Stock Market Prices Prediction Using Machine Learning

B.Rikitha
Department of E.C.E Snist
Hyderabad, India

B.Varunteja
Department of E.C.E Snist
Hyderabad, India

Sai kiran
Department of E.C.E Snist
Hyderabad, India

ABSTRACT: *In Stock Market Prediction, the aim is to predict the future value of the financial stocks of a company. The recent trend in stock market prediction technologies is the use of machine learning which makes predictions based on the values of current stock market indices by training on their previous values. Machine learning itself employs different models to make prediction easier and authentic. The technical and fundamental or the time series analysis is used by most of the stockbrokers while making the stock predictions. Programming language used to predict the stock market using machine learning is python. Here in this project we propose a machine learning approach that will be trained from available stocks data and gain intelligence and then use acquired knowledge for an accurate prediction.*

Date of Submission: 14-02-2022   Date of acceptance: 28-02-2022

I. INTRODUCTION

We all heard about the word stock one way or the other. Particularly stock is related to the associates and companies which are commercialized.An another word used for stock is share which is prominently used in day to day life. People even call it as an investment plan and its something people see as a long term investment that gives hope of security. Getting a company stock is buying a small share of it. People invested money on the same to get a future benefit. Its an investment that gives benefits in the future. Market is a place you cannot predict anything, so are the resources and factors. Some sort of prediction method is derived and it gives approximate values hoping for results ,but the resources cannot be trusted and are still not predictable in real. They predict the values and advise ,but the cost and the charge is high and the stock is never less the same. Market is changing in instantaneous rate. Even in a day there are so many   highs and lows in market. Stock is the best resource to start with. Stock in other word is said as the fair share or   ownership telling about the security measures. Stock is there from beggining and due to its uncertainty it been a word of fancy. There are so many agents available in the market to make you understand and invest the same.The main resources of the company is fund to carry out the daily work and create a profit from it. In need for an higher budget prediction and to overgrow from the resources they need the finance and approval. Stock is other way for company to collect money and cheer up the production for upper yield and to gain most out of the business plan for the bigger betterment. This is found to be a best way to invest and grow in this commercial field and it is the better alternative to reduce the financial crisis for the requirement .For investor its a risk where they invest their savings and hope it gives more yield in the future.. In Indian Society it is even considered as a point of business and people believe it as luck. When a person buys a company stock and then they're called as a shareholder and they will get a share from the same as they have invested in profit or the gain.. They can share their stock to other respective brokers, as there are many brokers available in the market.

1.1 PROBLEM DEFINITION

Stock is an unpredictable curve from ever since. It became popular along with time. People are interested even in the back days. Organization declared it as a better source of revenue generation rather than investing and taking a loan approval from the bank. It's way better and very less hectic from the point of view of firm. Stock is not predictable and it was the same from beginning. It has its superiority in the case of changes in the market along the time. Its easily said in some words but understanding it takes more time and it is harder.

The values fluctuate. Getting the closest and accurate value is the biggest problem itself. Considering all the data together and trying to predict correct values to increase efficiency is tough,

So taking all these points into consideration we implement the best. We thought of all the things and tried to reduce the risk. Now this is related to machine learning algorithm. Using machine learning how to get it is the problem we solved.

1.2 PROJECT PURPOSE

Stock market prediction is prediction system software that explains the risk that comes in the investment stock market. It predicts the stock rates. Data is taken as the digital fuel that gives the possibilities of higher accurate values. Knowledge is power and same holds correct with the stock. Stock is not predictable. The rise and fall is also not predictable. The stock is tough and hectic in real. The main theme of the project is to predict the rise and fall and bring the predictability method and undergo the process and use algorithms to conclude the values. Everything flows as a pattern. Pattern is the way of deriving. Stock follows a pattern in day to day life. Increase in some resource can increase the price of some whereas decrease the price rate for the others, The source and the outcome are derived on the basis of polarity which can either be positive, neutral or an negative flow. This project helps in getting all the resources together and giving information to the people to know and get the most out of stock and understand the generation and the problems that has to be seen and predicted. This is done by the analysis and the graphs we get from the analysis. The Predicting graph itself is a challenge and that's the main purpose of the report.

1.3 PROJECT FEATURES

Features deals with the flexibilities and the top marks that one can present. The project was headed with available resources and the most that the company demands and that is finance. Talking about finance and learning on the same gave an idea on the stocks. So the featuring of the idea came with handling and automating the resource which other agents are making fortune out of it.

Knowledge and learning is the curiosity where as outcome is the expectation so the resource deals with the getting multiple machine learning algorithms to learn, process and get the result to get correct outputs. There are various ways people ar trying to get the most accurate values. The efficiency percentage changes as there is a change in the stock market and it's prediction. The project is whole based on the values of data set and getting the values from them using the algorithms. By reducing all the risk factors.

1.5 MODULES DESCRIPTION
1.5.1 DATA SET

This is basic module before starting of project. The dataset is a group of data that are grouped together to give the data variations in a period of time, to estimate and the source of the resources and its outcome for the later time of evaluation. It generates the result optimization and gives a feasible time period to customize and get the flow to the derivation. Finding the best resource out of all resources and getting correct outputs from the resource data set.

1.5.2 DATA ABSTRACTION

Abstraction is finding the best resource and learn about it .Abstraction of the data is main part of the flow. .Abstraction of the data set is to set the data set and find the best constraint from it to take into consideration and the unwanted resources are waste data set which will be thrown away and the main dataset is created to get the outputs.Data are cleared on this level for the beginning of the process. This is all done to get a better and effective values . This is a feature abstraction module to extract the featuring of the dataset. This is a feature model process where all the feasible resources are categorized and the same will be in use for the featuring.

1.5.3 TRAINING DATASET

After the Abstraction of data. The data which is abstracted should be sent to machine. Machine should be trained for which we use training data. There are number of machine learning algorithms. Here we are using supervised learning. There are different datasets sent as input, every resource will be taken into consideration. After sending the data set machine learns all about the data. After learning, it integrates the information and it creates a pattern.

1.5.4 TEST DATASET

After learning from the data, these sets of data are obtained as result. At every testing phase results are generated. It is called as testing phase. Now new set of data sets are taken like training data and the efficiency is

calculated. After the process the efficiency is measured and calculated with others. To get accuracy various batches are implemented. Efficiency of the system is must for the prediction.

1.5.5 RESULT EVALUATION

Evaluation is the main part of any implementation of the project. Evaluation is the key point to the success. All calculations and implementation are compulsory to get the results. The evaluation, utilization and implementation undergoes a various level of extraction and evaluation. The main aim is to get accurate and perfect outputs without any problems. From the starting to the final the process is categorized, supervised and efficiency is check and the working is undergone. Testing is done and it's evaluation are done. The process undergoes the same for various time and phase.The remarks are to be noted and further work is done on the same with the implementation. After the evaluation the results are calculated and from the obtained results the graphs are drawn for more understanding of the results.

II.   Related Work

The literature attempting to prove or disprove the efficient market hypothesis can be classified into three strands, according to the choice of variables and techniques of estimation and forecasting. The first strand consists of studies using simple regression techniques on cross-sectional data (Enke et al., 2011; Ma & Liu, 2008; Khan et al., 2018; Ivanovski, 2016; Sen & Datta Chaudhuri, T, 2016c. The second strand of the literature has used time series models and techniques to forecast stock returns following economic tools like autoregressive integrated moving average (ARIMA), Granger causality test, autoregressive distributed lag (ARDL), and quantile regression (QR) to forecast stock prices (Ariyo et al., 2014; Jammalamadaka et al., 2019; Jarrett & Kyper, 2011; Mondal et al., 2014; Sen & Datta Chaudhuri, 2017; Xiao et al., 2014). The third strand includes work using machine learning, deep learning, and natural language processing for the prediction of stock returns (Mostafa, 2010; Dutta et al., 2006; Mehtab & Sen, 2019; Mehtab & Sen, 2020a; Mehtab & Sen, 2020b; Mehtab & Sen, 2020c; Mehtab et al., 2020d; Mehtab et al., 2020e; Mehtab & Sen, 2021; Porshnev et al., 2013; Obthong et al., 2020; Sen, 2018d; Tang & Chen, 2018; Wang et al., 2018; Zhou & Fan, 2019; Wu et al., 2008). Among some of the recent propositions in the literature on stock price prediction, Mehtab and Sen have demonstrated how machine learning and long- and short-term memory (LSTM)-based deep learning networks can be used for accurately forecasting NIFTY 50 stock price movements in the National Stock Exchange (NSE) of India (Mehtab & Sen, 2019). The authors used the daily stock prices for three years from January 2015 till December 2017 for building the predictive models. The forecast accuracies of the models were then evaluated based on their ability to predict the movement patterns of the close value of the NIFTY index on a time horizon of one week. For testing, the authors used NIFTY 50 index values for January 2018 till June 2019. To further improve the predictive power of the models, the authors incorporated a sentiment analysis module for analyzing the public sentiments on Twitter on NIFTY 50 stocks. The output of the sentiment analysis module is fed into the predictive model in addition to the past NIFTY 50 index values for building a very robust and accurate forecasting model. The sentiment analysis module uses a self-organizing fuzzy neural network (SOFNN) for handling non-linearity in a multivariate predictive environment.

2.1 MACHINE LEARNING

Machine Learning is the most finest word which is heard these days. Machine learning has the most important place in today's technology . It is much developed and evaluated but still it is not the end its still in the progress. machine learning bought a drastic change in this generation because of its automation. It's an aspiring term in todays time.Its a leading way to reduce the labor and increase and speed up the work and even we get the accurate values of whatever we want. Since Machine is considered most efficient and the level of mistakes are kept at the minimum .but also it became a threat because unemployment rate increased. ML is the abbreviation for Machine Learning. In other word it is making a human mind fitting inside a machine which uses the same to perform the task of thousands. Machine Learning consists of different algorithms. ML is a part of an AI(Artificial Intelligence).In this checking of test cases plays a important role. Multi-tasking and processing is also initiated by the same giving a dual output which a human can never ever possibly be able to. Statistics is the major key role in driving the machine learning in figure. Using machine learning we get the perfect efficiency. Its completely digital Various computational fields like Data Mining, Statistical Analysis, Optimization of resources, Automation are a major part of it. Machine can process the results on its own like human. If we cannot the exact value we can get the closest value. The classification of the same can be listed as follows:

2.1.1 SUPERVISED LEARNING
Supervised Learning deals with the supervision of the machine to derive the
Required input. It's a model where all the inputs and outputs are already known and its passed to the machine to get the expected output so that the efficiency is determined and this is the learning phase for the machine. Compute it and stores it into its memory for further process and if found a matching pattern it uses the same and learns from it and plot a result out of the same.

This is a dependent process. The machine totally depends on the user who has to feed the inputs and has to check the efficiency of the same and correct it with the flow of iteration. It's an ANN network. Up in the above figure There's an input vector and the output vector. The input vector derives and gives an output flow of the output vector. If the error signal is generated then the iteration is undergone some wrong process where as lacking of the same means the output field is derived and the output result is accurate and no modification needs to be undergone for same.

2.1.2 UNSUPERVISED LEARNING
Unsupervised learning deals with learning by itself. It is also known as Self
Learning Algorithm. Here only the input vector is passed. So the results deals with the input factors. Here the input factors are put together. Now they are formed as cluster .Cluster is the main thing in this technique. Test Data are passed and with the iteration of the same it learns from it derives itself more closer to the conclusion part. Machine itself should label the data set and classify, categorize all of them. Cluster and Communalization is the main essence of it.

As described in the above figure, In this ANN network when the input is
Given and the input is processed by the function ,the output had to be self derived and it should be matched with the cluster set to get results. If the result lacks the interpretation then it undergoes the iteration. All the data sets are formed and combined in a cluster set for the effective uses of the same in further cases. Feedbacks are not reciprocated incase of such it responds to commonalties. If the commonalties is found between the dataset then it applies the previous functionalities and derive the data. If not set then it learns and identifies for the others.

2.1.3 REINFORCEMENT LEARNING
In this type of learning a reinforced strategy is used. It's neither Supervised nor Unsupervised form of learning. Here in this technique they use dynamic techniques to get outputs . In these sort of algorithm set they don't assume the environmental set. These are even used in higher and complex mechanism finding likes genetic algorithm. These algorithms are mostly used in automation and also are used in Games and Automation of the vehicle resources.
As described in the figure the input vector is passed to a ANN model where the functionalities of the same are stored. If the accurate output is derived then a reward is given to the user making it go to the next level for further task of completion. If not then the Error signal is generated for the same. Then accuracy level is calculated and passed down to the user stating the same. Percentage of success is checked and success ladder is built. Machine iterates the same and to the error signal an add on of reinforced signal is passed which the machine learn and iterates on the same to get closer to the actual results.

2.1.4 CLASSIFICATION ALGORITHMS
2.1.5 TREE CLASSIFIERS
●   This classifier is mostly the graphic depiction of all probable resolutions to make a decision.
●   Choices are based on some situations. Decisions made can be easily explained.
For e.g. Let us make some imaginable decisions for playing badminton

2.1.6 LOGISTIC REGRESSION
●   Logistic regression is certainly one of the extreme devices gaining knowledge of algorithms, which comes under the supervised studying technique. It is used for awaiting the specific established variable using a given set of impartial variables.
●   Logistic regression expects the output of a definite structured variable. Therefore, the final results have to be a categorical or discrete fee. It may be both yes or no, 0 or 1, proper or false, and so forth. But in its location of giving the precise fee as 0 and 1, it offers the probabilistic values which lie between 0 and 1.
●   Logistic regression can be used to categorize the observations the usage of distinct sorts of records and can without difficulty manage the maximum actual variables used for the class.

2.1.7 SUPPORT VECTOR MACHINE

Aid vector system or SVM is one of the maximum commonplace supervised learning algorithms, that's cast-off for enterprise as well as regression problems. Nevertheless, principally, that is used for association difficulties in device studying.

The region of the SVM technique is to generate the finest mark or preference border that may separate n-dimensional interstellar into training so that we can definitely place the new facts opinion in the appropriate organization inside the future. This best choice boundary is known as a hyperplane.

SVM selects the intense factors/vectors that help in making the hyperplane. Those exciting cases are referred to as provision vectors, and therefore algorithm is called as guide vector machine.

SVM can be of two kinds:

O   Linear SVM: linear SVM is largely used for linearly divisible data, this means that that if we've got a dataset, it may be categorized essentially into two classes by the usage of a unmarried immediately line, then such facts is called as linearly divisible statistics, and the classifier that is used is called as linear SVM classifier.

O   Non-linear SVM: non-linear SVM is used for non-linearly divided information, which means that that if we have a dataset, it can't be categorised by using a directly line, then such data is known as non-linear data and classifier used is named as non-linear SVM classifier.

2.1.8 KNN

•   K-nearest neighbour is one of the meekest gadget mastering algorithms primarily based on supervised mastering techniques.

• K-NN algorithm undertakes the resemblance between the new case/information and current cases and put the new case into the organization that is finest similar to the existing organizations.

2.2 Deep Learning Models

        In this section, we discuss two deep learning-based regression methods: (i) the long- and short-term memory

        (LSTM) network, and (ii) the convolutional neural networks (CNNs).

        Long- and Short-Term Memory Network: LSTM is a variant of recurrent neural networks (RNNs) - neural networks with feedback loops (Geron, 2019). In such networks, output at the current time slot depends on the current inputs as well as the previous state of the network. However, RNNs suffer from the problem that these networks cannot capture long-term dependencies due to vanishing or exploding gradients during backpropagation in learning the weights of the links (Geron, 2019). LSTM networks overcome such problems, and hence such networks are quite effective in forecasting in multivariate time series. LSTM networks consist of memory cells that can maintain their states over time using memory and gating units that regulate the information flow into and out of the memory. There are different variants of gates used. The forget gates control what information to throw away from memory. The input gates are meant for controlling the new information that is added to the cell state from the current input. The cell state vector aggregates the two components - the old memory from the forget gate, and the new memory from the input gate. In the end, the output gates conditionally decide what to output from the memory cells. The architecture of an LSTM network along with the backpropagation through time (BPTT) algorithm for learning provides such networks a very powerful ability to learn and forecast in a multivariate time series framework. We use Python programming language and the Tensorflow and Keras deep learning frameworks for implementing LSTM networks. While building the LSTM models, we use the open price of the stock as the response variable, and the variables high, low, close, volume and NIFTY, are used as the predictors. Unlike the machine learning techniques, for the LSTM models, we don't compute the differences between successive slots. Rather, we forecast the open value of the next slot based on the values of the response and the predictor variables in the previous slots

2.2.1 Convolutional Neural Networks:

        CNNs emerged from the study of the brain's visual cortex, and they have been used in image recognition since the 1980s. In the last few years, thanks to the increase in computational power, the amount of available training data, and the tricks for training deep neural networks. CNNs havemanaged to achieve superhuman performance on some complex visual tasks. They power image search services, self-driving cars, automatic video classification systems, and more. Moreover, CNNs are not restricted to visual perception: they are also successful at many other tasks, such as voice recognition, natural language processing, and complex time series analysis of financial data (Binkowski et al., 2017;Lahmiri, 2014).In the present work, we exploit the power of CNN in forecasting the univariate and multivariate time series data of Godrej Consumer Products stock. CNNs have two major types of processing layers – convolutional layers and pooling or subsampling

layers. The convolutional layers read an input such as a 2-dimensional image or a one-dimensional signal using a kernel (also referred to as the filter) by reading the data in small segments at a time, and scan across the input data field. Each read result is an interpretation of the input that is projected onto a filter map and represents an interpretation of the input. The pooling or the subsampling layers take the feature map projections and distill them to the most essential elements, such as using a signal averaging (average pool) or signal maximizing process (max pool). The convolution and pooling layers are repeated at depth, providing multiple layers of abstraction of the input signals. The output of the final pooling layer is fed into one or more fully-connected layers that interpret what has been read and maps this internal representation to a class value. We use the power of CNN in multi-step time series forecasting in the following way. The convolutional layers are used to read sequences of the input data, and automatically extract features. The pooling layers are used for distilling the extracted features, and in focusing attention on the most salient elements. The fully connected layers are deployed to interpret the internal representation and output a vector representing multiple time steps. The benefits that CNN provides in our time series forecasting job are the automatic feature learning, and the ability of the model to output a multi-step vector directly.

We build three different types of CNN models for multi-step time series forecasting of stock prices. They are: (i) Multi-step time series forecasting with univariate input data, (ii) Multi-step time series forecasting with multivariate input data via channels – in this case, each input sequence is read as a separate channel, like different channels of an image (e.g., red, green, and blue), (iii) multi-step time series forecasting with multivariate input data via sub-models – in this case, each input sequence is read by a different CNN sub- model and the internal representations are combined before being interpreted and used to make a prediction. In the first case, we design a CNN for multi-step time series forecasting using only the univariate sequence of the open values. In other words, given some number of prior days of open values, the model predicts the next standard week of stock market operation. A standard week consists of five days – Monday to Friday. The number of prior days used as the input defines the one-dimensional (1D) data of open values that CNN will read and learn for extracting features.

The multi-step time series forecasting approach is essentially an autoregression process. Whether univariate or multivariate, the prior time series data is used for forecasting the values for the next week

2.2.2 Long- and Short-Term Memory Network:

LSTM is a variant of recurrent neural networks (RNNs) - neural networks with feedback loops (Geron, 2019). In such networks, output at the current time slot depends on the current inputs as well as the previous state of the network. However, RNNs suffer from the problem that these networks cannot capture long-term dependencies due to vanishing or exploding gradients during backpropagation in learning the weights of the links (Geron, 2019). LSTM networks overcome such problems, and hence such networks are quite effective in forecasting in multivariate time series. LSTM networks consist of memory cells that can maintain their states over time using memory and gating units that regulate the information flow into and out of the memory. There are different variants of gates used. The forget gates control what information to throw away from memory. The input gates are meant for controlling the new information that is added to the cell state from the current input. The cell state vector aggregates the two components - the old memory from the forget gate, and the new memory from the input gate. In the end, the output gates conditionally decide what to output from the memory cells. The architecture of an LSTM network along with the backpropagation through time (BPTT) algorithm for learning provides such networks a very powerful ability to learn and forecast in a multivariate time series framework. We use Python programming language and the Tensorflow and Keras deep learning frameworks for implementing LSTM networks. While building the LSTM models, we use the open price of the stock as the response variable, and the variables high, low, close, volume and NIFTY, are used as the predictors. Unlike the machine learning techniques, for the LSTM models, we don't compute the differences between successive slots. Rather, we forecast the open value of the next slot based on the values of the response and the predictor variables in the previous slots. We use the mean absolute error (MAE) as the loss function and the adaptive moment estimation (ADAM) as the optimizer for evaluating the model performance in all three cases. ADAM computes adaptive learning rates for each parameter in the gradient descent algorithm. In addition to storing an exponentially decaying average of the past squared gradients, ADAM also keeps track of the exponentially decaying average of the past gradients, which serves as the momentum in the learning process. Instead of behaving like a ball running down a steep slope like momentum, ADAM manifests itself like a heavy ball with a rough outer surface. This high level of friction results in ADAM's preference for a flat minimum in the error surface. Due to its ability to integrate adaptive learning with a momentum, ADAM is found to perform very efficiently in optimizing the performance of large-scale networks. This was the reason for our choice of ADAM as the optimizer in our LSTM modeling. We train the LSTM networks using different epoch values and batch sizes for the three different cases. The sequential constructor in the Tensorflow framework is used in building the LSTM model. The performance results of the LSTM models

2.3 EXISTING SYSTEM

As many people have put their time and effort in this world trade for getting developed and changing their lifestyle. In the past few years various methods and the plans are derived and deployed ever since it's continuation and the topic is still a point of research where people are coming up with new ideas to solve. Intelligence fascinates mankind and having one in a machine and integrating on the same is the hot key of research. There are various people giving so much effort on the same research. All the learning system from the past are limited and are simplest in nature but it cannot be done with the human brain.The main motto of learning was limitized and the learning model was not efficient. The existing models can't cope up with the vulnerabilities and remove the rarest information that they can't process causing it a major data loss which creates a problem. Observation is the integral part in the resource and prediction management. If the outcome can't be observed then it's a matter of time for the existing system to be wasted. The existing system is biased. Before the prediction of the data set a simple data retrieval should be generated and tested on the training data set which are more flexible and versatile in nature. As stock varies day to day ,the loss of sights is the major problem in the existing system.

2.4 PROPOSED SYSTEM

Stock is unpredictable and liberal in nature. The following is impressive and reluctant in nature. Finding the predictability and getting the nearest and closest value is the best goal. It is not at all possible to get the exact value. There are various other things that stock prices depend upon.Those constraints had to be taken in consideration before jumping to the conclusion and report derivation.

Here as described in the figure above, the proposed system will have an input from the dataset which will be extracted, featured wise and Classified. The classification technique used here is supervised and the various techniques of machine level algorithms are implemented on the same. Training Dataset are created for training the machine and the test cases are derived and implemented to carry out the visualization and the plotting. The result generated are passed and visualized in the graphical form.

III.   REQUIREMENT ANALYSIS

3.1 FUNCTIONAL REQUIREMENTS

Functional requirements deal with the functionality of the software in the engineering view. The component flow and the structural flow of the same is enhanced and described by it. Functional requirements needs raw datasets that are categorized. Later the datasets are grouped as clusters and efficiency is checked. After the checking of efficiency the pattern is formed and then the outputs are obtained.

3.2 NON-FUNCTIONAL REQUIREMENTS

Non- functional requirement deals with the external factors which are non- functional in nature. These are required for analysis purpose. Here only the performance is checked.Stock is feasible and it always changes its values,so these are used and here to check. Efficiency is checked. The usability of the code for the further effectiveness and to implement and look for the security.The System is reliable and the performance is maintained with the support of integration and portability of the same.

3.3 HARDWARE REQUIREMENTS

Processor : Intel i5 or above
RAM : Minimum 225MB or more.
Hard Disk : Minimum 2 GB of space
Input Device : Keyboard
Output Device : Screens of Monitor or a Laptop

3.4 SOFTWARE REQUIREMENTS

• Operating system : Windows
• IDE : Google Collaborator
• Data Set : .csv file
• Visualization : mat plot lib, pandas.
• Server : Web Server with HTTP process.

IV.   DESIGN

4.1 DESIGN GOALS

To make the project run smoothly it's required that we make plan and design some aspects like flowcharts and system architecture which are defined below.

4.1.1 Data Collection
Data collection is one of the important and basic thing in our project. The right dataset must be provided to get right results. Our data mainly consists of previous year stock prices. We will be taking these values and analyzing data using the code. After that seeing the accuracy we will use the data in our model.

4.1.2 Data Pre-processing
Humans can understand any type of data but machines cannot understand. Machine readable data must be made so the machine can learn from the beginning. Raw data is generally incomplete and inconsistent. Data pre-processing involves checking missing values, splitting the dataset and training the machine etc.

4.1.3 Training Model
Similar to feeding some things, machine/model should also learn by feeding and learning on data. The data set extracted will be used to train the model. The training model uses a raw set of data as the undefined dataset which is collected from the previous fiscal year and from the same dataset a refine view is presented which is seen as the desired output. For the refining of the dataset various algorithms are implemented to show the desired output.

4.2 SYSTEM ARCHITECTURE
The dataset we use for the proposed project is been taken from Apple, Google, Tata. But, this data set is in raw format. The data set is a collection of valuation of stock market information about some companies. The first step is to convert raw data into processed data. It is done by feature extraction, as we need only some data and other is not used. Feature extraction is a reduction process. The structure, behaviour and views of a system is given by structural model.
The above figure 4.1 gives the demonstration on the dataset extraction and refining the raw dataset by categorizing into two phase of training and testing data.
From the given dataset a well modified categorization is extracted and a graph set is plotted to gain the required output which gives the stock prediction range.

4.3 USE CASE DIAGRAM
      A dynamic and behavioral diagram in UML is use case diagram. Use cases are basically set of actions, services which are used by system. To visualize the functionality requirement of the system this use case diagram are used. The internal and external events or party that may influence the system are also picturized. Use case diagram specify how the system acts on any action without worrying to know about the details how that functionality is achieved. For the project we have created the below mentioned use case diagram.
      The above figure 4.2 shows the use-case diagram of the entitled project and it's flow. From the diagram it's seen that the user gives the raw dataset as input and with the flow of the input in the system. The system evaluates and the dataset and process it ,then train itself with the provided dataset and extract the meaningful dataset to process and refine the cluster data and from the given cluster of the data, the plotting of the data values are shown and with the given range the system plots the data gives a figurative output as prediction and display the same as the refined output in the display screen.

V. IMPLEMENTATION
These are the Machine Learning Algorithms implemented during the building of the project.

5.1 LONG SHORT TERM MEMORY (LSTM)
      In complete data science industry , sequence prediction problems are the hardest problems to solve. These problems includes things like finding patterns in stock markets and predicting sales. Also different things like language translations to predicting next word on keyboard. For all these problems LSTM has been observed as most effective solution. LSTM has capacity of remembering all patterns for long duration of time. LSTM may also make some additions in the patterns. LSTM flows through mechanism known as cell states. According to this LSTM can remember or forget things selectively. It uses this mechanism move information around.

VI. TESTING
      The purpose of testing is to find errors. Testing is process of getting every possible error during process .It looks at the functionality of the algorithms in the machine. There are different forms of testing. Each test type have different requirement. The various types of testing that follows are listed as below.

6.1 UNIT TESTING

Unit testing involves testing weather the programming logic is functioning correctly or not which program inputs produce valid outputs. All the things should be validated. It is about testing individual units and structural units. Unit tests make sure that every path of the process performs
correctly according to the specifications and have correct inputs and outputs.

6.2 INTEGRATION TESTING

Integration tests checks the integrated software stuff which are used to run the program. Testing is more concerned about the outcomes. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the mixture of components is correct and consistent. Integration testing tells about the issues faced due to mixing of components.

6.3 VALIDATION TESTING

Validation testing is the process which validates the functionality and checks weather it reaches the demands of humans. The business requirement logic or scenarios need to be tested intimately. All applications functionality is tested correctly.As a tester, it's always important to understand the way to verify the business logic or scenarios that are given to you. One such method that helps intimately evaluation of the functionalities is that the Validation Process.

6.4 SYSTEM TESTING

System testing of software or hardware is testing conducted on an entire , integrated system according to the requirements. All the logics are checked. As a rule, system testing takes, as its input, all of the "integrated" software components that have successfully passed integration testing and also the software itself integrated with any applicable hardware system(s).

System testing may be a more limited sort of testing; it seeks to detect defects both within the "inter-assemblages" and also within the system as an entire . System testing is performed on the whole system within the context of a Functional Requirement Specification(s) (FRS) and/or a System Requirement Specification (SRS)

System testing tests not only the planning , but also the behaviour and even the believed expectations of the customer.It's also intended to check up to and beyond the bounds defined within the software/hardware requirements specification(s).

VII. CONCLUSION AND FUTURE ENHANCEMENT

7.1 CONCLUSION

To conclude stock is an unpredictable mechanism which follows the segments of chain and the dependencies of the same are unpredictable. It is defined to be an curve which keeps on changing and turning the price from low to high and vice-versa. As the integration of the same is higher with other dependencies so leaving one dependencies compromises the level of accuracy. Accuracy is not the term used over in stock as the actual prediction is not possible for any fiscal days it keeps on changing and turning the tables day and night. Having higher component assets and the dependencies makes it more feasible and flexible in nature causing it even harder to predict. The approx value are taken into consideration and the hit or profit or the gain rate is calculated for the same.In the project various high level machine learning algorithms are implemented and integrated and the output is generated from the same making a user visible with the outputs in the form of graph which makes it easier for them to see and interpret what's the scenario and they can decide on the same to invest and get the benefit out of it, The proposed software takes the raw set of data from the dataset or the .csv file and process it. The cleaning and cleansing of data is done and then further processed to gain the effective outcomes. After the computational mean the output is displayed in the screen in the form of graph.

7.2 FUTURE ENHANCEMENT

Stock Market are the best alternative for business to grow and it's a side way income for the individuals who are ready to invest and earn from the same. The term stock had been in picture ever since and it's growing in bulk everyday. There are thousands of investors investing on the same and making the fortune out of it.

There are middle level agents and stock vendors who learn and invest on the same. The cost for the consultation on the stock is bulky and expensive. So when it comes to people they think a lot and invest and there's no chance and certainty for the same to produce a yieldful result. So stock being unpredictable and the tendency of its growth is higher than ever. If the stock market and its prediction can be done accurate than it's going to be a gain for both the individuals and the organization. The risk factor have to be mitigated so the efficiency of the system should be high and people can be certain about their investment in time.The project can be further continued to gain the effectiveness of the prediction with addition implementations of the content that

can involve real time scenario and the way of executing and processing the real time scenario. Various constrains has to be added and performance of the same can be acylated in the future time for the effective results. The expected form of the display is graph where as from the same the more appearance and setting of the display can be integrated and a pie-chart and a custom graph can further me implemented on the same

Reference

[1]. Ariyo, A. A., Adewumi, A. O. and Ayo, C. K. (2014) 'Stock price prediction using the ARIMA model', 16th International Conference on Computer Modelling and Simulation (UKSIM'14), Cambridge, UK, 26-28 March, pp. 106 – 112. DOI:10.1109/UKSim.2014.67.

[2]. Asghar, M. Z., Rahman, F., Kundi, F. M. and Ahmed, S. (2019) 'Development of stock market trend prediction system using multiple regression', Computational and Mathematical Organization Theory, Vol. 25, pp. 271-301, DOI: 10.1007/s10588-019-09292-7.

[3]. Basalto, N., Bellotti, R., De Carlo, F., Facchi, P. and Pascazio, S. (2005) 'Clustering stock market companies via chaotic map synchronization', Physica A: Statistical Mechanics and its Applications, Vol. 345, Nos. 1 – 2, pp. 196-206, DOI: 10.1016/j.physa.2004.07.034.

[4]. Bao, W., Yue, J. and Rao, Y. (2017) 'A deep learning framework for financial time series using stacked autoencoders and long-and-short-term memory', PLOSE ONE, Vol. 12, No. 7, DOI: 10.1371/journal.pone.0180944

[5]. Bentes, S. R., Menezes, R. and Mendes, D. A. (2008) 'Long memory and volatility clustering: Is the empirical evidence consistent across stock markets?', Physica A: Statistical Mechanics and its Applications, Vol. 387, No. 15, pp. 3826-3830, DOI: 10.1016/j.physa.2008.01.046.

[6]. Binkowski, M., Marti, G. and Donnat, P. (2018) 'Autoregressive convolutional neural networks for asynchronous time series', 35th International Conference on Machine Learning (ICML'18), 10-15 July, Sydney, Australia.

[7]. Chen, A.-S., Leung, M. T. and Daouk, H. (2003) 'Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index', Computers and Operations Research, Vol. 30, No. 6, pp. 901– 923, DOI: 10.1016/S0305-0548(02)00037-0.

[8]. Chen, Y., Dong, X. and Zhao, Y. (2005) 'Stock index modeling using EDA based local linear wavelet neural network', International Conference on Neural Networks and Brain (ICNNB'15), Beijing, China, 13 – 15 October, pp. 1646 – 1650, DOI: 10.1109/ICNNB.2005.1614946.

[9]. de Faria, E. L., Albuquerque, M. P., Gonzalez, J. L., Cavalcante, J. T. P. and Albuquerque, M. P. (2009) 'Predicting the Brazilian stock market through neural networks and a daptive exponential smoothing methods', Expert Systems with Applications, Vol. 36, No. 10, pp. 12506-12509, DOI: 10.1016/j.eswa.2009.04.032.

[10]. Dutta, G., Jha, P., Laha, A. K. and Mohan, N. (2006) 'Artificial neural network models for forecasting stock price index in the Bombay Stock Exchange', Journal of Emerging Market Finance, Vol 5, No 3, pp. 283- 295, DOI: 10.1177/097265270600500305.