

SHARE MARKET ANALYSIS & PREDICTION USING MACHINE LEARNING

A report submitted in partial fulfillment of the requirements for

The award of the degree of

B. Tech.

In

Computer Science & Engineering



Submitted by:

Akshat Gupta (160102135)

Abhishek Sharma (160102143)

Aayush Raj (160102140)

Rajat Jaiswal (160102103)

Under the Guidance of:

Dr. Amit Kumar Mishra

Associate Professor

Department of CSE

**COMPUTER SCIENCE AND ENGINEERING
DEPARTMENT**

DIT UNIVERSITY, DEHRADUN

CERTIFICATE

This is to certify that the project report entitled “SHARE MARKET ANALYSIS & PREDICTION USING MACHINE LEARNING” being submitted by “**AKSHAT GUPTA, ABHISHEK SHARMA, AAYUSH RAJ & RAJAT JAISWAL**” in partial fulfillment for the award of the Degree of Bachelor of Technology in Computer Science and Engineering to the Dehradun Institute of Technology (An Autonomous Institution of Uttarakhand Technical University) is a record of bona fide work carried out by them under my guidance and supervision.

The results embodied in this project report have not been submitted to any other University or Institute for the award of any Degree or Diploma.

Dr. Amit Kumar Mishra

Associate Professor

Computer Science & Engineering

Dr. Vishal Bharti

Head of Department

Computer Science & Engineering

CANDIDATES/DECLARATION

We hereby certify that the work, which is being presented in the report/ project report, entitled “Share market analysis & prediction using Machine learning”, in partial fulfillment of the requirement for the award of the Degree of **Bachelor of Technology** and submitted to the institution is an authentic record of our own work carried out during the period *February-2019* to *April-2019* under the supervision of Dr. Amit Kumar Mishra.

Date:

Signature of the Candidate(s)

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.

Date:

Signature(s) of the Supervisor(s)

ACKNOWLEDGEMENT

We have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly indebted to Dr. Amit Kumar Mishra for his guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

I would like to express my gratitude towards my parents for their kind co-operation and encouragement which help me in completion of this project.

I would like to express my special gratitude and thanks to the institute persons for giving us such attention and time. Our thanks and appreciations also go to our colleagues in developing the project and people who have willingly helped us out with their abilities.

ABSTRACT

Stock Market Analysis and Prediction is the project on technical analysis, visualization and prediction using data of few S&P 500 Companies..The successful prediction of a stock's future price could yield significant profit. In this report we analyze existing and new methods of stock market prediction. We take three different approaches at the problem: Fundamental analysis, Technical Analysis, and the application of Machine Learning. We find evidence in support of the weak form of the Efficient Market Hypothesis, that the historic price does not contain useful information but out of sample data may be predictive. We show that Fundamental Analysis and Machine Learning could be used to guide an investor's decisions. We demonstrate a common flaw in Technical Analysis methodology and show that it produces limited useful information. Based on our findings, algorithmic trading programs are developed and simulated using Jupyter Notebook .

Another part of this project deals with the prediction of future stock price using three different Machine Learning models which can be useful for the investors to check how the stock will perform in the future. Accuracy and efficiency are two main measures that tells us about the quality of the model.

The project is created from scratch by using the basic libraries.

The features of the project include:

- i) Moving Averages of Stock.
- ii) Daily Returns of Stock.
- iii) Comparison of Stocks of different companies.
- iv) Stock Price Prediction.

TABLE OF CONTENTS

Title	Page No.
CHAPTER 1 -INTRODUCTION	
1.1 Purpose.....	1
1.2 Objective.....	1
1.3 Motivation.....	2
1.4 Definition and Overview.....	2
CHAPTER 2- ANALYSIS	
2.1 Data Exploration.....	3
2.2 Data Visualization.....	4
2.3 Algorithms and Techniques.....	5
CHAPTER3 –METHODOLOGY	
3.1 Data Preprocessing.....	9
3.2 Implementation.....	9
3.3 Results.....	10
CHAPTER4 -SYSTEM REQUIREMENTS	
3.1 External Interface Requirement.....	12
3.1.1 Hardware Requirement.....	12
3.1.2 Software Requirement.....	13
CHAPTER 4 CONCLUSION AND FUTURE WORK	
4.1 Conclusions.....	14
4.2 Scope for Future Work.....	14
REFERENCES.....	15

LIST OF FIGURES

1. Dataset Visualization
 - 1.1 Closing Price Vs Trading Days
 - 1.2 Volume Traded Vs Trading Days
2. Technical Analysis
 - 2.1 Moving Averages of Stock
 - 2.2 Daily Returns of Stock
 - 2.3 Daily Return Comparison of different companies
3. Implementation Model
4. Predicted values Vs Actual Values
 - 4.1 Linear Regression Model
 - 4.2 Support Vector Regression Model
 - 4.3 Decision Tree Regression Model

LIST OF TABLES

1. Stock Data
 - 1.1 Dataset Values
 - 1.2 Dataset Description(mean, std, min, max)

CHAPTER 1

INTRODUCTION

1.1 PURPOSE

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

1.2 OBJECTIVE

Despite its prevalence, Stock Market prediction remains a secretive and empirical art. Few people, if any, are willing to share what successful strategies they have. A chief goal of this project is to add to the academic understanding of stock market prediction. The hope is that with a greater understanding of how the market moves, investors will be better equipped to prevent another financial crisis. The project will evaluate some existing strategies from a rigorous scientific perspective and provide a quantitative evaluation of new strategies. It is important here to define the scope of the project. Although vital to any investor operating in the real world, no attempt is made in this project at portfolio management. Portfolio management is largely an extra step done after an investor has made a prediction on which direction any particular stock will move. The investor may choose to allocate funds across a range of stocks in such a way to minimize his or her risk.

Major objective of making this project are listed below:

- To predict the future price of stocks.
- To help investors use this predicted value to invest in different companies.

1.3 MOTIVATION

Since the advent of stock markets, people have been trying to multiply their profits by investing in trusted ventures and tried to find strong patterns in share prices over time. But the prediction of the price

of a stock has always been a tough task at hand. This problem has been an interesting one as multiple consultancy firms and individuals have their own theories and methods.

This problem occurs because the fluctuations in the price of a stock is depended on many factors such as the overall market trends, economy, news about the company, financial issues or scandals which we have no control on and are completely unpredictable.

When machine learning came into picture, it changed the way people looked at this problem as people began to find patterns and used different machine learning models to predict future stock price using previous stock price data. One major improvement to this solution was by using neural networks.

As machine learning and artificial intelligence is constantly advancing, we can hope for much better and accurate methods in the future.

1.4 DEFINITION AND OVERVIEW

Machine learning is a subset of artificial intelligence in the field of computer science that often uses statistical techniques to give computers the ability to "learn" (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed.

The name machine learning was coined in 1959 by Arthur Samuel. Evolved from the study of pattern recognition and computational learning theory in artificial intelligence, machine learning explores the study and construction of algorithms that can learn from and make predictions on data such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or infeasible; example applications include email filtering, detection of network intruders or malicious insiders working towards a data breach, optical character recognition (OCR), learning to rank, and computer vision.

CHAPTER 2

ANALYSIS

1. Data Exploration

The data used in this project is of the **Apple Inc** from **08-02-2013 to 07-02-2018**, this is a series of data points indexed in time order or a time series. Our goal was to analyze the data using various technical analysis methods like moving averages, daily return and predict the closing price for any given date after training. For ease of reproducibility and reusability, all data was taken from S&P 500 stock data from Kaggle.

The prediction has to be made for Closing (Adjusted closing) price of the data. Since Google Finance already **adjusts the closing prices for us** , we just need to make prediction for “CLOSE” price.

The dataset is of following form :

Date	Open	High	Low	Close	Volume
08-02-13	67.7142	68.4014	66.8928	67.8542	158168416
11-02-13	68.0714	69.2771	67.6071	68.5164	129029425
12-02-13	68.5014	68.9114	66.8205	66.8428	151829363

Table1.1: The whole data can be found out in 'AAPL_data.csv' in the project root folder

Note- On observation of the dataset, there was not any abnormality, missing values or negative values in the dataset.

The mean, standard deviation, maximum and minimum of the data was found to be following:

Feature	Open	High	Low	Close	Volume
Mean	129.625	130.595	128.704	129.713	3.302348e+07
Std	26.9703	27.0413	26.8127	26.9313	1.568011e+07
Max	179.37	180.10	178.25	179.26	1.333697e+08
Min	90.00	91.67	89.47	90.34	1.147592e+07

Table1.2: Description of the Apple Inc. Dataset

2. Data Visualization

For visualization, I have used Matplotlib and Seaborn libraries of python. We have plotted the closing price of the Apple Inc stocks from 04-01-2016 to 07-02-2018.

Following is the graph of Closing Prices of the stock Vs Trading Days:

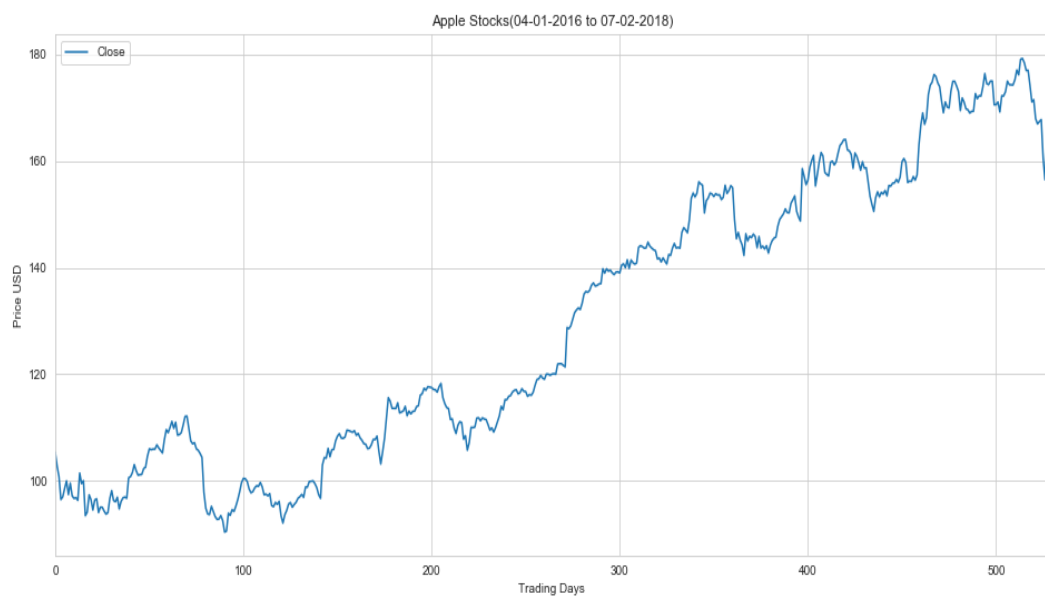


Fig 1.1: Closing Prices Vs Trading Days

The next graph is the representation of the volumes of stock trades over the time.

Following is the graph of Volume Traded Vs Trading Days:-

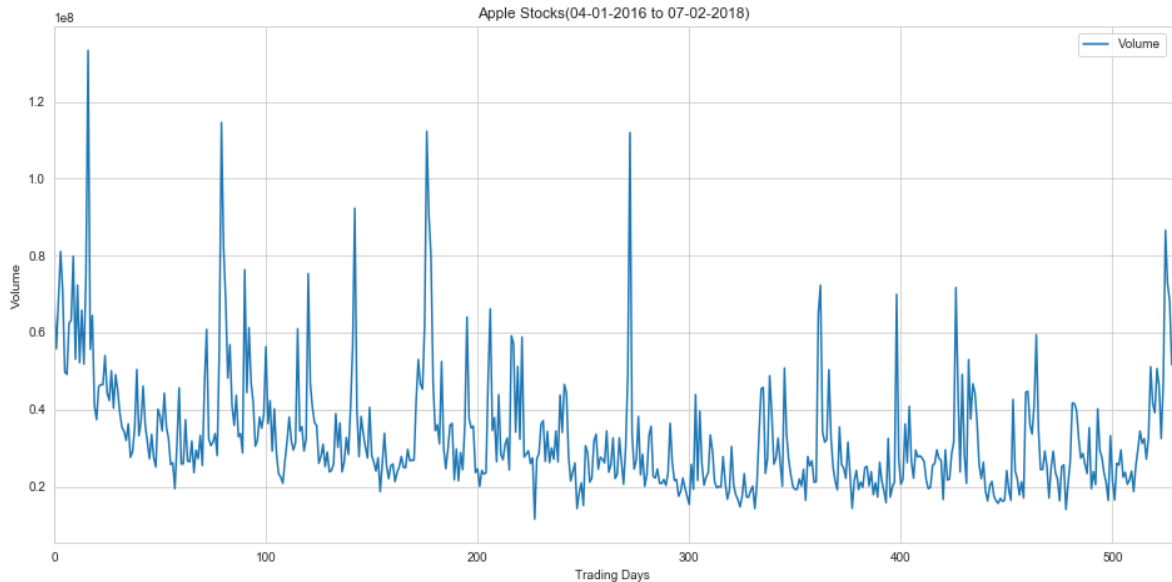


Fig 1.2: Volume Traded Vs Trading Days

3. Algorithms and Techniques

Moving Averages of Stock

Moving Average(MA) is a simple technical analysis tool which is generally used with time-series data in order to smooth out short-term fluctuations and highlight long-term trends or cycles. The average is taken over a specific period of time like, 10 days, 30 –minutes, 20 weeks or any period of time the trader chooses. It can only be used for historical data it cannot be used for prediction.

There are various benefits of applying moving averages to our trading and it is also one of the most popular strategies which can be fitted to any time frame, suiting both long-term investors and short-term traders.

- Moving average helps in cutting down the amount of “noise” on the price chart.
- It can also act as a support or resistance (like, a ceiling the price hits the level and then starts to drop again).

A moving average can be measured or calculated in various ways:

1. Simple Moving Average(SMA)
2. Exponential Moving Average(EMA)

Following is the graph of Moving Averages of Stock:-

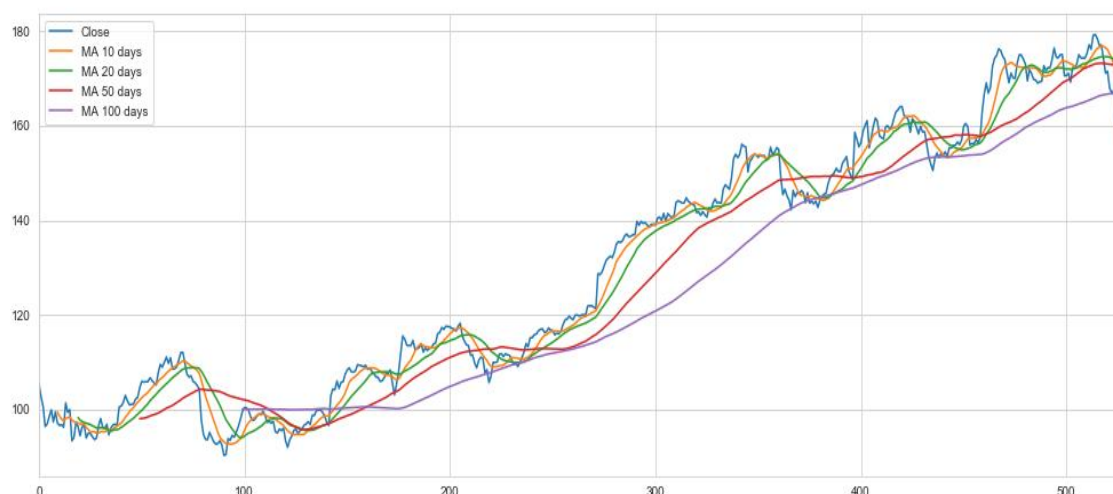


Fig 2.1: Moving Averages of Stock

Daily Return of Stock:

Stock Market provides us the facility to generate **Passive Income**(It is the money which we can earn by doing minimal activity through a variety of risk or undertaking which requires little bit of daily effort).

Once we invested the money in the stock market ,we get to share the profits or the losses of the company.

It is usually recommended to do the investment for a long term but, It can be a fun task to calculate the daily profits or not so much fun task to calculate the daily losses of the company after having a good or bad day in the market.

In order to calculate how much a company gained or loss per day of the stock firstly, we need to subtract opening price from the closing price and then, multiply the result by the number of the shares we own in the company.

Following is the graph of Daily Return of Stocks-

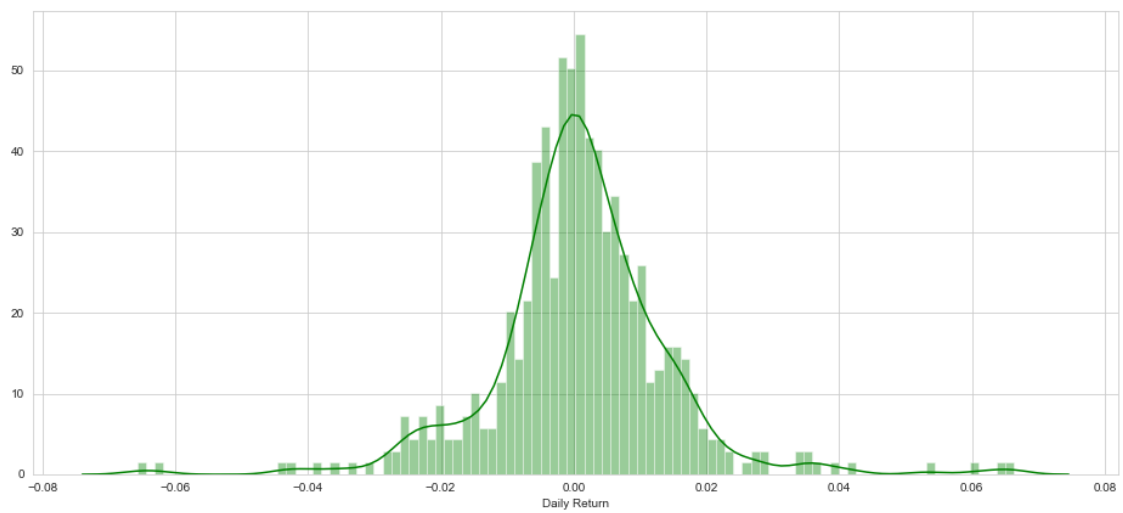


Fig 2.2: Daily Return of Stock

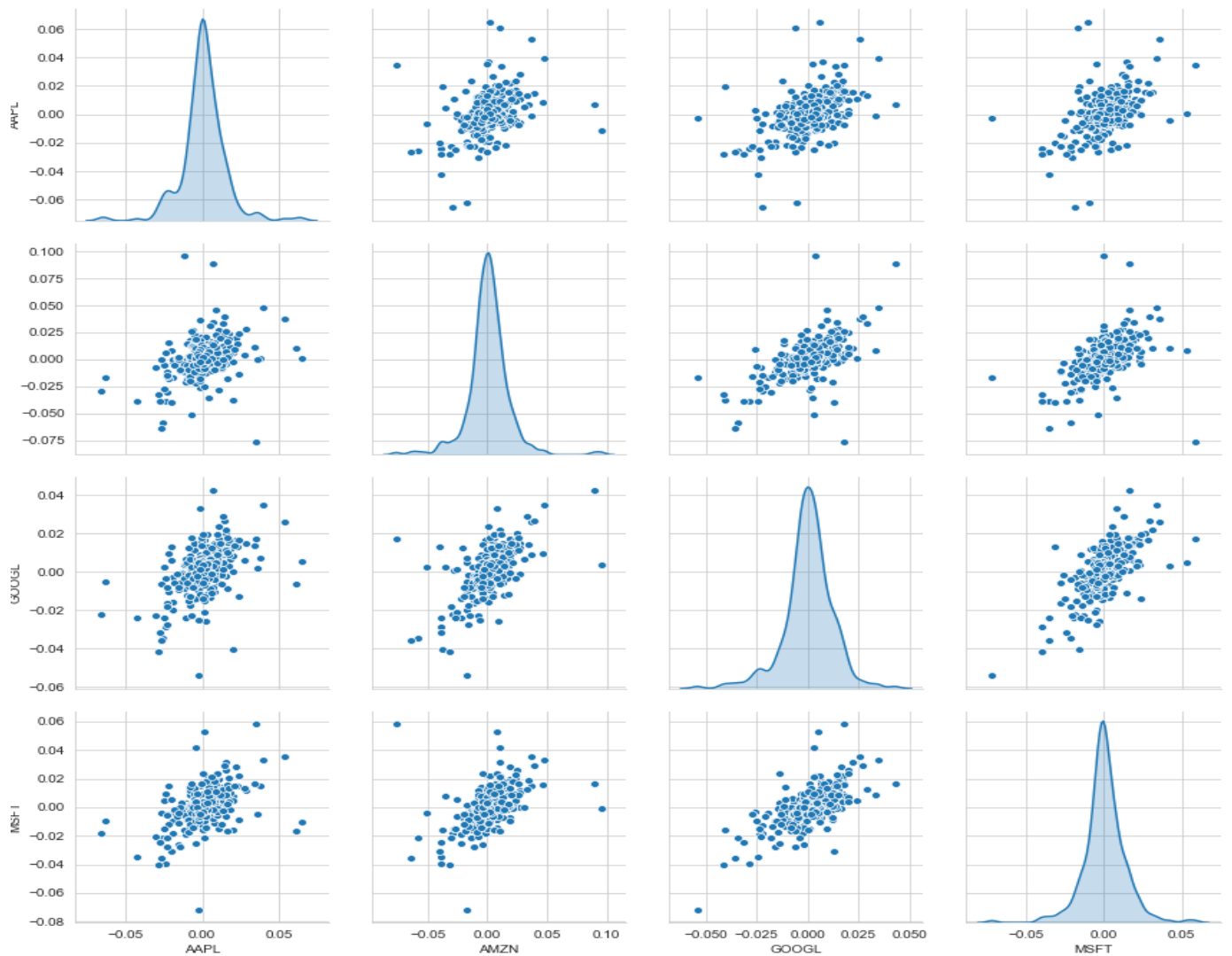


Fig 2.3: Daily Return of Stock of different companies

Linear Regression

Linear regression is one of the most well-known and well understood algorithms in statistics and machine learning. It is a basic and commonly used type of predictive analysis

Simple linear regression is a statistical method that allows us to summarize and study relationship between two continuous variables.

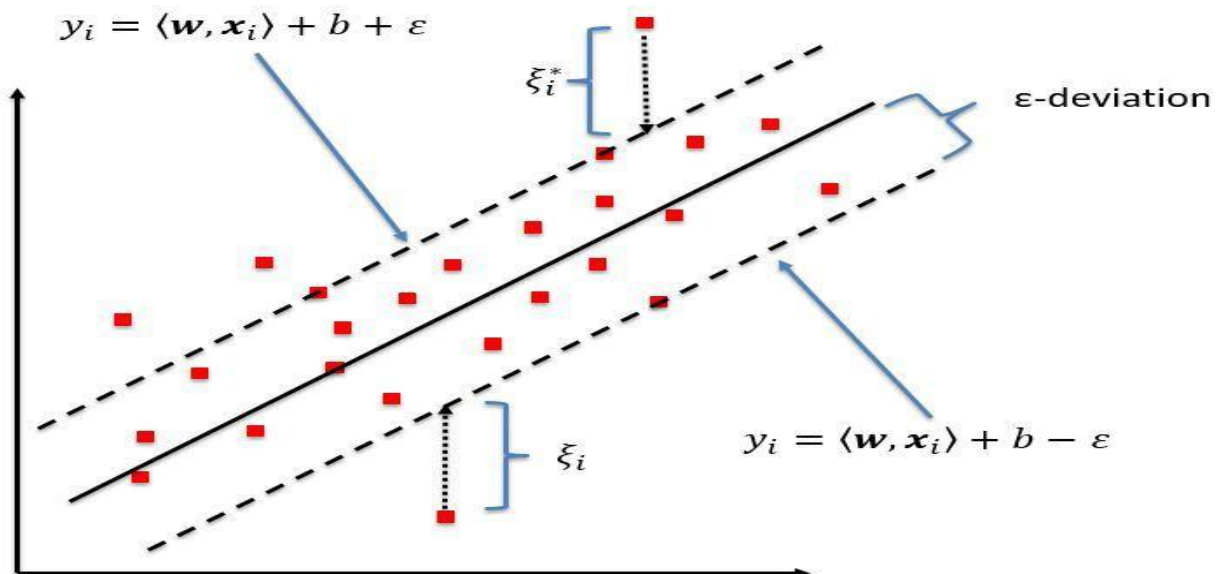
In simple linear regression, we predict scores on one variable from the scores on a second variable. The variable we are predicting is called the *criterion variable* and is referred to as Y. The variable we are basing our predictions on is called the *predictor variable* and is referred to as X. When there is only one predictor variable, the prediction method is called *simple regression*.

Support Vector Regression

This regression model is based on Support Vector Machine. SVM is not only used for classification problems but can also be used for Regression problems.

Our objective when we are moving on with SVR is to basically consider the points that are within the boundary line. Our best fit line is the line hyper plane that has maximum number of points.

1. **Hyper plane:** In SVM this is basically the separation line between the data classes. Although in SVR we are going to define it as the line that will help us predict the continuous value or target value
2. **Boundary line:** In SVM there are two lines other than Hyper Plane which creates a margin . The support vectors can be on the Boundary lines or outside it. This boundary line separates the two classes. In SVR the concept is same.
3. **Support vectors:** This are the data points which are closest to the boundary. The distance of the points is minimum or least.



CHAPTER 3

METHODOLOGY

Data Preprocessing

Reading Dataset

We used Pandas library to read the csv file.

Standardization

The idea behind Standard Scaler is that it will transform your data such that its distribution will have a mean value 0 and standard deviation of 1. Given the distribution of the data, each value in the dataset will have the sample mean value subtracted, and then divided by the standard deviation of the whole dataset.

Train Test Split

Here, we split our data into training and testing sets. We train our model on the training set data and the predictions are further done on the test set. We use Scikit Learns `train_test_split` for the same.

Implementation Model

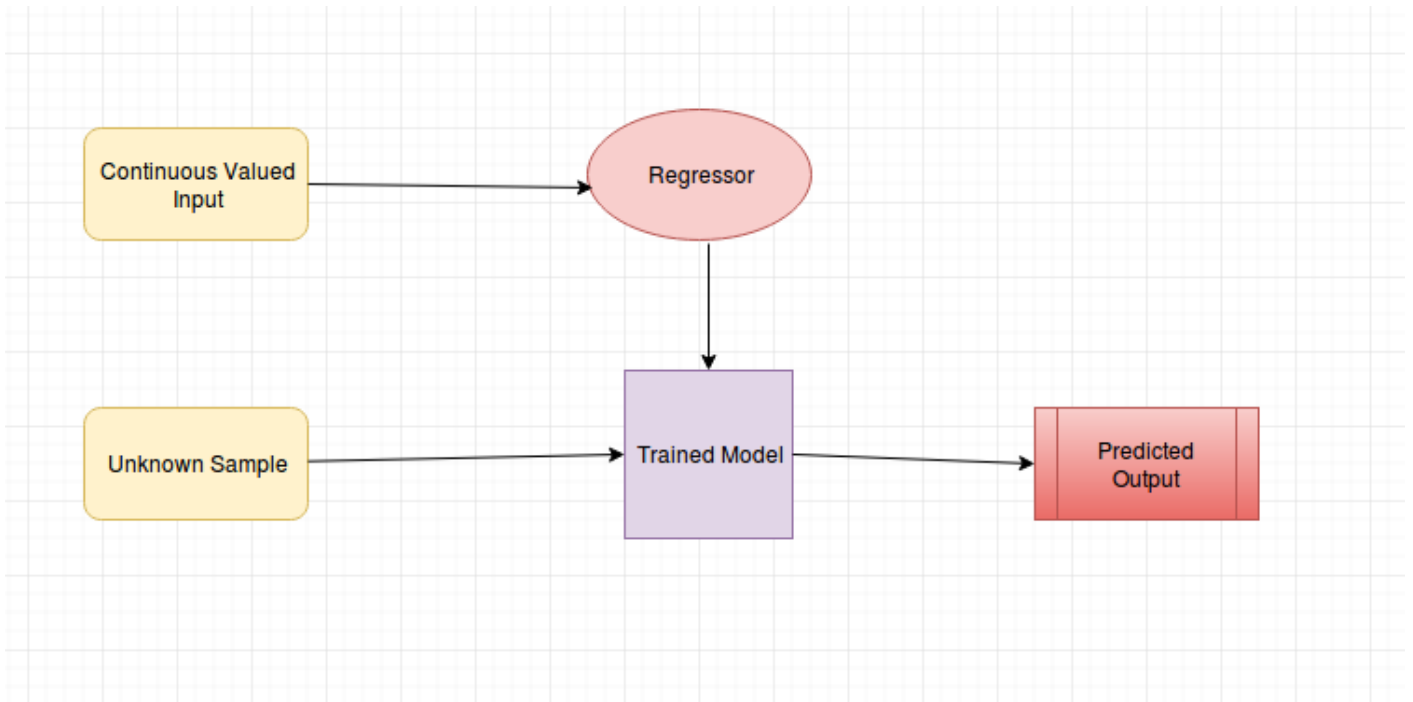


Fig 3:Implementation of Regression Models

Prediction Results

We predicted our stock prices using three of the following regression models:-

1. Linear Regression
2. Support Vector Regression
3. Decision Tree Regression

After the data preprocessing, we fit our training set in the above models and then predict the values on the test set. Then we plotted the predicted values and the actual values on the graph using Matplotlib.

Following are the graphs of the Predicted Values Vs Actual Values of different models-

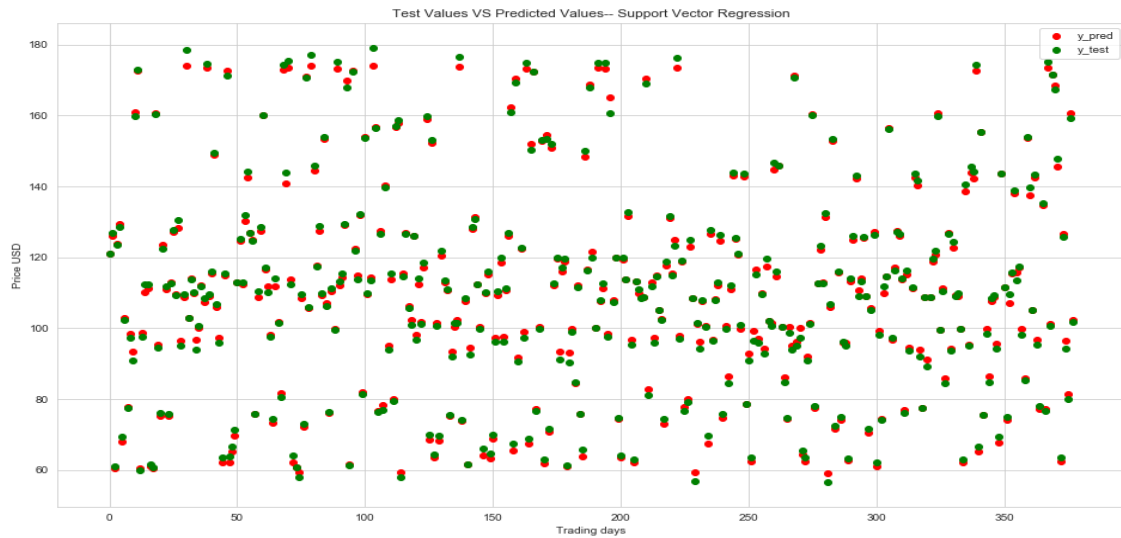


Fig 4.1:Support Vector Regression

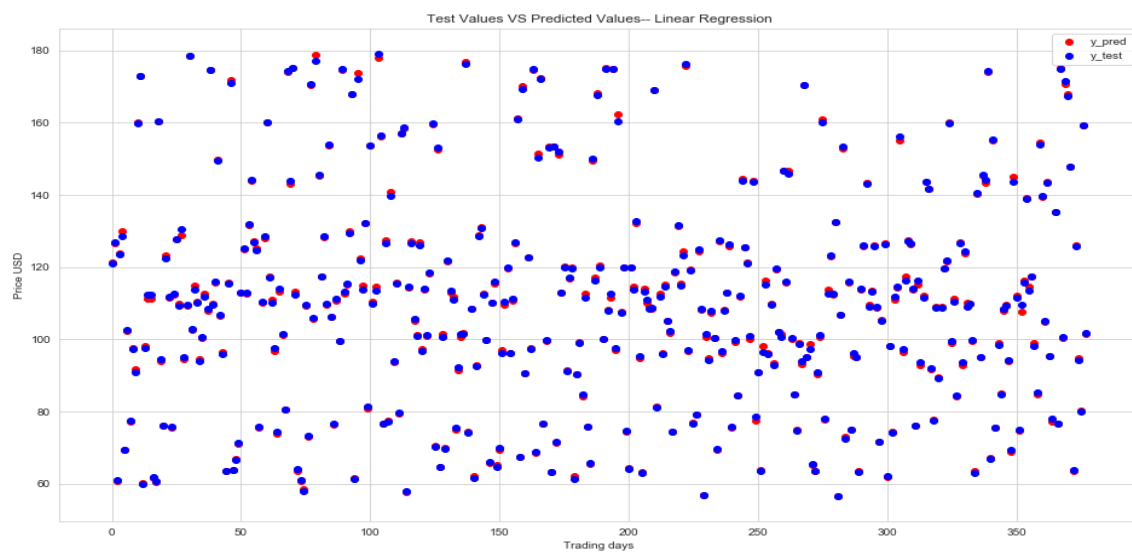


Fig 4.2:Linear Regression

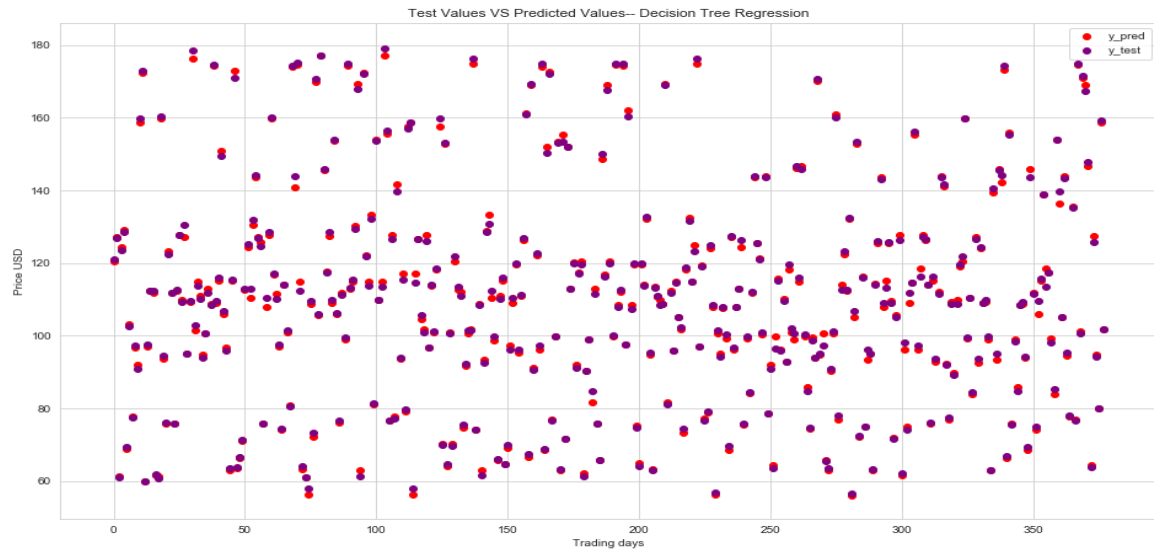


Fig 4.2: Decision Tree Regression

R2 Score

R^2 is a statistic that will give some information about the goodness of fit of a model. In regression, the R^2 coefficient of determination is a statistical measure of how well the regression predictions approximate the real data points.

An R^2 of 1 indicates that the regression predictions perfectly fit the data.

Following are the R^2 scores of different models:

Linear Regression : 0.9996938628208057

Support Vector Regression: 0.9983546794613578

Decision Tree Regression : 0.9988577962885473

CHAPTER 4

SYSTEM REQUIREMENTS

4.1 HARDWARE REQUIRMENTS

Device :	Laptop or Desktop
Processor :	Minimum i3 processor or above
RAM :	Minimum 4GB or above
Hard Disk :	Minimum 500GB or above

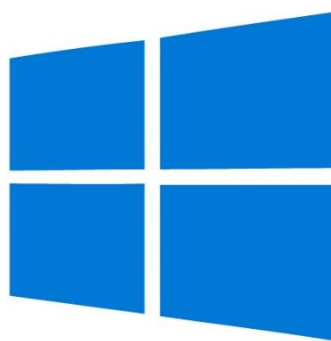
4.2 SOFTWARE REQUIRMENTS

Operating System :	Windows 10/ iOS/ Ubuntu
Language :	Python
Platform :	Pycharm IDE
Backend :	Pandas, Numpy, Matplotlib, Tkinter

4.2.1 OPERATING SYSTEM

Windows 10

Windows 10 is a series of personal computer operating system developed and released by Microsoft, as part of the Windows NT family of operating systems. It was released on July 29, 2015. Windows 10 receives new releases on an ongoing basis, which are available at no additional cost to users. Devices in enterprise environments can receive these updates at a slower pace, or use long-term support milestones that only receive critical updates, such as security patches, over their ten-year lifespan of extended support.



Windows 10 introduces what Microsoft described as "universal apps"; expanding on Metro-style apps, these apps can be designed to run across multiple Microsoft product families with nearly identical code—including PCs, tablets, smartphones, embedded systems, Xbox One, Surface Hub and Mixed Reality. The Windows user interface was revised to handle transitions between a mouse-oriented interface and a touchscreen-optimized interface based on available input devices—particularly on 2-in-1 PCs; both interfaces include an updated

Start menu which incorporates elements of Windows 7's traditional Start menu with the tiles of Windows 8. The first release of Windows 10 also introduces a virtual desktop system, a window and desktop management feature called Task View, the Microsoft Edge web browser, support for fingerprint and face recognition login, new security features for enterprise environments, and DirectX 12 and WDDM 2.0 to improve the operating system's graphics capabilities for games.

4.2.2 LANGUAGE

Python

Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales.



Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open source software and has a community-based development model, as do nearly all of its variant implementations. CPython is managed by the non-profit Python Software Foundation.

Jupyter Notebook

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.



CHAPTER 5

CONCLUSION

This document presents a detailed description of the ‘**Share Market Analysis & Prediction using Machine Learning**’ which is used by data scientists around the globe to predict and analyze the price of stocks of different companies.

Since, predicting the value of stocks has always been a tough problem at hand, this project will certainly prove beneficial for investors in share markets.

The project is easy to understand and requires moderate knowledge of Python. It has been depicted as simple as possible.

We had a great experience while developing the project and working on Machine learning using Python. It gave us the experience of working with different libraries as well.

Future work

The project can be extended to add several other features some of which are listed below:

1. Potential improvement can be made to improve our data collection and analysis method.
2. We can implement this project as an Android app which will be easy to access on Android phones.
3. We can also implement this project on a web server so that anyone can predict stock price from anywhere.
4. We can add different functionalities like user friendly interface, graphics and real-time comparison.
5. We can implement much better machine learning models to improve efficiency and increase accuracy.

References

<https://www.ryerson.ca/content/dam/science/ryscimatch/Guide%20for%20Writing%20Project%20Proposals.pdf>

<https://ieeexplore.ieee.org/document/7373006>

<https://www.thebalance.com/the-difference-between-a-stock-s-value-and-price-3140860>

<https://www.slideshare.net/anilsth91/stock-market-analysis-and-prediction>

https://en.wikipedia.org/wiki/Stock_market_prediction