A PRELIMINARY REPORT ON

STOCK MARKET PREDICTION USING MACHINE LEARNING

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)

SUBMITTED BY

ASHWIN GADUPUDI
MOUSHMI JAIN
RUSHIKESH KHAMKAR
RUSHIKESH KADAM

Sinhgad Institutes

DEPARTMENT OF COMPUTER ENGINEERING

STES'S SMT. KASHIBAINAVALECOLLEGE OF ENGINEERING

VADGAON BK, OFF SINHGAD ROAD, PUNE 411041

SAVITRIBAI PHULE PUNE UNIVERSITY
2018 -2019

Sinhgad Institutes

CERTIFICATE

This is to certify that the project report entitles

"STOCK MARKET PREDICTION USING MACHINE LEARNING"

Submitted by

ASHWIN GADUPUDI
MOUSHMI JAIN
RUSHIKESH KHAMKAR
RUSHIKESH KADAM

is a bonafide student of this institute and the work has been carried out by him/her under the supervision of **Prof. Mohasin B. Tamboli** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering** (Computer Engineering).

(Prof. M.B. Tamboli)
Guide ,
Department of Computer Engineering

(Dr. P. N. Mahalle)
Head,
Department of Computer Engineering

(Dr. A. V. Deshpande)
Principal,
Smt. Kashibai Navale College of Engineering Pune – 41

Place: Pune

Date:

ACKNOWLEDGEMENT

It gives us great pleasure in presenting the preliminary project report on 'STOCK MARKET PREDICTION USING MACHINE LEARNING'.

We would like to take this opportunity to thank our internal guide Prof. M. Tamboli for giving us all the guidance and help needed. We are really grateful to him for his kind support. His valuable suggestions were very helpful.

We are also grateful to Dr. P. N. Mahalle, Head of Computer Engineering Department, STES' Smt. Kashibai Navale College of Engineering for his indispensable support and suggestions.

We would also like to thank our colleagues to have directly or indirectly guided and helped us in the preparation of this report. Lastly, we would like to thank our parents who have always been very supportive.

Ashwin Gadupudi
Moushmi Jain
Rushikesh Khamkar
Rushikesh Kadam

ABSTRACT

Stock Market is a place where shares of companies are traded. Companies issue their shares to general public through an IPO to raise capital. The main objective of the proposed model is to detect the patterns and predict the price of stocks. The data of previously traded prices of stocks is vast and very difficult to detect patterns manually, thus in this model we use **Machine Learning Techniques** to do the job for the investor. The proposed model gives future prices of a particular stock based on the previous day-closing prices of the stock. The different attributes considered in the model are Sensex Index, Commodity prices, Crude-oil prices, National average interest rates and Foreign Exchange rate.

In this model, we have used 4 Machine Learning algorithms to predict the prices of the stock which are **i) Support Vector Machine**, **ii) Single-level perceptron**, **iii) Multi-level perceptron**, and **iv) Radial-basis function**. Our goal is to calculate the accuracy of all the four models and then provide the best model to the user with its predictions for the future of the particular stock. This model will save a lot of time and effort of the investors where they spend hours researching about a particular stock.

Keywords: Sensex Index, Stock prediction, Beta, Moving average, Portfolio, Neural networks.

INDEX

CHAPTER **TITLE** **PAGE NO.**

LIST OF FIGURES

1.INTRODUCTION

For a new investor, the share market can feel a lot like legalized gambling. Randomly choose a share based on gut instinct . If the value of your share goes up you're in profit else you're in loss. The share market can be intimidating, but the more you learn about shares, and the more you understand the true nature of stock market investment, the better and smarter you'll manage your money.

Terms:

• A stock of a company constitutes the equity stake of all shareholders.

• A share of stock is literally a share in the ownership of a company. When investor purchases a share of stock, the investor is entitled to a small fraction of the assets and profits of that company.

• Assets include everything the company owns (real estate, equipment, inventory)

• Shares in publicly traded companies are bought and sold at a stock market or a stock exchange.

These are some examples of popular stock exchanges:

• NYSE - New York Stock Exchange

• NASDAQ - National Association of Securities Dealers

• NSE – National Stock Exchange(India)

• BSE – Bombay Stock Exchange

There is no way to predict the accurate trends in stock market. Many factors affect rises the share prices of different companies .The best way to understand stock markets is to analyze and study how the market movements have been in the past.

Share market trends tend to repeat themselves overtime. After you study the cycle of a particular stock, you can make predictions about how it will change over the course of time. Some stocks might be truly arbitrary in which case the movement is random but in most of the cases there is a particular trend that repeats itself. Recognizing these patters will enable you to predict the future trend.

The project goal is to build a system where the machine learning algorithms try to predict the prices of stocks based on their previous closing prices and other attributes that influence its price like Interest rates, Foreign exchange and Commodity prices.

1.1 MOTIVATION

Stock market movements make headlines every day. In India, 3.23 crore individual investors trade stocks. Maharashtra alone accounts for one-fifth of these investors. However, a report from Trade Brains shows that 90% of these investors lose money in due to various reasons like insufficient research, speculation, trading with emotions etc.

Higher inflation rate and lower interest rate makes it ineffective to put one's money into savings account or fixed deposits. Thus, many people look up to stock market to keep up with the inflation. In this process of multiplying their money many investors have made a fortune while, some have lost a lot of money due to unawareness or lack of time to research about a stock.

There are lots of contradicting opinions in the news and an individual may not have the time or may not know how to research about a stock. Most importantly, it is very difficult to manually predict the stocks prices based on their previous performance of that stock. Due to these factors many investors lose a lot of money every year.

A system that could predict the stock prices accurately is highly in demand. Individuals can know the predicted stock prices upfront and this may prevent them from investing in a bad stock. This would also mean a lot of saved time for many of the investors who are figuring out weather a particular stock is good or not.

1.2 PROBLEM DEFINITION

The price of a stock depends highly on the previously traded prices and patterns repeat themselves. Stock market has a lot of data which is tough for an individual to analyse. This project aims at detection of patterns and prediction of future prices of a particular stock. Four different machine learning algorithms are used to predict the stock prices among which the most accurate one is chosen and presented to the user.

Fig. 1.1 System Overview

2.LITERATURE SURVEY

1. Comparative analysis of data mining techniques for financial data using parallel processing

[2014] [IEEE] Do the comparative analysis of several data mining classification techniques on the basis of parameters accuracy, execution time, types of datasets and applications. Simple Regression and multivariate analysis used, Regression analysis on attributes is used. No use of machine learning. Does not provide the algorithm used.

2. Stock market prices do not follow random walks: Evidence from a simple specification test

[2015] [IEEE]Test the random walk hypothesis for weekly stock market returns by comparing variance estimators derived from data sampled at different frequencies. Simple trading rules extraction and Extraction of Trading Rules from Charts and Trading Rules. No alternative provided for human investing. Show only the flaws on manual investments.

3. A Machine Learning Model for Stock Market Prediction

[2017] [IJAERD] Support Vector Machine with Regression Technology (SVR), Recurrent Neural networks(RNN). Regression analysis on attributes using simple Regression and multivariate analysis used. It is not tested in real market. Shows how social media affects share prices. Does not account for other factors.

4. Twitter mood predicts the stock market

[2010] [IEEE] Analyse the text content of daily Twitter feeds by two mood tracking tools, namely Opinion Finder that measures positive vs. negative mood and Google-Profile of Mood States . These results are strongly indicative of a predictive correlation between measurements of the public mood states from Twitter feeds. Difficult to scan each every text extraction from large set of data, difficult Text mining.

5.Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets

[2017] [Research] Propose a generic framework employing Long Short-Term Memory (LSTM) and convolutional neural network(CNN)for adversarial training to forecast high frequency stock market. This model achieves prediction ability superior to other benchmark methods by means

of adversarial training, minimizing direction prediction loss, and forecast error loss. It Can't predict Multi scale Conditions and live data

6. Stock Market Prediction Using Machine Learning

[2016] [IEEE] Uses different modules and give different models and give best accuracy using live streaming data. Predict Real Market Data and calculate Live data using single and multilevel perspective, SVM, Radial Bias . It Couldn't work Textual Data form different Browsing Data (Web Crawling)

7. Stock Market Prediction by Using Artificial Neural Networks

[2014] [IEEE] This model takes help of Artificial Intelligence and uses only neural networks to predict the data. Predicting data using single and multi-level perceptron. It uses 10 hidden layers with the learning rate of 0.4, momentum constant at 0.75 and Max Epochs of 1000. This model doesn't use machine learning algorithms like SVM and radial basis function to determine their accuracy.

8. Price trend prediction Using Data Mining Algorithm

[2015][IEEE] This paper presented a data mining approach to predict the long term trend of the stock market. The proposed model detects anomalies in data according to the volume of a stock to accurately predict the trend of the stock. This paper only provides long term predictions and does not give predictions to the immediate trends.

3.SOFTWARE REQUIREMENTS SPECIFICATION

3.1 INTRODUCTION

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.

A stock exchange facilitates stock brokers to trade company stocks and other securities. A stock may be bought or sold only if it is listed on an exchange. Thus, it is the meeting place of the stock buyers and sellers. India's premier stock exchanges are the Bombay Stock Exchange and the National Stock Exchange.

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.

3.1.1. Project Scope

The software will be a web-portal that contains the live stock predictions Also the future stock market values will be predicted based on recent values. Rather than going for an stock broker this software will act as intermediate source between clients and company for better outcome and future scope.

The proposed System intends to assist in prediction which will minimize the considerable amount of time spent by the consumers in seeing the history of the previous stocks as well as they will be able to make better choices.

The customer just needs to select the stock that they want to predict and by using the algorithms we can predict if the stock value will increase or decrease.

3.1.2 User Classes And Characteristics

PRIMARY USER:

Stock holders are the primary user of this system as they will be able to buy stock and predict the accurate upgoing stock based on recent analysis shown by the system. Public as well as private sector can easily get any detail records about stock after authentication to the system.

DEVELOPER:

The role of developer who is server-side controller is to verify the live API working. There are many investors who will be looking for better prediction as prices can fluctuate any time because of this it is an responsibility of server side to deploy more accuracy with specific algorithms and techniques.

3.1.3 Assumptions and Dependencies

Assumptions and dependencies are the main aspect of the project.

Certain assumptions to be made in this project can be :

1.The Indian Market have seen many ups and downs in the last decade .There have been many predictions over the period of time. We are assuming that there is no external factor present in the market which will affect the stock price.

2.We are assuming that the future stock market price will be based only on previous years analysis and will remain unaffected by factors like social media ,political turmoil ,policy change ,etc.

Also the project outlines certain dependencies which include the dependency of the previous data analysis.

The more the data better will be the prediction and will be beneficial for our analysis.

The accuracy and limitations of machine learning classifiers used in the algorithms are also the dependencies to be taken into consideration while building the project.

3.2. FUNCTIONAL REQUIREMENTS

The Functional requirements define the functionality of the system as a whole or as a collection of subsystems. The functional requirements are the functions or features that must be included in a system to satisfy the business needs and be acceptable to the users .

1. The system should be able to generate an approximate share price.
2. Dataset: The system should collect accurate data from the NPESE website in consistent manner.
3. User GUI: The user is required to select which company is he interested in amongst the various companies that have been provided. The client will be provided with the proper graphs and rates according to the prediction.
4. Result Subsystem: The subsystem displays the dashboard for displaying the companies stocks. It also displays the stock price analysis graph .The result is displayed to the user using the same User GUI from where the input was taken by the system

3.3 EXTERNAL INTERFACE REQUIREMENTS

3.3.1 User Interfaces

- The UI consists of three main parts:
- The Registration of client
- The Login client
- Company stock dashboard
- Stock prize analysis graph.

3.3.2 Hardware Interfaces

- PC having active Internet connection
- Any Web browser

3.3.3 Software Interfaces

- Python3
- Machine learning tools
- Jupyter notebook
- Ipython library
- HTML for front end

3.3.4 Communications Interfaces

- The communication between web-browser and server will be taken through live API which will use TCP/IP protocol for better efficiency.
- Machine learning algorithm will pickup the best accurate data from server through linear regression and various algorithms.

3.4.NON-FUNCTIONAL REQUIREMENTS

3.4.1. Performance Requirements

The proposed system that we are going to develop will be used for monitoring stock prizes.

1. The system should be easy to handle for end user.
2. System should give expected results i.e predictions.
3. System should provide output within time.

At any given instance there will be a great number of users interacting with the application. The application should be able to handle all the interactions and should respond in the minimum amount of time as per the response time

The application server should be highly scalable. This means that it should scale to high as well as low number of users easily. The application server should also use resources like memory, processors, bandwidth, etc as per the number of users interacting with the application. There shouldn't be a case that high number of resources are being utilized for a small number of users as it will increase the cost for the application server.

The performance of the system is greatly determined by the correct prediction using most efficient algorithms for current as well as future stock prices.

3.4.2. Safety Requirements

- Authorized user login.
- No third-party interference
- An unauthorized user will not have access to modify the content of system.

3.4.3. Security Requirements

The user account should be protected from outside attack or hacking. For this a login system has been implemented in which user has to enter his username and password to login and access the application. This can be termed as client-level security.

There is also a security concern for the server and the database as the attacker can modify the datasets sent to the user which due to which wrong predictions can be shown to the user and can cause financial losses to the user. This can be avoided by employing a good and trusted storage partners.

3.4.4. Software Quality

Various Software Quality Attributes are-

1. Availability: The server should be available all the time as there is no specific time when a user will interact with the application. The server should listen for the incoming connections all the time and should accept the connection after successful authentication.

2. Isolation: When multiple users are online at the same time then their connections should be isolated from each other and behave as if they are the only ones interacting with the server.

3. Correctness: The application should be able to properly run the algorithms and do proper analysis on the datasets.

4. Learnability: The application should be able to learn well and generalize from the trained models and apply the classifier correctly on the new unseen data.

5. Readability: The interface and text attributes should be in away by which the user can easily read his input and the response generated by the application.

6. Robustness: The application should be able to cope with errors in the input and should be able to process this erroneous input and generate an accurate response.

7. Usability: The interface and the users guide should be designed in such a way which is easy to understand and easy to use.

3.5. SYSTEM REQUIREMENTS

3.5.1. Database Requirements

The project requires many datasets to be stored and processed. The standard way to deal with this is in the file system, and links/pointers to these images in the relational database.

The Database platform required: MySQL

3.5.2. Software Requirements

Platform Choice

1. Operating System: Windows
2. IDE: Android Studio.
3.Libraries: NumPy, skit.
4. Programming Language: Python, java

3.5.3. Hardware Requirements:

1. Laptop
2 .Wifi Connection.

3.6. ANALYSIS MODELS: SDLC MODEL TO BE APPLIED

3.6.1 SDLC Model used – Agile Model

Agile SDLC model is a combination of iterative and incremental process models with focus on process adaptability and customer satisfaction by rapid delivery of working software product. Thus, this approach is suitable for the system as it works iteratively to achieve higher accuracy.

Agile Methods break the product into small incremental builds. These builds are provided in iterations. Each iteration typically lasts from about one to three weeks. Every iteration involves cross functional teams working simultaneously on various areas like -

• Planning

• Requirements Analysis

• Design

• Coding

• Unit Testing

• Acceptance Testing.

At the end of the iteration, a working product is displayed to the customer and important stakeholders. Agile model believes that every project needs to be handled differently and the existing methods need to be tailored to best suit the project requirements. In Agile, the tasks are divided to time boxes (small time frames) to deliver specific features for a release.

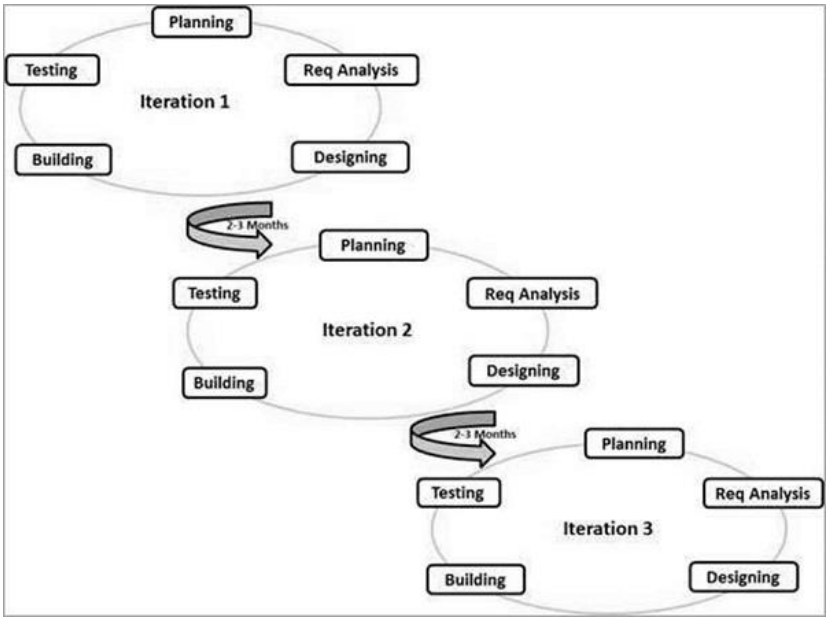


Fig. 3.1 Agile Model

3.7. SYSTEM IMPLEMENTATION PLAN

3.7.1. Introduction

Stock Market Prediction Using Machine Learning can be a challenging task. The process of determining which indicators and input data will be used and gathering enough training data to training the system appropriately is not obvious. The input data may be raw data on volume, price, or daily change, but also it may include derived data such as technical indicators (moving average, trend-line indicators, etc.) or fundamental indicators (intrinsic share value, economic environment, etc.). It is crucial to understand what data can be useful to capture the underlying patterns and integrate into the machine learning system. The methodology used in this work consists on applying Machine Learning systems, with special emphasis on Genetic Programming. GP has been considered one of the most successful existing computational intelligence methods and capable to obtain competitive results on a very large set of real-life application against other methods. Section 3.7.2 introduces the problem statement as well as the software used. 3.7.3 Different Algorithms used in algorithm .

3.7.2. Tools and Technologies Used

- Python
- Usage of libraries like – OpenCV , scikit , pandas , numpy
- Machine Learning techniques -classifiers
- Linear regression techniques
- Jupyter IDE

3.7.3 Methodologies / Algorithms

3.7.3.1 Support Vector Machine (SVM) :-

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high dimensional feature spaces.

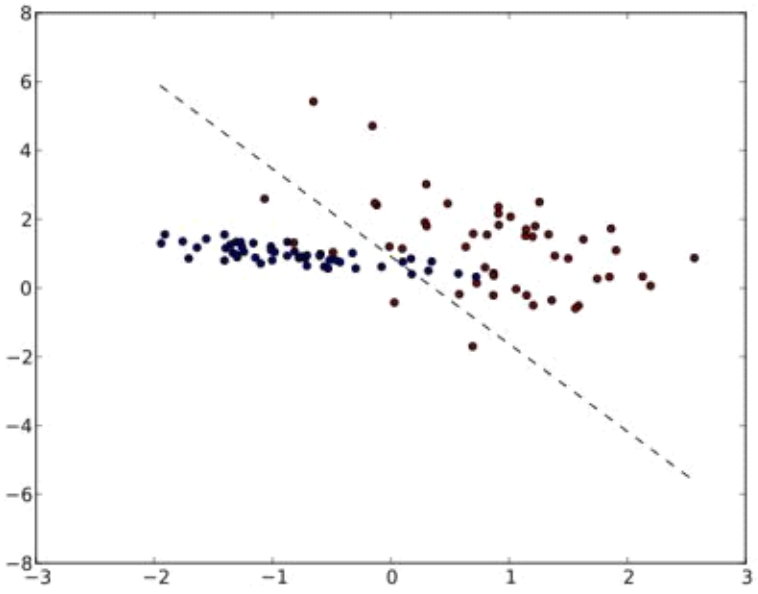


Figure 3. 1: demonstration of SVM

3.7.3.2 Radial Bias

Radial basis function network is an artificial neural network which uses radial basis functions as activation functions. These networks are feed forward networks which can be trained using supervised training algorithms .These networks are used for function approximation in regression, classification and time series predictions . Radial basis function networks are three layered networks where the input layer units does no processing, the hidden layer units implement a radial activation function and the output layer units implement a weighted sum of the hidden unit outputs. Nonlinearly separable data can easily be modeled by radial basis function networks .To use the radial basis function networks we have to specify the type of radial basis activation function, the number of units in the hidden layer and the algorithms for finding the parameters of the network.

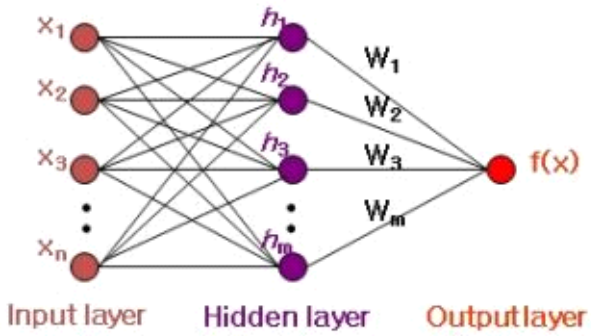


Figure 3.2: An demonstration of Radial Bias

3.7.3.3 Single Layer and Multi-layer Perceptron

A single layer perceptron (SLP) is a feed-forward network based on a threshold transfer function. SLP is the simplest type of artificial neural networks and can only classify linearly separable cases with a binary target (1, 0). The single layer perceptron does not have a priori knowledge, so the initial weights are assigned randomly. SLP sums all the weighted inputs and if the sum is above the threshold (some predetermined value), SLP is said to be activated (output=1). The input values are presented to the perceptron, and if the predicted output is the same as the desired output, then the performance is considered satisfactory and no changes to the weights are made. However, if the output does not match the desired output, then the weights need to be changed to reduce the error.

A multi-layer perceptron (MLP) has the same structure of a single layer perceptron with one or more hidden layers. The backpropagation algorithm consists of two phases: the forward phase

where the activations are propagated from the input to the output layer, and the backward phase, where the error between the observed actual and the requested nominal value in the output layer is propagated backwards to modify the weights and bias values. 2 Propagation: Forward and Backward

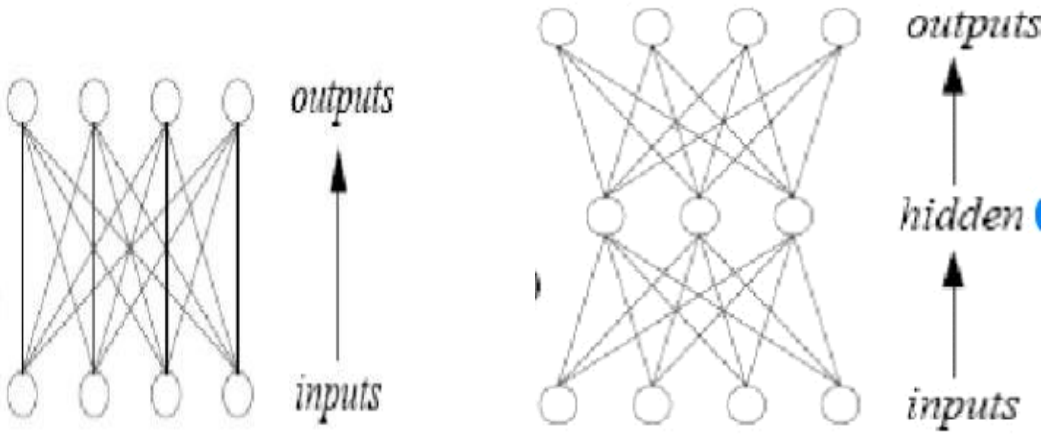


Figure 3.3: An demonstration of Single Level And Multi Level Perceptron

4.SYSTEM DESIGN

4.1 SYSTEM ARCHITECHTURE

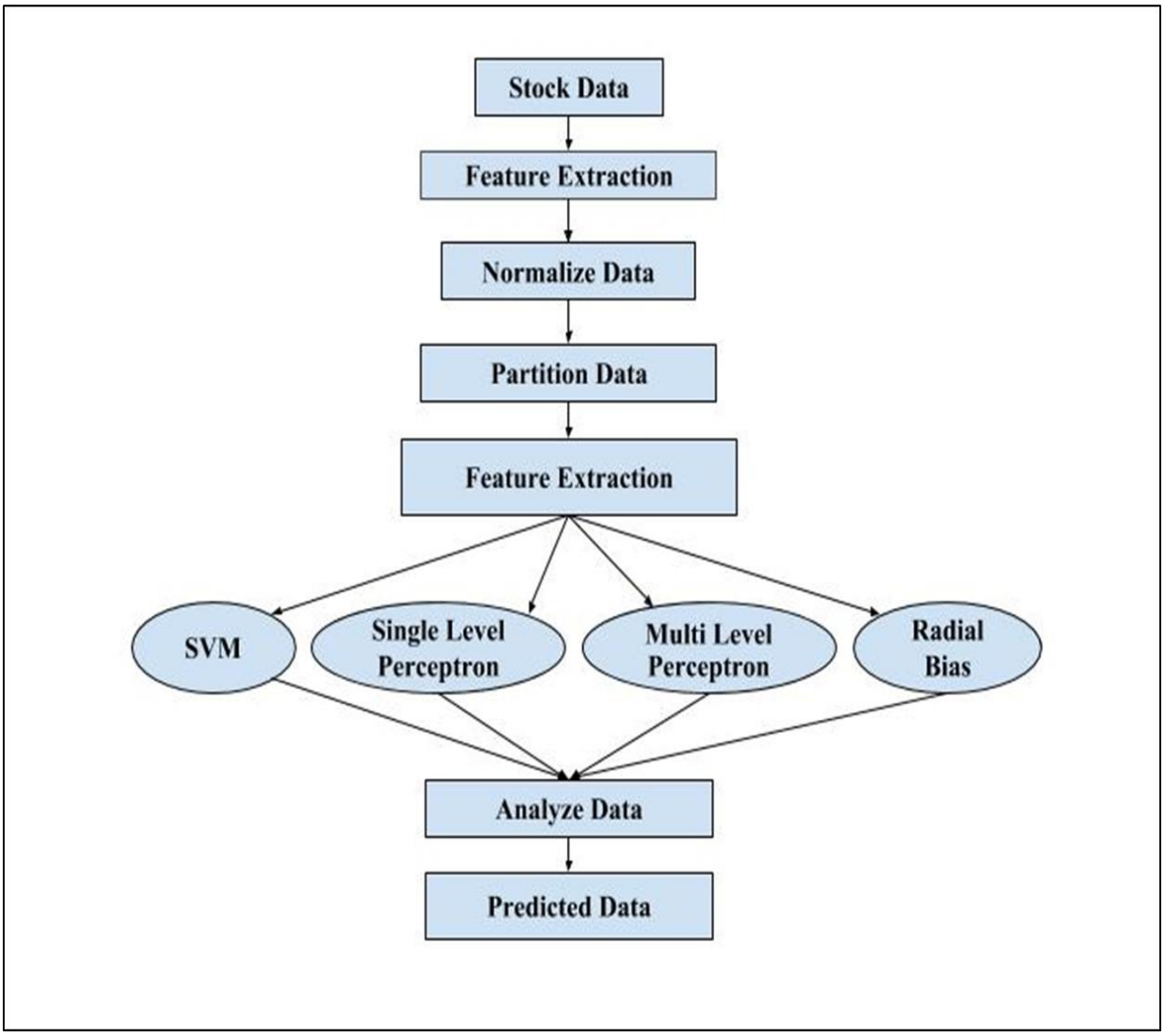


Fig 4.1 System Architecture

4.2 Use case Diagram

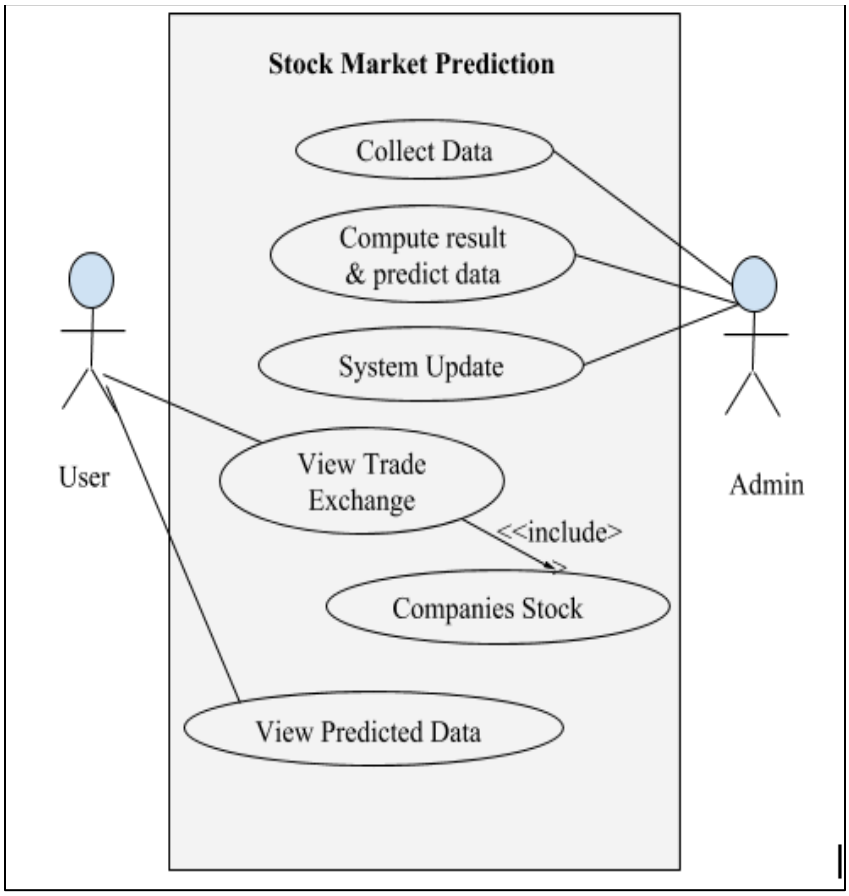


Fig 4.2 . Use case Diagram

4.3 DATA FLOW DIAGRAMS

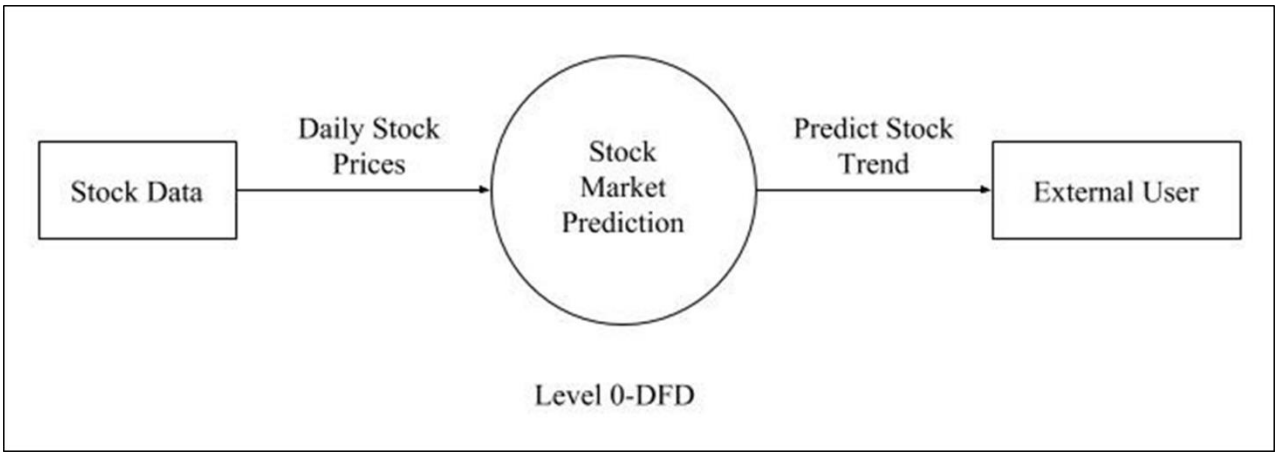


Fig 4.3: Data Flow Diagram Level 0

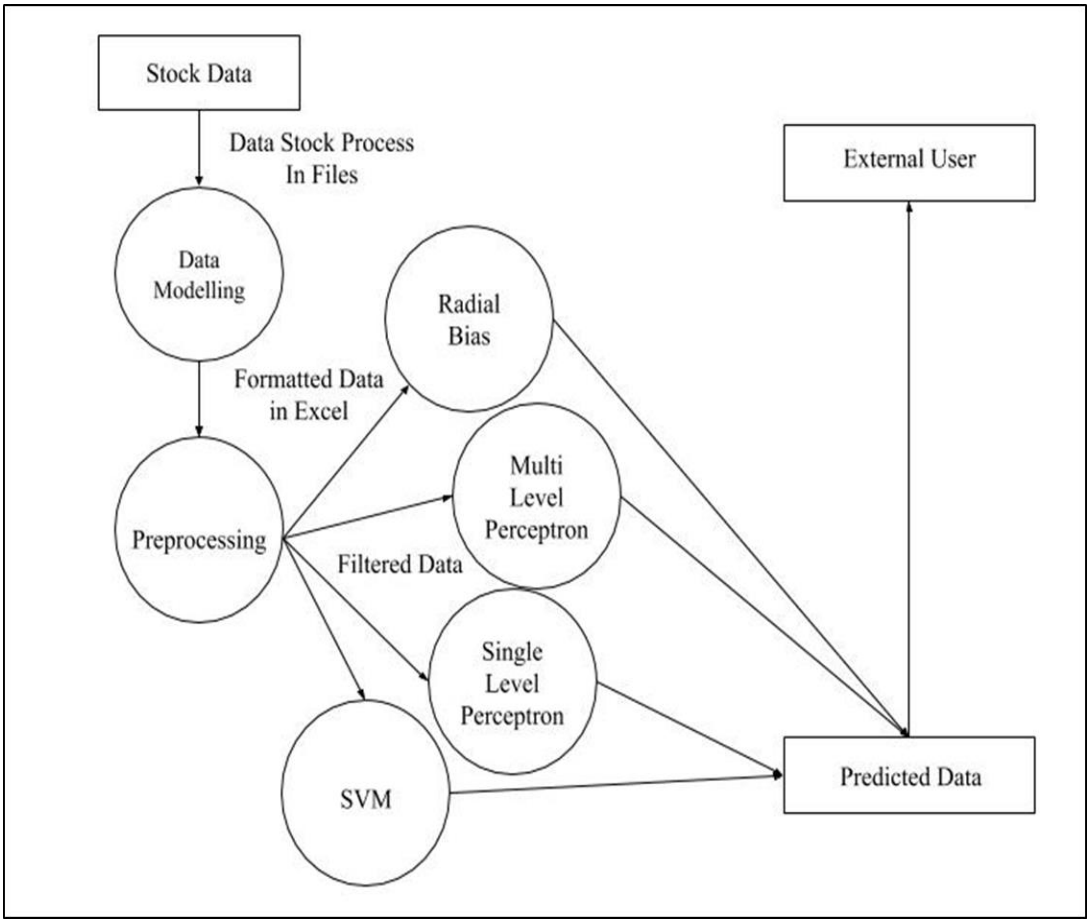


Fig 4.4: Data Flow Diagram Level 1

4.4 ENTITY RELATIONSHIP DIAGRAM

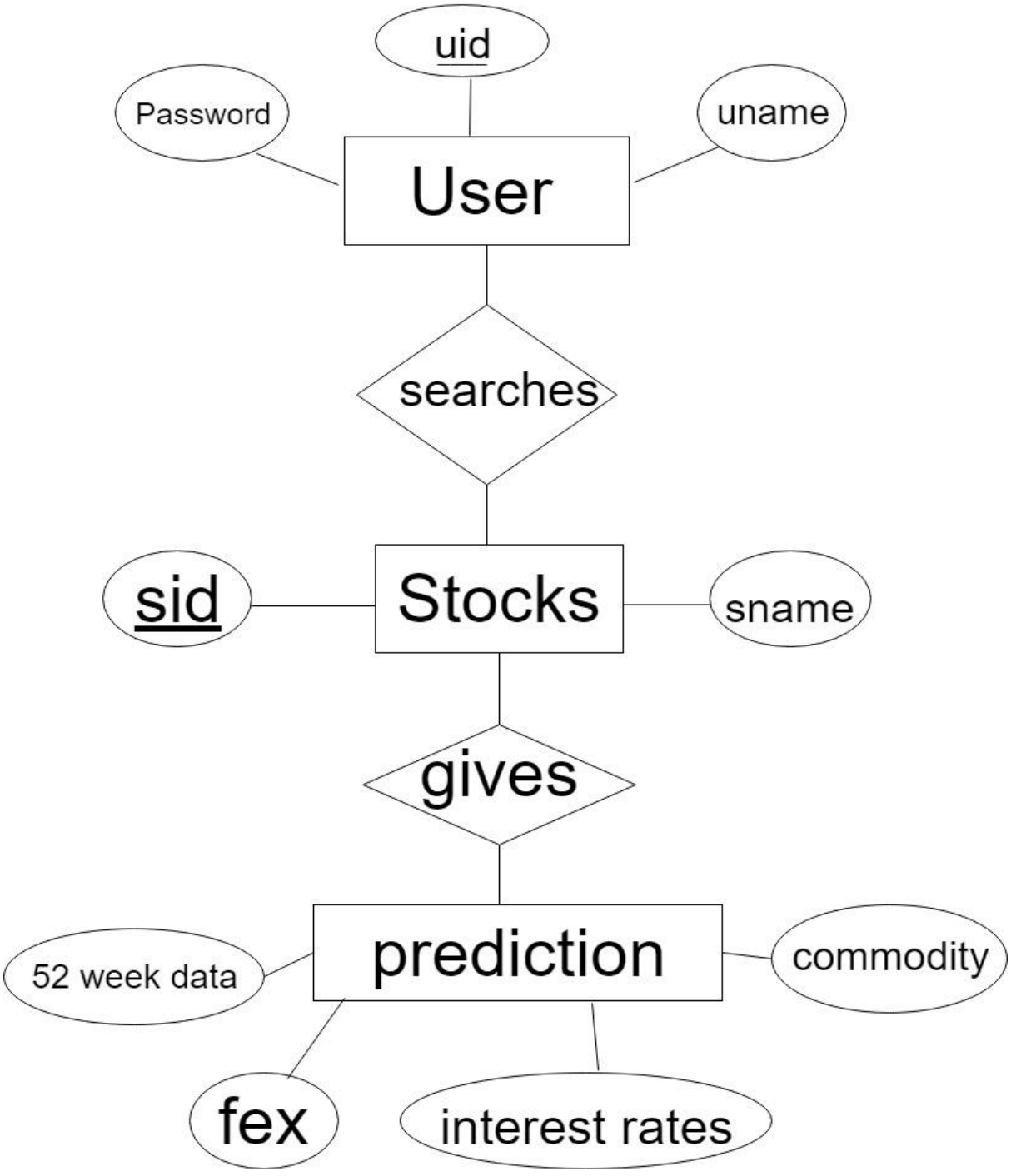


Fig. 4.5 E-R Diagram

4.5 UML DIAGRAMS

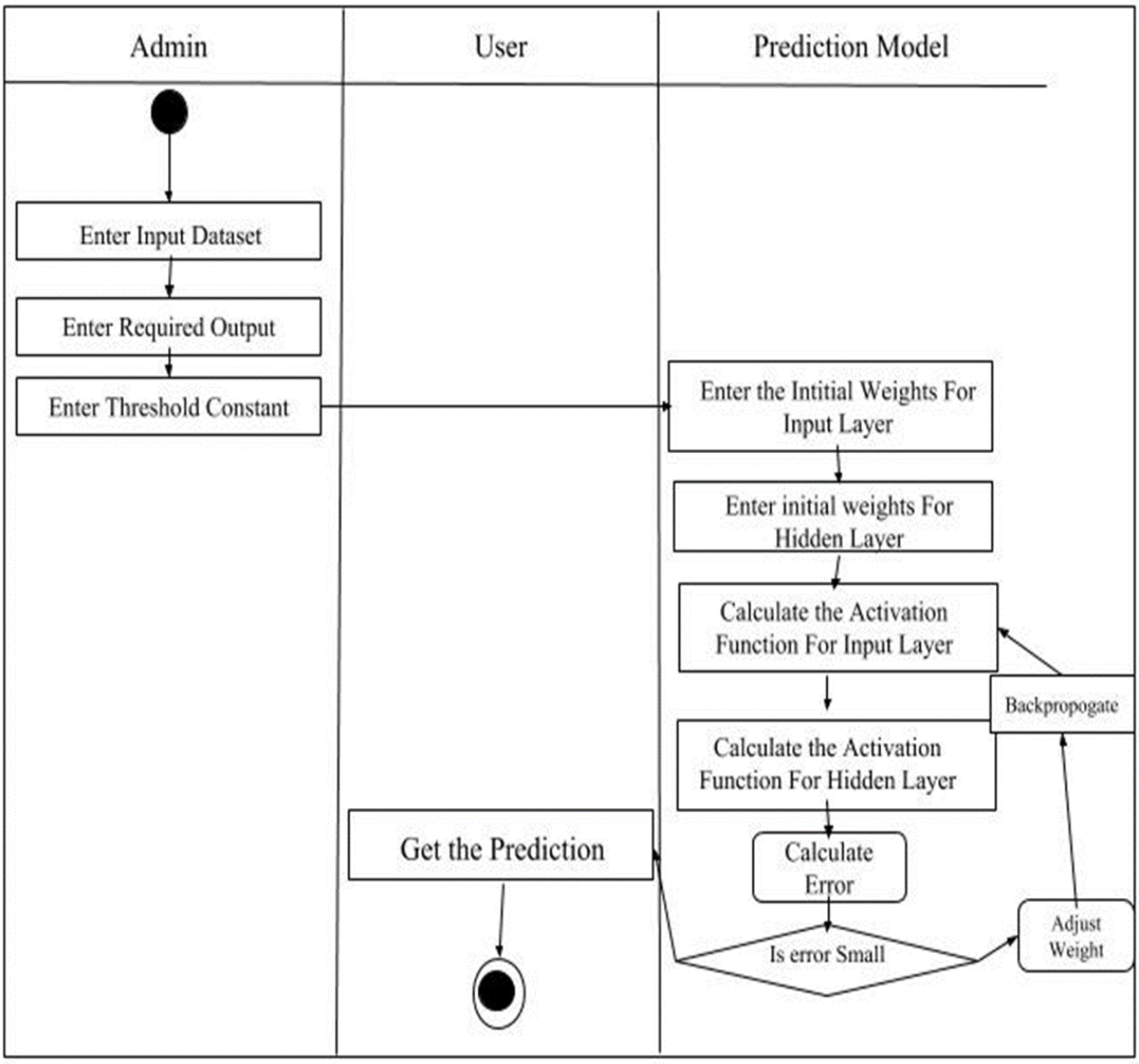


Fig 4.6 : Activity Diagram

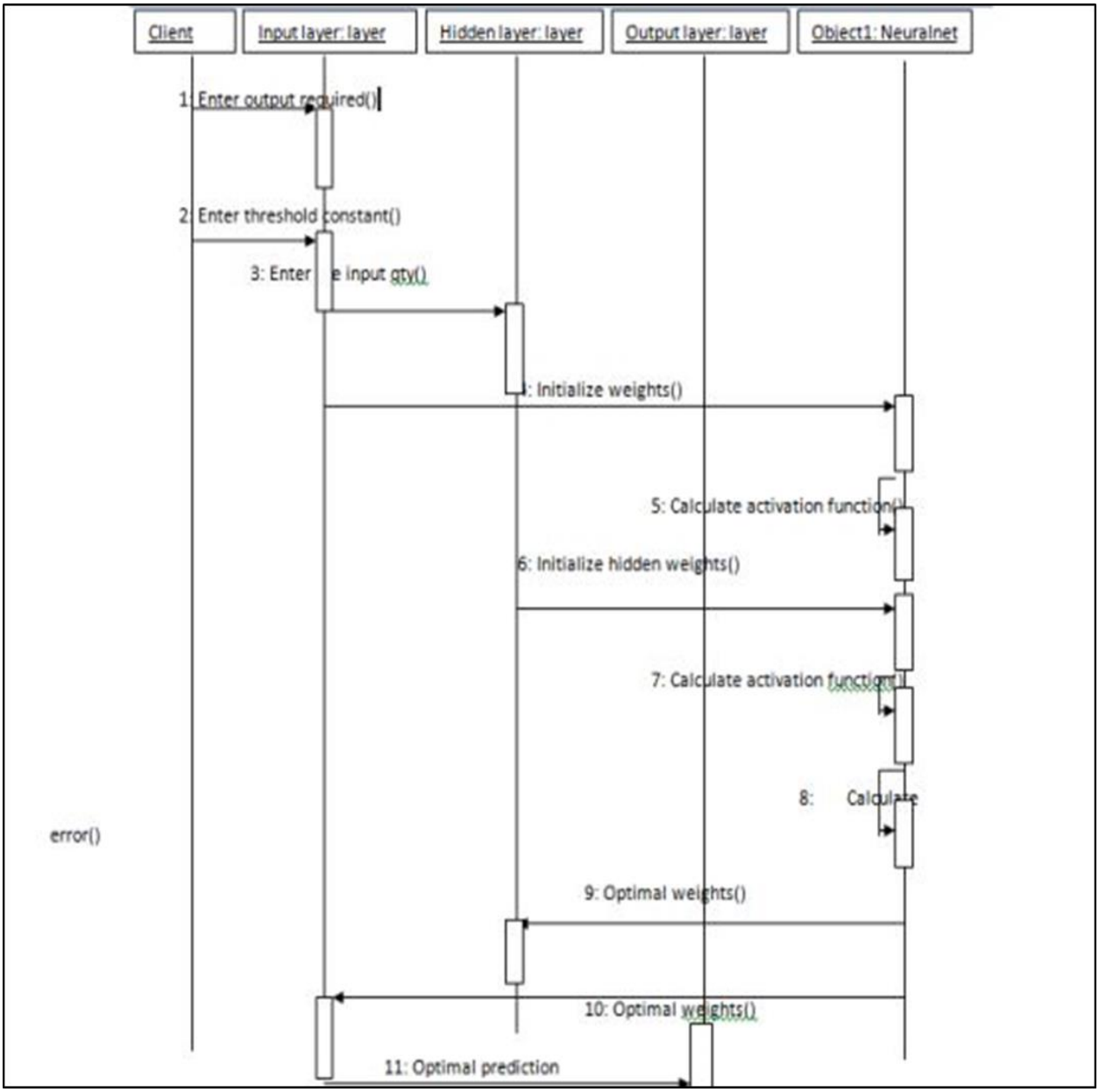


Fig 4.7 : Sequence Diagram

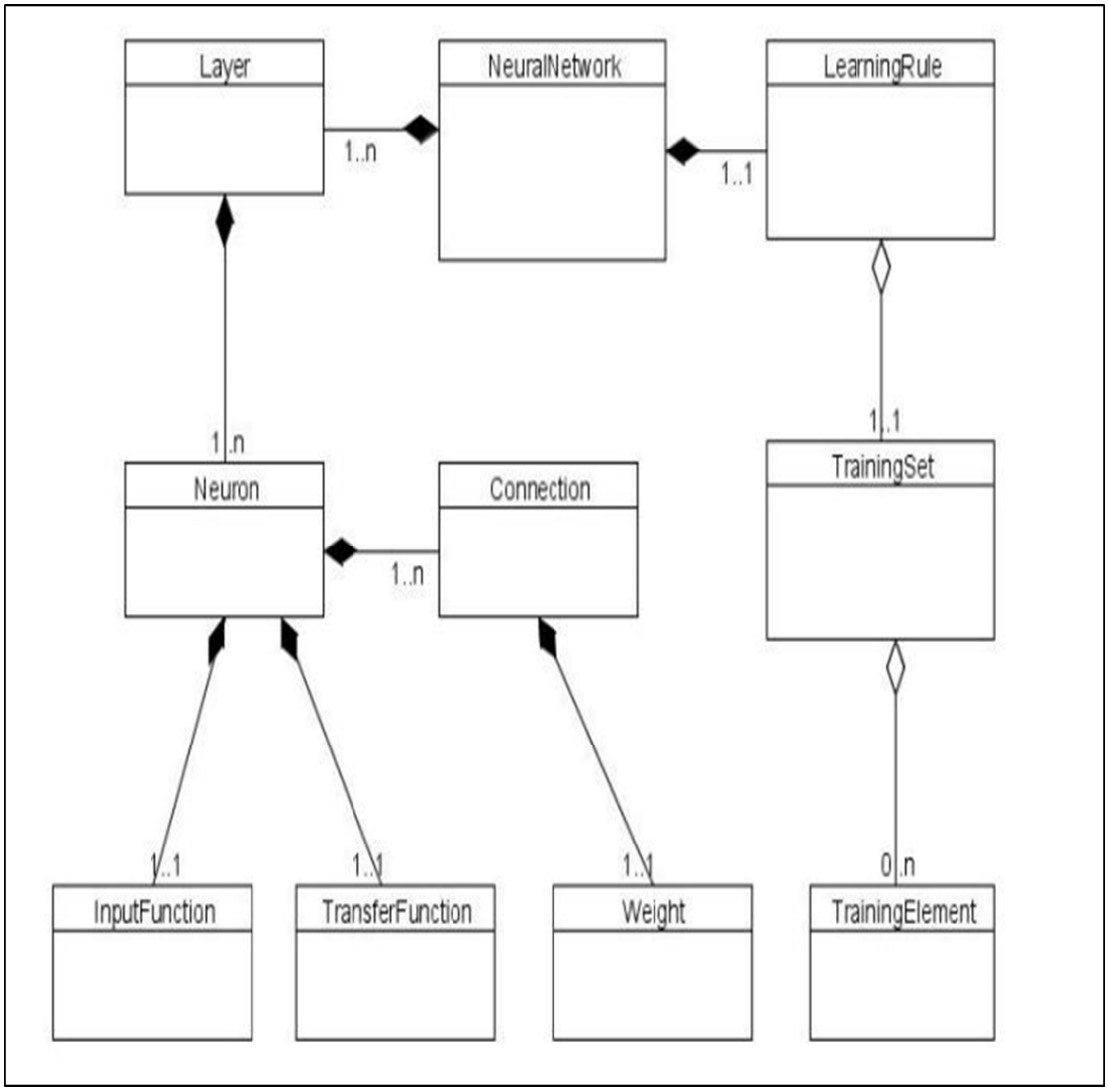


Fig 4.8 : Architectural Diagram

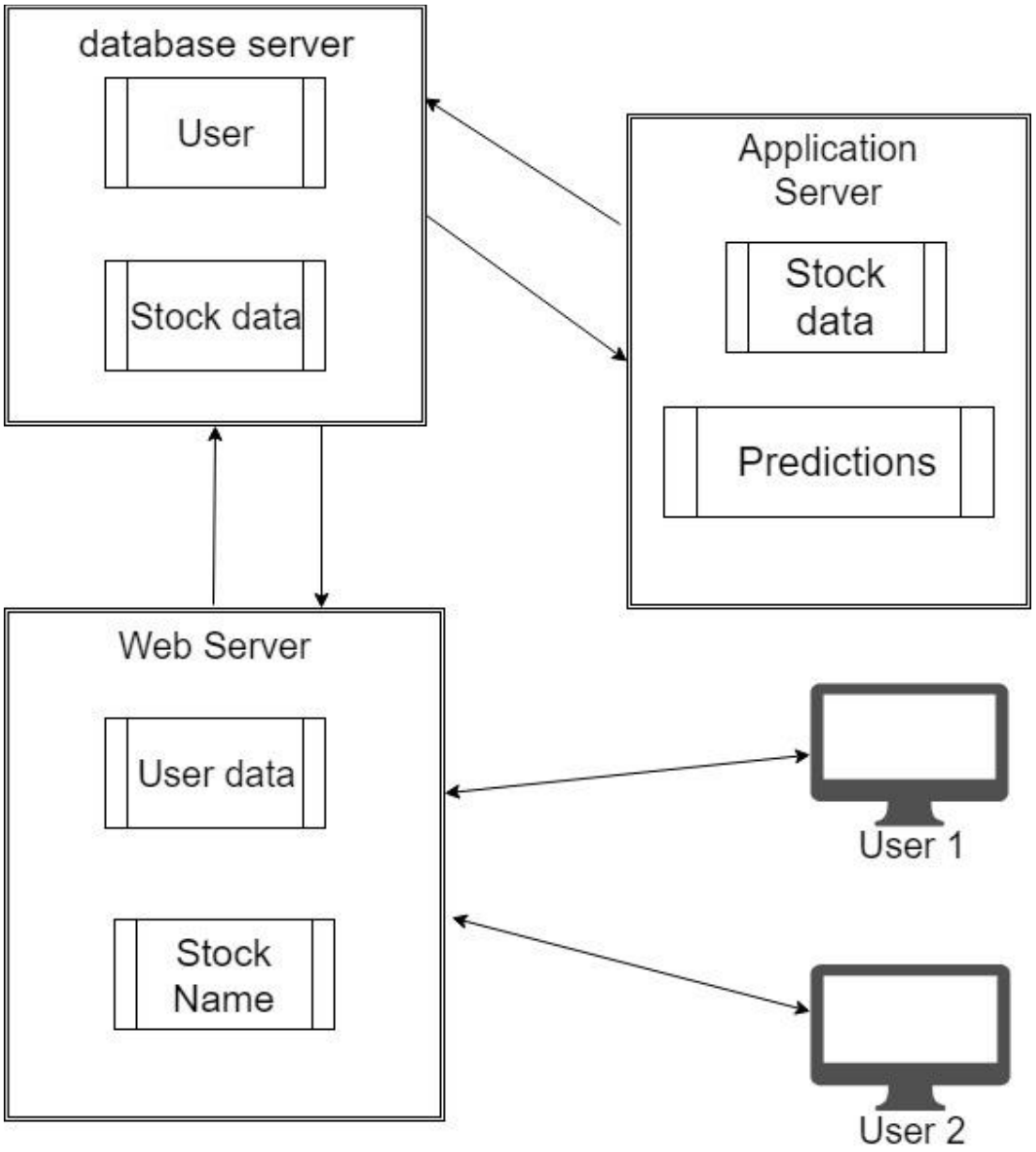


Fig 4.9: Deployment Diagram

5 . SPECIFICATION

5.1 ADVANTAGES

- The system uses four Machine learning algorithms to decide which is the most accurate
- The system can dig through a lot of data to predict the stock price.
- A reliable and accurate program will predict the market movement for a particular stock.

5.2 LIMITATIONS

- This project does not take into account various external factors due to which the predictions can go wrong.
- The system works only on previous datasets to find the prediction and does not know the current trends of the market.

5.3. APPLICATIONS

An automated Stock Market Prediction System is of great use for everyone who want to buy stocks and want to know about their predictions. This project has the following applications

- This system uses four machine learning algorithms (single-layer perceptron, multi-layer perceptron, SVM and radial basis) and then compare the results with actual results to find out the most accurate ML algorithm.
- Stock broker can give advice and suggestions to his clients.
- This system saves opening, closing, intraday low values and intraday peak values of the stocks.
- It provides tabular as well as graphical representation of the closing rates.

6. CONCLUSION AND FUTURE WORK

Conclusion:

In this thesis, we looked at the problem of forecasting stock performance. Although a substantial volume of research exists on the topic, very little is aimed at long term forecasting while making use of machine learning methods and textual data sources. We prepared over ten years worth of stock data and proposed a solution which combines features from textual yearly and quarterly filings with fundamental factors for long term stock performance forecasting. Additionally, we developed a new method of extracting features from text for the purpose of performance forecasting and applied feature selection aided by a novel evaluation function. Problems Overcome. To produce effective models, there were two main problems we were faced with and had to overcome. The first was that of market efficiency, which places theoretical limits on how patterns can be found in the stock markets for the purpose of forecasting. This property can become a concrete problem by patterns being exhibited in the data which are useless or even detrimental for predicting future values. The way we tried to deal with this was by carefully splitting our data into training, validation, and testing data with expanding windows so as to make maximum use of it while trying to avoid accidental overfitting. The second way we dealt with this was by using a tailored model performance metrics , which aimed to ensure good test performance of models by not only maximizing model validation , but also minimizing the variation across validation years of this value. The third way we dealt with market efficiency was by performing feature selection using the Algorithm, so as to remove those features which performed poorly or unreliably. The second set of problems came from putting together a dataset to use for experimentation and testing. Due to the large volume of the data, care had to be taken when cleaning and preparing it, and the inevitable mistakes along the way required reprocessing of the data. Using expert knowledge, we determined how to deal with the various problems in the data and ended up using mean substitution and feature deletion.

Future Work:

1. Model Updating Frequency:-

They are trained once and then used for predicting stock performances over the span of a year. Since we use a return duration of 120 trading days, there is a necessary wait of half a year before data can be used to train models, which means that models end up making predictions using data which is over a year old. One way to make use of data as soon as it become available is to completely retrain the model every week (or less). A faster way to improve model performance may be through updating using incremental machine learning algorithms, which can update model parameters without re-training on all data.

2. Explore More Algorithms :-

Although many different models were considered in this thesis, including various linear regression methods, gradient boosting, random forests, and neural networks, there is always more room to explore.

3. Improve Feature Extraction :-

In this thesis, a few methods for extracting features from filings with textual data were explored. The problem of extracting features from text and determining text sentiment in particular are well studied, and other natural language processing methods may perform better. Our approach of using autoencoders to extract features may also benefit from further exploration. In particular, when using the auxiliary loss, a more accurate method for estimating the financial effect corresponding to a given filing would be useful.

4. Utilize Time Series Information. :-

Similar to the idea of updating model frequency, another area for exploration includes utilizing the time series aspect of the data. Our current models are not aware that the samples occur in any temporal order, and thus are not able to spot patterns in stock performance that depend on knowing the order of samples. One type of model that is often used to find and make use of these type of patterns are recurrent neural networks.

7. References

1- Johan Bollen,Huina Mao, Xiao-Jun Zeng "Twitter mood predicts the stock market."

2- Xingyu Zhou , Zhisong Pan , Guyu Hu , Siqi Tang,and Cheng Zhao1,2 Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets

3- Raut Sushrut Deepak 1 , Shinde Isha Uday 2,Dr. D. Malathi 31,2B.Tech Student
SRM University, Kattankulathur, Chennai. - MACHINE LEARNING APPROACH IN STOCK MARKET PREDICTION

4- Ritu Sharma1, Mr. Shiv Kumar2, Mr. Rohit Maheshwari3 - Comparative Analysis of Classification Techniques in Data Mining Using Different Datasets

5- Andrew W. Lo, A. Craig MacKinlay "NATIONAL BUREAU OF ECONOMIC RESEARCH STOCK MARKET PRICES DO NOT FOLLOW RANDOM WALKS: EVIDENCE FROM A SIMPLE SPECIFICATION TEST"

6- Prof. S .P. Pimpalkar, Jenish Karia, Muskaan Khan, Satyam Anand, Tushar Mukherjee
Department of Computer Engineering, AISSMS IOIT,- Stock Market Prediction using Machine Learning

7- Osman Hegazy - Faculty of Computers and Information Cairo University and Omar S. Soliman Cairo University - A Machine Learning Model for Stock Market Prediction

8- Yash Omer Department , Nitesh Kumar Singh of Information Technology Bharati Vidyapeeth University, College of Engineering - Stock Prediction using Machine Learning

8. Appendix

Appendix A

Mathematical Model

1. SVM

linear discriminant function:

$$f(x) = w.x+b$$

In this function, **x** refers to a training dataset vector, **w** is referred to as the weight vector and the value **b** as the bias term. The term **w.x** refers to the dot product (inner product, scalar product), which calculates the sum of the products of vector components.

Classification hyper-plane equations:

Positive margin hyper-plane equation: $w.x - b = 1$

Negative margin hyper-plane equation: $w.x - b = -1$

Middle optimum hyper-plane equation: $w.x - b = 0$

2. Single and Multi-level perceptron

Single and multi-level perceptrons have multiple inputs and a single output. Consider **x1,x2,…xn** be input vectors and **w1,w2,…wn** be the weights associated with them.

Output $a = x1.w1 + x2.w2 + \ldots xn.wn$

3. Radial basis function

$$h(x) = \Phi((x - c)^T R^{-1} (x - c))$$

Where Φ is the function used, **c** is the center and **R** is the metric. The term $(x - c)^T R^{-1} (x - c)$ is the distance between the input **x** and the center **c** in the metric defined by **R**. There are several common types of functions used such as Gaussian $\Phi(z) = e^{-z}$, the multi-quadratic $\Phi(z)=(1+z)^{1/2}$, the inverse multi-quadratic $\Phi(z) = (1+z)^{-1/2}$ and the Cauchy $\Phi(z) = (1+z)^{-1}$.

Appendix B

1- Johan Bollen1,Huina Mao1,Xiao-Jun Zeng Twitter mood predicts the stock market.
Do the comparative analysis of several data mining classification techniques on the basis of parameters accuracy, execution time, types of datasets and applications.

2- Xingyu Zhou ,1 Zhisong Pan ,1 Guyu Hu ,1 Siqi Tang,1 and Cheng Zhao1,2 Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets
Test the random walk hypothesis for weekly stock market returns by comparing variance estimators derived from data sampled at different frequencies.

3- Raut Sushrut Deepak 1 , Shinde Isha Uday 2,Dr. D. Malathi 31,2B.Tech Student
SRM University, Kattankulathur, Chennai. - MACHINE LEARNING APPROACH IN STOCK MARKET PREDICTION

4- Ritu Sharma1, Mr. Shiv Kumar2, Mr. Rohit Maheshwari3 - Comparative Analysis of Classification Techniques in Data Mining Using Different Datasets

5- Andrew W. Lo, A. Craig MacKinlay NATIONAL BUREAU OF ECONOMIC RESEARCH
STOCK MARKET PRICES DO NOT FOLLOW RANDOM WALKS: EVIDENCE FROM A SIMPLE SPECIFICATION TEST

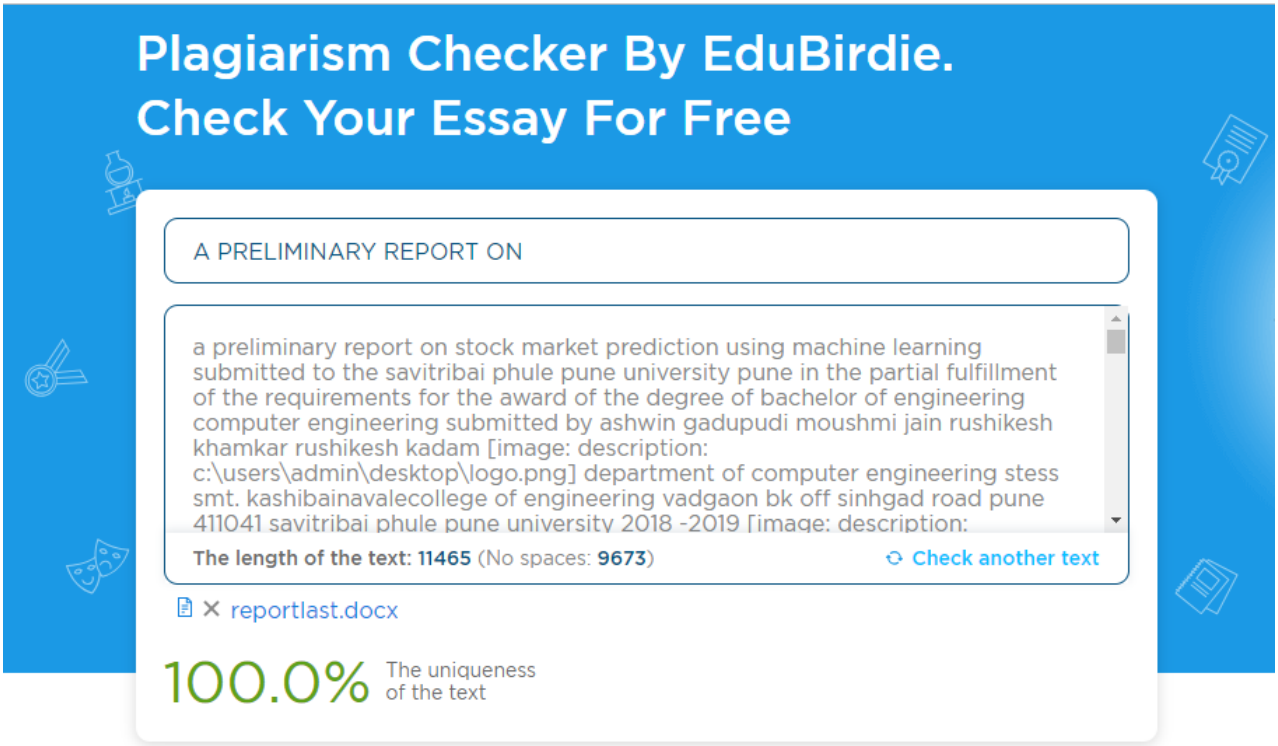
6- Prof. S .P. Pimpalkar, Jenish Karia, Muskaan Khan, Satyam Anand, Tushar Mukherjee
Department of Computer Engineering, AISSMS IOIT,- Stock Market Prediction using Machine Learning

7- Osman Hegazy - Faculty of Computers and Information Cairo University and Omar S. Soliman Cairo University - A Machine Learning Model for Stock Market Prediction

8- Yash Omer Department , Nitesh Kumar Singh of Information Technology Bharati Vidyapeeth University, College of Engineering - Stock Prediction using Machine Learning .
Support Vector Machine with Regression Technology (SVR), Recurrent Neural networks(RNN).
Regression analysis on attributes using simple Regression and multivariate analysis used.

Appendix-C

Plagiarism Report



7.1 Types of Testing

Unit Testing:

A unit is the smallest testable part of any software. It usually has one or a few inputs and usually a single output. Unit testing basically focuses on unit parts of the software design. In this we test an individual unit or group of inter related units. It is often done by programmer by using sample input and observing its corresponding outputs. The purpose is to validate that each unit of the software performs as designed. For ex checking if loop in a program or any method in a program.

Integration Testing:

The objective is to take unit tested components and build a program structure that has been dictated by design. Integration testing is testing in which a group of components are combined to produce output. The purpose of this level of **testing** is to expose faults in the interaction between integrated units. The purpose of integration testing is to verify the functional, performance, and reliability between the modules that are integrated. Example is white box testing and black box testing.

Regression Testing:

Regression testing is a type of software testing in which testing is performed on the whole system to check whether the system is working properly after addition of new code in the existing code and the older code is working properly. Before a new version of a software product is released, the old test cases are run against the new version to make sure that all the old capabilities still work. The reason they might not work is because changing or adding new code to a program can easily introduce errors into code that is not intended to be changed.

Alpha Testing:

Alpha Testing is defined as a type of software testing performed to identify bugs before releasing the product to real users or the public. It is a type of acceptance testing. Alpha testing takes place at the developer's site by the internal teams, before release to external customers.

7.2 Test Cases and Test Results

ID	PURPOSE	DESCRIPTION	EXPECTED RESULTS
1.	Verify login with valid login credentials.	Step 1: Open app	App opens
		Step 2: Go to the login page after reading instructions.	Login page opens
		Step 3: Enter valid username and Password	
		Step 4: Click the login button.	The app logs you in and you can get access to the app.

ID	PURPOSE	DESCRIPTION	EXPECTED RESULTS
2.	Verify login with invalid credentials	Step 1: Open app	App opens
		Step 2: Go to the login page after reading instructions.	Login page opens
		Step 3: Enter invalid username or password.	
		Step 4: Click the login button.	Popup appears showing invalid username or password.

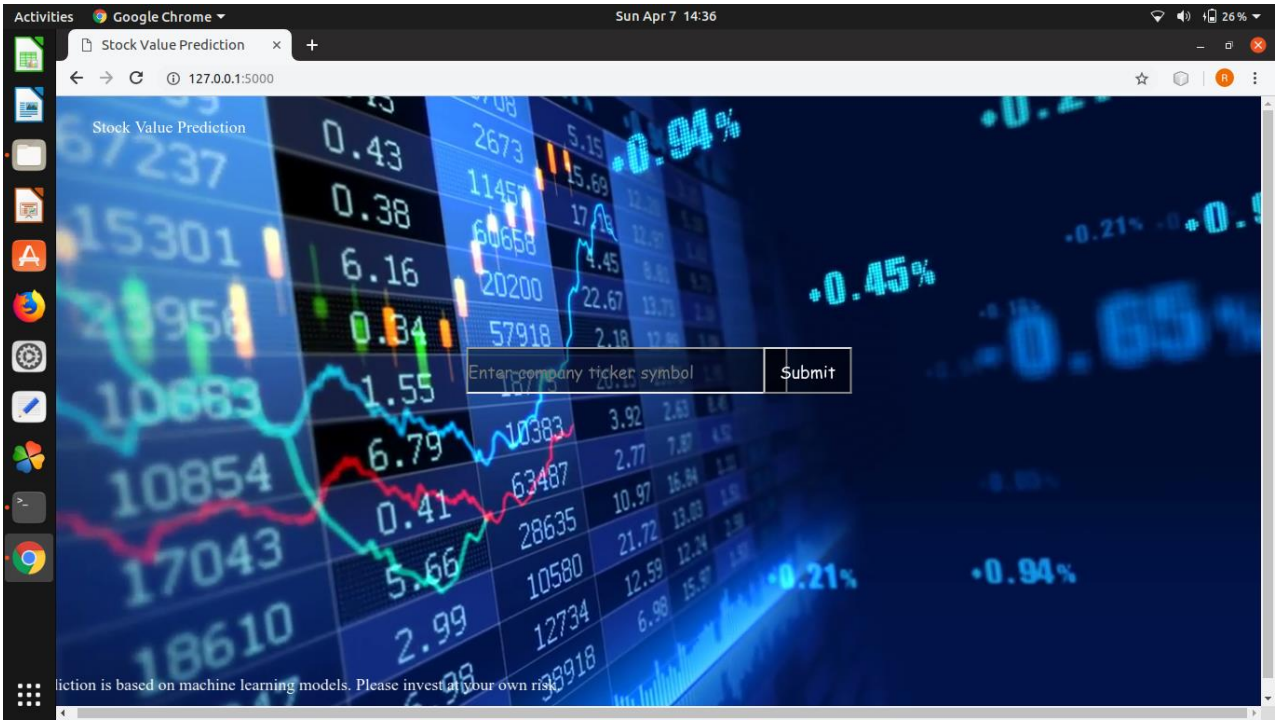
ID	PURPOSE	DESCRIPTION	EXPECTED RESULTS
3.	Forecasting of Stock Prices	Step 1: Open app.	App opens.
		Step 2: Go to login page after reading instructions	Login page opens.
		Step 3: Enter valid username and password details.	
		Step 4: Click the login button.	After login it will redirect you to the next page.
		Step 5: Enter the ticker of the stock that you want to forecast	
		Step 6: Click on submit	'Fbprophet' algorithm will forecast the said stock price and show you the results.

8. RESULTS

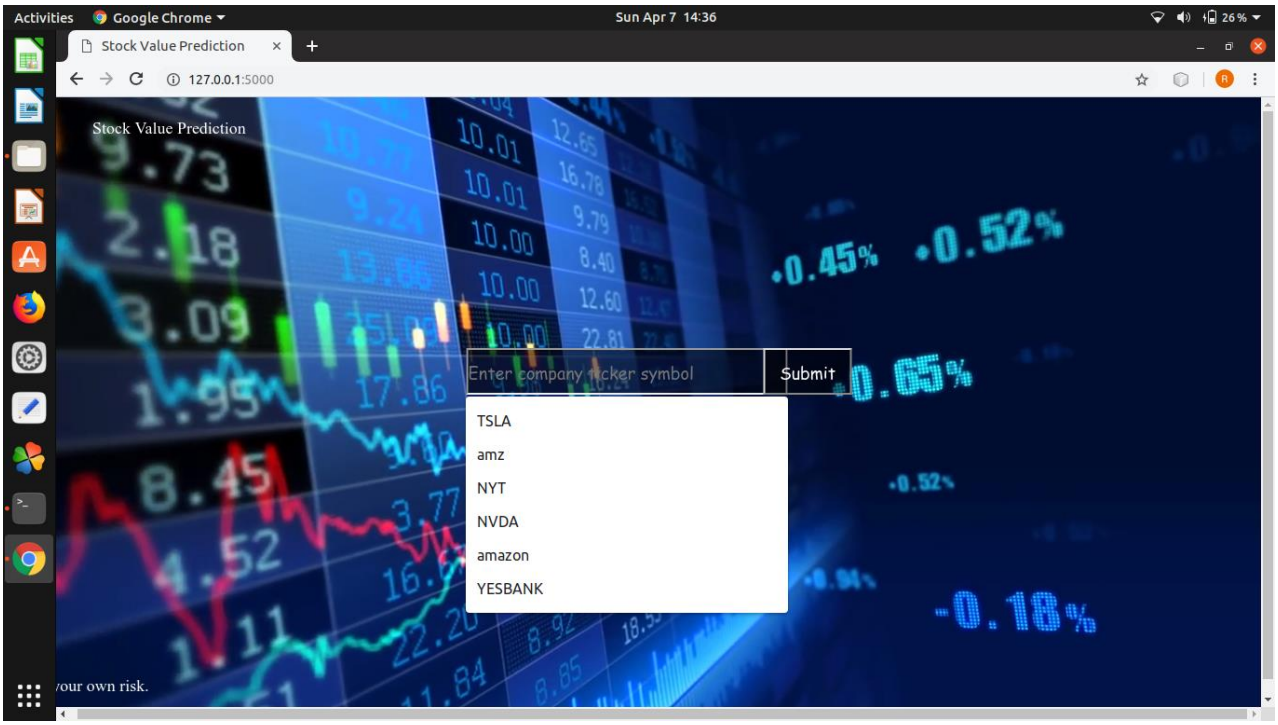
8.1 Outcomes

8.2 Screenshots

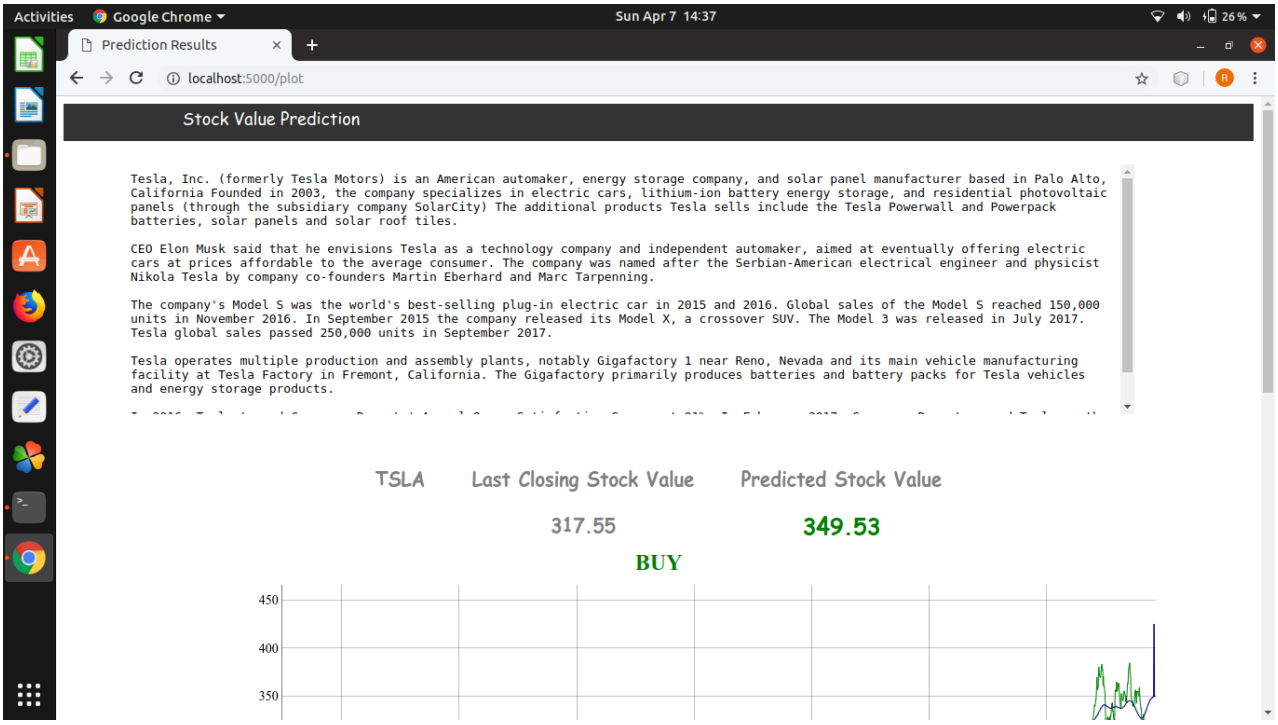
Screenshot-1



Screenshot-2

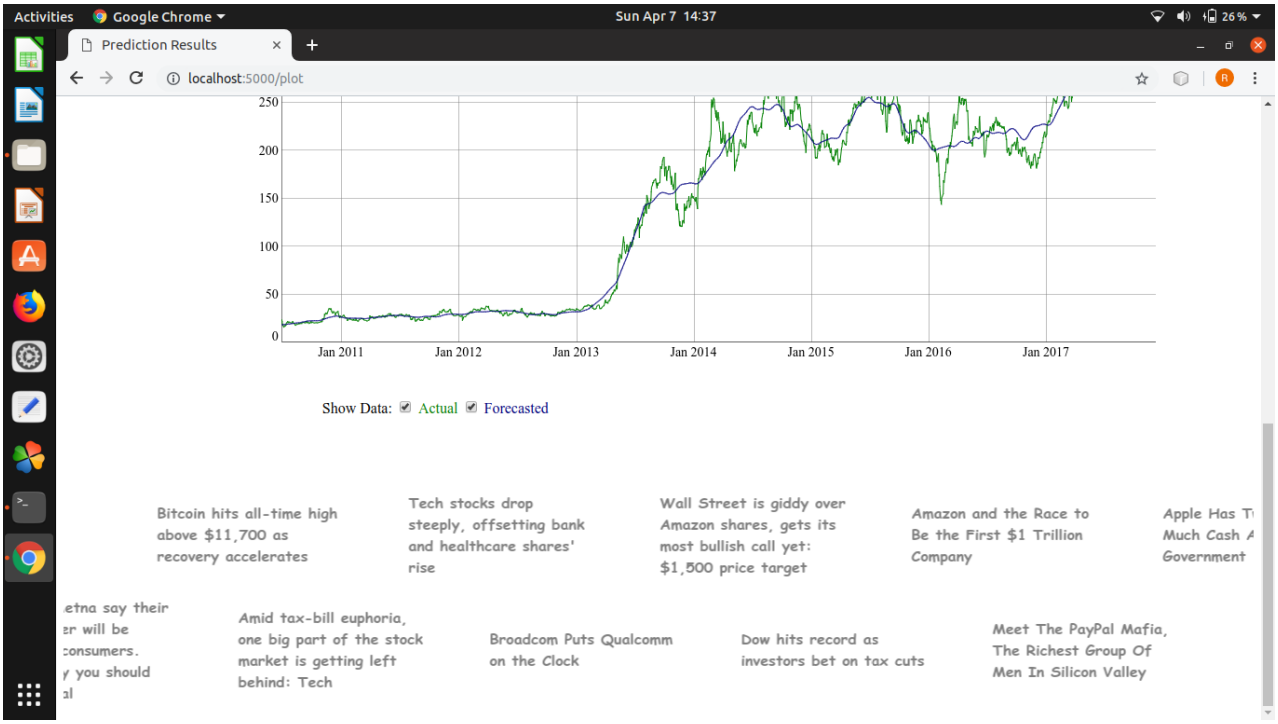


Screenshot-3



Screenshot-4

Screenshot-5



Screenshot-6