STOCK PRICE PREDICTION

A PROJECT REPORT

Submitted by

Supervisor:– Ms. Harsh Sharma

Co Supervisor:– Mr. Siroj Kumar Singh

NAME OF THE CANDIDATE(S)

NAME	-	UID
ANMOL SHAKYA	-	20BCS5146
BUNTY PRASAD NAYAK	-	20BCS1160
DEVANSH TIWARI	-	20BCS1235

in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

CHANDIGARH UNIVERSITY

Discover. Learn. Empower.

Chandigarh University

OCTOBER & 2022

STOCK PRICE PREDICTION

BONAFIDE CERTIFICATE

Certified that this project report **"STOCK PEICE PREDICTION"** is the
bonafide work of "**ANMOL SHAKYA , BUNTY PRASAD NAYAK and
DEVANSH TIWARI "** who carried out the project work under my/our
Ms. Harsh Sharma supervision.

SIGNATURE	SIGNATURE
HEAD OF THE DEPARTMENT	**Ms.Harsh Sharma**
CSE	**SUPERVISOR**
	CSE

Submitted for the project viva-voce examination held on

ACKNOWLEDGEMENT

I 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 "CHANDIGARH UNIVERSITY" for their guidance and constant supervision as well as for providing necessary information regarding the industrial training project & also for their support in completing the project.

I would like to express my gratitude towards my project supervisor Harsh Sharma 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 industry persons for giving me such attention and time. My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities.

TABLE OF CONTENTS

ABSTRACT

The main objective of this paper is to find the best model to predict the value of the stock market. During the process Of considering various techniques and variables that must be taken into account, we found out that techniques like random forest, support vector machine were not exploited fully. In, this paper we are going to present and review a more feasible method to predict the stock movement with higher accuracy. The first thing we have taken into account is the dataset of the stock market prices from previous year. The dataset was pre-processed and tuned up for real analysis. Hence, our paper will also focus on data preprocessing of the raw dataset. Secondly, after preprocessing the data, we will review the use of random forest, support vector machine on the dataset and the outcomes it generates. In addition, the proposed paper examines the use of the prediction system in real-world settings and issues associated with the accuracy of the overall values given. The paper also presents a machine-learning model to predict the longevity of stock in a competitive market. The successful prediction of the stock will be a great asset for the stock market institutions and will provide real-life solutions to the problems that stock investors face.

Keywords: Machine Learning, Data Pre-processing, Data Mining, Dataset, Stock, Stock Market.

CHAPTER 1.

INTRODUCTION

1.1. Client Identification/Need Identification/Identification of relevant Contemporary issue

• **Stock Market -** Refers to several exchanges in which shares of publicly held companies are bought and sold.

• **"Stock market"** and **Stock exchange"** are often used interchangeably. Traders in the stock market buy or sell shares on one or more of the stock exchages .

• **Stock market**: The process and facilitation of investors buying and selling stocks with one another.

• **Stock exchange**: The actual intermediary that connects buyers with sellers.

• **Stock Price :** Factors that can effect the price.

      1- An earnings base.

      2- A valuation multiple.

1-An earnings base : we use EPS, an accounting measure, to illustrate the concept of earnings base.The way earnings power is measured may also depend on the type of company being analyzed.

2-A valuation multiple : The valuation multiple expresses expectations about the future. As we already explained, it is fundamentally based on the discounted present value of the future earnings stream. Therefore, the two key factors here are:

      1- The expected growth in the earnings base

      2- The discount rate.

1.1.2 Identification of relevant Contemporary issue -

•   Commodity Price Risk.

•   Headline Risk.

•   Inflationary Risk and Interest Rate Risk.

•   Model Risk.

1.2. Identification of Problem

• Stock market is one of the fastest way of making money but there is risk also . Like if we purchase Stock of any company than there is no guarantee that in future he/she can gain big profit , there is more chance that he/she in loss.

• Stock market of company that user spending money that a higer chances that price of that company stock is not stable .so it is diffect to predict the price of Stock market . It always go up and down quickly.

1.2.1 DISADVANTAGES OF THE EXISTING SYSTEM :

• The existing system fails when there are rare outcomes or predictors, as the algorithm is based on bootstrap sampling.

• The previous results indicate that the stock price is unpredictable when the traditional classifier is used.

• The existence system reported highly predictive values, by selecting an appropriate time period for their experiment to obtain highly predictive scores.

• The existing system does not perform well when there is a change in the operating environment.

• It doesn"t focus on external events in the environment, like news events or social media.   It exploits only one data source, thus highly biased.

• The existing system needs some form of input interpretation, thus need of scaling.

• It doesn"t exploit data pre-processing techniques to remove inconsistency and incompleteness of the data.

1.3. Identification of Tasks

Stock market prediction and analysis are some of the most difficult jobs to complete. There are numerous causes for this, including market volatility and a variety of other dependent and independent variables that influence the value of a certain stock in the market. These variables make it extremely difficult for any stock market expert to anticipate the rise and fall of the market with great precision.

1.4. Timeline

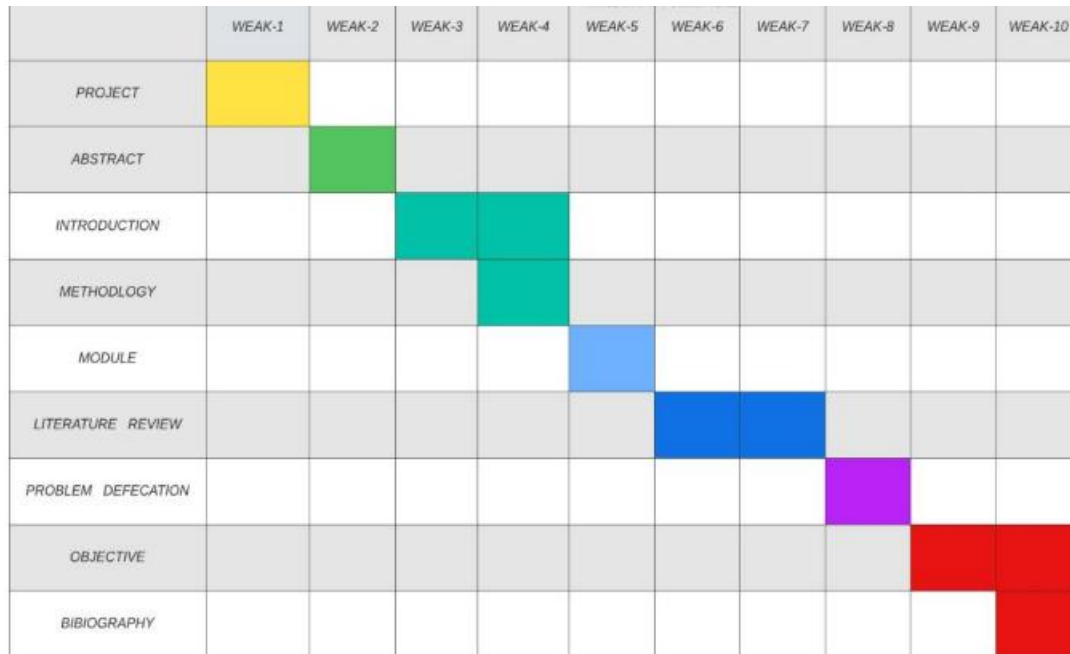|	WEAK-1	WEAK-2	WEAK-3	WEAK-4	WEAK-5	WEAK-6	WEAK-7	WEAK-8	WEAK-9	WEAK-10
PROJECT	■									
ABSTRACT		■								
INTRODUCTION			■	■						
METHODLOGY				■						
MODULE					■					
LITERATURE REVIEW						■	■			
PROBLEM DEFECATION								■		
OBJECTIVE									■	■
BIBIOGRAPHY										■

Fig 1. Gantt Chart

1.5. Organization of the Report

The material presented at the report is organized into five chapters. After this introductory chapter 1, chapter 2 describes the "Literature survey" by reading previous year papers for the project "Stock Market Prediction" using Machine Learning.

Chapter 3 describes the Design flow/Process that are followed in our project for the development of project stock market prediction. It defines the various steps and processes in which our project is divided. Chapter 4 provides Results analysis and validation of our project. In this we will evaluate our project on the basics of the final outcome that we received after the proper deployment of the project.

Chapter 5 presents the Conclusion and the future work of our project. In this chapter we will discuss the future of our project and also conclude our project on the various results that we got from the previous Chapter 4. In future scope we will also define the various sectors where we can deploy our project and how it will benefit the user.

CHAPTER 2.

LITERATURE REVIEW/BACKGROUND STUDY

2.1 Timeline of the reported problem -

•   The stock of a corporation constitutes the equity stake of its owners. it represents the residual assets of the company that would be due to stockholders after discharge of all senior claims such as secured and unsecured debt.

•   Stock market prediction is the act of trying to determine the duture 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.

•   Whenever new information comes up the market absorbs it by correcting itself , thus there is no space for prediction . Example of crashes of market unexpected.

2.1.1 The Biggest Stock Market Crashes in History-

1. The Wall Street Crash of 1929-

The stock market began right around 1600, and the first stock market crash was soon to follow. However, the Black Tuesday stock market crash that took place in 1929 remains the worst stock market crash in US history.

2. The Tech Bubble Crash of 1999-

The 1990s were a period of rapid technological development, and the commercialization of the internet caused valuations of internet-based companies to soar.

3.The Housing Market Crash of 2008-

This is one you probably remember — the housing market collapse of 2008. Over the course of 2008, the Dow fell almost 34%, and it wasn't until early 2009 that it began to climb again.

4.The Stock Market Crash of 2020-

This leads us to our most recent example: the stock market crash of 2020, which disrupted a ten-year all-time-high bull market. Though the crash was largely due to the impact of the coronavirus, many weaknesses within the market built up for years, creating a massive stock market bubble.

That being said, it doesn't take much for a market to crash. It only requires a specific set of circumstances to manifest within the economy and manifest they did. By April 2020, unemployment climbed to 14.8%.

Stock market prediction mechanisms currently being used-

1.Survey of Stock Market Prediction Using Machine Learning Approach:

The stock market prediction has become an increasingly important issue in the present time. One of the methods employed is technical analysis, but such methods do not always yield accurate results. So it is important to develop methods for a more accurate prediction. Generally, investments are made using predictions that are obtained from the stock price after considering all the factors that might affect it. The technique that was employed in this instance was a regression. Since financial stock marks generate enormous amounts of data at any given time a great volume of data needs to undergo analysis before a prediction can be made. Each of the techniques listed under regression hasits own advantages and limitations over its other counterparts. One of the noteworthy techniques that were mentioned was linear regression. The way linear regression models work is that they are often fitted using the least squares approach, but they may alternatively be also be fitted in other ways, such as by diminishing the "lack of fit" in some other norm, or by diminishing a handicapped version of the least squares loss function. Conversely, the least squares approach can be utilized to fit nonlinear models.

2.Impact of Financial Ratios and Technical Analysis on Stock Price Prediction Using Random Forests:

 The use of machine learning and artificial intelligence techniques to predict the prices of the stock is an increasing trend. More and more researchers invest their time every day in

coming up with ways to arrive at techniques that can further improve the accuracy of the stock prediction model. Due to the vast number of options available, there can be n number of ways on how to predict the price of the stock, but all methods don"t work the same way. The output varies for each technique even if the same data set is being applied. In the cited paper the stock price prediction has been carried out by using the random forest algorithm is being used to predict the price of the stock using financial ratios form the previous quarter. This is just one wayof looking at the problem by approaching it using a predictive model, using the random forest to predict the future price of the stock from historical data. However, there are always other factors that influence the price of the stock, such as sentiments of the investor, public opinion about the company, news from various outlets, and even events that cause the entire stock market to fluctuate. By using the financial ratio along with a model that can effectively analyze sentiments the accuracy of the stock price prediction model can be increased.

3.Stock Market Prediction via Multi-Source Multiple Instance Learning :

Accurately predicting the stock market is a challenging task, but the modern web has proved to be a very useful tool in making this task easier. Due to the interconnected format of data, it is easy to extract certain sentiments thus making it easier to establish relationships between various variable and roughly scope out a pattern of investment. Investment pattern from various firms show sign of similarity, and the key to successfully predicting the stock market is to exploit these same consistencies between the data sets. The way stock market information can be predicted successfully is by using more than just technical historical data, and using other methods like the use of sentiment analyzer to derive an important connection between people"s emotions and how they are influenced by investment in specific stocks. One more important segment of the prediction process was the extraction of important events from web news to see how it affected stock prices.

4. Stock Market Prediction: Using Historical Data Analysis -

The stock market prediction process is filled with uncertainty and can be influenced by multiple factors. Therefore, the stock market plays an important role in business and

finance. The technical and fundamental analysis is done by sentimental analysis process. Social media data has a high impact due to its increased usage, and it can be helpful in predicting the trend of the stock market. Technical analysis is done [6] using by applying machine learning algorithms on historical data of stock prices. The method usually involves gathering various social media data, news to extract sentiments expressed by individuals. Other data like previous year stock prices are also considered. The relationship between various data points is considered, and a prediction is made on these data points. The model was able to make predictions about future stock values.

5. A Survey on Stock Market Prediction Using SVM :

The recent studies provide a well-grounded proof that most of the predictive regression models are inefficient in out of sample predictability test. The reason for this inefficiency was parameter instability and model uncertainty. The studies also concluded the traditional strategies that promise to solve this problem. Support vector machine commonly known as SVM provides with the kernel, decision function, and sparsity of the solution. It is used to learn polynomial radial basis function and the multi-layer perceptron classifier. It is a training algorithm for classification and regression, which works on a larger dataset. There are many algorithms in the market but SVM provides with better efficiency and accuracy. The correlation analysis between SVM and stock market indicates strong interconnection between the stock prices and the market index.

6. Predicting Stock Price Direction Using Support Vector Machines :

Financial organizations and merchants have made different exclusive models to attempt and beat the market for themselves or their customers, yet once in a while has anybody accomplished reliably higher-than-normal degrees of profitability. Nevertheless, the challenge of stock forecasting is so engaging in light of the fact that the improvement of only a couple of rate focuses can build benefit by a large number of dollars for these organizations. [

7. A Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA) :

The time series prediction problem was researched in the work centers in the various financial institution. The prediction model, which is based on SVM and independent analysis, combined called SVM-ICA, is proposed for stock market prediction. Various time series analysis models are based on machine learning. The SVM is designed to solve regression problems in non-linear classification and time series analysis. The generalization error is minimized using an approximate function, which is based on risk diminishing principle. Thus, the ICA technique extracts various important features from the dataset. The time series prediction is based on SVM. The result of the SVM model was compared with the results of the ICA technique without using a preprocessing step.

8. Machine Learning Approach In Stock Market Prediction:

The vast majority of the stockbrokers while making the prediction utilized the specialized, fundamental or the time series analysis. Overall, these techniques couldn't be trusted completely, so there emerged the need to give a strong strategy to financial exchange prediction. To find the best accurate result, the methodology chose to be implemented as machine learning and AI along with supervised classifier. Results were tried on the binary classification utilizing SVM classifier with an alternate set of a feature list. The greater part of the Machine Learning approach for taking care of business [2] issues had their benefit over factual techniques that did exclude AI, despite the fact that there was an ideal procedure for specific issues. Swarm Intelligence [2] optimization method named Cuckoo search was most easy to accommodate the parameters of SVM. The proposed hybrid CS-SVM strategy exhibited the performance to create increasingly exact outcomes in contrast with ANN. Likewise, the CS-SVM display [2] performed better in the forecasting of the stock value prediction. Prediction stock cost utilized parse records to compute the predicted, send it to the user, and autonomously perform tasks like buying and selling shares utilizing automation concept. Naïve Bayes Algorithm was utilized.

9. Corporate Communication Network and Stock Price Movements: Insights from Data Mining-

This paper tries to indicate that communication patterns can have a very significant effect on an organization‟s performance. This paper proposed a technique to reveal the

performance of a company. The technique deployed in the paper is used to find the relationships between the frequencies of email exchange of the key employees and the performance of the company reflected in stock values. In order to detect association and non-association relationships, this paper proposed to use a data mining algorithm on a publicly available dataset of Enron Corp. The Enron Corporation was an energy, commodities, and services company based in Houston, Texas whose stock dataset is available for public use.

2.2 Proposed solutions-

Stock market prediction seems a complex problem because there are many factors that have yet to be addressed and it doesn't seem statistical at first. But by proper use of machine learning techniques, one can relate previous data to the current data and train the machine to learn from it and make appropriate assumptions. Machine learning as such has many models but this paper focuses on two most important of them and made the predictions using them.
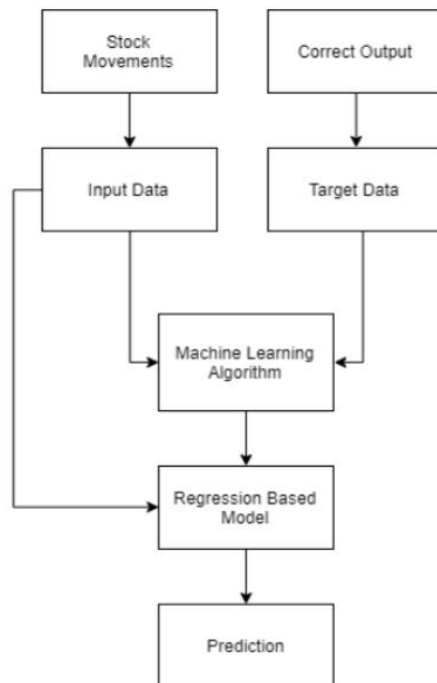

Fig 2. Solution Proposed

2.3 Review Summary-

Two techniques have been utilized in this paper.LSTM and Regression, on the Yahoo finance dataset. Both the techniques have shown an improvement in the accuracy of predictions, thereby yielding positive results. Use of recently introduced machine learning techniques in the prediction of stocks have yielded promising results and thereby marked the use of them in profitable exchange schemes. It has led to the conclusion that it is possible to predict stock market with more accuracy and efficiency using machine learning techniques.

In the future, the stock market prediction system can be further improved by utilizing a much bigger dataset LSTM is the advanced version of Recurrent-Neural- Networks (RNN) where the information belonging to previous state persists. These are different from RNNs as they involve long term dependencies and RNNs works on finding the relationship between the recent and the current information. This indicates that the interval of information is relatively smaller than that to LSTM.

The main purpose behind using this model in stock market prediction is that the predictions depends on large amounts of data and are generally dependent on the long term history of the market . So LSTM regulates error by giving an aid to the RNNs through retaining information for older stages making the prediction more accurate .

Since stock market involves processing of huge data, the gradients with respect to the weight matrix may become very small and may degrade the learning rate of the system. This corresponds to the problem of Vanishing Gradient. LSTM prevents this from happening. The LSTM consists of a remembering cell, input gate, output gate and a forget gate. The cell remembers the value for long term propagation and the gates regulate them .

In this paper, a sequential model has been made which involves stacking two LSTM layers on top of each other with the output value of 256. The input to the layer is in the form of two layer [0] and layer[1]. A dropout value of 0.3 has been fixed which means that 0.3 out of total nodes will be frozen during the training process to avoid over-fitting of data and increase the speed of the training process. At last, the core dense layer where each neuron is connected to every other in the next layer is added providing input of 32 parameters to the next core layer which gives output as 1. The model is compiled with a

mean square cost function to maintain the error throughout the process and accuracy is chosen as a metric for the prediction .

2.4 Problem Definition-

Stock Price Prediction by Machine Learning present to estimate the stock future value and machine learning technique like LSTM for existing work. This machine-learning algorithm is to perform the best predicting result of the stock future price. LSTM is capable to catching the modifications in the behaviour of the stock price for the indicated period in this proposed system. Propose a machine learning-based normalization for stock price prediction.. The data reflected the stock price at some time intervals for every day of the year. It contains various data like date, symbol, open price, close price, low price, high price and volume. Here, the data for only one company was considered. All the data was available in a file of CSV format which was first read and transformed into a data frame using the Pandas library in Python. The normalization of the data was performed through the sklearn library in Python and the data were divided into training and testing sets. The experiment set was kept as 20% of the available dataset. This paper focuses on two architecture Regression-based Model and LSTM. The Regression-based Model is employed for predicting unbroken values through some given autonomous values Regression uses a given linear function for predicting continuous values of the most important amongst them and made the predictions using these. LSTM architecture is able to identify the changes in trends which show evident from the result. LSTM is identified as the best model for the proposed methodology. This shows that the proposed system is capable of identifying some interrelation within the data. In the stock market, there may not always follow the same cycle or may not always be in a regular pattern for the changes that are occurred. The period of the existence will differ and the existence of the trend is based on the companies and the sectors. For investors, this type of analysis of trends and cycles will obtain more profit. We must use networks like LSTM as they rely on the current information to analyse various information.

Proposed Work-
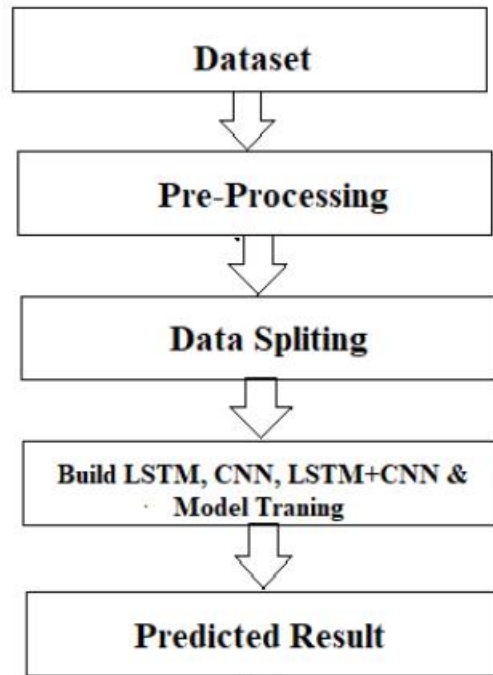
The system presented here composes of five modules:-

1. Input as Dataset

 2. Pre processing

3. Data splitting

4. Build & Model train Lstm

5. Output as Predicted Result.

2.5 Goals/Objectives-

1- A stock market prediction is described as an action of attempting to classify the future value of the company stock or other financial investment traded on the stock exchange. The forthcoming price of a stock of the successful estimation is called the Yield significant profit. This helps you to invest wisely for making good profits.

2 - The future price of a stock is the main motivation behind the stock price prediction. In various cases like business and industry, environmental science, finance and economics motivation can be useful. The future value of the company's stock can be determining.

CHAPTER 3.

DESIGN FLOW/PROCESS

3.1 Evaluation & Selection of Specifications/Features -

• A stock market prediction is described as an action of attempting to classify the future value of the company stock or other financial investment traded on the stock exchange. The forthcoming price of a stock of the successful estimation is called the Yield significant profit. This helps you to invest wisely for making good profits.

• The future price of a stock is the main motivation behind the stock price prediction. In various cases like business and industry, environmental science, finance and economics motivation can be useful. The future value of the company's stock can be determining.

3.2 Analysis and Feature finalization subject to constraints:

3.2.1 Stock Market

Analysis of stocks using data mining will be useful for new investors to invest in stock market based on the various factors considered by the model.

Stock market includes daily activities like sensex calculation, exchange of shares. The exchange provides an efficient and transparent market for trading in equity, debt instruments and derivatives.

Our model will be analyzing sensex based on company's stock value. The stock values of company depend on many factors, some of them are:

1> *Demand and Supply:*

Demand and Supply of shares of a company is a major reason price change in stocks. When Demand Increase and Supply is less, price rises. and vice versa.

2> *Corporate results:* This will be regarding to the profits or progress of the company over a span of time say 3 months.

3> *Popularity:* Main Strength in hands of share buyer. Popularity of a company can effect on buyers. Like if any good news of a company, may result in rise of stock price.

And a bad news may break dreams.

The stock value depends on other factors as well, but we are taking into consideration only these main factors.

3.2.2 Prominent features of the Project:

A. Analyzing stock data.

We need to provide data of a particular company, and its Monthly Sales / Profit report with Months High and Low points of its Stock.

B. Analyzing the factors.

We have to obtain the data in the same period for the following factors.

1. Demand and Supply: We will obtain by the previous data entered.

2. Corporate results: Companies declare their performance results and profit at the end of each quarter.

3. Popularity: If any news about a company is about to come and is it bad or good.

We have to analyze the variations in the stock value of the companies with respect to these factors using some data mining algorithms.

3.2.3 Fundamental analysis:

Fundamental Analysts are concerned with the business that reasons the stock itself. They assess a company's historical performance as well as the reliability of its accounts. Different performance shares are created that aid the fundamental forecaster with calculating the validity of a stock, such as the P/E ratio. Warren Buffett is probably the foremost renowned of all Fundamental Analysts. What fundamental analysis within the stock market is making an attempt to reach, is organizing the true value of a stock, that then will be matched with the worth it is being listed on stock markets and so finding out whether or not the stock on the market is undervalued or not. Find out the correct value will be completed by numerous strategies with primarily a similar principle. The principle is that an organization is price all of its future profits. Those future profits has to be discounted to their current value. This principle goes on the theory that a business is

all about profits and nothing else. Differing to technical analysis, the fundamental analysis is assumed as further as a long approach. Fundamental analysis is created on conviction that hominoid society desires capital to make progress and if the company works well, than it should be rewarded with an additional capital and outcome in a surge in stock price. Fundamental analysis is usually used by the fund managers as it is the maximum sensible, objective and prepared from openly existing data like financial statement analysis. One more meaning of fundamental analysis is on the far side bottom-up business analysis, it discusses the top-down analysis since initial analysing the world economy, followed by country analysis and also sector analysis, and last the company level analysis.

3.2.4 Technical analysis :

Chartists or the technical analysts are not involved with any other of the fundamentals of the company. The long run price of a stock based generally exclusively on the trends of the past value (a form of time series analysis) that is set by them. The head and shoulders or cup and saucer are various numerous patterns that are employed. Also the techniques, patterns are used just like the oscillators, exponential moving average (EMA), support and momentum and volume indicators. Candlestick patterns, believed to have been initial developed by Japanese rice merchants, are nowadays widely used by technical analysts. For the short-term approaches, the technical analysis is used compare to long-run ones. So, in commodities and forex markets it is more predominant wherever traders target short-term price movements. There are basic rules are used in this analysis, first all significant about a company is already priced into the stock, another being that the value changes in trends and finally that history (of prices) tends to repeat itself that is especially due to the market science.
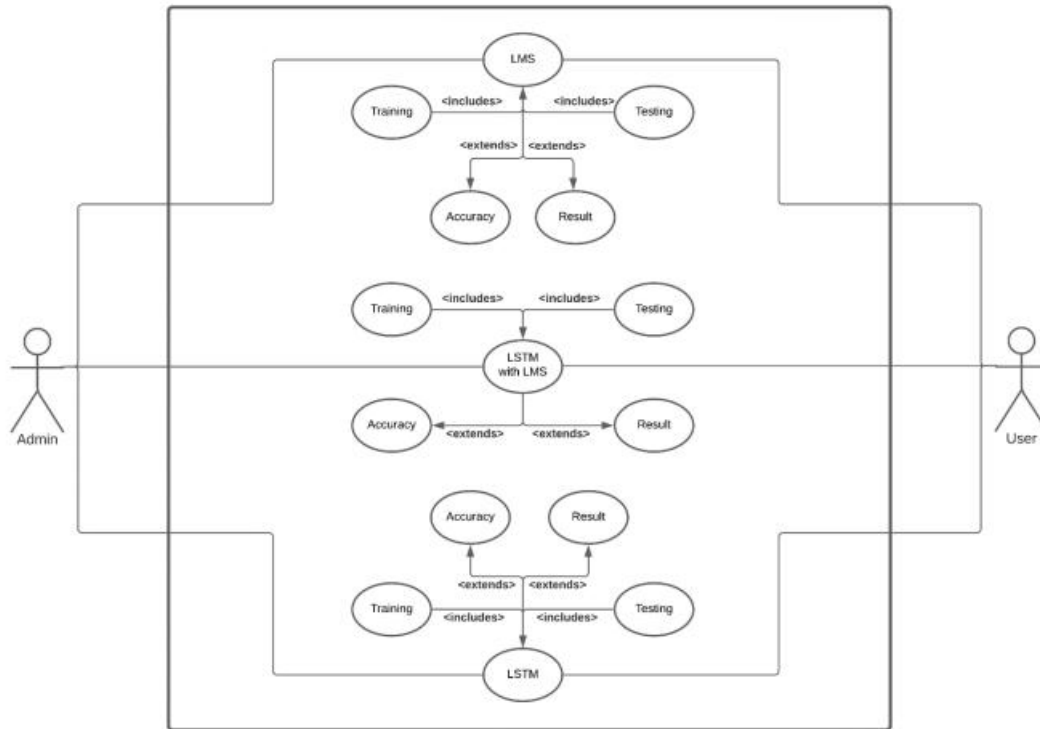
3.3 Design Flow :

MODEL USER BASED-

Fig 4. Model based on the user interface

3.4 Design selection :

Stock market prediction seems a complex problem because there are many factors that have yet to be addressed and it doesn't seem statistical at first. But by proper use of machine learning techniques, one can relate previous data to the current data and train the machine to learn from it and make appropriate assumptions. Machine learning as such has many models but this paper focuses on two most important of them and made the predictions using them.

3.5 Implementation plan/methodolog

1 PYTHON :

The language of select for this project was Python. This was a straightforward call for many reasons.

1. Python as a language has a vast community behind it. Any problems which may be faced is simply resolved with visit to Stack Overflow. Python is the foremost standard language on the positioning that makes it is very straight answer to any question.

2. Python is an abundance of powerful tools ready for scientific computing Packages. The packages like NumPy, Pandas and SciPy area unit freely available and well documented. These Packages will intensely scale back, and variation the code necessary to write a given program. This makes repetition fast.

3. Python is a language as forgiving and permits for the program that appear as if pseudo code. This can be helpful once pseudo code give in tutorial papers should be required and verified. Using python this step is sometimes fairly trivial.

However, Python is not without its errors. The python is dynamically written language and packages are area unit infamous for Duck writing. This may be frustrating once a package technique returns one thing that, for instance, looks like an array instead of being an actual array. Plus the standard Python documentation did not clearly state the return type of a method, this can't lead without a lot of trials and error testing otherwise happen in a powerfully written language. This is a problem that produces learning to use a replacement Python package or lbrary more difficult than it otherwise may be.

2 NUMPY- Numpy is python package which provide scientific and higher level mathematical abstractions wrapped in python. It is [20] the core library for scientific computing, that contains a provide tools for integrating C, strong n-dimensional array object, C++ etc. It is also useful in random number capability, linear algebra etc. Numpy's array type augments the Python language with an efficient data structure used for numerical work, e.g., manipulating matrices. Numpy additionally provides basic

numerical routines, like tools for locating Eigenvectors.

3 JUPITER NOTEBOOK - The Jupyter Notebook is an open-source web application that enables to making and sharing documents that contain visualizations, narrative text, live code and equations. Uses include: data , data visualization, data transformation, statistical modelling, machine learning, numerical simulation, data cleaning and much more.

4 STREAMLIT - Streamlit is an open source app framework in Python language. It helps us create web apps for data science and machine learning in a short time. It is compatible with major Python libraries such as scikit-learn, Keras, PyTorch, SymPy(latex), NumPy, pandas, Matplotlib etc. With Streamlit, no callbacks are needed since widgets are treated as variables. Data caching simplifies and speeds up computation pipelines. Streamlit watches for changes on updates of the linked Git repository and the application will be deployed automatically in the shared link.

5 DATETIME - In Python, date and time are not a data type of their own, but a module named **datetime** can be imported to work with the date as well as time. **Python Datetime module** comes built into Python, so there is no need to install it externally.
Python Datetime module supplies classes to work with date and time. These classes provide a number of functions to deal with dates, times and time intervals. Date and datetime are an object in Python, so when you manipulate them, you are actually manipulating objects and not string or timestamps.

The DateTime module is categorized into 6 main classes –
I.   **date** – An idealized naive date, assuming the current Gregorian calendar always was, and always will be, in effect. Its attributes are year, month and day.
II. **time** – An idealized time, independent of any particular day, assuming that every day has exactly 24*60*60 seconds. Its attributes are hour, minute, second, microsecond, and tzinfo.
III. **datetime** – Its a combination of date and time along with the attributes year, month,

day, hour, minute, second, microsecond, and tzinfo.

IV. **timedelta** – A duration expressing the difference between two date, time, or datetime instances to microsecond resolution.

V. **tzinfo** – It provides time zone information objects.

VI. **timezone** – A class that implements the tzinfo abstract base class as a fixed offset from the UTC (New in version 3.2)

6 PANDAS - Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license.

7 YFINANCE - Python library that gives you current and historical stock market price data from Yahoo Finance, and so much more. The cool thing is that it is entirely free, doesn't require an API key, and allows for a relatively high throughput of 2,000 requests per hour.

8 PLOTLY - The plotly Python library is an interactive, open-source plotting library that supports over 40 unique chart types covering a wide range of statistical, financial, geographic, scientific, and 3-dimensional use-cases.

9 PROPHET - Prophet is a forecasting procedure implemented in R and Python. It is fast and provides completely automated forecasts that can be tuned by hand by data scientists and analysts.

The various modules of the project would be divided into the segments as described:

I. Data Collection :

Data collection is a very basic module and the initial step towards the project. It generally deals with the collection of the right dataset. The dataset that is to be used in the market prediction has to be used to be filtered based on various aspects. Data collection also

complements to enhance the dataset by adding more data that are external. Our data mainly consists of the previous year stock prices. Initially, we will be analyzing the Kaggle dataset and according to the accuracy, we will be using the model with the data to analyze the predictions accurately.

II. Pre Processing :

Data pre-processing is a part of data mining, which involves transforming raw data into a more coherent format. Raw data is usually, inconsistent or incomplete and usually contains many errors. The data pre-processing involves checking out for missing values, looking for categorical values, splitting the data-set into training and test set and finally do a feature scaling to limit the range of variables so that they can be compared on common environs.

III. Training the Machine :

Training the machine is similar to feeding the data to the algorithm to touch up the test data. Thetraining sets are used to tune and fit the models. The test sets are untouched, as a model should not be judged based on unseen data. The training of the model includes cross-validation where we get a well-grounded approximate performance of the model using the training data. Tuning models are meant to specifically tune the hyperparameters like the number of trees in a random forest. We perform the entire cross-validation loop on each set of hyperparameter values. Finally, we will calculate a cross-validated score, for individual sets of hyperparameters. Then, we select the best hyperparameters. The idea behind the training of the model is that we some initial values with the dataset and then optimize the parameters which we want to in the model. This is kept on repetition until we get the optimal values. Thus, we take the predictions from the trained model on the inputs from the test dataset. Hence, it is divided in the ratio of 80:20 where 80% is for the training set and the rest 20% for a testing set of the data.

IV. Data Scoring

The process of applying a predictive model to a set of data is referred to as scoring the data. The technique used to process the dataset is the Random Forest Algorithm. Random forest involves an ensemble method, which is usually used, for classification and as well as regression. Based on the learning models, we achieve interesting results. The last module thus describes how the result of the model can help to predict the probability of a stock to rise and sink based on certain parameters. It also shows the vulnerabilities of a particular stock or entity. The user authentication system control is implemented to make sure that only the authorized entities are accessing the results.

CHAPTER 4.

RESULTS ANALYSIS AND VALIDATION

4.1 Implementation of solution -

import streamlit as st

Streamlit is an open-source python framework for building web app. we are using in the program to build web app and show the result of ML algorithms.

We can rapidly build the tools you need. Build apps in a dozen lines of Python with a simple API. Streamlit is a tool in the Machine Learning Tools category of a tech stack.

Using it we can also works with TensorFlow, Keras, PyTorch, Pandas, Numpy, Matplotlib, Seaborn, Altair, Plotly, Bokeh, Vega-Lite, and more.

from datetime import date

Python Datetime module , Using it we can get current local date and time .

import pandas as pd

Pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

Pandas is the best tool for handling this real-world messy data. And pandas is one of the open-source python packages built on top of NumPy.

import yfinance as yf

Yfinance is a python package that enables us to fetch historical market data from Yahoo Finance API in a Pythonic way.

It becomes so easy for all the Python developers to get data with the help of yfinance.

We can easily download historical stock data from yfinance.

import plotly.express as px

#The plotly.express module (usually imported as px) contains functions that can create entire figures at once, and is referred to as Plotly Express or PX.

#Plotly Express is the easy-to-use, high-level interface to Plotly, which operates on a

variety of types of data and produces easy-to-style figures.

#Plotly Express provides functions to visualize a variety of types of data. Most functions such as px. bar or px.

from prophet import Prophet

#Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects.

#It works best with time series that have strong seasonal effects and several seasons of historical data.

from plotly import graph_objects as go.

from prophet.plot import plot_plotly, plot_components_plotly

Get the current local date.

START = "2016-01-01"
# TODAY = "2017-01-01"
TODAY= date.today().strftime("%Y-%m-%d")

# Main title in web app.
st.title("Stock Price Prediction App")

# displaying all the possible stock
stocks =pd.read_csv('https://raw.githubusercontent.com/kaushikjadhav01/Stock-Market-Prediction-Web-App-using-Machine-Learning-And-Sentiment-Analysis/master/Yahoo-Finance-Ticker-Symbols.csv')

# creating a select box in web app.
selected_stocks = st.selectbox("Select Stock for Prediction",stocks)

Loading data .
@st.cache #Cache the data so we don't have to download the data again and again.
def load_data(ticker):
  data = yf.download(ticker,START,TODAY)

```python
    data.reset_index(inplace=True)
    return data

data_load_state = st.text("Loading data.........")
data = load_data(selected_stocks)
data_load_state.text("Data Loaded ... done ! ")

st.subheader('Stock data')
st.text("(All prices are in USD)")
st.text('Fixed width text')
# last 5 column will be printed of the selected stock
st.write(data.tail())

# Plot the data on graph
# #df = px.data.stocks()
fig = px.line(data, x='Date', y=['Open','Close'])
fig.layout.update(title_text = "Time Series Data" , xaxis_rangeslider_visible = True)
fig.update_layout(
    margin=dict(l=10, r=20, t=50, b=40),
    )
st.plotly_chart(fig)

# Sliding bar
n_years = st.slider("Years of Prediction: ", 1,7)
period = n_years*365

#Forecasting by train the dataset
df_train = data[['Date','Close']]
df_train  = df_train.rename(columns={"Date": "ds","Close": "y"})
m = Prophet()
m.fit(df_train)
```

```
future = m.make_future_dataframe(periods=period)
forecast = m.predict(future)

# forecast[['ds', 'yhat', 'yhat_lower', 'yhat_upper']].tail()
# st.write(forecast)
st.subheader('Predict Stock')
st.write(forecast.tail())
st.write('Predict Stock')
# fig1 = go(m,forecast)
fig1 = plot_plotly(m, forecast )
st.plotly_chart(fig1)
st.write('forecast Components')
fig2 = plot_components_plotly(m, forecast)
st.write(fig2)
```

4.2 OUTPUT:

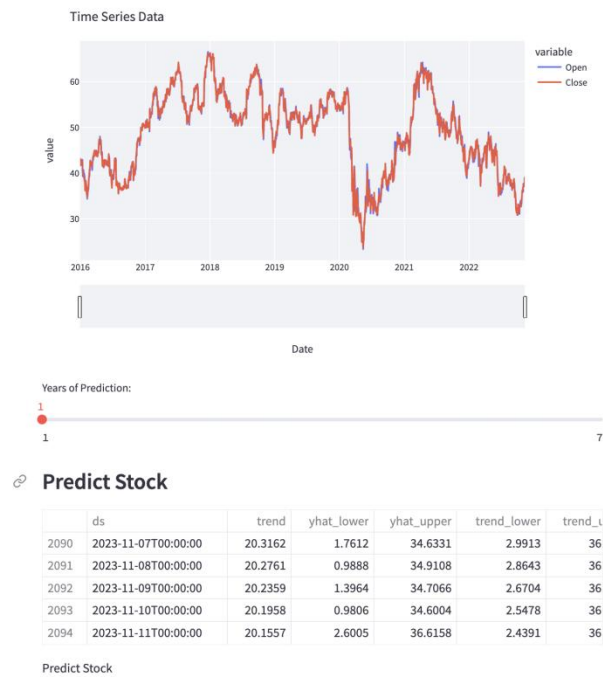**Fig.1 Dashboard Web App to predict Stock Price**

Fig.2 Slidebar to predict according to year
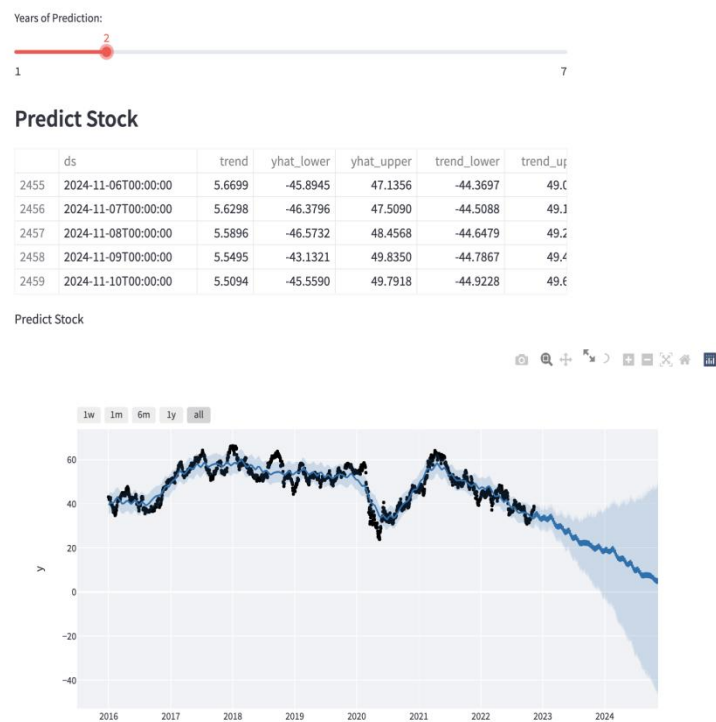
Fig.3 Plotting graph for prediction

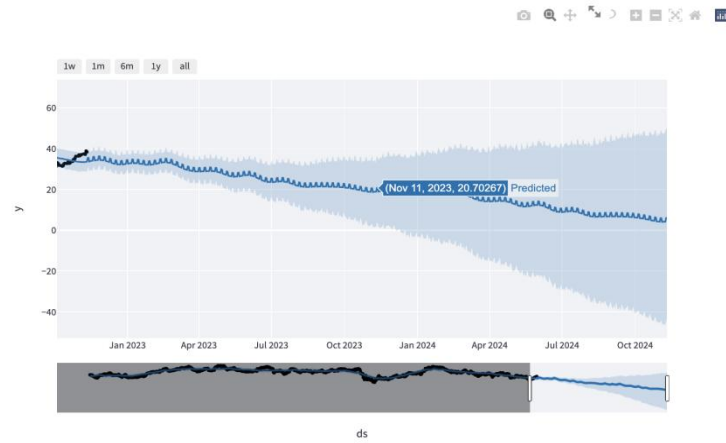|	ds	trend	yhat_lower	yhat_upper	trend_lower	trend_up
2455	2024-11-06T00:00:00	5.6699	-45.8945	47.1356	-44.3697	49.0
2456	2024-11-07T00:00:00	5.6298	-46.3796	47.5090	-44.5088	49.1
2457	2024-11-08T00:00:00	5.5896	-46.5732	48.4568	-44.6479	49.2
2458	2024-11-09T00:00:00	5.5495	-43.1321	49.8350	-44.7867	49.4
2459	2024-11-10T00:00:00	5.5094	-45.5590	49.7918	-44.9228	49.6

Predict Stock

Fig.4. Prediction of stock in the months.

Fig.5 Forecast components on trend, yearly, and weekly basis.

CHAPTER 5.
CONCLUSION AND FUTURE WORK

5.1 Conclusion :

The project lays the foundation for democratizing machine learning technologies for retail investors, connecting predictions made by machine learning models to retail investors through a mobile application. It helps investors navigate through the stock markets with additional analysis and help them make more informed decisions. The findings demonstrated that the application provides significance in trend prediction. When compared to the baseline, the prediction shows useful trend tendency with the real stock trend. Through the application interface, the user can easily compare the predictions and model scores from different machine learning models, then choosing the one that fits their preference. The models used in the application will continue to improve itself by searching for a better model topology, structure and hyperparameters through evolution algorithm. The findings concluded the usefulness of evolution algorithm in lowering the mean squared error when predicting stock prices, which is helpful for improving the trend prediction for retail investors. Therefore, with the application and research findings, to large extent the project team achieved the aim of creating an user-friendly system for retail investors whom does not have previous technical knowledge to navigate the machine model predictions result with useful benchmarks. There are 4 possible further improvements building upon the findings of this project. First, multiple approaches to framing the problems could be explored in the future, such as Page 80 of 124 predicting whether the stock price goes up or down (binary classification) based on the previous stock prices. Other features could be incorporated, such as market news and sentiment. Combined with the development of more advanced machine learning techniques, the accuracy of the information provided to retail investors might be improved significantly. Second, a larger scale of evolution with larger population size and more iterations could also be tested for achieving better results. Model inputs can also be included into the evolution algorithm as a variable to optimize. Regularized evolution [38] can be tested to

eliminate old models regardless of their accuracy, which could allow the algorithm to search for more distant models in the search space. Third, it is also possible to use more finance-specific scores, like those introduced, as the objective function instead of simple mean squared errors to achieve better results. Fourth, mobile applications with better presentation of stock price predictions could be developed to help investors understand the implications of the stock price predictions, e.g. when to buy or sell. This would allow investors to make more informed decisions based on the machine learning models and truly democratize machine learning technologies, which were believed to be only in the hands of very few people.

5.2 Future work :

we found that the most suitable algorithm for predicting the market price of a stock based on various data points from the historical data is the random forest algorithm. The algorithm will be a great asset for brokers and investors for investing money in the stock market since it is trained on a huge collection of historical data and has been chosen after being tested on a sample data. The project demonstrates the machine learning model to predict the stock value with more accuracy as compared to previously implemented machine learning models.

5.2.1 FUTURE ENHANCEMENT :

Future scope of this project will involve adding more parameters and factors like the financial ratios, multiple instances, etc. The more the parameters are taken into account more will be the accuracy. The algorithms can also be applied for analyzing the contents of public comments and thus determine patterns/relationships between the customer and the corporate employee. The use of traditional algorithms and data mining techniques can also help predict the corporation s‟ performance .

REFERENCES

1. N. K. Chowdhury and C. K.–S. Leung. Improved travel time prediction algorithms for intelligent transportation systems. In *Proc. KES 2011, Part II*, pp. 355––365.

2. A. Cuzzocrea, C. K.–S. Leung, and R. K. MacKinnon. Mining constrained frequent itemsets from distributed uncertain data. *FGCS*, 37: 117––126, July 2014.

3. D. Hodges. Is your fund manager beating the index? *MoneySense*, 15(7):10, Dec. 2013/Jan. 2014.

4. D. E. King. Dlib–ml: a machine learning toolkit. *JMLR*, 10: 1755––1758, July

5. C. K.–S. Leung, F. Jiang, and Y. Hayduk. A landmark–model based system for mining frequent patterns from uncertain data streams. In *Proc. IDEAS 2011*, pp. 249––250. ACM.

6. Quandl Stock Index. http://www.quandl.com/stocks

7. S&P 500 information technology. S&P Dow Jones Indices. http://ca.spindices.com/indices/equity/sp–500–information–technology–sector

8. C. Sammut and G. I. Webb (eds.) *Encyclopedia of Machine Learning*. Springer, 2010.

9. Yahoo! Finance – Canada. http://ca.finance.yahoo.com