SAVITRIBAI PHULE PUNE UNIVERSITY

A PROJECT REPORT ON
Stock Market Prediction System

BY

Harshali Bedmutha **(B150028507)**
Ria Garg **(B150028525)**
Ritika Jain **(B150028535)**

Under the Guidance of

Dr. Mrs. Himangi Pande

In partial fulfillment of

MAEER's

MAHARASHTRA INSTITUTE OF TECHNOLOGY

B.E (INFORMATION TECHNOLOGY)

Savitribai Phule Pune University

2018-2019

MAHARASHTRA INSTITUTE OF TECHNOLOGY

Department of Information Technology

CERTIFICATE

This is to certify that the Project entitled

"Stock Market Prediction System"

Submitted by

Name Of Candidate	Exam Seat No:
Harshali Bedmutha	B150028507
Ria Garg	B150028525
Ritika Jain	B150028535

Is record of bonafide work carried out by him/her, during the second semester of academic year 2018-2019 under my guidance, towards the partial fulfillment of the degree of Bachelor of Engineering in Information Technology of Savitribai Phule Pune University under the Department of Information Technology, Maharashtra Institute of Technology, Pune.

Date: 1-6-2019

Place: Maharashtra Institute of Technology

Dr. Mrs. Himangi Pande
(Internal Guide)

Prof.
(External Examiner)

Mrs. Sumedha Sirsikar
(***Head, Dept. of Information Technology***)

ACKNOWLEDGEMENT

We take this opportunity to express our sincere appreciation for the cooperation given by our guide Dr. Mrs. Himangi Pande (Department of Information Technology) and need a special mention for all the motivation and support. We are deeply indebted to our guide for completion of this project for which she has guided and helped us going out of the way.

We take this opportunity to express our sincere appreciation for the cooperation given by Prof. Mrs. Sumedha Sirsikar, HOD (Department of Information Technology) and all the teachers of our department.

For all efforts behind the Seminar report, we would also like to express our sincere appreciation to staff of Department of Information Technology, Maharashtra Institute of Technology Pune, for their extended help and suggestions at every stage.

We would also like to thank all our friends who have helped us to understand the concepts related to the project.

Content

Abstract

Stock market or equity market have a profound impact in today's economy. A rise or fall in the share price has an important role in determining the investor's gain. Predictions on stock market prices are a great challenge due to the fact that it is an immensely complex, chaotic and dynamic environment. There are many examples of Machine Learning algorithms been able to reach satisfactory results when doing prediction on stock market data. Different Deep Learning algorithms for prediction of stock market data include Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), Convolutional Neural Network (CNN), Support Vector Machine (SVM) and many more.

The idea of the project is to develop a system for Stock Market Prediction for various companies listed in National Stock Exchange (NSE). We use CNN and SVM to construct a model to effectively predict stock prices .Different layers used in CNN are Convolution , Max Pooling , Dropout, Fully Connected and Dense. The optimizer used is adam. Support Vector Regression with linear kernel model to give the predicted value.

The project takes input the historical and real-time stock data from Yahoo Finance. The parameter taken into consideration are Open, High, Low, Close, Adjacent Close . We use the above listed features to predict the Closing Price for next day. We also develop a User Interface with functionalities to Get Quote of a company stock, Get predicted Close Value and visualize the data using graphs.

Keywords: *Stock market, Stock market prediction, Machine Learning, RNN, CNN, LSTM, SVM, NSE, day-wise prediction,real-time*

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

	Abbreviation	Full-text
1	RNN	Recurrent Neural Network
2	CNN	Convolutional Neural Network
3	LSTM	Long Short Term Memory
4	ETF	Exchange Trade Fund
5	NSE	National Stock Exchange
6	SVM	Support Vector Machine

CHAPTER 1

INTRODUCTION TO STOCK MARKET

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.

Fig 1.1

Fig 1.2

A stock exchange is an exchange where stock brokers and traders can buy and sell shares of stock, bonds, and other securities. Many large companies have their stocks listed on a stock exchange. This makes the stock more liquid and thus more attractive to many investors. The exchange may also act as a guarantor of settlement. Other stocks may be traded "over the counter", that is, through a dealer. Some large companies will have their stock listed on more than one exchange in different countries, so as to attract international investors.

1.1 Overview of Stock Market Prediction

Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit. The efficient-market hypothesis suggests that stock prices reflect all currently available information and any price changes that are not based on newly revealed information thus are inherently unpredictable. Others disagree and those with this viewpoint possess myriad methods and technologies which purportedly allow them to gain future price information.

Prediction methodologies fall into two broad categories which can (and often do) overlap. They are fundamental analysis and technical analysis.

a. Fundamental analysis

Fundamental Analysts are concerned with the company that underlies the stock itself. They evaluate a company's past performance as well as the credibility of its accounts. Many performance ratios are created that aid the fundamental analyst with assessing the validity of a stock, such as the P/E ratio. Warren Buffett is perhaps the most famous of all Fundamental Analysts.

b. Technical analysis

Technical analysts or chartists are not concerned with any of the company's fundamentals. They seek to determine the future price of a stock based solely on the (potential) trends of the past price (a form of time series analysis). Numerous patterns are employed such as the head and shoulders or cup and saucer. Alongside the patterns, techniques are used such as the exponential moving average (EMA). Candle stick patterns, believed to have been first developed by Japanese rice merchants, are nowadays widely used by technical analysts.

1.2 Motivation behind project topic

Stock price prediction is a classic and important problem. With a successful model for stock prediction, we can gain insight about market behaviour over time, spotting trends that would otherwise not have been noticed. With the increasingly computational power of the computer, machine learning will be an efficient method to solve this problem. We can try to improve the accuracy of prediction by modifying the existing algorithms.

1.3 Aim and Objectives of the work

The aim of this project is to create a prediction system for stock market and accurately predict the future value of stock based on the past data.

Objectives:

- To develop a system for accurately predicting stock price using Convolutional Neural Network Architecture(CNN) and Support Vector Machine(SVM).
- To determine the best values for pool length, no. of filters, no of hidden layer while construction of the CNN model.
- To determine appropriate parameters for SVM model.
- To achieve higher efficiency in stock prediction.

CHAPTER 2

LITERATURE SURVEY

[1] Sreelekshmy Selvin, Vinayakumar R, Gopalakrishnan E.A, Vijay Krishna Menon, 'STOCK PRICE PREDICTION USING LSTM,RNN AND CNN-SLIDING WINDOW MODEL', International Conference on Advances in Computing, Communications and Informatics (ICACCI),IEEE, Udupi, India, Sept 2017 . In [1] the proposed method is a model independent approach. They are not fitting the data to a specific model, rather they are identifying the latent dynamics existing in the data using deep learning architectures. In this work they use three different deep learning architectures for the price prediction of NSE listed companies and compares their performance. [1] uses a sliding window approach for predicting future values on a short term basis. The performance of the models were quantified using percentage error. There are at least four types of stock data: (a) the historic price data of the stock (e.g. volume, high, low, open); (b) the technical analysis data that is calculated from (a) (e.g. moving average convergence / divergence (MACD)); (c) the historic price data of market indexes and/or other related stocks; (d) the economic fundamentals (e.g. gross domestic product (GDP), oil price). [1] uses time series data or historic price data for analysis. While [2] used the (a) and (c) but not (b) type of stock data, in an effort to avoid the co-founding pitfalls, with also a limit of not including the (d) type of data.

Strengths:

- This paper makes use of the sliding window model.
- Plots graphs for comparing results.

Limitations:

- Short term prediction is done, prediction for the next 10 minutes.

[2] Luca Di Persio , Oleksandr Honchar 'Artificial Neural Networks architectures for stock price prediction: comparisons and applications , International Journal of Circuits, Systems and signal processing, Verona,Italy ,2016. In [2], they presented an Artificial Neural Network (ANN) approach to predict stock market indices, particularly with respect to the forecast of their trend movements up or down. they considered the Multi-layer Perceptron

(MLP), the Convolutional Neural Networks(CNN), and the Long Short-Term Memory (LSTM) recurrent neural networks techniques. They focus on the importance of choosing the correct input features, along with their preprocessing, for the specific learning algorithm one wants to use. They also compared results trained on a daily basis, for the S&P500 index, showing, in particular, that convolutional neural networks (CNN) can model financial time series better then all the other considered architectures. They also implemented a novel Wavelet + CNN algorithm which outperforms other neural network approaches.

Strengths:

- Ensembles produce better results.
- Compares the results of MLP, CNN and LSTM.

Limitations:

- PCA produces poor results.

[3] Avraam Tsantekidis, Nikolaos Passalis, Anastasios Tefas, Juho Kanniainen, Moncef Gabbouj and Alexandros Iosifidis 'Forecasting Stock Prices from the Limit Order Book using Convolutional Neural Networks',2015 . In [3], they proposed a deep learning methodology, based on Convolutional Neural Networks (CNNs), that predicts the price movements of stocks, using as input large-scale, high-frequency time-series derived from the order book of financial exchanges. This includes an intelligent normalization scheme that takes into account the differences in price scales between different stocks and different time periods of an individual stock. To measure the performance of our model we use Kohen's kappa which is used to measure the concordance between sets of given answers, taking into consideration the possibility of random agreements happening. We compare our results with those of a Linear SVM model and an MLP model with Leaky Rectifiers as activation function. The proposed method significantly outperforms all the other evaluated models on the presented metrics, showing that the convolutional neural network can better handle the sequencial nature of the LOB data and better determine the microstructure of the market in order to detect mid-price changes that occur.

Strengths:

- Calculates recall, precision and F1.

- Uses an intelligent normalization scheme that takes into account the differences in price sales between different stocks.

Limitations:

- Data used for the model is less and can be increased for more accurate results.

[4] Anastasia Borovykh Sander,Bohte, Cornelis W. Oosterle,'Conditional time series forecasting with convolutional neural networks' June 2018. [4] present a method for conditional time series forecasting based on an adaptation of the recent deep convolutional WaveNet architecture. The proposed network contains stacks of dilated convolutions that allow it to access a broad range of history when forecasting, a ReLU activation function and conditioning is performed by applying multiple convolutional filters in parallel to separate time series which allows for the fast processing of data and the exploitation of the correlation structure between the multivariate time series. We analyze the performance of the network on the S&P500 data in combination with the volatility index and the CBOE 10 year interest rate to analyze the ability of the model to extract both unconditionally as well as conditionally. We define a training period of 750 days (approximately three years) and a testing period of 250 days (approximately one year) on which we perform the one-day ahead forecasting. The data from 01-01-2005 until 31-12-2016 is split into nine of these periods with non-overlapping testing periods. We conclude that even though time series forecasting remains a complex task and finding one model that fits all is hard, we have shown that the WaveNet is a simple, efficient and easily interpretable network that can act as a strong baseline for forecasting.

Strengths:

- Use of convolutional wavenets gives efficient system with easy forecasting.
- Addresses the problem of overfitting.

Limitations:

- This approach can be improved by making CNN learn non-linear dependencies.

[5]Meesad, P., & Rasel, R. I. (2013). Predicting stock market price using support vector regression. 2013 International Conference on Informatics, Electronics and Vision (ICIEV). In [5] paper, support vector regression (SVR) analysis is used as a machine learning

technique in order to predict the stock market price as well as to predict stock market trend. Moreover, different types of windowing operators are used as data preprocess or input selection technique for SVR models. Support vector regression is a useful and powerful machine learning technique to recognize pattern of time series dataset. It can produce good prediction result if the value of important parameters can be determined properly. This study is done on a well known company of Dhaka stock exchange (DSE), named ACI group of company Limited. Four year's historical time series dataset are collected from the DSE from 2009 to 2012, as daily basis for experimentations. Finally, predicted results from WinSVR models are compared with actual price values of DSE to evaluate the model prediction performance.

Strengths:

- It makes use of different kinds of windowing functions for data pre-processing.
- This approach can be used for predicting values of 1-day ahead, 5-day ahead and 22-day ahead also.

Limitations:

- Only one dataset of DSE Bangladesh was used for training and testing.

[6] R., A., & M.S., V. (2012). Stock Price Prediction Using Support Vector Regression. Global Trends in Computing and Communication Systems, 588–597. This paper demonstrates the applicability of support vector regression, a machine learning technique, for predicting the stock price by learning the historic data. The stock data for the period of four years is collected and trained with various parameter settings. The performance of the trained model is evaluated by 10-fold cross validation for its predictive accuracy. It has been observed that the support vector regression model with RBF kernel shows better performance when compared with other models. The historical stock data include opening price of the stock, previous closing price, highest price of stock, lowest price of stock, last price of the stock in a day and average price; these are all the input parameters or predictors. The input parameters are applied to learn the model, which is then used to predict the closing price of the stock.

Strengths:

- The performance of the models has been analysed based on the evaluation criteria such as predictive accuracy, mean squared error and correlation coefficient.

[7] M. Karazmodeh, S. Nasiri, and S. Majid Hashemi, 'Stock Price Forecasting using Support Vector Machines and Improved Particle Swarm Optimization', Journal of Automation and Control Engineering, North Cyprus, Turkey, June ,2013

[7] paper proposed the improved Particle Swarm Optimization (PSO) based on Support Vector Machines (SVM) for efficient prediction of various stock indices. Some experimental results show that PSO has greater" global search" ability, but the "local search" ability around the optimum is not very good. In order to enhance "local search" ability of PSO, an improved particle swarm optimization was introduced in this paper, which was PSO with mutation. In this paper, a set of technical indicators, obtained from the stock to be predicted, and also from the stocks exhibiting high correlation with that stock were used as input features. The approach is done on the three most well-known stock market indices, DJI, S&P 500 and Nasdaq-100. The used data is fetched from yahoo finance, Google finance and NYSE that starts in 10.02.2008 till 9.10.2012. 60% of the data was used for training, 20% for validation and 20% for testing the system. The calculated results have been compared with the solo SVM model. Results state that mutation in particles led the accuracy being higher, since the particle always searches the whole state space which prevents the mistakes of not finding the other best options in other possible states. By taking the political & economical factors into account even more accurate results could be observed.

Strengths:

- This approach is done on three most well-known stock market indices.
- PSO used with SVM gives very high accuracy of prediction, of up to 95%.

Limitations:

- Does not take into account the political and economic factors affecting the stock value.

CHAPTER 3

SPECIFICATION

3.1 System Requirements

- The system should collect data from web sources like Yahoo Finance and update its database from a set of 3 NSE companies (Infosys, TCS, Cipla).
- The system should train itself from a set of past data and simulate on a remaining set of test data to improve its precision.
- The system should aggregate all indications given by technical indicators and try to provide a recommendation that supports the more robust neural network system.
- A friendly user interface should be designed for customers and access to specific services should be restricted to only registered users to dissuade hackers from destabilizing the system.
- Various customer profiles should be maintained by the system and confidential customer data should be protected at all costs from prying softwares

3.2 Software Requirements

Technology/Tool	Platform	Use
HTML,CSS, Javascript	Windows	Web UI Development
XAMPP(MYSQL)	Windows	Backend Database
Python 3.0	Windows	To run ML algorithms
Keras	Windows	For Deep Learning Algorithms
Matplotlib	Windows	To visualize the data properly
Scikit-learn	Windows	Used for data-mining
Numpy	Windows	Used for scientific computing

Table 3.1 : Software Requirements Table

3.2.1 HTML/CSS/JavaScript:

HTML is at the core of every web page, regardless the complexity of a site or number of technologies involved. It's an essential skill for any web professional. It's the starting point for anyone learning how to create content for the web. And, luckily for us, it's surprisingly easy to learn. CSS stands for Cascading Style Sheets. This programming language dictates how the HTML elements of a website should actually appear on the frontend of the page. JavaScript is a more complicated language than HTML or CSS, and it wasn't released in beta form until 1995. Nowadays, JavaScript is supported by all modern web browsers and is used on almost every site on the web for more powerful and complex functionality

.

- **HTML** provides the basic structure of sites, which is enhanced and modified by other technologies like CSS and JavaScript.
- **CSS** is used to control presentation, formatting, and layout.
- **JavaScript** is used to control the behavior of different elements.

3.2.2 XAMPP (MySQL):

XAMPP is an open source software developed by Apache friends. It is basically a local host or a local server. This local server works on your own desktop or laptop computer. The use of XAMPP is to test the clients or your website before uploading it to the remote web server. This XAMPP server software gives you the suitable environment for testing MYSQL, PHP, Apache and Perl projects on the local computer.

The full form of XAMPP is X stands for Cross-platform, (A) Apache server, (M) Maria DB, (P) PHP and (P) Perl. The Cross-platform usually means that it can run on any computer with any operating system.

MYSQL is an open source software. It is actually a relational database management system (RDBMS). This SQL stands for Structured Query Language. It is the most popular and best RDBMS used for developing a variety of web-based software applications. With the help of MYSQL, it is possible to organize the information, manage, retrieve and update the data whenever you wish to do.

3.2.3 Python 3.0:

Python's standard library is very extensive, offering a wide range of facilities The library contains built-in modules (written in C) that provide access to system functionality such as file I/O that would otherwise be inaccessible to Python programmers, as well as modules written in Python that provide standardized solutions for many problems that occur in everyday programming. Some of these modules are explicitly designed to encourage and enhance the portability of Python programs by abstracting away platform-specifics into platform-neutral APIs. The Python installers for the Windows platform
usually include the entire standard library and often also include many additional components. For Unix-like operating systems Python is normally provided as a collection of packages, so it may be necessary to use the packaging tools provided with the operating system to obtain some or all of the optional components.

3.2.4 Keras:

Keras is an open-source neural-network library written in Python. It is capable of running on top of Tensorflow, Microsoft Cognitive Toolkit, Theano, or PlaidML. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible. Keras contains numerous implementations of commonly used neural-network building blocks such as layers, objectives, activation functions, optimizers, and a host of tools to make working with image and text data easier. The code is hosted on GitHub, and community support forums include the GitHub issues page, and a Slack channel. In addition to standard neural networks, Keras has support for convolutional and recurrent neural networks. It supports other common utility layers like dropout, batch normalization, and pooling.

3.2.5 Matplotlib:

It is a Python 2D plotting library which produces publication quality figures in avariety of hard copy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits. Matplotlib can be used generate plots, histograms, power spectra, bar charts, error charts, scatterplots, etc., with just a few lines of code. For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with IPython.

3.2.6 Scitkit-learn:

Simple and efficient tools for data mining and data analysis. Accessible to everybody and reusable in various contexts built on Numpy, SciPy and matplotlib. Open source, commercially usable – BSD license

3.2.7 Numpy:

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. Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

3.3 Hardware Requirements

Name	Use
PC/Laptop/mobile	To use the web based system

Table 3.2 : Hardware Requirements Table

CHAPTER 4

DESIGN

4.1 Architectural Diagram

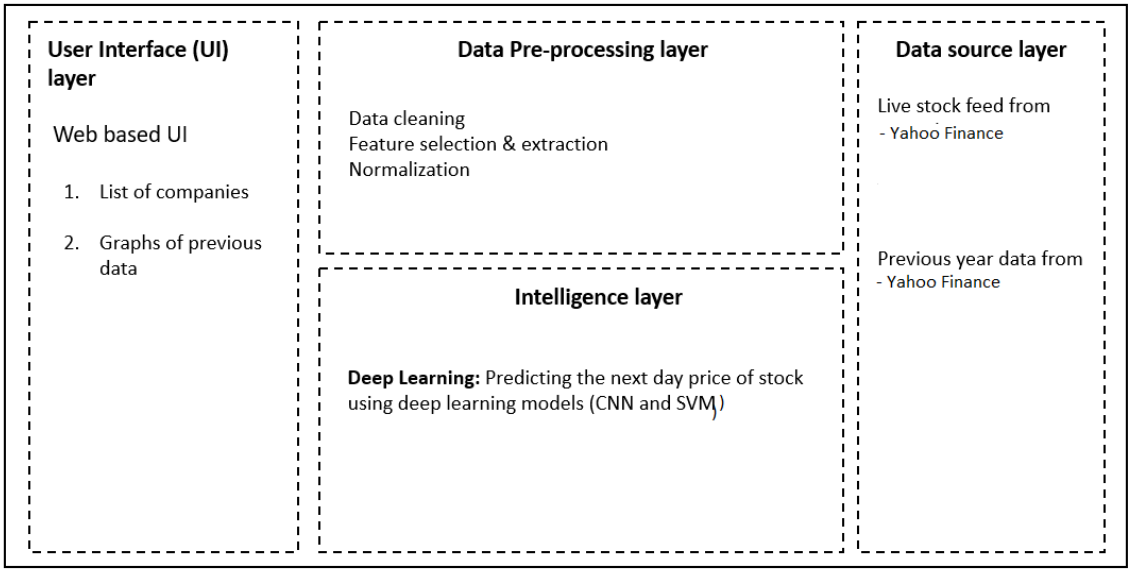

Fig 4.1 : Architectural Diagram

The given diagram has four layers:

1) **User Interface Layer**

The user interface (UI), in the industrial design field of human–computer interaction, is the space where interactions between humans and machines occur. The goal of this interaction is to allow effective operation and control of the machine from the human end, whilst the machine simultaneously feeds back information that aids the operators' decision-making process. Examples of this broad concept of user interfaces include the interactive aspects of computer operating systems, hand tools, heavy machinery operator controls, and process controls. The design considerations applicable when creating user interfaces are related to or involve such disciplines as ergonomics and psychology. Basically it will provide a list of companies whose data has been analysed by our project.

2) **Data Processing Layer**

Data pre-processing is an important step in the data mining process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values (e.g., Income: −100), impossible data combinations (e.g., Sex: Male, Pregnant: Yes), missing values, etc. Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running an analysis.

If there is much irrelevant and redundant information present or noisy and unreliable data, then knowledge discovery during the training phase is more difficult. Data preparation and filtering steps can take considerable amount of processing time. Data pre-processing includes cleaning, Instance selection, normalization, transformation, feature extraction and selection, etc. The product of data pre-processing is the final training set.

3) **Intelligence Layer**

Neural networks exhibit striking capacity to extract context from convoluted or estimated information. They are utilized to derive patterns and identify trends that are too convoluted to be in any way be comprehended by either humans or other conventional computer processes. A convolutional neural network (CNNs) is a biologically-inspired type of deep neural network(DNN) that has recently gained popularity due to its success in classification problems (e.g. image recognition or time series classification). The CNN consists of a sequence of convolutional layers, the output of which is connected only to local regions in the input.

Support vector machines are supervised learning algorithms, used for the purpose of classification and regression analysis. The basic SVM is a non-probabilistic binary linear classifier. The SVM takes into consideration the training data tied to a specific class and builds a model that assigns new data into one category or other. The algorithm constructs a hyperplane which can be used for the purpose of classification or regression.

4) **Data Source Layer**

Taking live stocks from platforms like Yahoo Finance. Also previous data is collected directly from Yahoo Finance. News updates are generated from Google Alerts API.

4.2 Component Diagram

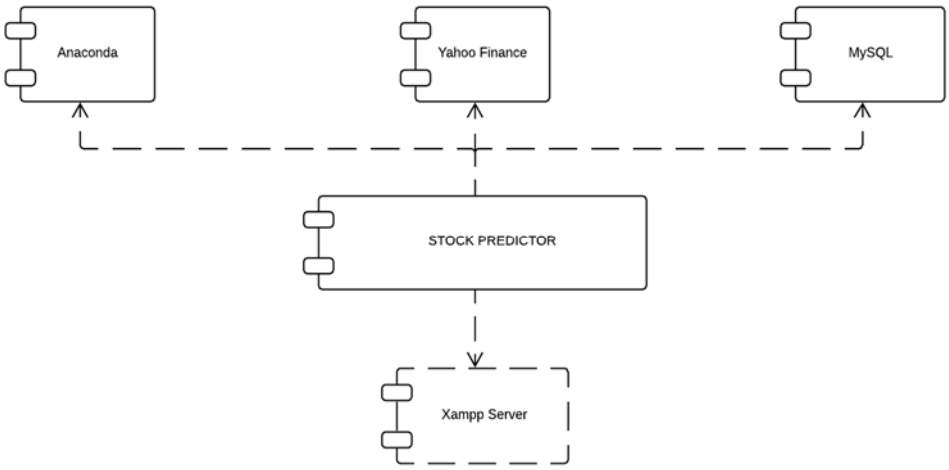

Fig 4.2 : Component Diagram

The above Figure 4.2, shows the component of the system. The component diagram in the figure shows all the components that together forms the whole system. It shows how various smaller components together form a larger component such as the interface component. The components Database and Authentication also contain smaller components.

4.3 Deployment Diagram

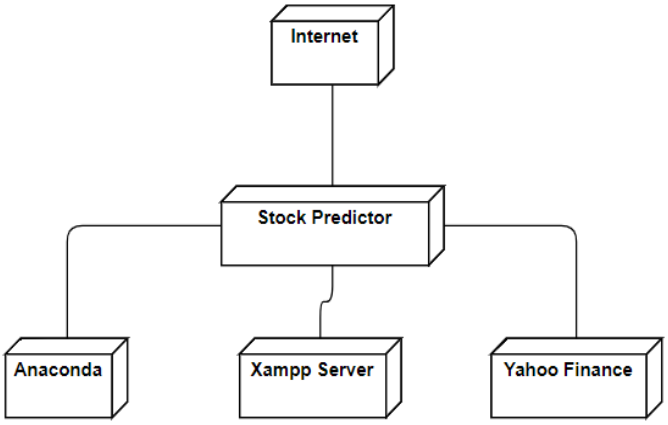

Fig 4.3 : Deployment Diagram

4.4 Sequence Diagram

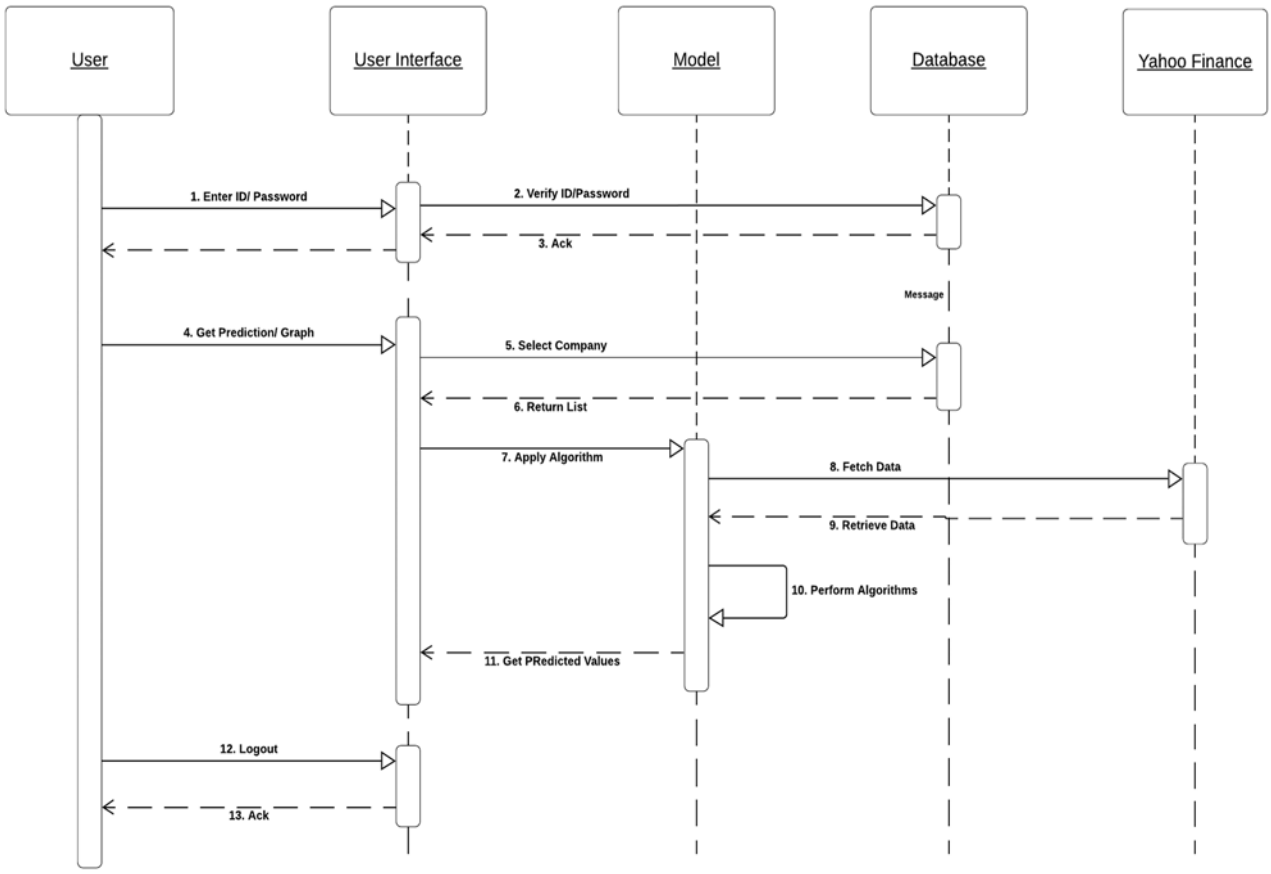

Fig 4.4 : Sequence Diagram

The figure shows the sequence diagram for getting the predicted values for the selected company.

4.5 Use Case Diagram

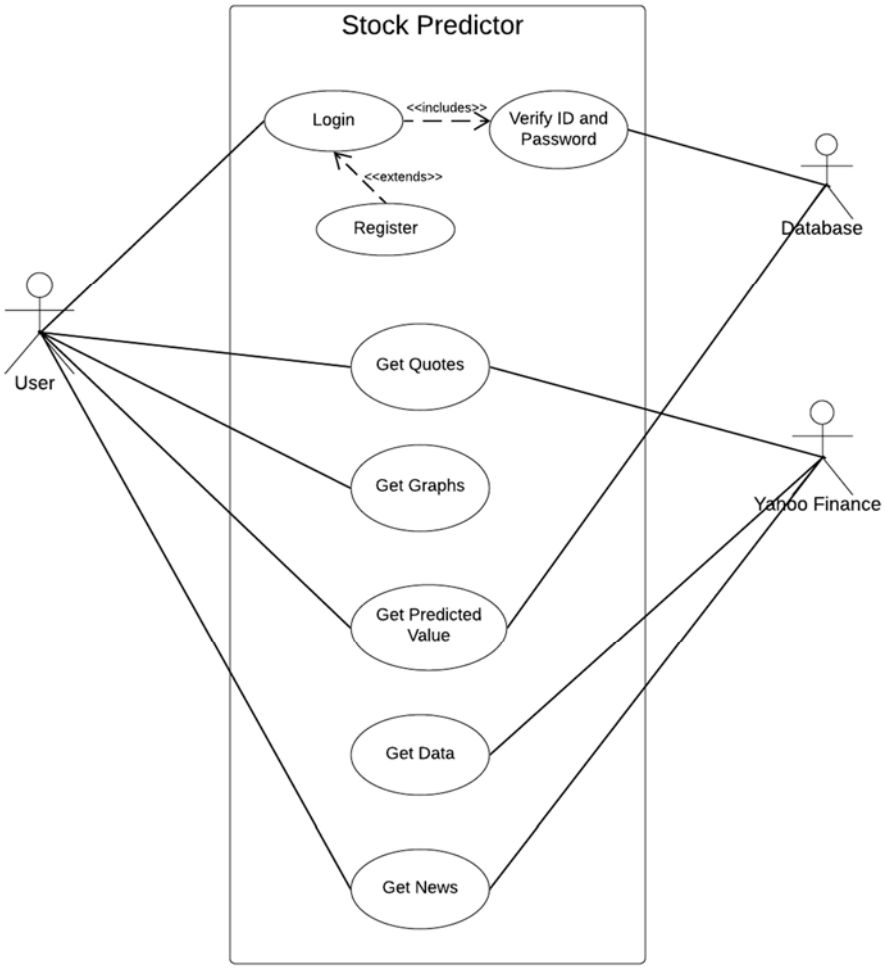

Fig. 4.5 : Use Case Diagram

4.6 Activity Diagram

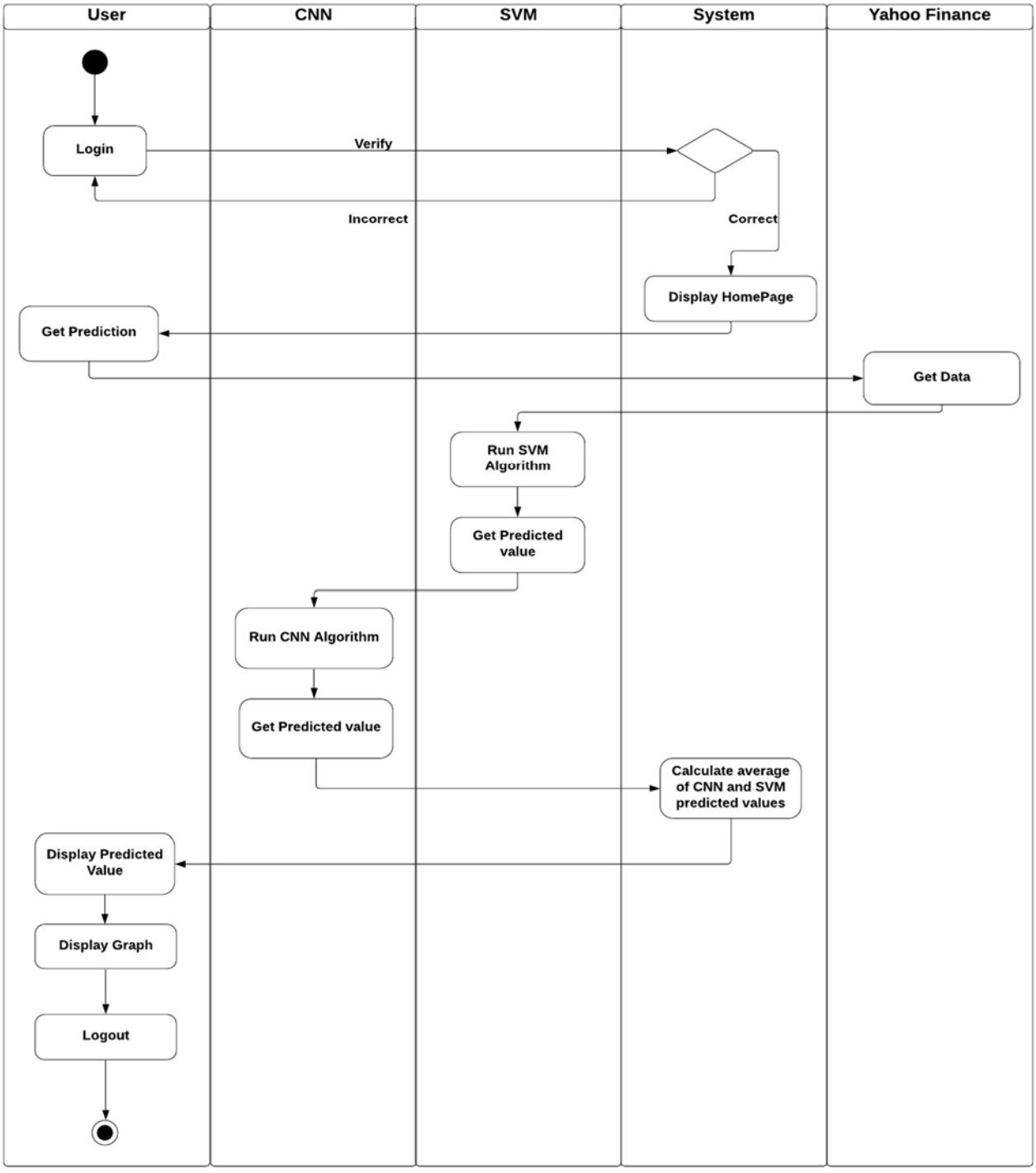

Fig. 4.6 : Activity Swim Lane Diagram

4.7 Data Flow Diagrams

4.7.1 Data Flow Diagram Level 0

Fig 4.7.1 : DFD Level 0

The figure, shows the data flow of the system. A context diagram is a top level (also known as Level 0) data flow diagram. It only contains just the main entities that generalizes the function of the entire system in relationship to external entities.

4.7.1 Data Flow Diagram Level 1

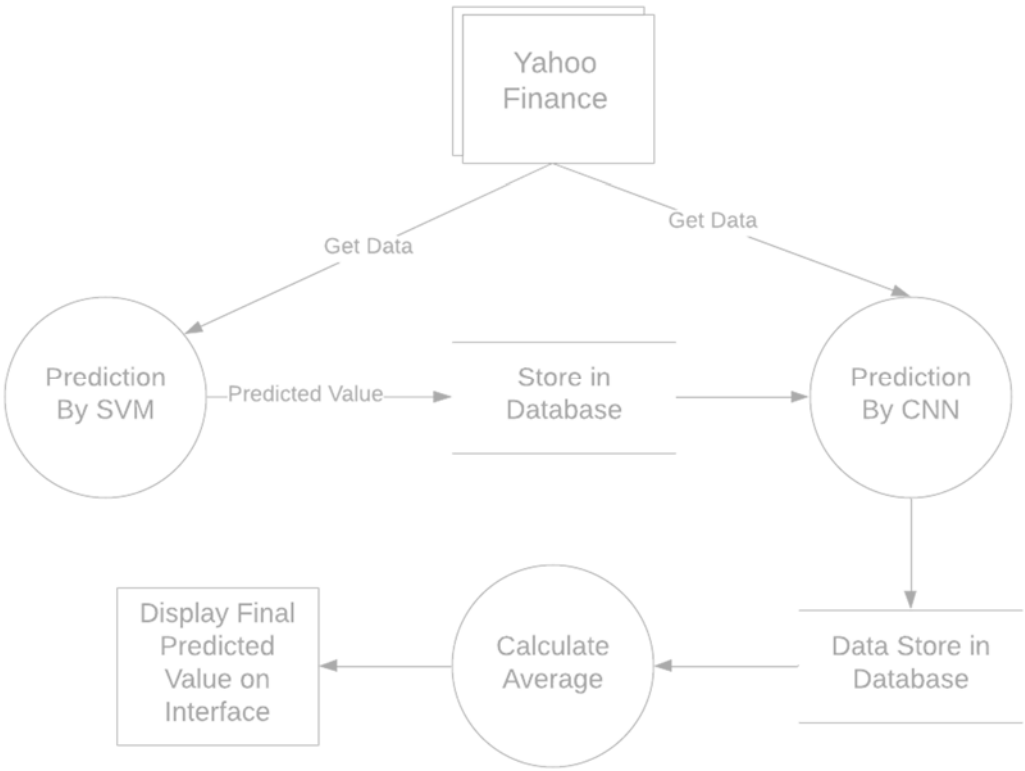

Fig. 4.7.2 : DFD Level 1

It expands the DFD 0 and shows the detailed flow in the proposed system. It shows the different processes that take place to perform the prediction.

4.8 Class Diagram

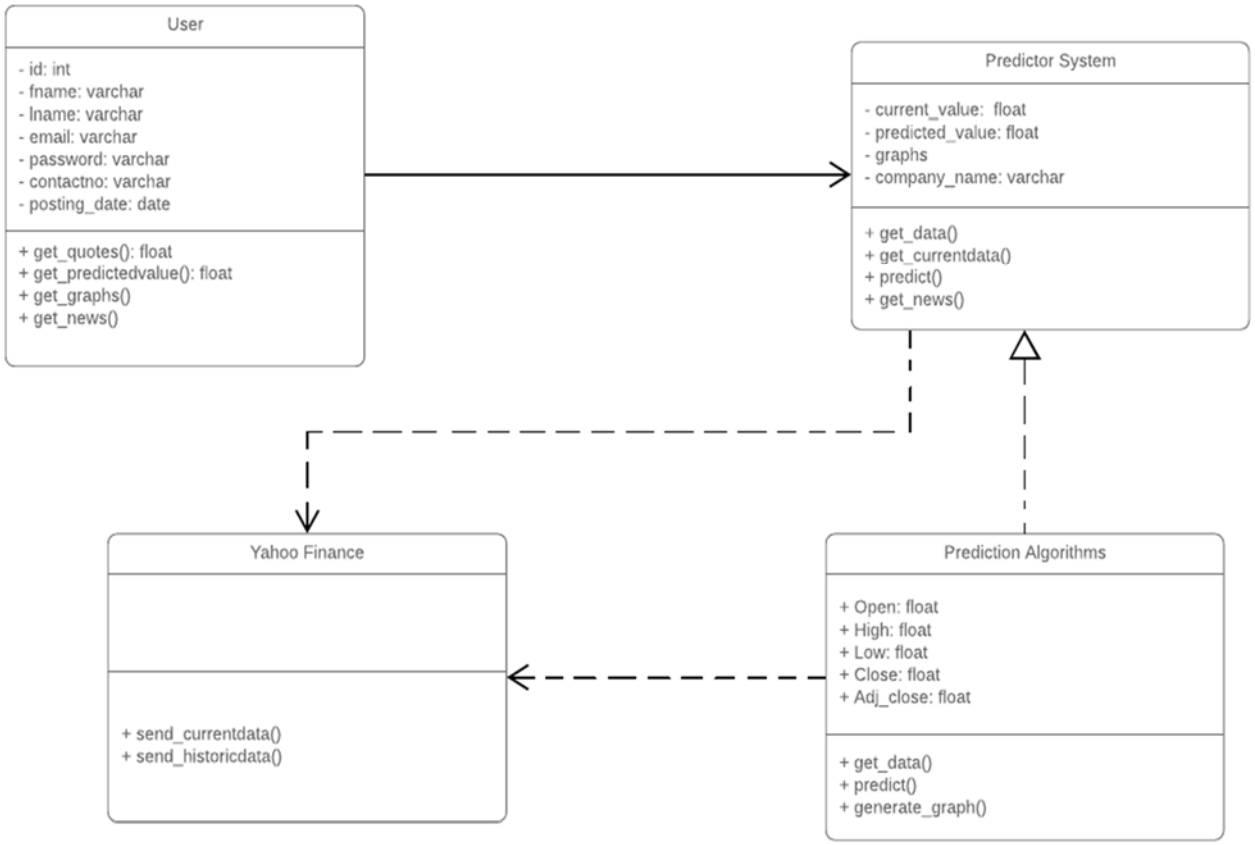

Fig. 4.8 : Class Diagram

Class diagram is a type of structure diagram that shows the structure of the classes, attributes, operations and relationship among them. Given below is the class diagram of the proposed system. The class diagram showed in figure 5.13 below, includes in all 4 classes which describe the structure of proposed system.

CHAPTER 5

IMPLEMENTATION

5.1 IMPLEMENTATION DETAILS

Dataset:

The dataset consists of data of 3 companies TCS, Cipla and Infosys for theperiod 1st September 2013 to today. The data is obtained from Yahoo Finance using pandas_datareader. The features used are Open, High, Low, Close, Adjacent Close. Other features are discarded. The data is cleaned for missing values. The data is fetched everytime the python script for prediction is run and is saved to a .csv file. The dataset obtained from Yahoo Finance is split into training and testing set. 70% of data is used in training and 30% for testing. The metrics for evaluation is accuracy. The results after application of algorithms are stored into excel file(.xlsx) for further use of plotting graphs.

Date	High	Low	Open	Close	Volume	Adj Close
02-09-2013	1034.5	1008.75	1010	1028.43	2154034	872.158
03-09-2013	1033.5	993.075	1025.05	1000.55	2355064	848.5145
04-09-2013	1039.4	996.975	1000	1034.55	2626516	877.3481
05-09-2013	1034.5	990.525	1034.5	995.375	3927492	844.1257
06-09-2013	1012.5	987.5	1004.5	994.4	2759596	843.299
10-09-2013	1008.35	990.5	993	993.95	4300972	842.9173
11-09-2013	1009	991.75	1004	996.9	3008494	845.4191
12-09-2013	999.275	972.9	997.5	983.95	1905568	834.4369
13-09-2013	995	973.275	982.5	976.875	2096320	828.437
16-09-2013	992.5	947.025	975	951.275	2460506	806.727
17-09-2013	976.275	948.55	950	973.725	3327816	825.7656
18-09-2013	982.4	968.95	971.025	977.25	2109818	828.755
19-09-2013	995	963.25	974	983.15	3907240	833.7587
20-09-2013	994.95	970.5	975	975.15	3591858	826.974
23-09-2013	988.6	970	975	977.55	2988758	829.0092
24-09-2013	988.95	963	977.5	969.75	4940300	822.3947
25-09-2013	985.475	960.1	970.75	979.525	4333858	830.6842
26-09-2013	987.05	967.55	977.4	971.45	4919704	823.8365
27-09-2013	980.975	968.725	971.55	974.05	1784666	826.0411
30-09-2013	980.925	960	977	963.9	1987350	817.4335
01-10-2013	977.425	958.35	966	974.525	1591994	826.444
03-10-2013	1018.45	980.2	985.9	1012.3	4672354	858.4791

Fig. 5.1(a) : Dataset Snapshot

Backend Module:

The system uses Python 3.7 with idle integrated development environment for python. The system uses XAMPP webserver for Apache server to host all the UI files and MYSQL

database to store data. The database (Name of Database: stock) consists of tables as follows: 1) Users: It stores the username, id, password, e-mail and contact number. 2) Prediction: It stores the predicted close value for a company by CNN and SVM. The python script evaluates the predicted close price by CNN and SVM and it is inserted into the prediction table of MYSQL database for further use.

User Interface Module:

The system uses HTML, CSS, PHP and Java-script for user interface. HTML and CSS are used in creating GUI layouts. PHP is used for database connectivity to update or fetch data from database. The User Interface consists of Login/Register page. A new user must register to interact with the system. The user must enter all relevant details asked and the details are verified and inserted into database. After registration only a user can log in to the system. User must provide his email-id and password to log into the system. The system verifies the details from the database and on successful verification the user is taken to the homepage. The homepage provides real-time feed of stock market related news. User can then navigate to fetching current stock quotes, getting predicted values or visualizing trends in stock value via graphs plotted. To get the stock quote user clicks on Get Quote tab in the homepage. Once the Get Quote tab is clicked real-time stock data is fetched from yahoo finance. We parse the html content of yahoo finance website to get a stock quote. This operation requires an active internet connection to function. After getting stock quote user can get predicted value by clicking on Get Predicted tab in the page. The predicted closing values are displayed to the user. This values change every time the python script for prediction is run in the backend. We display a average of predicted value obtained from the two algorithms CNN and SVM. The user can also visualize trends in the stock value of a company through graphs by clicking on the Graph tab in UI. The Graph shows actual and predicted close values for a company, User must log out of the system on finishing his work.

Prediction module:

The code for prediction of stock closing price are written in python. The code must run once a day after the market closes. Everyday the values are updated into the csv file used

for storing data. The prediction module applies two algorithms SVM and CNN for prediction of close price of next day.

CNN uses layers : Conv1D, Flatten, Max-Pooling, Dropout and Dense. The loss function used is mean squared error for both the algorithms. The optimizer is ADAM for CNN. CNN uses 64 filters.

Code snippet for CNN algorithm:

```
def conv1D(x_train,y_train,x_test):
def build_regressor():
verbose, epochs, batch_size = 0, 150, 16
n_timesteps, n_features, n_outputs = x_train.shape[1], x_train.shape[2],
y_train.shape[1]
model = Sequential()
model.add(Conv1D(filters=64, kernel_size=3,
activation='relu', input_shape=(n_timesteps,n_features)))
model.add(Conv1D(filters=64, kernel_size=3, activation='relu'))
model.add(Dropout(0.5))
model.add(MaxPooling1D(pool_size=2))
model.add(Flatten())
model.add(Dense(100, activation='relu'))
model.add(Dense(n_outputs, activation='softmax'))
model.summary()
model.compile(loss='mean_squared_error',
batch_size=batch_size,
optimizer='adam',
metrics=['accuracy'])
```

SVM prediction model uses linear Support Vector Regression for predicting the stock value for the next day. The cost value which is used for our prediction system is of 1, because it gives better results with this value and it gives the lowest value to our testing data.

Code for SVM model:

```
def SVM(x_train,y_train,x_test):
    #Training the train data
    from sklearn import svm
    print ('----------------------------------------------------------------------')
    print ('Training the SVM Model...')
    clr = svm.SVR(C=1,kernel='linear')
```

```
clr.fit(x_train, y_train)
print ('Training Completed.')
print ('------------------------------------------------------------------------')
#Making predictions
y_predict=clr.predict(x_test)
return y_predict
```

Snapshots of the training process:

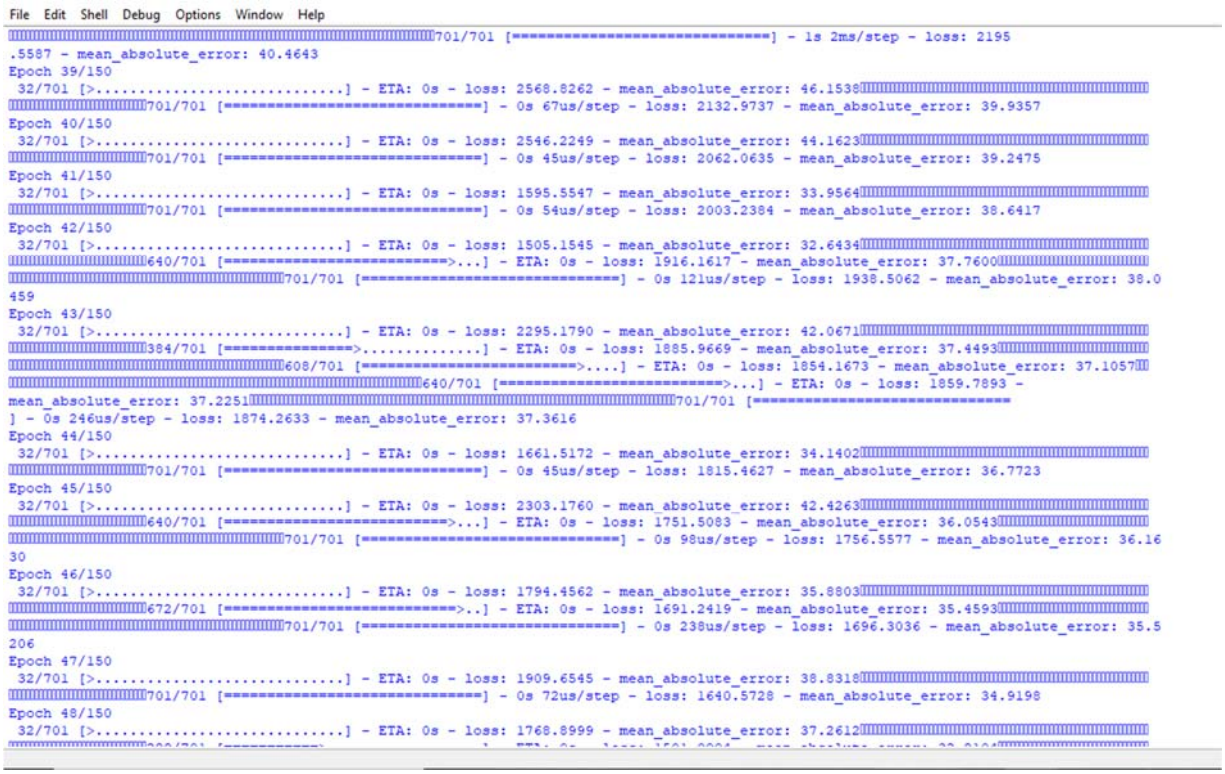

Fig. 5.1(b) : Training Data Screenshot 1

Fig. 5.1(c) : Training Data Screenshot 2

CHAPTER 6

PERFORMANCE EVALUATION AND TESTING

Performance Evaluation functions in the area of modelling, measurement, and evaluation of performance aspects of computing and communication systems. Software testing is the process of evaluating a software item to detect differences between given input and expected output. Also to assess the feature of a software item.

6.1 EVALUATION METRICS :

Evaluation is the process of assessment or making a judgement about the performance of the model. We have assessed the accuracy of our prediction models for both CNN and SVM. The accuracy of our predicted value and the observed closing value of the day is calculated. For evaluation, the summaries generated by both the models is observed.

Evaluation metrics explain the performance of a model. An important aspects of evaluation metrics is their capability to discriminate among model results. Different types of evaluation metrics available are:

6.1.1 Accuracy

Accuracy is the ratio of number of correct predictions to the total number of input samples. It works well only if there are equal number of samples belonging to each class. We have used the sklearn library for calculating accuracy for both the algorithms of prediction. R2_score library is used which is the R-squared formula for accuracy calculation. The parameters passed to the R-squared library is the test data and the predicted data. The formula for calculating the R-squared is:

$$R2\ Score = \left(1 - \frac{(sum\ of\ true\ values\ for\ each\ row - predicted\ value)\text{^}2}{(sum\ of\ true\ for\ each\ prediction - mean)\text{^}2}\right) \times 100$$

6.2 TESTING :

Testing can be stated as the process of validating and verifying that an application or a product meets the requirements and satisfy the needs of the stakeholders. The metrics for comparing testing results are covered in this section.

6.2.1 Snapshots of the Test Cases:

Test Case ID	**Test Case Description**	**Test Steps**			**Test Status (P/F)**
		Step	**Expected**	**Actual**	
T1	Login	Enter Login ID and Password	If ID and password are correct, then the user is directed to the next page.	Login Successful	P
		Enter Invalid ID/Password	If ID or Password are incorrect or in any of them is blank then error message is shown	Display Error Message	P
T2	Get quotes	Click get quotes. Current data is fetched from Yahoo Finance	Current stock value is displayed for the selected company.	Current value is displayed	P
T3	Get Predicted Value	Click get predicted value Predictions algorithms are run and the value is returned.	Predicted value is displayed.	Predicted value is displayed.	P

Table 6.3.1 : Test Cases Table

CHAPTER 7
RESULT AND ANALYSIS

7.1 Results:

The results which are obtained are the predicted values of the particular Stock for the next day. Both the algorithms give the predictions for the following day and the average of these values is displayed as the final predicted value. The user also gets a prediction graph, which includes the predicted value for the next day along with the previous stock data.

Snapshots of the GUI :

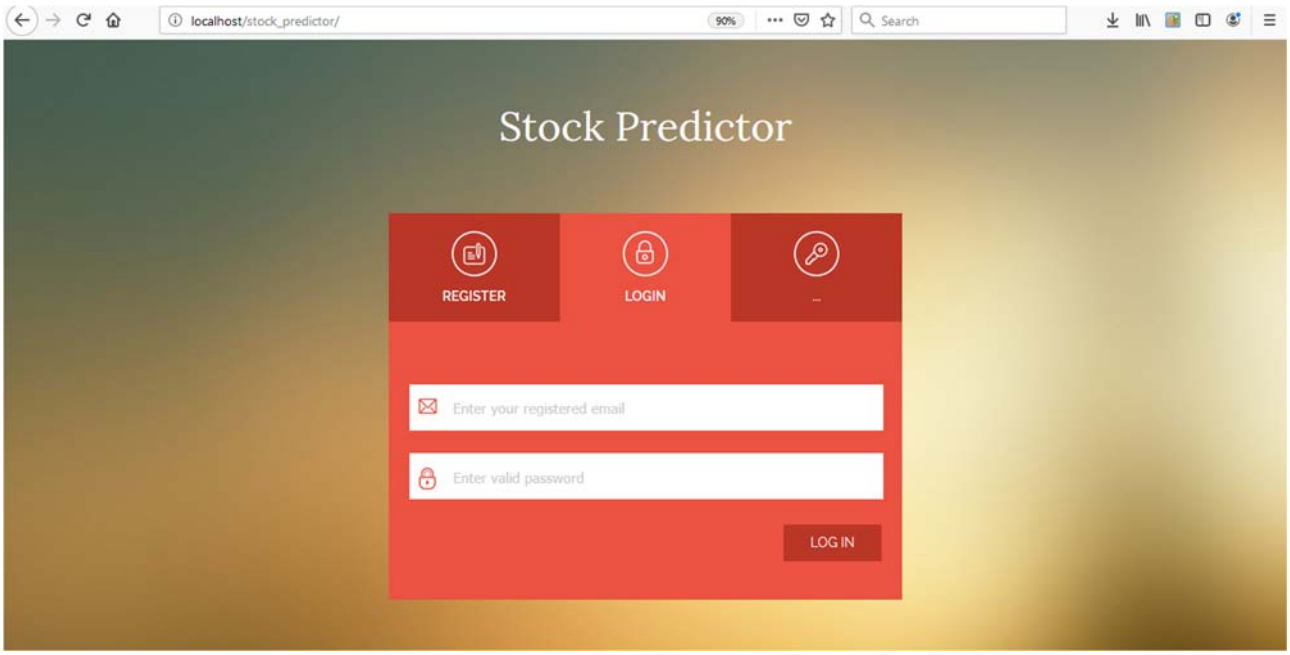

Fig. 7.1 (a) : Login Page

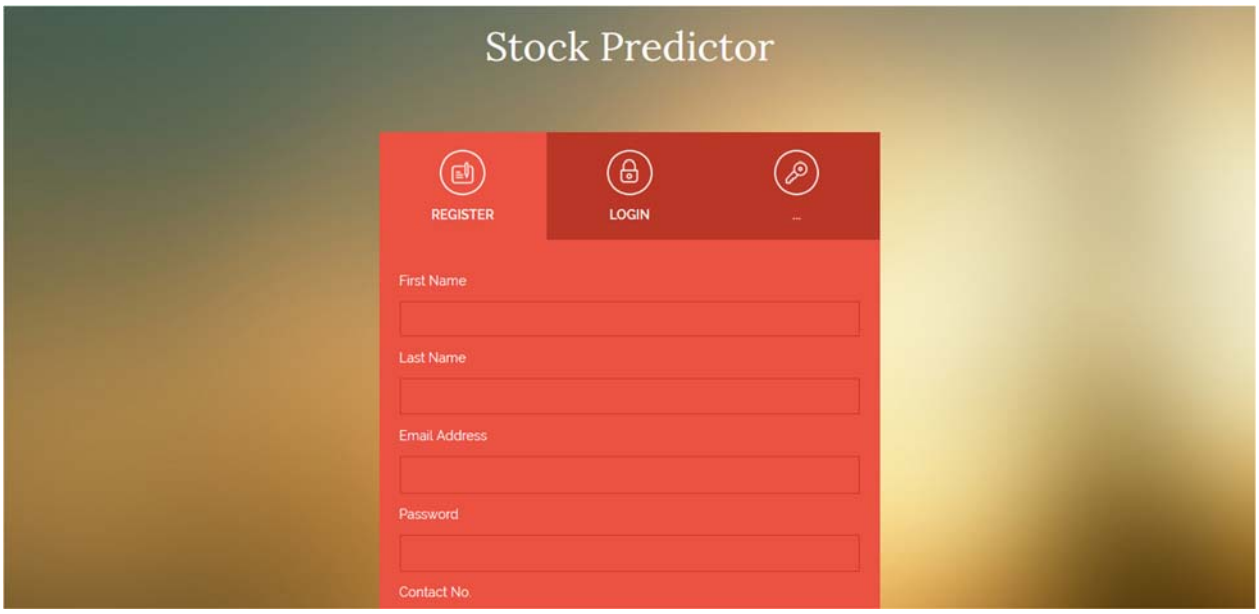

Fig. 7.1(b) : Registration Page

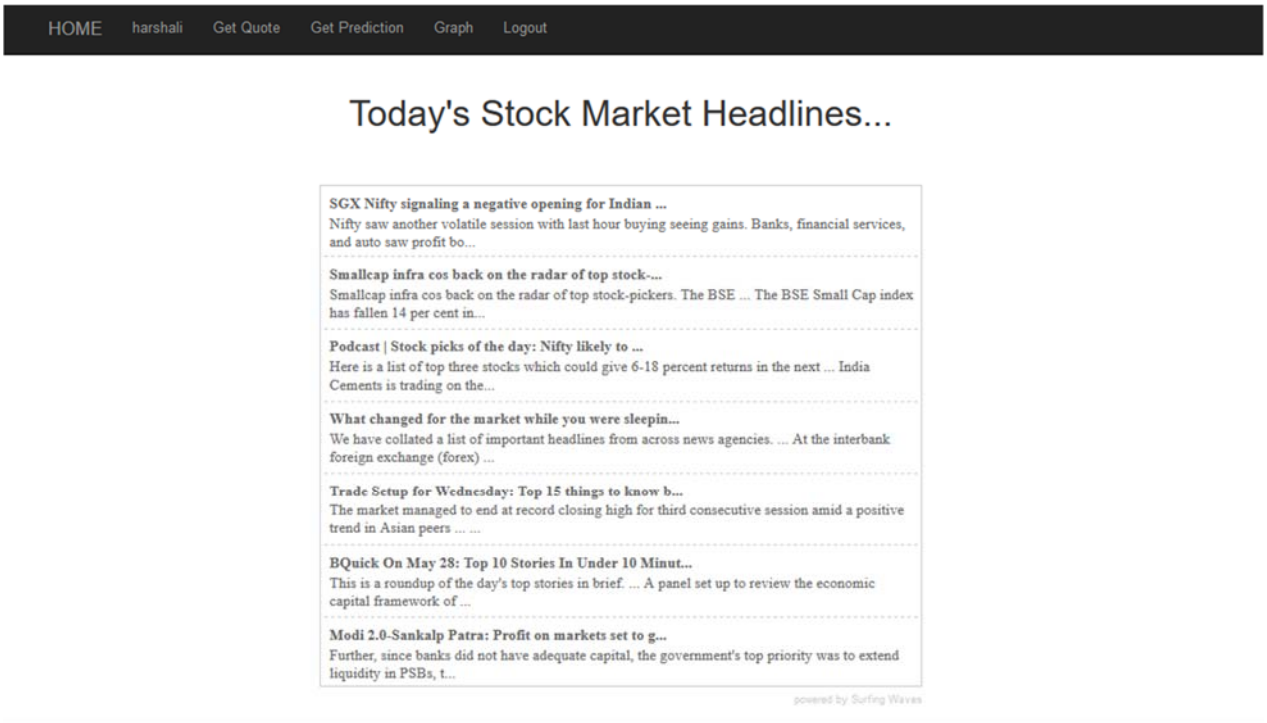

Fig. 7.1(c) : HomePage

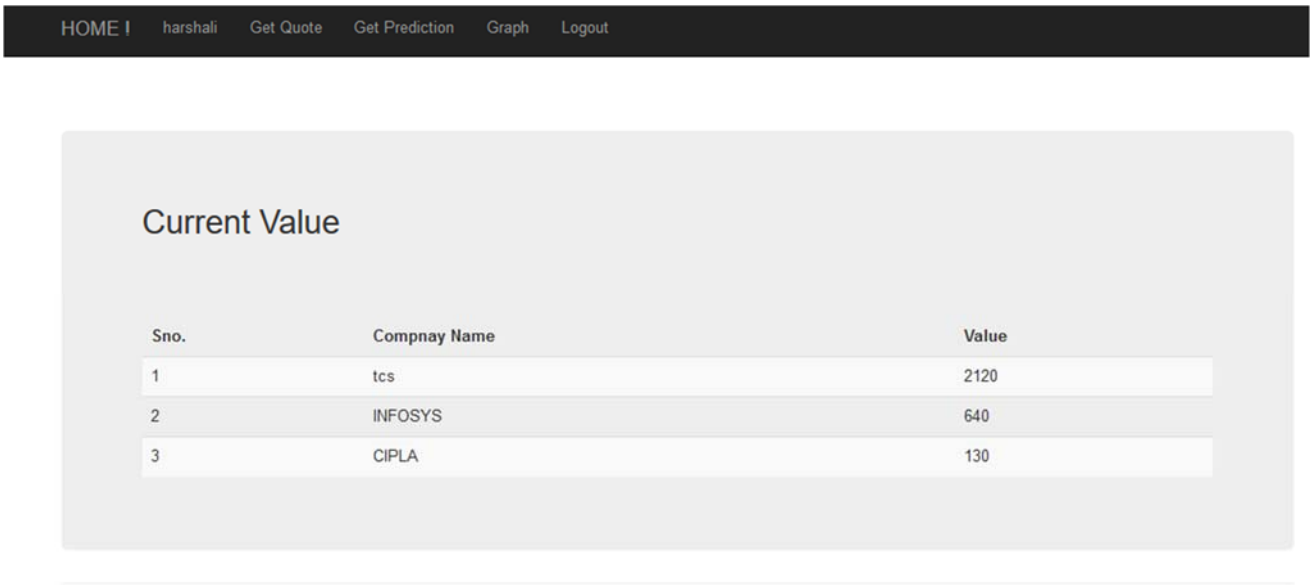

Fig. 7.1(d) :Get Quote Snapshot

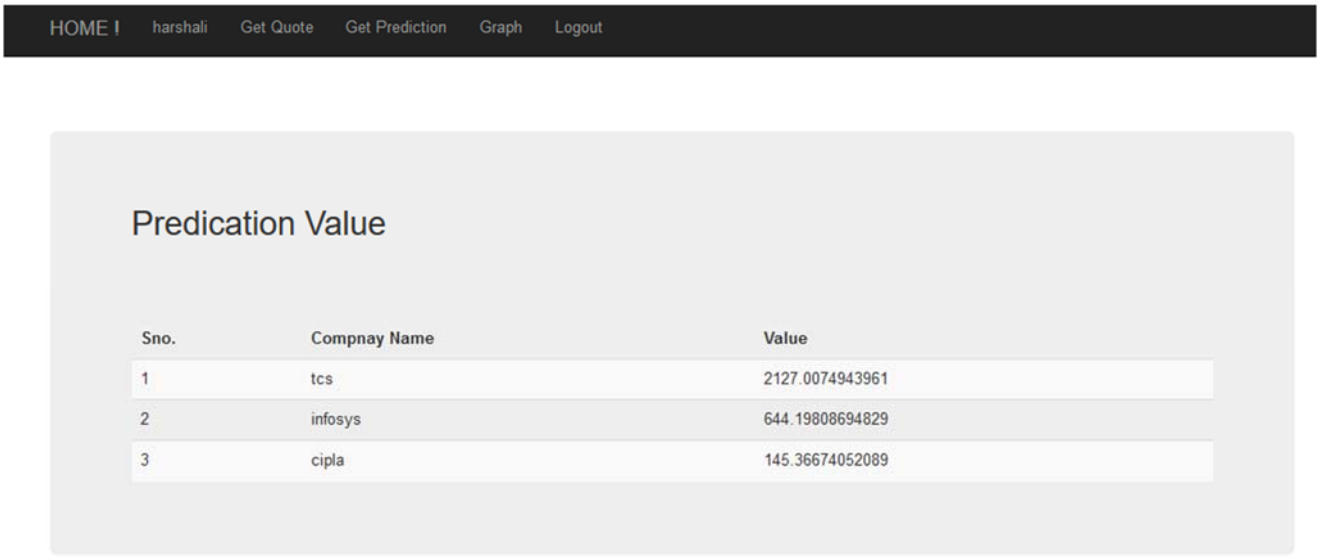

Fig. 7.1(e) : Get Prediction Snapshot

7.2 Analysis:

Analysis of the predicted values by both the algorithms shows that, the value obtained by taking average of both the values generated individually by them, is much closer to the actual value of the particular Stock. The accuracy of the proposed system is above 99.5 %.

A snapshot of the actual and the predicted values obtained by the system is:

Close_Actual	Close_predict	Date
1246.550049	1247.179093	2016-07-13
1260.949951	1263.358515	2016-07-14
1222.680054	1231.412862	2016-07-15
1216.719971	1226.477695	2016-07-18
1233.099976	1237.141518	2016-07-19
1247.469971	1251.380126	2016-07-20
1253.030029	1255.318821	2016-07-21
1257.550049	1262.658984	2016-07-22
1279.469971	1285.640377	2016-07-25
1276.300049	1283.922041	2016-07-26
1288.780029	1295.526415	2016-07-27
1309.619995	1319.0698	2016-07-28
1309.650024	1319.849184	2016-07-29
1354.119995	1364.902484	2016-08-01
1349.599976	1362.238816	2016-08-02
1328.150024	1342.414045	2016-08-03
1326	1337.459397	2016-08-04
1324.550049	1335.569401	2016-08-05
1324.75	1336.283578	2016-08-08
1325.099976	1321.681047	2016-08-09
1337.699951	1347.656104	2016-08-10
1354.949951	1364.538178	2016-08-11

Fig. 7.2 : Actual and Prediction Stock Values

The graph obtained by the results of the prediction algorithms gives us the overview of the actual and predicted value of the stock for the particular day.

Graph for CIPLA :

Fig. 7.2(b) : CIPLA SVM and CNN Graph

Graph for Infosys :

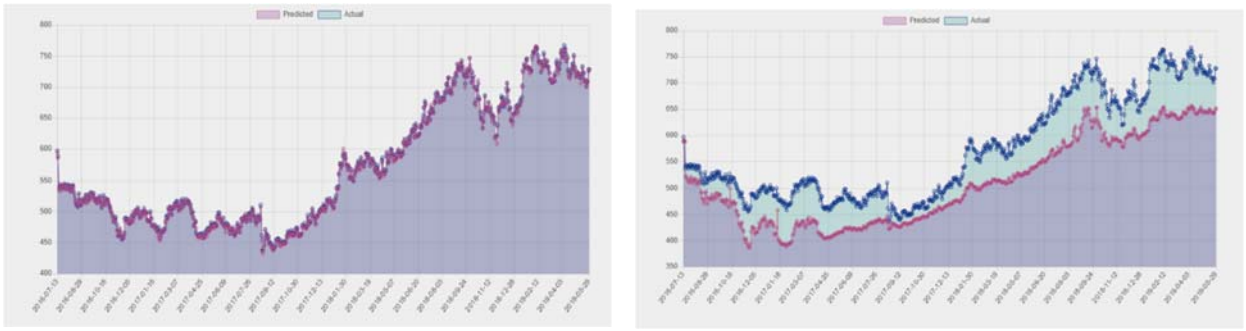

Fig. 7.2(c) : INFY SVM and CNN Graph

Graph for TCS :

Fig. 7.2(d) : TCS SVM and CNN Graph

CHAPTER 8

APPLICATIONS

Prediction of Stock Market is a classic and important problem. The applications for this prediction model are:

- Prediction value of the stock for the next day gives us an insight as to how the company stocks are trending, helping us make decision whether to buy or sell a company's stock.
- The system also gives us the current value of the particular company, keeping the user up to date with the company stock prices.
- The system also includes updated Stock market news to help interested people gain knowledge of the trends in the Stock Market.

CHAPTER 9

CONCLUSION

Existing systems to predict stock prices make use of a single machine learning algorithm. In order to provide a more accurate prediction, our project makes the use of two machine learning algorithms; Convolutional Neural Network and Support Vector Machine. The same dataset is applied to both the algorithms which gives the predicted value for the next day. This hybrid approach gives results which has more accuracy as compared to individual prediction models. From our results, we have observed that CNN is used because it makes use of information given at a particular moment. Support Vector Machine model along with linear regression does not lose accuracy outside the training sample. Our model makes use of the R-squared score to give us accuracy of the model. Both these algorithms give an accuracy of above 95%, but the combined result is observed to be much closer to the actual value of the stock. Hence, the objective of building a stock market prediction system which gives accurate results has been achieved.

CHAPTER 10

REFERENCES

[1] Sreelekshmy Selvin, Vinayakumar R, Gopalakrishnan E.A, Vijay Krishna Menon, 'STOCK PRICE PREDICTION USING LSTM,RNN AND CNN-SLIDING WINDOW MODEL', International Conference on Advances in Computing, Communications and Informatics (ICACCI),IEEE, Udupi, India, Sept 2017

[2] Luca Di Persio , Oleksandr Honchar 'Artificial Neural Networks architectures for stock price prediction: comparisons and applications , International Journal of Circuits, Systems and signal processing, Verona,Italy ,2016

[3] Avraam Tsantekidis, Nikolaos Passalis, Anastasios Tefas, Juho Kanniainen, Moncef Gabbouj and Alexandros Iosifidis 'Forecasting Stock Prices from the Limit Order Book using Convolutional Neural Networks',2015

[4] Anastasia Borovykh Sander,Bohte, Cornelis W. Oosterle,'Conditional time series forecasting with convolutional neural networks' June 2018

[5] Meesad, P., & Rasel, R. I. (2013). Predicting stock market price using support vector regression. 2013 International Conference on Informatics, Electronics and Vision (ICIEV).

[6] R., A., & M.S., V. (2012). Stock Price Prediction Using Support Vector Regression. Global Trends in Computing and Communication Systems, 588–597.

[7] M. Karazmodeh, S. Nasiri, and S. Majid Hashemi, 'Stock Price Forecasting using Support Vector Machines and Improved Particle Swarm Optimization', Journal of Automation and Control Engineering, North Cyprus, Turkey, June ,2013