the system is measured as the percentage of the predictions that were correctly determined by the system. For instance, if the system forecasts an upward trend and the index indeed goes up, it is supposed to be correct, otherwise, if the index goes down or remains stable for an uptrend, it is assumed to be wrong.

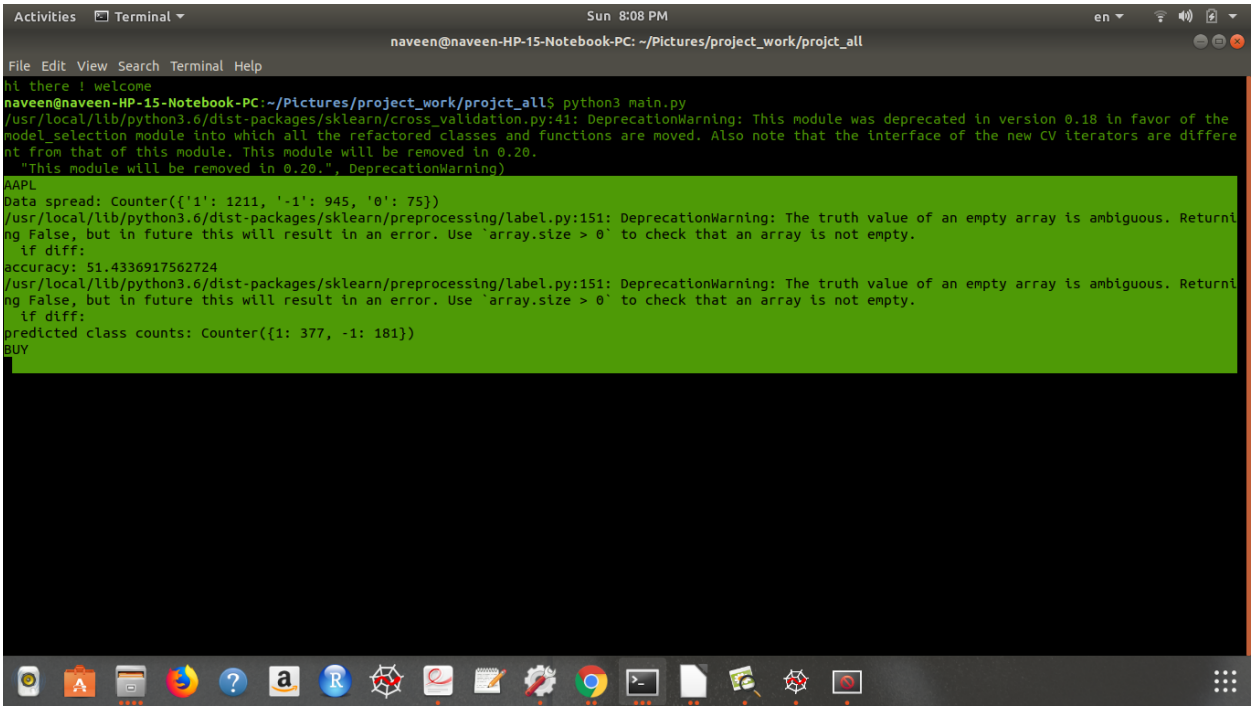
Following stock dataset is taken as sample training data of Apple(AAPL) over the period of 22 days. Corresponding rates file is also provided along with this. Predictions using stock quotes are shown .Whenever the desired predictions using quotes are varying from actual one, we rebuilt neural network by considering the



	Date	Close	High	Low	Open	Volume
2	2010-01-01	30.1046	30.4786	30.08	30.4443	0
3	2010-01-04	30.5729	30.6429	30.34	30.5	123432050
4	2010-01-05	30.6257	30.7986	30.4643	30.6843	150476004
5	2010-01-06	30.1386	30.7471	30.1071	30.6257	138039594
6	2010-01-07	30.0829	30.2857	29.8643	30.24	119282324
7	2010-01-08	30.2829	30.2857	29.8657	30.0571	111969081
8	2010-01-11	30.0157	30.4286	29.7786	30.4243	115557365
9	2010-01-12	29.6743	29.9671	29.4886	29.9029	148614774
10	2010-01-13	30.0929	30.1329	29.1571	29.7314	151472335
11	2010-01-14	29.9186	30.0657	29.86	30.0057	108288411

news data of that day





5.1 CONCLUSION

Evaluating the Stock market prediction has at all times been tough work for analysts. Thus, we attempt to make use of vast written data to forecast the stock market in dices. If we join both techniques of textual mining and numeric time series analysis the accuracy in predictions can be achieved. Artificial neural network is qualified to forecast BSE market upcoming trends. Financial analysts, investors can use this prediction model to take trading decision by observing market behavior.

5.2 FUTURE WORK

1. More work on refining key phrases extraction will definitely produce better results. Enhancements in the preprocessor unit of this system will help in improving more accurate predictability in stock market.

2. Twitter feeds message board, Extracting RSS feeds and news.