Practical Training Report

on

Analysis of Stock Prices using Machine Learning

(BTC – 550)

submitted in partial fulfillment of the requirements

for the award of the degree

of

Bachelor of Technology

in

Computer Science & Engineering

By

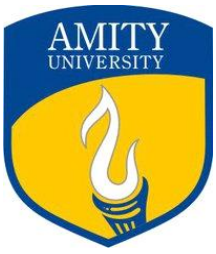
Shreyash Mehta

Enrollment No. A60205216031

Under the guidance of

Ms. Madhavi Dhingra

Assistant Professor



Department of Computer Science & Engineering
Amity School of Engineering & Technology
Amity University Madhya Pradesh, Gwalior

November 2018

Department of Computer Science and Engineering
Amity School of Engineering and Technology
Amity University Madhya Pradesh, Gwalior

Declaration

I, **Shreyash Mehta**, student of Bachelor of Technology in Computer Science & Engineering hereby declare that the practical training report entitled **"ANALYSIS OF STOCK PRICES USING MACHINE LEARNING"** which is submitted by me to Department of Computer Science & Engineering, Amity School of Engineering & Technology, Amity University Madhya Pradesh, in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science & Engineering, has not been previously formed the basis for the award of any degree, diploma or other similar title or recognition.

Shreyash Mehta

Date: (Enrollment No. – A60205216031)

Department of Computer Science & Engineering
Amity School of Engineering and Technology
Amity University Madhya Pradesh, Gwalior

CERTIFICATE

This is to certify that **Shreyash Mehta (Enrollment No. A60205216031),** student of B.Tech(C.S.E) V semester, Department of Computer Science & Engineering, ASET, Amity University Madhya Pradesh, has done his practical training entitled **"ANALYSIS OF STOCK PRICES USING MACHINE LEARNING"** under my guidance and supervision during **"28 May 2018 – 10 July 2018"**

The work was satisfactory. He has shown complete dedication and devotion to the given project work.

Date:

(Ms. Madhavi Dhingra)
Assistant Professor
Supervisor

External Examiner

(Dr. Venkatadri Marriboyina)
Head of the Department

Be Practical and Knowledgeable

Binary SYSTEMS

Address : 214-215, 1 Floor Garima Arcade, Shinde Ki Chhawani, Lashkar, Gwalior (M.P.)
Phone : 0751-4004328
Mob.: +91-9009054508, 9630883466
E-Mail : priyank.binarysystems@gmail.com
Website : http://binarysystems.co.in/

Ref.: BS-2018/Training/P-13 **Date:**23-JULY-2018

To whom so ever it may concern

This is to certify that **MR. SHREYASH MEHTA** (Enroll No. A60205216031) AMITY SCHOOL OF ENGINEERING & TECHNOLOGY (AUMP), B.Tech (CSE) has been accepted as a trainee at our organization **BINARY SYSTEMS. Gwalior (M.P.)**. His training details:

Period- **28 MAY 2018 to 10 JULY 2018**
Technology- **Python with Machine Learning**

All of us at **BINARY SYSTEMS** are pleased to have him in our team. As such his internship includes training orientation and focus primarily on learning and developing new skills and gaining a deeper understanding of concepts through hands in application of the knowledge he learned.

We take this opportunity to wish you a long, happy and successful career.

For Priyank
Authorized Signatory
BINARY SYSTEMS

BINARY SYSTEMS BS GWALIOR

Acknowledgement

I am very much thankful to our honorable Vice Chancellor **Lt Gen. V. K. Sharma AVSM (Retd)** for allowing us to carry out my practical training. I would also like to thank **Prof. (Dr.) M. P. Kaushik**, Pro-Vice Chancellor, Amity University Madhya Pradesh for his support.

I extend my sincere thanks to **Maj. Gen. (Dr.) S. C. Jain, VSM** (Retd),** HOI, Amity School of Engineering and Technology, Amity University Madhya Pradesh, Gwalior for his guidance and support for the selection of appropriate industry for my practical training. I would also like to thank **Prof.(Dr.) Venkatadri Marriboyina** Head of Department (CSE), for his kind concern throughout the practical training.

I am also very grateful to **Ms. Madhavi Dhingra,** Assistant Professor, Department of Computer Science & Engineering, Amity School of Engineering and Technology, Amity University Madhya Pradesh and **Mr. Priyank Gupta**, Binary System my internal and external supervisor respectively, for their constant guidance and encouragement provided in this endeavor.

I am also thankful to the whole staff of Binary Systems for the co-operation and giving me friendly environment which made very comfortable for learning and that of ASET, AUMP for teaching and helping me always. Last but not the least I would like to thank my parents and friends for their constant support.

Shreyash Mehta

Enrollment No- A60205216031

ABSTRACT

"Prediction of Stock Market Using Machine Learning" is basically designed for Artificial Intelligence (AI) based Application. The main purpose of this project is to save the time of the investors by providing all the information at one platform.

It has been tedious task for the investors to get the information about the Future Prediction of Stock Prices of different companies listed on the Stock Exchange.

To combat form problem like this, it was necessary to design a system or application where the workforce required would be decreased and also process of searching, analyzing the records will be easy.

In wake of this ideology, a "Prediction of Stock Market Using Machine Learning" has been implemented in Python 3.6 IDLE or Anaconda Jupyter Notebook. This project contains all information about the Historical data of company, IPO Price of company and company details. Investors can register their self for a long time and can search all the information regarding with the companies. Through this investors can get a detailed information of the companies and their historical data.

Stock Market Using Machine Learning is implemented in two modules first one is Admin module and another one is Investors module. Admin module will perform various functions like add company, upload company details and update database so that investors can get information properly.

So Stock Market Using Machine Learning is the perfect online arena, where investors can find their perfect company to invest without wasting their time.

Keywords: Stock, Machine Learning, Prediction.

LIST OF FIGURES

CONTENTS

Chapter 1

INTRODUCTION

The Analysis of Stock Market using Machine Learning is the software to predict the stock prices. Over the past years people have been going to invest into or buy the shares of the company where sometimes people fail to invest in right company and suffer a great loss. So, this project is developed to help the administrator to manage Companies and even to save and respect the time of users. This project is useful because many people who faces a difficulty to manage the details of the company they want to invest into or to buy shares of that company. Other than this, this project is designed in order to upgrade the manual system and make such a system which is easy to access and quite systematic. With this system software the users can see the company's historical data their IPO Price and their future predictions and after analyzing the whole scenario then user can invest and buy the shares of the company that can be less risky.

1.1 Objective

The main objective of this project to design system software that will automate the whole Procedure of predicting the stock prices and save time. The software that would have facilities like viewing the details of the company seeing the historical data of the company viewing their IPO Price and their future predictions.

The main feature of this project is that it is time saving, as there is high risky in investing in stock market or buying the shares of the company so by this project users can analyze the company details and their future predictions and then invest or buy the shares which can reduce the amount of risk.

Chapter 2

REVIEW OF LITERATURE AND DEFINITION OF PROBLEM

- **Existing System:**

  In the existing system, people have to analyze the market by themselves. They don't have mode to analyze the stock prices of the specific company. Due lack of analyze of the company people suffer a great loss in their shares.

- **Problem with the Existing System:**

  Since the existing system is a long process to find the best company to invest in and uploading the data separately like the procedure of modification, addition and deletion with the new needs. This causes delay in the process of working and getting details about it.

- **Need for Advanced System:**

  To eliminate this existing problem, this system software has been designed and it is a advanced system. This type of system is needed because of the lack of a précises to predict the prices in the existing system. This system provides speed with accuracy and efficiency. The new implementation can be done with in a seconds. The flow of information process becomes faster.

- **What is Stock Market:** Stock market is a place where the shares of listed companies are traded. The primary market is the place where companies can bring the shares for the public in a initial public offering (IPO) to raise the company's capital.

  When the new securities is sold in the primary market, then they are traded in the secondary market, in secondary market an investor buys shares of another investor's at the price prevailing on the market or at any price which is accepted by the buyer and the seller. The secondary market is regulated by the governing authority. In India, primary and secondary markets are controlled by the Security and Exchange Board of India (SEBI).

  Stock exchange presence has made the work for stock brokers simpler to trade shares in the company and other securities. An action can must only be bought or sold only if it is listed in the stock exchange. Hence, it has become the meeting point for buyers and

sellers of securities. The major leading stock exchanges present in India are the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE).

Fig 2.1 Stock Market

- **What is Share:** The stock market is certainly not a risk free investment like a government bond, it is certainly not a gambling game, nor is it a get rich quick scheme. For anybody to succeed in the stock market there is one and only one key attribute that separates a success from a failure – PATIENCE!

  Most day traders claim to earn several hundred or thousands of dollars in a single day and propose to make you a millionaire following their only strategy. A day trader might have had some good days in his career, but in the long run, day trading is not a strategy that must be adopted by a retail investor as it is bound to fail. Not only will you lose your money, but also lose the faith to ever invest and attain your goals of generating a passive income.

  Our goal is to enhance our understanding of how the market works in general and how we can invest to generate a source of passive income. I do not encourage anybody to quit their day job and invest full time into the stock market, but use their savings to maintain another stream of income. In order to make the right choices, you need to take matters into your own hands as opposed to listening to countless analysts on television, deciding how you must spend your savings on his/her list of stocks. If Warren Buffett relied on analysts, he would not be the third richest man on this planet, or the greatest investor known to us. Doing your own research (DYOR) is the first step in this journey.

Fig 2.2 Share Terminologies

- **Capital Market:** Among all the investment options available, the capital market. It is considered the most challenging and the most rewarding. The capital market is a stock market (capital and debt), where companies (and the government) re-launch long-term funds of public investors, and in which investors can subsequently exchange between them in these values.

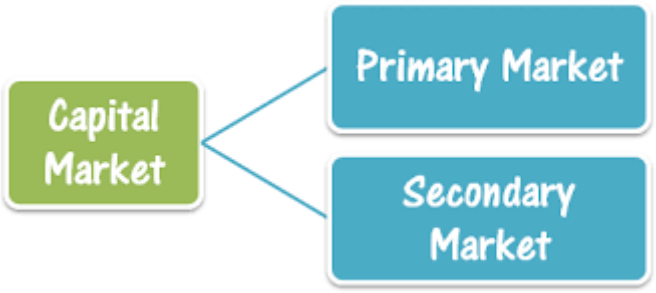


Fig 2.3 Capital Market

- **Equity Shares:** Typically, personal savings of an entrepreneur and if required then the contributions of friends / relatives are the source of funding to start new business For a big project, however, because the fund's requirements are great, these Not only will it require term loans, but it will go further. Therefore, availability of capital

  It is an important voice to establish or expand large-scale enterprises. There's a way increase equity beyond oneself or a limited group of a small circle of friends and relatives. This is collecting money from the public across the country through

  sell company shares.

Fig 2.4 Equity Shares

- **National Stock Exchange:** The National Stock Exchange of India (NSE) is the one of the main stock exchange of India, situated in Mumbai. The NSE was founded in 1992. NSE was the first exchange in the country to provide a modern, fully automated, screen-based e-commerce system which offered investors simple and widespread trading ease breadth of the country.

Fig 2.5 National Stock Exchange

- **Bombay Stock Exchange:** Bombay Stock Exchange (BSE) is Second major Indian stock exchange located in Dalal Street, Kala Ghoda, Mumbai (formerly Bombay), Maharashtra, India. Became Recognized in 1875, BSE is the Asia's first stock exchange, dues to be the fastest stock in the world exchange, with an average exchange rate of 6 microseconds, BSE is the 11th in the world .It is the one of the Largest stock

with a market capitalization of around $ 2 billion starting from July 2017.Around 5500 companies are listed on the stock exchange in the BSE.

Fig 2.6 Bombay Stock Exchange

Chapter 3

MATERIALS AND METHODS

3.1 Python

Python is a robust and extensively used programming language. It is a high level programming language which is often used as a scripting language. Python can be used for various applications and is ranked in top 5-10 worldwide in terms of popularity.

Python is broadly used in the data security industry, and has been utilized in a number of business programming items, including 3D movement bundles, for example,

Maya and Blender, and 2D imaging programs like GIMP and Inkscape.

Python is a PC programming dialect intended for comprehensibility and usefulness and is powerfully composed essentially implies that the translator or on the other hand compiler will make sense of for you what compose a variable is at run-time, so you don't need to pronounce variable composes yourself.

Fig 3.1 Python

3.1.1. Object Oriented Programming:

Class: An item is an instantiation of a class. It contains genuine qualities rather than variables. The class is one of the characterizing thoughts of OOPs. It is logical entity and incorporates fields and methods.

- Object: An object is a real world entity. It can be a combination of function, variables. The utilization and implementation of objects makes the working of software program easy.

- Abstraction: It is used just to hiding the backend information of **how** and displaying the **what** is called abstraction.
- Encapsulation: This binds the data into a single unit. Data encapsulation prompted the essential OOP idea of data hiding.
- Inheritance: In this, the classes in programming can inherit their parent classes. This reduces the length of the code and makes the programs easy.
- Polymorphism: one method can perform a number of tasks "method overloading" is another advantage.

3.1.2. Anaconda: Anaconda is a free and open source conveyance of the Python and R programming dialects for information science and machine learning related applications (substantial scale information handling, prescient investigation, logical registering), that means to disentangle bundle administration and organization. Bundle variants are overseen by the bundle administration framework conda. The Anaconda appropriation is utilized by more than 6 million clients, and it incorporates in excess of 250 prevalent information science bundles reasonable for Windows, Linux, and MacOS.

With more than 6 million clients, the open source Anaconda Distribution is the quickest and most demanding approach to do Python and R information science and machine learning on Linux, Windows, and Mac OS X. It's the business standard for creating, testing, and preparing on a solitary machine.

Anaconda Enterprise is an AI/ML enablement stage that engages associations to create, oversee, and computerize AI/ML and information science from PC through preparing to generation. It gives associations a chance to scale from individual information researchers to community oriented groups of thousands, and to go from a solitary server to a huge number of hubs for model preparing and arrangement.

3.2 Machine Learning

Machine Learning is a Concept that is based on an Application of Artificial Intelligence (AI) that offers the systems with the ability to automatically learn and improve from experience without being programmed by the programmers. Machine Learning's main objective is on the development of building computer programs that are able to access data and can use them for learning by themselves.

Machine learning can also be defined as the connection between theoretically sound computer science and practically noisy data. Essentially, it's all about making the machines able to make sense out of data in much the same way that we humans do.

The primary objective of Machine Learning is to design models and using that model for building the appropriate Algorithms that can get the input data from the users and statistical analysis can be applied to predict an output value between an acceptable range.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

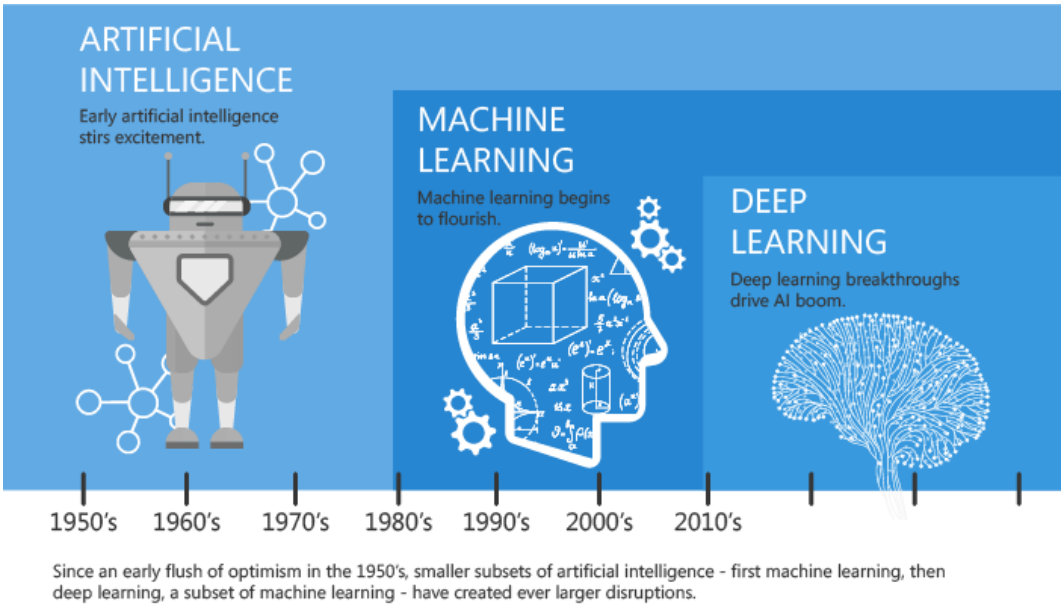


Fig 3.3 Artificial Intelligance VS Machine Learning Vs Deep Learning

Some Types of Machine Learning are as follows:

Machine Learning Algorithms can be divided into Supervised Machine Learning, Unsupervised Machine Learning and Reinforcement Learning.

Supervised Machine Learning Algorithms works on the concept of Historical Data and using them for Future Predictions. Supervised Machine Learning can be applied on that what has been learned by the computer from the past data to new data and using the labeled examples to predict future data. Beginning with the analysis of the known training dataset, the learning algorithm produces a conditional function called Model to make predictions about the output values.

Supervised algorithms are algorithms in which humans need to give the dataset that consists of the both input and desired output, in order to for supplying feedback about the accuracy of predictions done by the system during training. Once training is completed, then the algorithm is ready to be applied on was learned to new data.

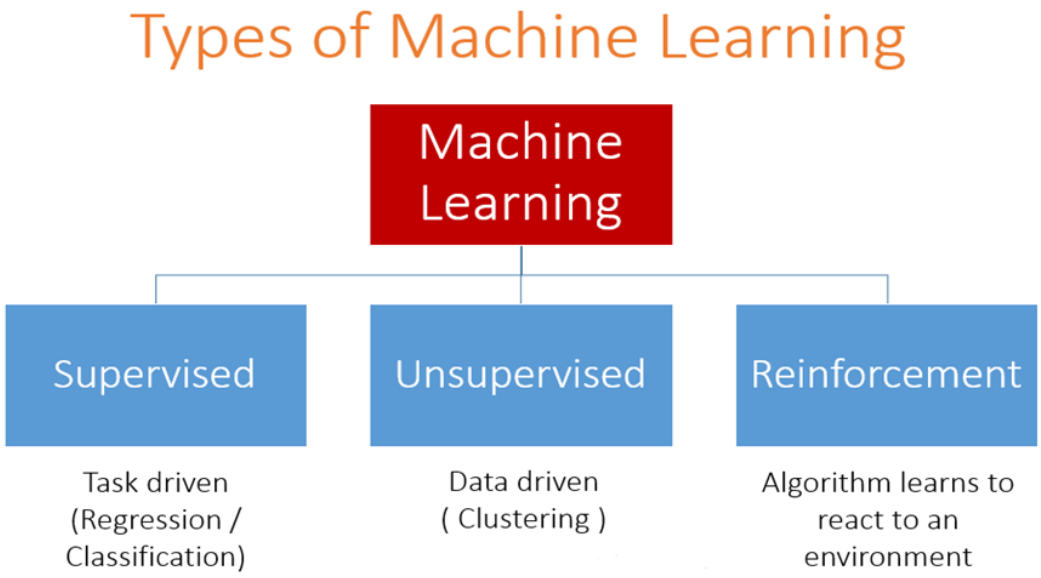


Fig 3.4 Types of Machine Learning

Some of the Supervised Machine Learning Algorithms are:

1. Linear Regression
2. Decision Tree
3. Logistic Regression
4. K- Nearest Neighbor

Now lets see the another type of learning that is, Unsupervised Machine Learning which is based on concept of learning by own that is computer is provided with only the inputs without the outputs which means computer learns by themselves. Unsupervised is used when the information that is used to train the model is neither the classified nor the labeled. Unsupervised learning studies how systems can surmise a function to describe the hidden structure derived from the unlabeled data.

Unsupervised Algorithms are based on the concept that they don't require to be trained with desired outcome data. In its place, they use an iterative approach called Deep Learning to review data and arrive at a conclusion. Unsupervised Learning Algorithms are also capable for used for more complex processing of large data than Supervised Learning Algorithms.

Some of the Unsupervised Machine Learning Algorithms are:

1. K-Means Clustering

Now at last but the least let discuss the concept of reinforcement learning.
Reinforcement Machine Learning Algorithms is a learning method that is related with the environment around it and then generating actions and learns from the punishments or rewards that they get. Rewards and punishments that are received the system are the most significant features of Reinforcement Learning. This concept of learning helps in allowing the machines and software agents to determine by themselves the ideal behavior surrounded by a specific context so that they can maximize their performance.
Here are the some Machine Learning Applications in Real World .

Machine Learning in Our Everyday Life

- **Google Maps:** These are able to keep check on the speed of drive of traffic at any given time instant, Maps can more easily to be incorporated as the user-reported traffic incidents such as road under construction and accidents. They also have the access to the large volumes of the data that is being given to its named algorithms. Maps can reduce travels going by advising the users with the fastest routes that is possible to and back from work.

- **Rides hiring Apps Like Ola and Uber:** These apps also uses the Machine Learning to predict the price of your ride, waiting time for your booked cab.

- **Banking/Personal Finance**: Machine Learning can be also be applied in Fraud Prevention and Credit Decisions making in Banking Sectors.

- **Online Shopping:** Online Shopping uses Machine Learning in a way that you can see the recommendations of the products that you have viewed and interested in you as the customers have viewed that item and you have bought this item, as well as via modified recommendations on the home page, bottom of item pages, and through email. Companies like Amazon uses the technique of the Artificial Neural Networks to make these product recommendations.

- **Smart Personal Assistants:** Google Assistant, Alexa and Microsoft Cortana are various personal assistants that also works on the concept of Machine Learning.

Machine learning is being able to utilized in an extensive variety of utilizations in today. A standout amongst the most surely understood precedents is Facebook's News Feed. The News Feed utilizes machine figuring out how to customize every part's channel. On the off chance that a part much of the time quits looking to peruse or like a specific companion's posts, the News Feed will begin to indicate a greater amount of that companion's action prior in the channel. In the background, the product is just utilizing factual investigation and prescient examination to distinguish designs in the client's information and utilize those examples to populate the News Feed. Should the part never again stop to peruse, as or remark on the companion's posts, that new information will be incorporated into the informational collection and the News Feed will alter in like manner.

Machine learning is likewise entering a variety of big business applications. Client relationship administration (CRM) frameworks utilize learning models to examine email and incite deals colleagues to react to the most essential messages first. Further developed frameworks can even suggest possibly compelling reactions. Business insight (BI) and investigation merchants utilize machine learning in their product to help clients naturally recognize possibly imperative information focuses. Human asset (HR) frameworks utilize learning models to recognize attributes of viable workers and depend on this information to locate the best candidates for open positions.

Machine adapting likewise assumes a vital job in self-driving autos. Profound learning neural systems are utilized to recognize protests and decide ideal activities for securely guiding a vehicle not far off.

1. Linear Regression

In Linear Regression, we are given with a set of input data (x) which are used to predict the output data (y). There exists a relationship between the input data and the output data. The goal of Linear Regression Algorithm of Machine Learning is to calculate this relationship.

In Linear Regression, the existed relationship between the input data (x) and output data (y) can be defined as a mathematical equation of type $y = a + bx$. Thus, the main objective of linear regression is to propose the values of coefficients a and b. Here, a represents the intercept and b represents the slope of the line.

The primary goal of Linear Regression is to design a line that is closest to most of the data in given dataset. This Algorithm helps in reducing the distance also known as error that occurs between the y value of a data point and the line.

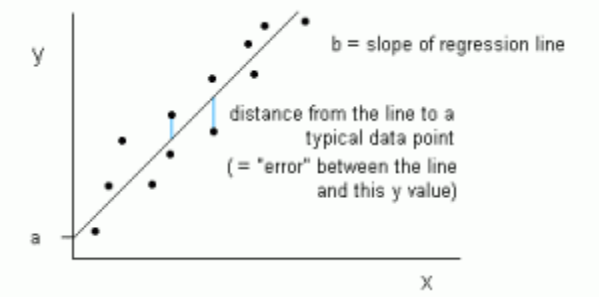


Fig 3.5 Linear Regression

2. Logistic Regression

Linear regression as above discuss is able to predict the continuous values for example: rainfall in cm, Logistic Regression is used for making predictions having the values are discrete values for example: whether a student passed/failed, after applying a Regression function.

The ultimate aim of Logistic Regression Algorithm is to use the given training dataset to find the best values for the coefficients b0 and b1 in order that it will be able to minimize the error between the predicted outcome and the actual outcome. These coefficients are calculated by using the method of Maximum Likelihood Estimation.

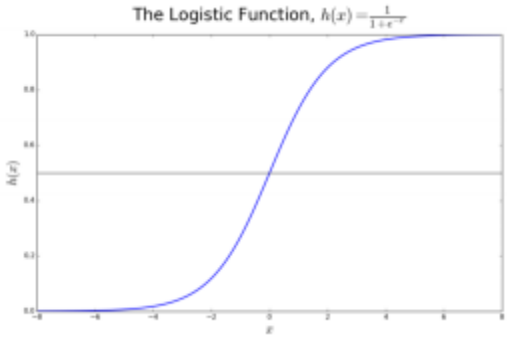


Fig 3.6 Logistic Regression

3. CART

Classification and Regression Trees (CART) is an application of Decision Trees, among others such as ID3, C4.5.

The decision tree in Figure define classifies whether the person will buy a sports car or a minivan which is dependent on their age and their marital status. If the age of person is over 30 years and his martial status is not married, we go through the tree as followed : ‘over 30 years?’ -> yes -> ’married?’ -> no. Therefore, the model gives an outputs of sportscar.

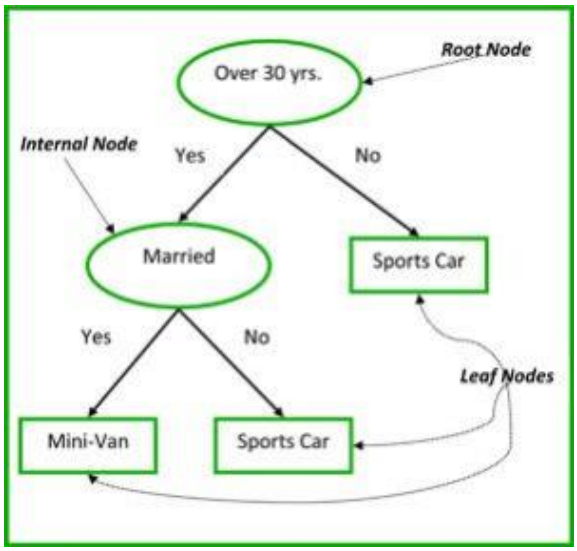


Fig 3.7 Classification and Regression Trees

4. KNN

KNN stands for K-Nearest Neighbour .The k-nearest neighbours algorithm is based upon using the entire given dataset as the training set, without dividing the given dataset into training set and test set.

When the prediction is made for a new data instance, the KNN algorithm moves through the whole dataset to find the k-nearest instances to the new instance.

5. K-means

K-means is Algorithm based on the iterative algorithm that based on the concept of grouping the similar data together in clusters. Then it calculates the centroids of k clusters and allocates a data point for that cluster having minimum distance between its centroid and the data point.

Step 1: k-means initialization:

Step 2: Linking each observation to a cluster:

Step 3: Recalculating the centroids:

Step 4: Repeat, then exit if unchanged

Chapter 4

RESULTS AND DISCUSSION

4.1 Screenshots

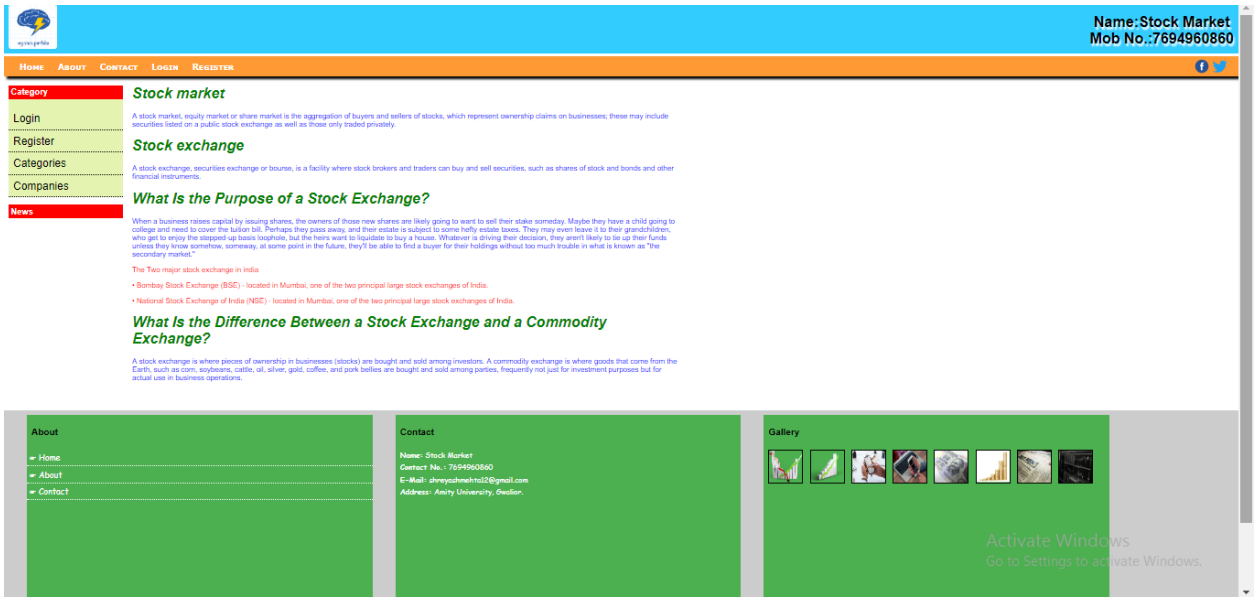


Fig 4.1 Screenshot for Website Interface.

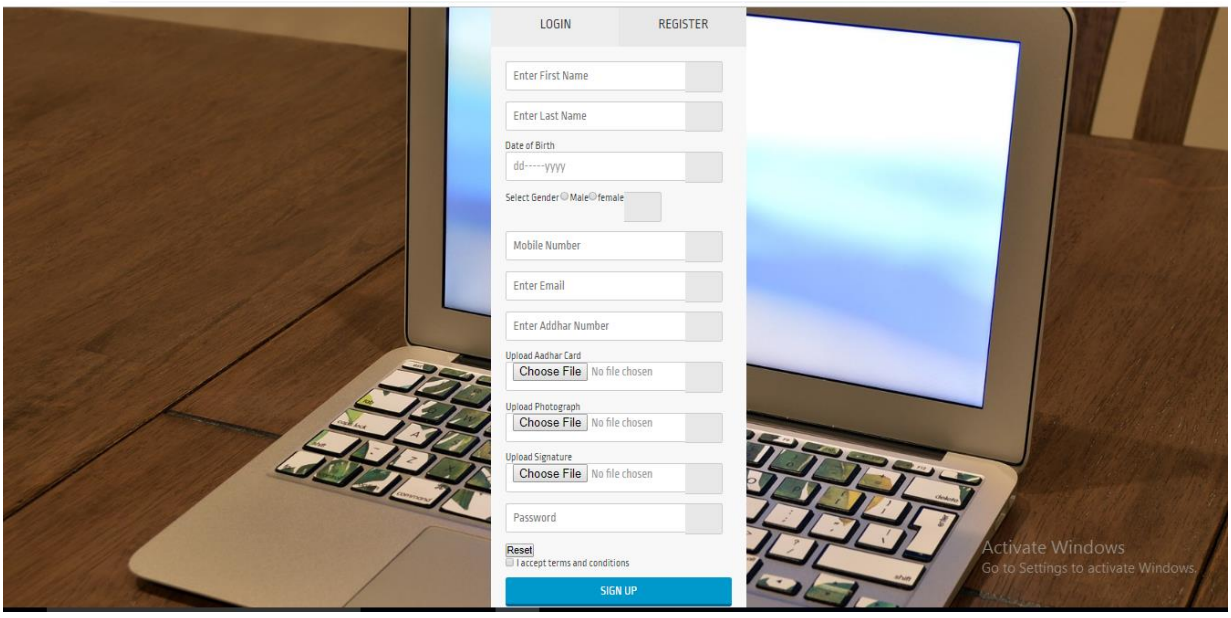


Fig 4.2 Screenshot for Registration.

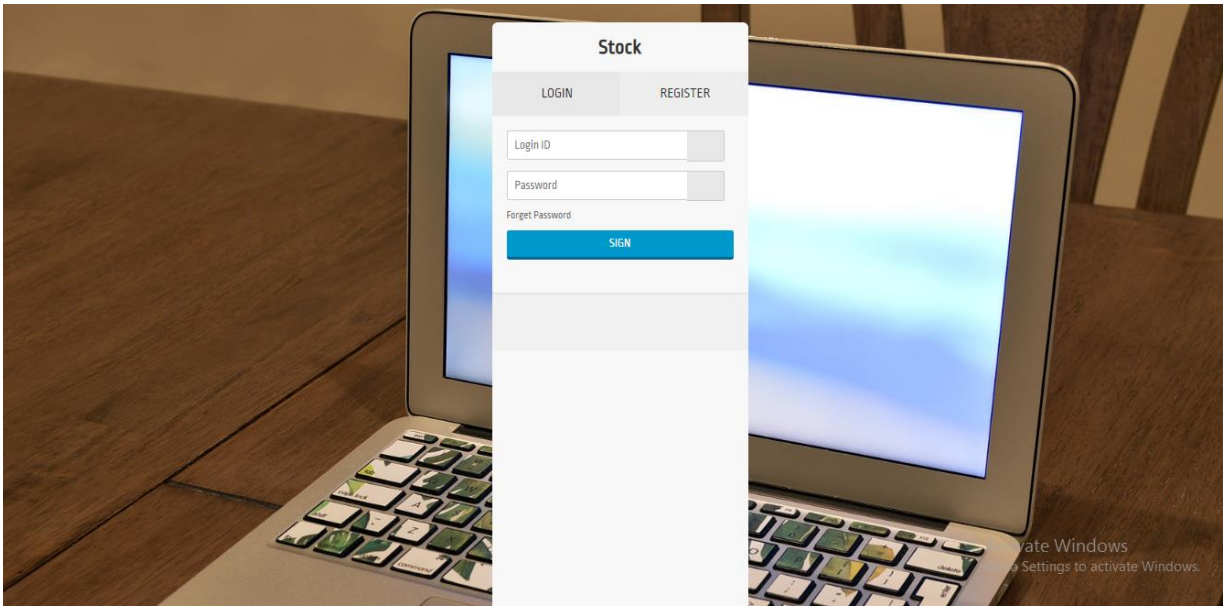


Fig 4.3 Screenshot for Login.

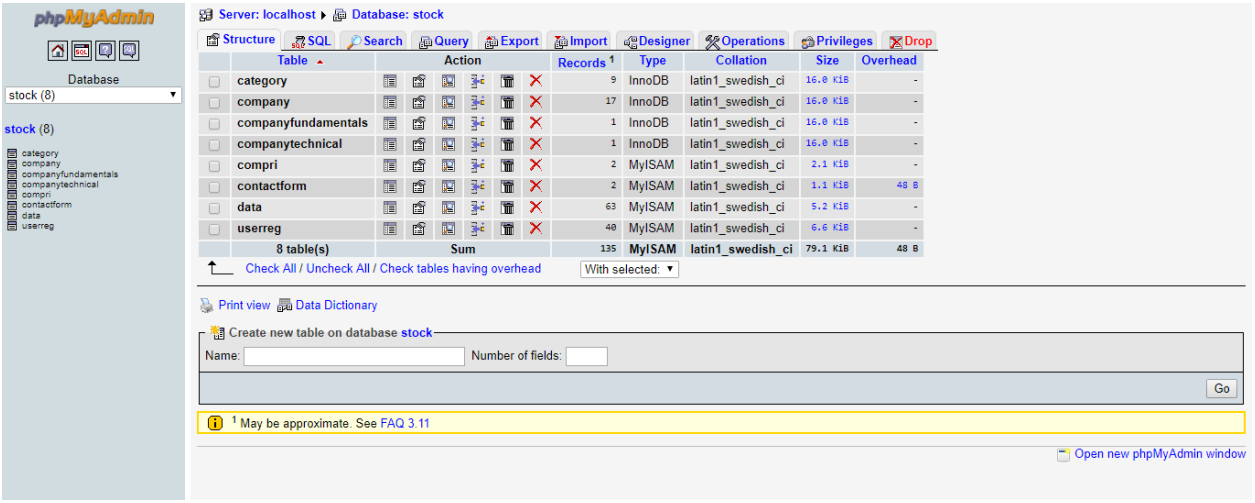


Fig 4.4 Screenshot for Database.

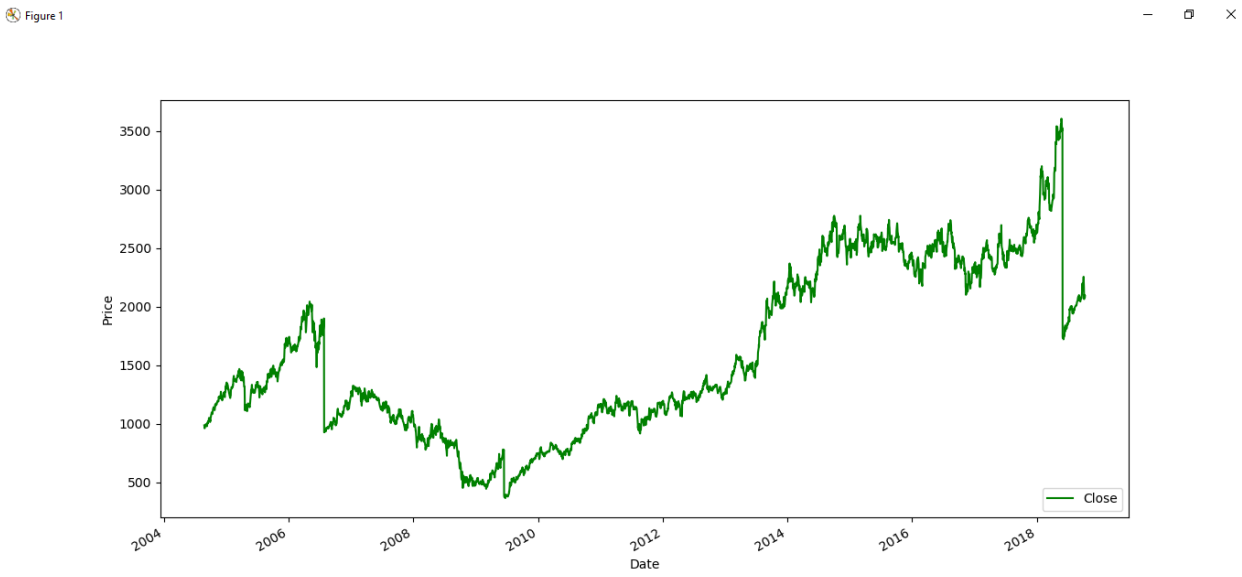


Fig 4.5 Graph of Historical Data

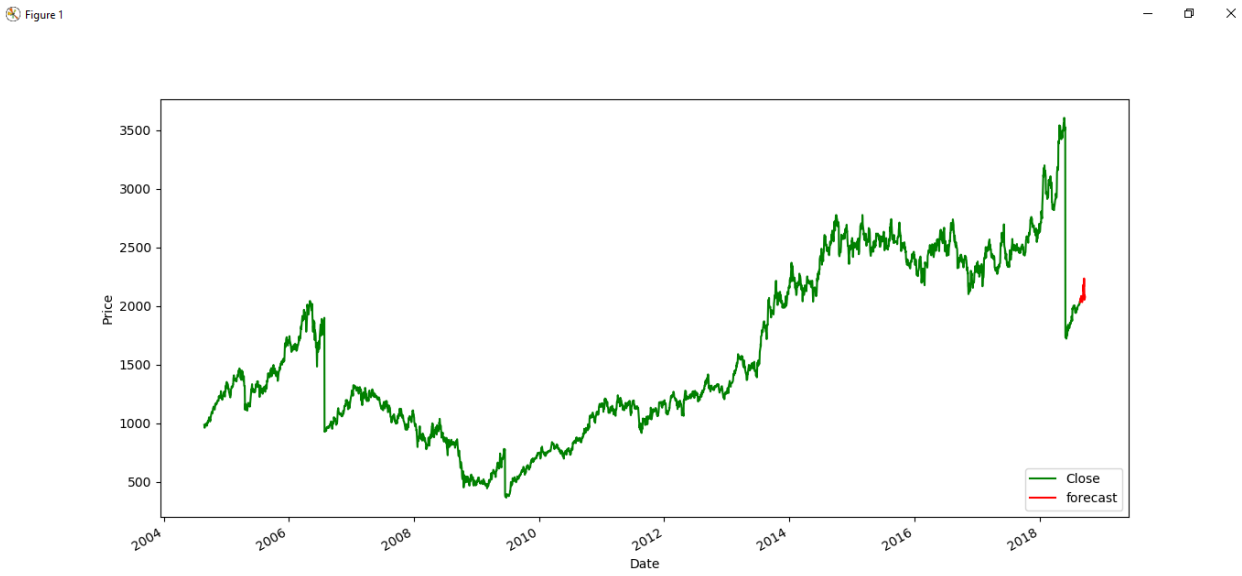


Fig 4.6 Graph of Predicated Data.

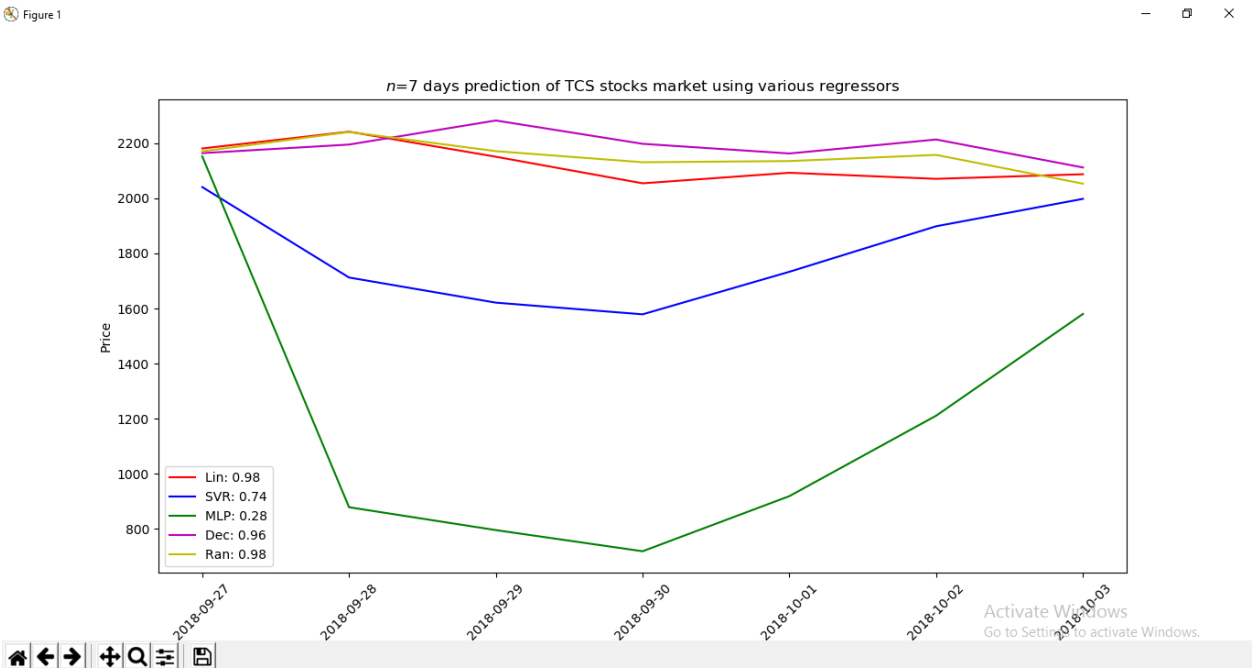


Fig 4.7 Comparison of Different Algorithm.

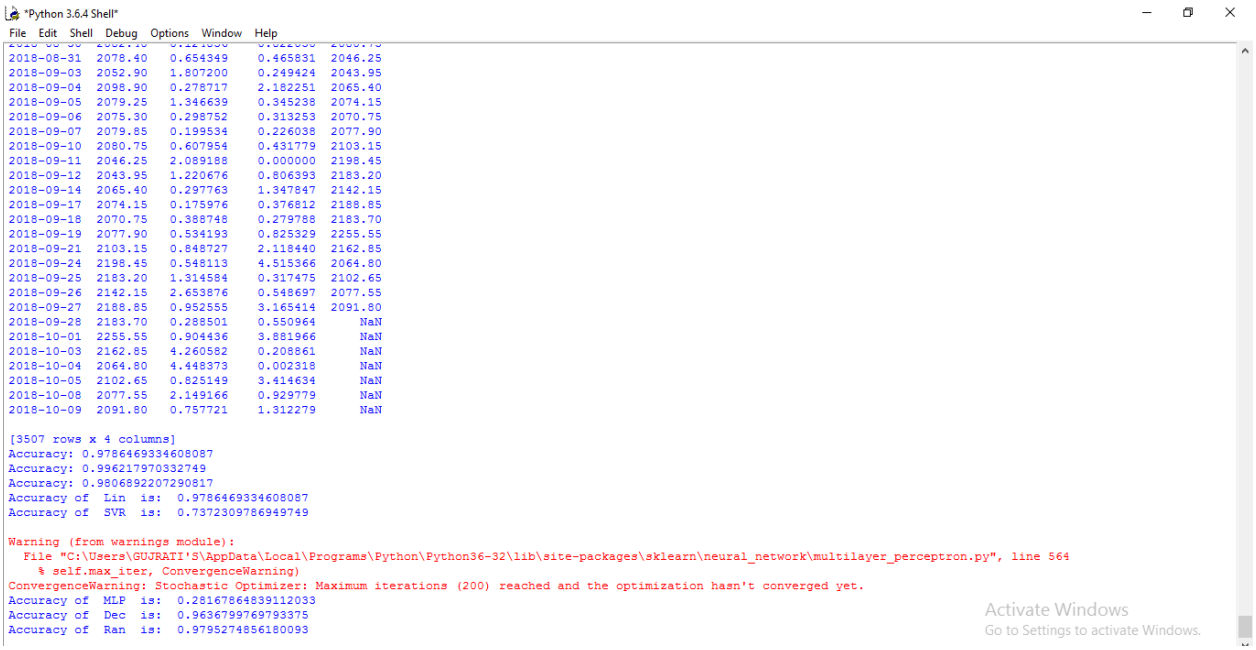


Fig 4.8 Accuracy of Different Algorithm.

4.2 Libraries Required:

- Sklearn.
- Scipy

- Pandas
- Numpy
- Matplotlib
- Quandl

4.3 Software requirement:

- Operating system-windows 7,8,10,
- Anaconda
- Jupyter Notebook
- Python 3.6
- Php MyAdmin

Chapter 5

CONCLUSION AND FUTURE PROSPECTS

The future scope of this system is that it can be extend easily when it will require. This system reduces the time of calculating and helps to manage the data in a proper way. It also saves the user`s time. This system has user friendly features and it is easy to understand. Through this system administrator can easily find whole details of the users easily at one place. There is flexibility in this project which makes it more friendly and attractive. This system fulfills all requirements of the users in a best possible way.

In future more features can be added in this project. Deep Learning and TensorFlow can be applied to the Historical data of the company to predict the prices more precisely and accurately.

GUI can be made more attractive using Django and other tools. Predicted values graph can be more interactive using the advanced MatPlotLib which makes it more interactive and user-friendly.

Chapter 6

SUMMARY

In this project, we made AI based application on Machine Learning to overcoming the existing problem. In this project, Python 3.6, PhpMyAdmin, and Jupyter Notebook software are used. In this project admin can register all the companies and upload all the details of the company. User has to register themselves to see the historical data of the company they supposed to find and see the predicted price of that company. This is basically the application which is used in the Stock Market field to predict the future prices of the companies beforehand.

Chapter 7

BIBLIOGRAPHY

1. Aurelien Geron, “Hands on Machine Learning with Scikit-Learn and Tensor Flow (2017)”
2. Christopher Bishop, “Pattern Learning and Machine learning (2006)”
3. Drew Conway, “Machine Learning for Hackers (2009)”
4. Jerome H Friedman, “The Elements of Statistical Learning (2001)”
5. Kevin Murphy, “Machine Learning (2012)”
6. Luis Pedro, “Building Machine Learning Systems with Python (2012)”
7. Sebastian Raschka, “Python Machine Learning (2009)”
8. Shai Ben-David, “Understanding Machine Learning (2014)”
9. Stephen Marsland, “Machine Learning: Algorithm Perspective (2009)”
10. Toby Segaran, “Programming Collective Intelligence (2007)”