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 first semester of academic year 2017 2018 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: 16-12-2018

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

A stock market, equity market or share market is the aggregation of buyers and sellers 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. 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. 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. 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 studies from various areas aiming to take on that challenge and Machine Learning approaches have been the focus of many of them. There are many examples of Machine Learning algorithms been able to reach satisfactory results when doing that type of prediction. 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 successful prediction of a stock's future price could yield significant profit. This project aims at developing a system for Stock Market Prediction for National Stock Exchange (NSE) data from various companies. We use CNN and SVM with Particle Swarm Optimization (PSO) to construct a model to effectively predict stock prices. The project takes input the historical stock data and creates a model for day-wise prediction of stock price.

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

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
7	PSO	Particle Swarm Optimization

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 develop a Sliding Window based approach for better prediction
- 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 using Particle Swarm Optimization (PSO).
- To achieve higher efficiency in stock prediction.

CHAPTER 2

LITERATURE SURVEY

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.

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.

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.

[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.

In [5] paper, three different classification algorithms will be compared for the purposes of maximizing capital while minimizing risk to the investor. For the purposes of short-term stock value prediction, the support vector machine model will be boosted by particle swarm optimization (PSO). The performance of the proposed method is compared to the naive Bayes classifier and artificial neural network models in terms of predictive accuracy and computation time required. In this, two well-known companies with public stock options are selected: Apple, Inc. (AAPL) and Google (GOOG.) are used. These two stock profiles will allow for practical tests of the predictive accuracy of the proposed methods. The stock data is available on websites such as Yahoo! Finance and Google Finance. The methods are implemented in MATLAB utilizing nine years worth of daily stock data (opening price and adjusted closing price), spanning from January 2005 to March 2014. From the results, NBC is shown to be no more reliable than performing a coin-flip for predicting if tomorrow's close will result in a gain or a loss. And tuning support vector machine parameters with particle swarm optimization leads to highly accurate (approximately 95%) and robust stock forecasting for historical datasets.

In [6], the proposed algorithm integrates Particle swarm optimization (PSO) and least square support vector machine (LS-SVM). The proposed model is based on the study of historical data, technical indicators and optimizing LS-SVM with PSO algorithm to be used in the prediction of daily stock prices. Levenberg-Marquardt (LM) algorithm is used as a benchmark for comparison with LS-SVM and LS-SVM-PSO models. The proposed model architecture contains six inputs vectors represent the historical data and derived technical indicators and one output represents next price. The technical indicators include relative strength index, money flow index, exponential moving average, stochastic oscillator and moving average convergence/divergence. The algorithm cover all stock sectors in S&P 500 stock market. These sectors are Information Technology (Adobe, Hp, and Oracle); Financials (American Express and Bank of New York); Health Care (Life Technologies, and Hospera); Energy (Exxon-Mobile and Duck energy); Communications (AT&T); Materials (FMC Corporation); Industrials (Honey Well). LS-SVM-PSO, LS-SVM and ANN algorithms were trained and tested with datasets form Jan 2009 to Jan 2012. All datasets are divided into training part (70%) and testing part (30%). The

performance of the proposed model is better than LS-SVM and compared algorithms. LS-SVM-PSO achieves the lowest error value followed by single LS-SVM, while ANN-BP algorithm is the worst one.

[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.

CHAPTER 3

SPECIFICATION

3.1 System Requirements

- The system should collect data from web sources like Yahoo Finance, Google finance at a time interval of 15 minutes and update its database from a set of 25 companies. http://finance.google.com http://finance.yahoo.com

- The system should maintain a host of web services that fairly link different service modules with client interface.

- Different prediction models need to be built that give a fair recommendation to the customer whether to buy/sell/hold stock.

- 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

Table 3.1

3.3 Hardware Requirements

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

Table 3.2

CHAPTER 4

DESIGN

4.1 Architectural Diagram

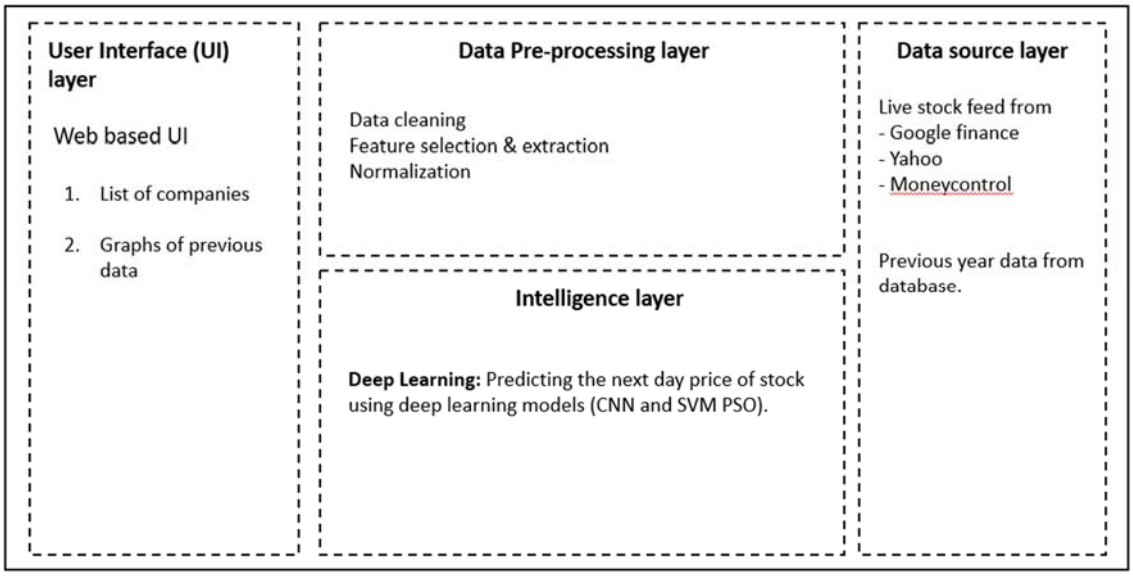

Fig 4.1

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 Google Finance, Yahoo Finance, Money Control. Also previous data is collected into the database from Google Finance, Kaggle , Money control.

4.2 Data Flow Diagrams

4.2.1 CNN DFD

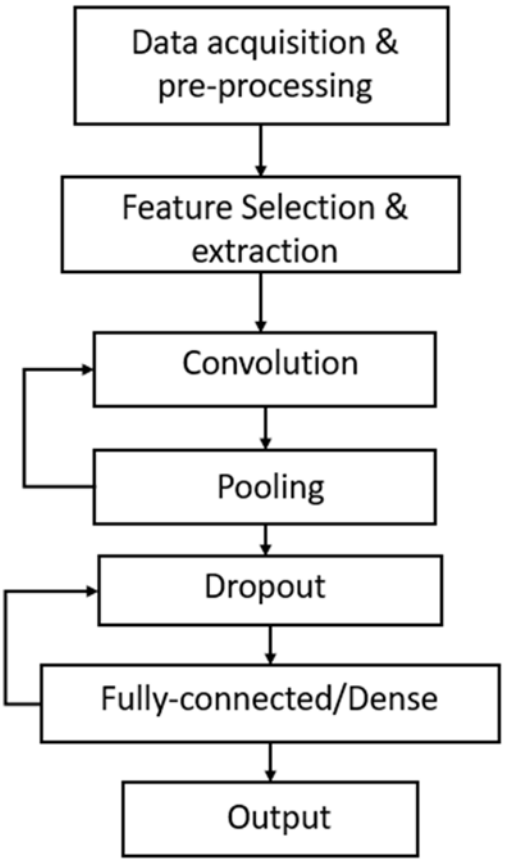

Fig 4.2

4.2.2 SVM-PSO DFD

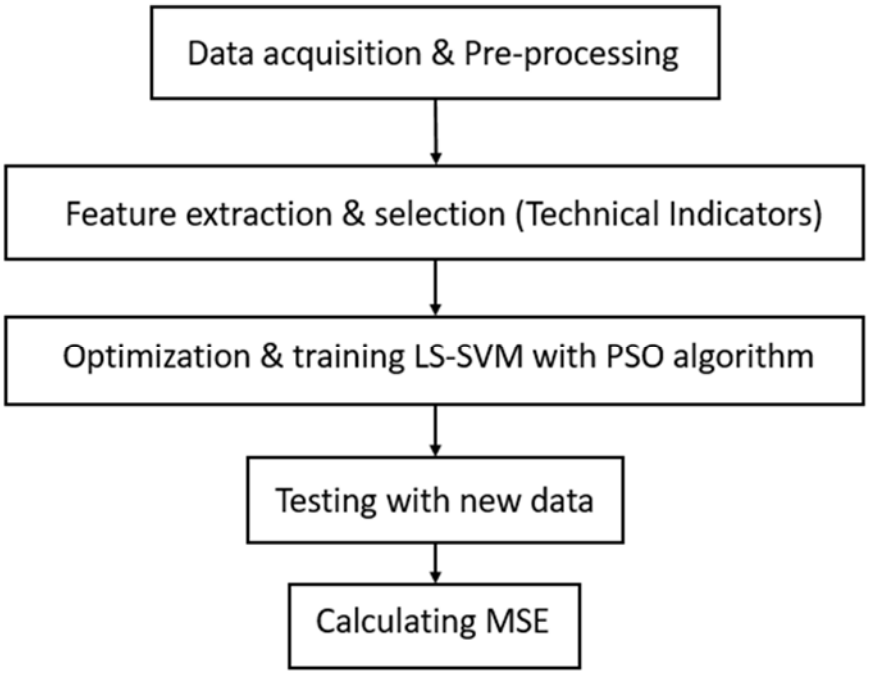

Fig 4.3

4.2.3 Combined DFD

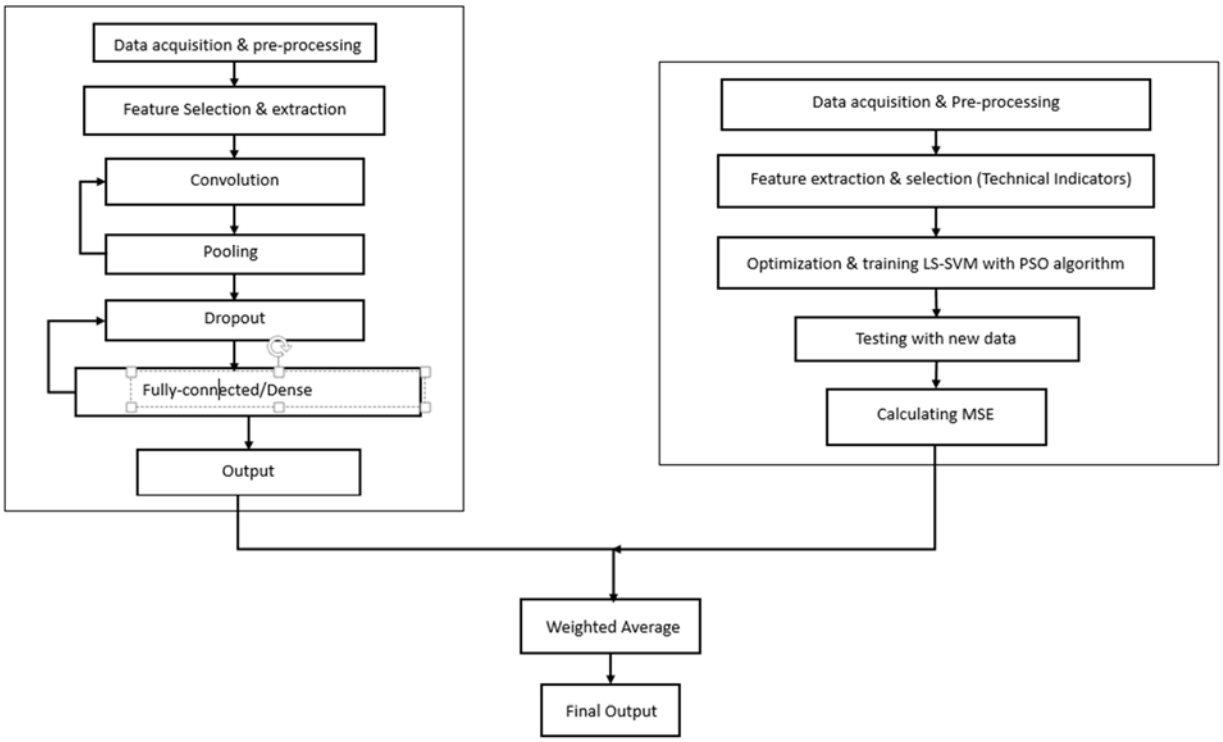

Fig 4.4

4.3 UML Diagrams

4.3.1 Use Case Diagram

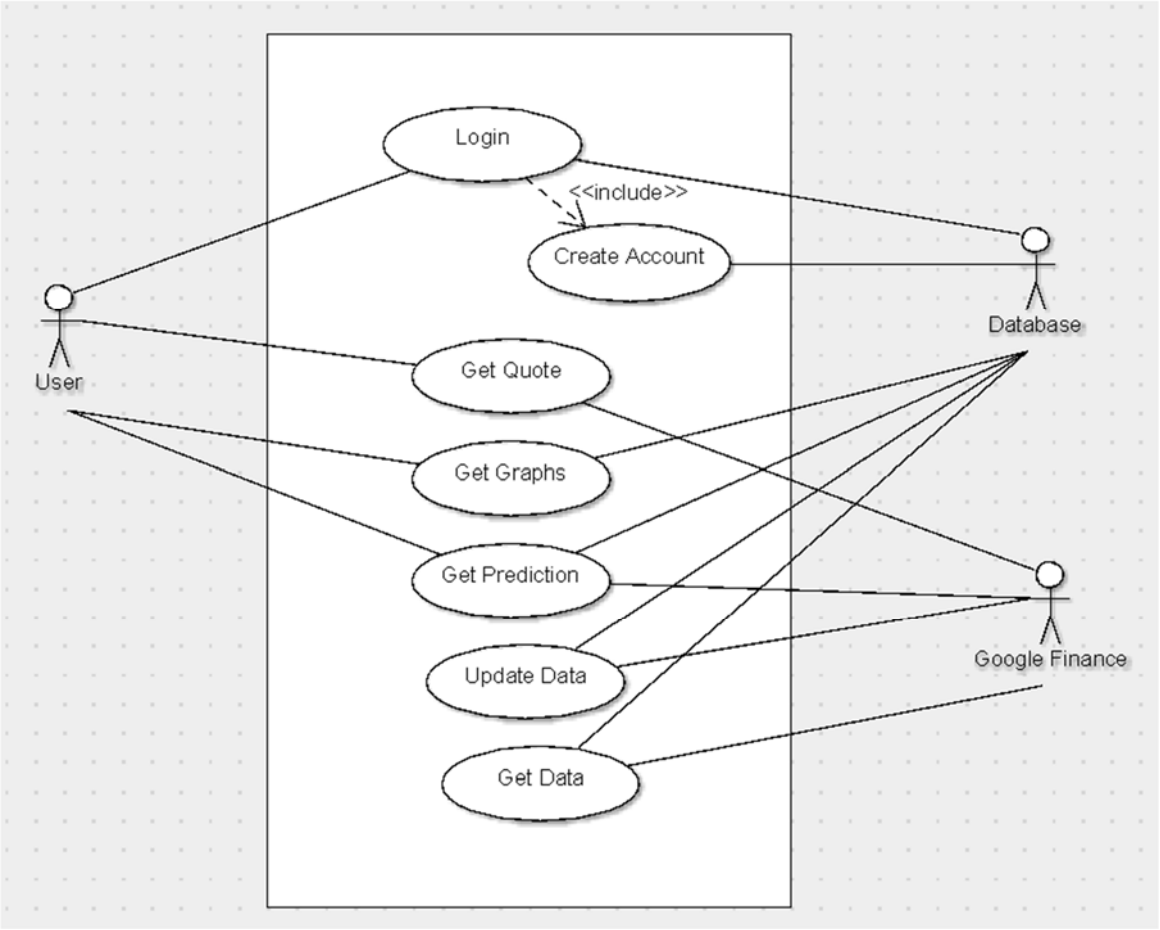

Fig. 4.5

Use Case Specification:

Name: Login

Identifier: UC1

Description:
The registered user need to login to the system using unique ID and password that would be checked by the Database Server.

Goal:
The registered user will login to their account successfully.

Preconditions:
1. The user must be already registered.
2. The user must have his own unique password and ID to login.

Basic Course:
1. The use case begins when the registered user opens the website login page.
2. The actor enters login details that is, Password and his unique ID.
3. The system checks the validity of the actor from the database.
4. The use case ends when the actor's details are verified and he successfully logs in to his homepage.

Actor:
User
Database

Include:
Create Account

Name: Registration

Identifier: UC2

Description:
The new user can register to the system by setting unique ID and password. This is the basic functionality to be used to get permanently registered into the system.

Goal:
The new user will enter his/her personal details and then verify those details. After this , they will generate their username and password.
Preconditions:
1. The user must be not be already registered.
Basic Course:
1. The use case begins when a new user enters the registration webpage of the system.
2. The actor has to enter his personal details like name, date of birth.
3. The actor will set the password and a unique id.
4. The system will check for validity of ID and password according to predefines specifications.
5. The use case ends when the new user is registered successfully.
Actors:
User
Database

Name: Get Quotes

Identifier: UC3

Description:

User will select the list of companies to get the quotes and graphs of stocks of those companies.

Goal:

To get the quotes and graphs of selected stocks.

Preconditions:

1. The user does not have any quotes for a particular stock.

Basic Course:

1. The use case begins when a user opens the login page.
2. User fills the user id and password on the page.
3. Select the Company for which he wants the quotes/graphs.
4. Select the get quoted option.
5. Check the quotes and graphs obtained.

Actor:

User

Name: Get Prediction

Identifier: UC4

Goal:

To get the prediction of selected stocks.

Preconditions:

1. The user does not have any prediction for a particular stock.

Basic Course:

1. The use case begins when a user opens the login page.
2. User fills the user id and password on the page.
3. Select the Company for which he wants the prediction.
4. Select the get prediction option.
5. Check the predictions obtained and make decisions accordingly.

Actor:

User

Database

4.3.2 Activity Diagrams

1) Login :

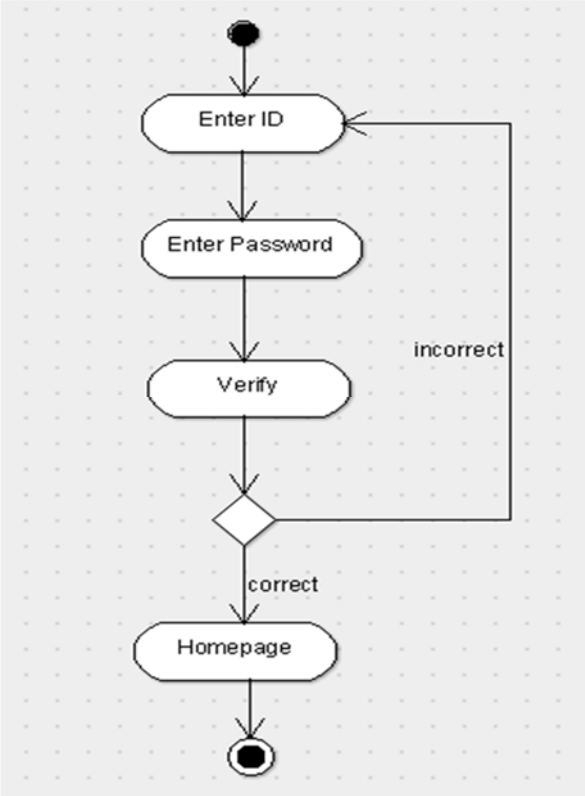

Fig 4.6

2) Register :

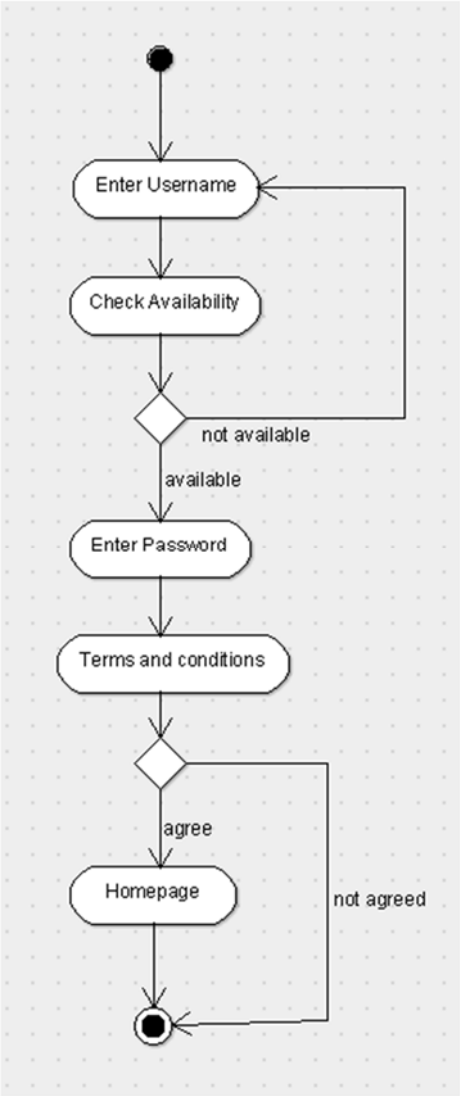

Fig 4.7

3) Get Quotes :

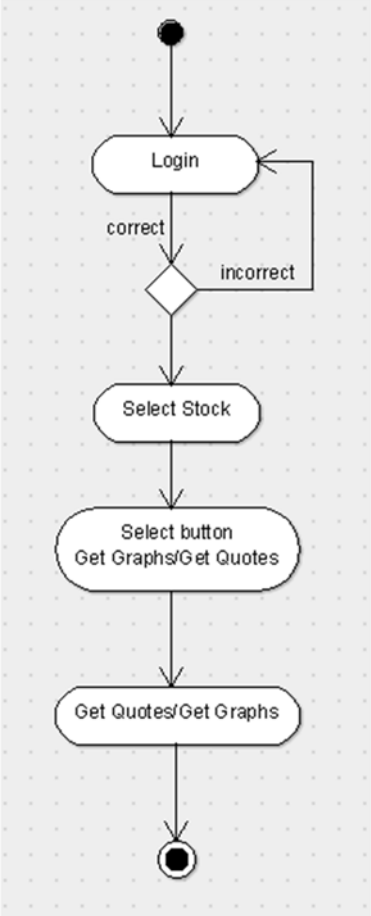

Fig 4.8

4) Get Predictions:

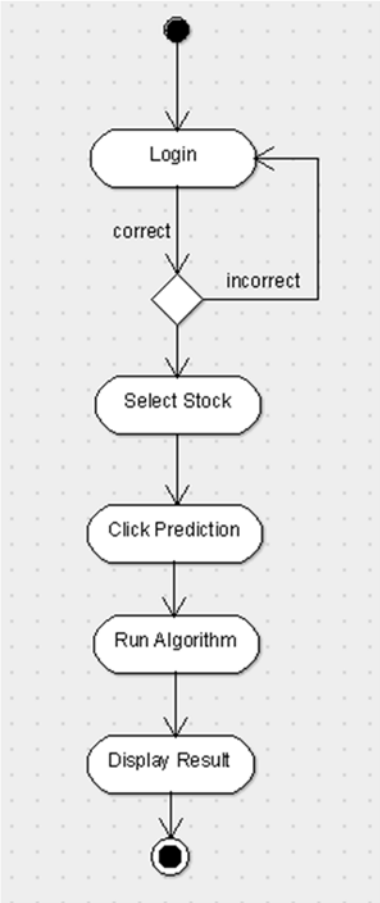

Fig 4.9

4.3.3. Sequence Diagram :

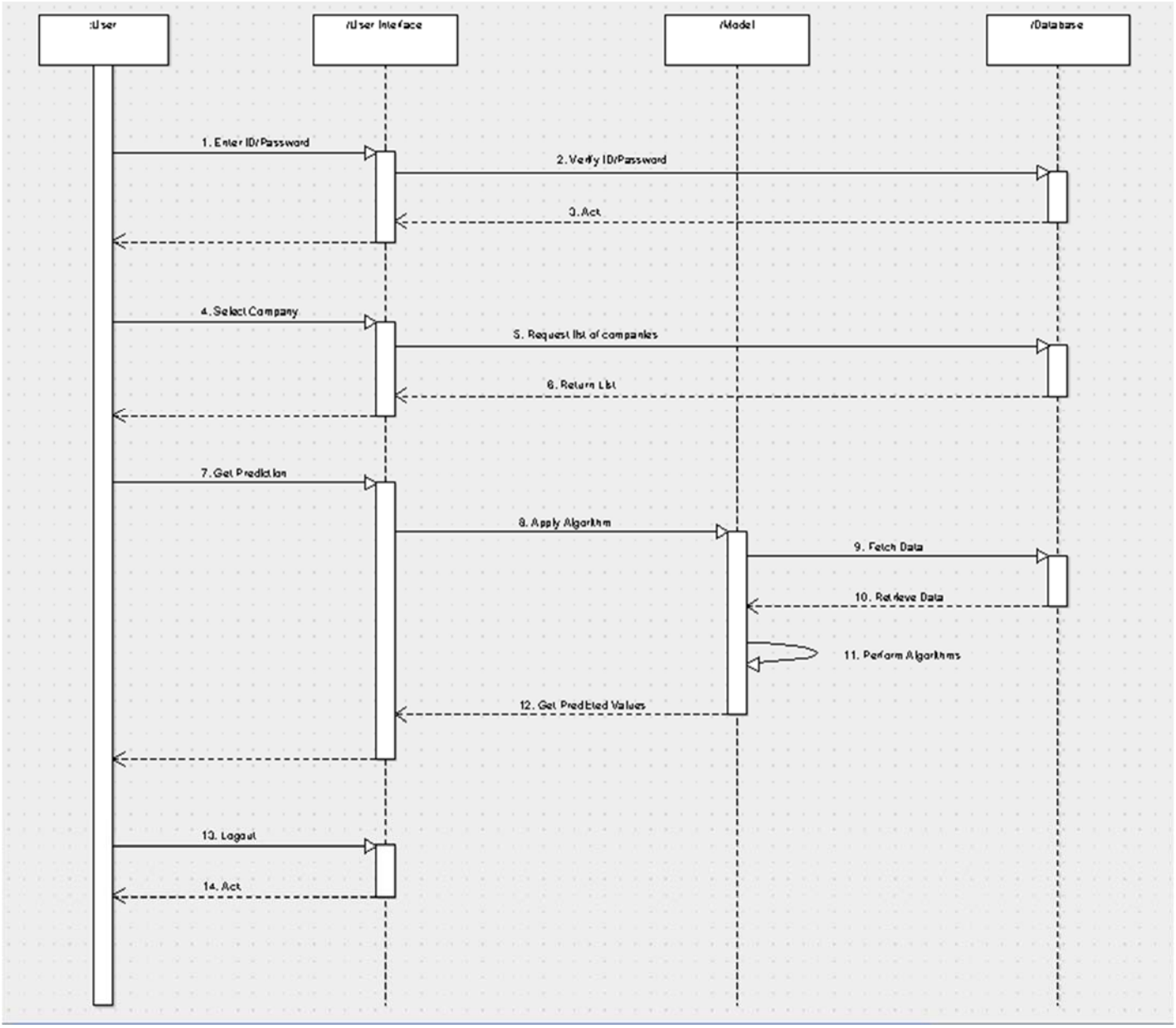

Fig 4.1

CHAPTER 5

CONCLUSION

Stock market is one of the topic in trend today. There is a huge amount of buying and selling of stocks on a regular basis. Thus accurate prediction of future price of a stock might be very useful to the investors. Results have also shown that many of the stock trade in countries like US are done through algorithmic analysis on the past data. There are many researches going in the field of stock market prediction. Many systems have been developed for prediction of stock price.

Deep Learning algorithms are extensively used for this purpose. Various methods like RNN, CNN and LSTM are continuously being researched on to apply in stock price prediction. A comparative study of the three above mentioned methods revealed that CNN is the best amongst all. CNN outperforms other algorithms because it uses information from particular instant. Along with the above mentioned techniques stock price prediction can also be done with the help of sentiment analysis. Thus the system can be made more accurate by combining result from technical as well as sentiment analysis

In particular, support vector machines boosted by particle swarm optimization provide a relatively easy system to understand and implement for the purposes classifying stock performance and in predicting future outcomes. The findings in this work show that the utilization of properly tuned particle swarms to optimize support vector machine parameters yields a highly accurate predictor for stock behavior prediction. These predictions are reliable for the short term, which is useful for casual interday traders.

CHAPTER 6

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] Trevor M. Sands, Deep Tayal, Matthew E. Morris, and Sildomar T. Monteiro,' Robust Stock Value Prediction using Support Vector Machines with Particle Swarm Optimization', 2015 IEEE Congress on Evolutionary Computation (CEC) **,** IEEE, Sendai, Japan, May 2015

[6] Osman Hegazy, Omar S. Soliman and Mustafa Abdul Salam, 'A Machine Learning Model for Stock Market Prediction' , International Journal of Computer Science and Telecommunications , Egypt, December 2013

[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