Stock Market Prediction Using Artificial Intelligence.

Mrs.Sumitra Samal[1], Prachi Agrawal[2], Vidya Laxmi Sahu[3]

[1]Assistant Professor, CSE Dept, [23]B.E,CSE Dept.,
SSIPMT, Raipur, Chhattisgarh

Abstract- Stock market prediction is noteworthy realm to find the capabilities of artificial intelligence network and machine learning. The nature of the stock market is volatile, refined, and sensitive to external factors, which makes it difficult to predict. High level precision and accuracy is required to predict the stock prices. Different machine learning are developed to forecast future stock prices. The artificial neural network ANN is a useful tool to predict the prices for a specific stock or multiple stocks. Using historical stock market data from the finance.google.com or yahoo, technical indicators are computed and used along with a stock's price as features associated with a target output, which is the future stock price. This provides the dataset to predict the data using the various neural network models. The models used are: linear regression, kNN regressor, Feedforward Neural Network (FFNN), and Long Short Term Memory (LSTM) Recurrent Neural Net-work (RNN). The prediction models are compared and evaluated using different metrics. Several case studies are performed to judge the performance of the ANN models. From the case studies, few insights are created. Firstly, the LSTM RNN outperformed all the opposite models. Secondly, the LSTM RNN model is capable of accurately predicting the next-day value unless a serious external event impacts the stock value suddenly. Lastly, the LSTM RNN model naturally lags on learning on external events that impact the stock value suddenly.

Index Terms- Market analysis, ANN machine learning, stock, stock prediction etc.

I. INTRODUCTION

A stock is a piece of a publicly-traded company. A person who owns one or more shares is called a shareholder. A shareholder is a partial owner of the underlying company. A shareholder is entitled to a set of rights including, but not limited to voting power on major issues, ownership in a portion of the company, the right to transfer ownership and entitlement to dividends. A dividend is a distribution of a portion of a company's earnings, decided by the board of directors, paid to a class of its shareholders. Dividends can be issued as cash payments, as shares of stock, or other property. Dividends are decided by a company's board of directors. Dividends are used in times of profit to attract and keep investors. The more stable the company, the more dividends it would offer. On the other hand, a higher growth company will rarely distribute dividends in order to reinvest in the company to maintain its high growth. Dividends are based on a company's earnings.

Stocks can be described in different ways. Some of the measures that can be used include:

Market Capitalization: Market capitalization refers to the total market value of a company's outstanding shares. Commonly referred to as market cap, it is calculated by multiplying a company's outstanding shares by the current market price of one share. The investment community [1-2]uses this figure to determine a company's size, as opposed to using sales or total asset. Usually companies with market cap fewer than 1 billion dollars, between 1-5 billion dollars, over 5 billion dollars are classified as small-cap, mid-cap, large-cap respectively. Large-cap companies tend to be much less vulnerable to the ups and downs of the economy, a large part due to their huge financial reserves. On the other hand, small-cap companies are greatly affected by turmoil, but can provide for a lot more growth versus large-caps.

Sectors and Industries : Companies are divided into sectors and industries. A sector represents a large part of the economy. Common sectors include electronics, finance, energy, etc. On the other hand, industries are much more specific and are considered to be a subset of some sector. In terms of stock price movement, industries within the same sector tend to exhibit similar reactions to changes in the economy. In order to minimize risks, it is suggested to make sure that a portfolio, which means an investor's collection of owned stocks, includes companies from different sectors and industries.

Seculars and Cyclical : Stocks [3-5] can be categorized into two different categories: seculars and cyclical. They mainly differ in the way they make profit. Secular stocks consist of stocks of companies that provide essentials. In other words, it provides people needs they cannot live without. For example: consumer staples, and healthcare. Cyclical stocks typically relate to companies that sell inessential items consumers can afford to buy more of in a booming economy and cut back on during a recession. For example: travel, and luxury goods. A cyclical stock price is highly reactive to the economy. It requires strong economy succeed.

Seculars, however, are much less volatile and are stable regardless of the economy.

Stock prediction is the major part in field of finance. Due to its financial gain in very short span of time, it has attracted a lot of people towards it. With the stock prediction tool, we can analyze the viability of the shares. So, let's think whether which shares will help to gain a maximum amount with a minimum investment in a minimum time span[6-7]. The answer to this question is complicated. From one to two decades, the debate is still going on whether we can predict the prices of the shares or not. A number of theories like time series hypothesis, random walk theory, efficient market hypothesis and other researchers have been made on it. From the few years back some application and software have developed like money control, yahoo finance and many more

which helps to predict the data and prices but it is not up to market. Nevertheless finding the best possible may to predict will remain impossible as the prices of the stocks keep fluctuating every hour in a particular day.

The prediction of the market is divided into the three parts:

1. Collecting the data- First of all, we have to collect the data or the prices of the stocks of over 2-3 decades we want to predict. In technical terms we can say it data mining.
2. Analyze the data- with the help to the tools like linear regressor artificial neural network (ANN), fuzzy logic, we need to calculate and find the data we need to plot the graph'
3. Plotting the graph- This is the last part of the stock prediction. Graph is the easiest manner to understand the thing in a least amount of time. A person who does not know the technical terms can also understand it in a glance In this some of the tools are needed to plot the graph of the share we want.

II. OUR WORK

From the above literature review, we can analyse that the stock prediction started from studying the behaviour of the stock to the artificial intelligence using machine learning, data mining etc. Nowadays, there have been different methods on stock prediction like stock movement analysis, stock behaviour analysis, price prediction, human sentiment analysis etc.

In our project we have project we have tried to predict the price of the stock with the help of the neural network, linear regressor, kNN regression and fuzzy logic.

The flow chart of predicting the prices of the stocks is as shown in figure 1
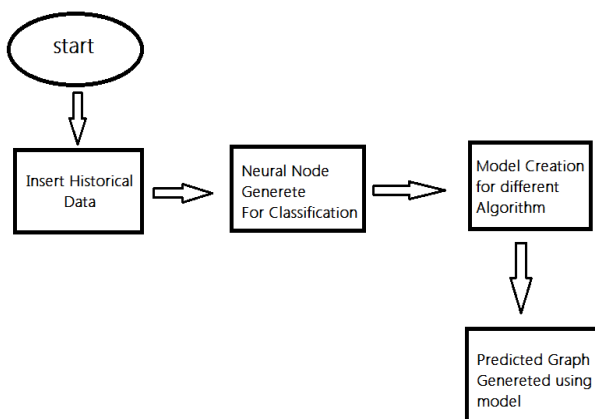

Fig.1: Flow chart

i. In this research, we are mainly interested in
A. Two different types of ANNs: Feedforward Neural Network (FFNN), and Long Short Term Memory (LSTM) Recurrent Neural Network (RNN).
B. FFNN
The FFNN is the first and simplest type of an ANN. It contains multiple neurons arranged in layers. There are connections or edges between nodes from neighboring layers. There are weights associated with all these connections. In a FFNN, the information moves only in the forward direction from the input nodes, through the hidden nodes and to the output nodes. There are no cycles or loops in the FFNN.

C. RNN
RNNs are ANNs with loops in them, allowing information to persist. A loop allows information to be passed from one step of the network to the next. A RNN can be thought of as multiple copies of the same network, each passing a message to a successor. This chain-like nature reveals that a RNN in its core design is related to sequences and lists. RNN's natural architecture makes it optimal for datasets that involve sequences, lists, or time series. One of the applications of RNNs is that they might be able to connect previous information of a previous task to a present task. For example, using previous neurons' inputs might help in the understanding of the present neuron input. However, one important thing to consider is how far are the related tasks from each other. In some cases, only recent information is needed to perform the present task. In such cases, where the gap between the relevant information and the place where it is needed is small, RNNs can learn to use the past information. However, there are cases where more context is needed. It is feasible for the gap between the relevant information and the point where it is needed to become very large. Unfortunately, as that gap grows, RNNs become unable to learn to connect the information.

D. LSTM
LSTMs are a special kind of RNNs. They are capable of learning long-term dependencies. LSTMs are capable of handling the long-term dependency problem. They specifically remember information for long periods of time. All RNNs have the structure of a chain of repeating modules of neural network. What is special for LSTMs is the cell state, the cell state runs straight through the entire chain, with only some minor linear interactions. It is very easy for information to just along it unchanged. An LSTM can remove or add information to the cell state. This is controlled by structures called gates. Gates are a way to optionally let information through. They are constructed out of a sigmoid neural net layer and a point wise multiplication operation. The sigmoid layer outputs numbers between zero and one, that controls how much of each component should be let through. A value of zero lets nothing through, while a value of one lets everything through. An LSTM has these gates to protect and control the cell state.

ii. Model Representation of Machine Learning
A machine learning algorithm requires one or more input variables (denoted as x_i), also called input features, and produces an output variable or prediction target variable (denoted as y_i). A pair (x_i; y_i) is called a training example, and the dataset to be used to learn is a list of n training examples (x_i; y_i); i = 1,…, n, this is called a training set. The subscript (i) in the notation is simply an index into the training set. With a supervised learning problem, the goal is given a training set,

learn a function h : X -> Y so that h(x) is a good predictor for the corresponding value of y. For historical reasons, this function, h, is called a hypothesis.
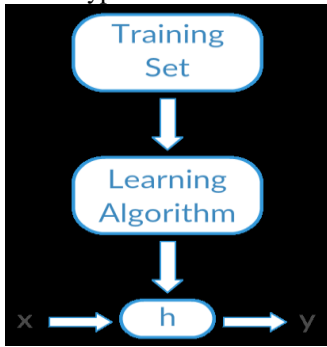

Fig.2: Model Representation of Machine Learning Algorithm

When the target variable to be predicted is continuous, the learning problem is called a regression problem. When the target variable can take on only a small number of discrete values, the learning problem is called a classification problem.

The cost function used in our proposed work is –
The accuracy of a hypothesis function can be measured by using a cost function. A cost function computes an average difference between the results of the hypothesis and the actual output. Mean Squared Error (MSE) is a popular cost function.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(h(x_i) - y_i)^2$$

where:

i = index into the training set
n = number of training pairs
$\hat{y}_i$ = predicted output of index i
y_i = actual output of index i
h = the hypothesis function
x_i = input features of index i

iii. Computational Model of Artificial Neural Network (ANN):
 An ANN is a computational model. It is artificial in the sense that it tries to process information in the same way biological neural networks in the human brain do. The basic unit of computation in an ANN is the neuron, also known as a node or unit. It receives input either from some other nodes, or from an external source, and a bias to provide every node with a trainable constant value. Each input has an associated weight, which is assigned based on its relative importance to other inputs. A node applies a function called the activation function to the weighted sum of its inputs. The activation function serves the purpose of introducing non-linearity into the output of a neuron. The output of the neuron is the output of its activation function. It is important to introduce non-linearity, because most real world data is non-linear and neurons should be capable of learning these non-linear representations. An

ANN consists of three types of nodes: input, hidden, and output.
1. Input Nodes:
The input nodes transmit the dataset into the ANN. Input nodes are referred to as the input layer. No computation is performed in any of the input nodes, they just pass on the dataset to the hidden nodes.
2. Hidden Nodes
The hidden nodes perform computations and transfer information from the input nodes to the output nodes. A group of hidden nodes forms a hidden layer.
3. Output Nodes
The output nodes are collectively referred to as the output layer. They are perform computations and transfer the output out of the ANN.

III. INDIAN STOCK MARKET
In the Indian stock market, there are two stock exchange that is National Stock Exchange (NSE) and Bombay Stock Exchange (BSE). Stock exchange means an organized market where the securities of the government, semi-government and private companies are sold and purchased.[4]
NSE is the 11th largest stock exchange with the 1952 listings and a market capital of US $2.27 trillion as on April 2018.[5]
BSE is the 10th largest stock exchange with the 5439 listings and a market capital of US $2.1 trillion as on April 2018.[6]

IV. CONCLUSION
The Stock Prediction project has been successfully able to plot the graph between the prices and year. This paper also survey on the various techniques being used for the stock prediction.
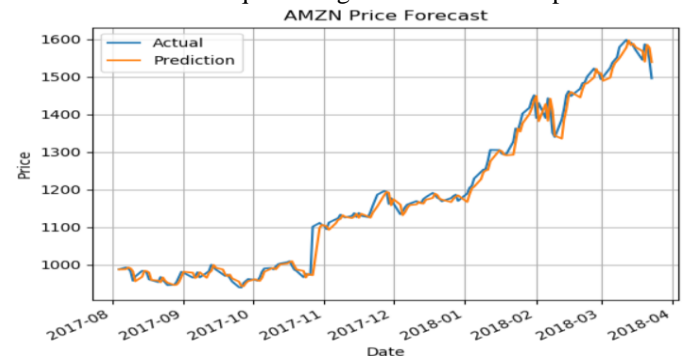

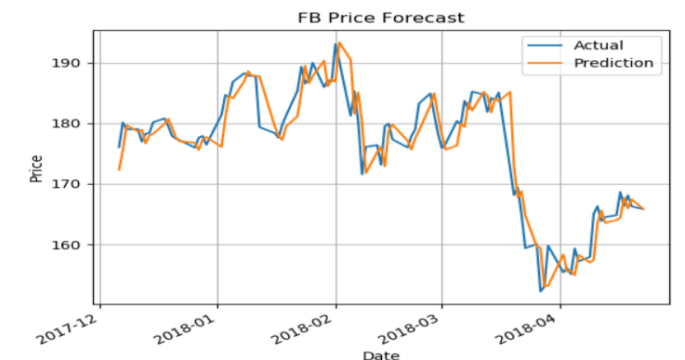Fig.3: Amazon Price Forecast

Fig.4 FAcebook Price Forecast

We can conclude that this paper has brought light to:

- Different methods used till now to develop the stock prediction.
- We can predict the price of stock accurately.
- The Indian stock market has reached to a high level of conjecture.
- There are various methods to predict the stock prices.

V.   REFERENCES

[1]. Stock Market volatility-A study of indain stock market by Sameer Yadav Research scholar Department of Commerce and Business Administration, University of Allahabad, Uttar Pradesh.

[2]. Survey On stock Market Prediction Techniques by Shyam Kute and Sunil Tamhankar.

[3]. An Analysis on Stock Market Intelligence and Research Approaches by Rajesh V. Argiddi, Dr.Mrs.S.S.Apte, Bhagyashri U.Kale

[4]. Stock Market Forecasting Techniques:Literature Survey by Vivek Rajput and Sarika Bobde

[5]. www.optionalpha.com/is-the-stock-market-random-19323.html

[6]. www.nseindai.com/about

[7]. www.bseindia.com/about