



# **Stock Price Prediction using Neural Network**

A Dissertation Proposal

Submitted By

**Pankaj Kumar Sood**

To

**Department of CSE/IT**

In fulfilment of the Requirement for the Award of the Degree of

**Master of Technology** in Computer Science and Engineering

**Under the guidance of**

**Virrat Devesar**

**May/2015**

**PAC Form**

## ABSTRACT

Stock prices prediction is a continuing area of research to get the better accurate result. The main part in this proposal is to determine the network architecture that can give better result. Many models have been proposed from the last few years, but they all suffer a confinement issues and poor performance due to its highly volatile movement. Before the neural network, fundamental and technical analysis, linear regression, fuzzy logic is all utilized to predict it, but none of the technique has got the level of accuracy that is required in the good forecasting tool. The research given in this proposal presents the neural network based approach to predict three month ahead closing prices of companies from the banking sector. A neural network which is artificial intelligence method as well as data mining method is becoming very popular to make stock market predictions because it's ability of self learning. Several research improvements have attempted in this proposal by varying the input parameters with respect to their effect on the target. After doing the various researches on parameter and the working of the neural model, an optimal model is presented in this proposal for the purpose of forecasting. The dataset used to develop the model consist of previous closing price value of Bank Nifty Index and SBI stock. The 1448 days of trading data were used to train the network and forecasted the closing prices up to next upcoming three months. The historical data available at nseindia.com is used to generate the predicted output. The two input configurations with previous closing of SBI bank equity (1-1-1) and with previous close price SBI bank equity as well as bank index (2-1-1) are used to determine the stock price and result of each of input configuration is compared with each other to evaluate the appropriate model. This Dissertation Proposal will attempt to get the accuracy in their predictions which many researchers have been attempted to do for longer period. The accuracy of the model in predicting the future trend of the SBI stock is reported at 78% over a three month period. The main focus in this proposal is on to use the all available learning algorithm, different parameter of a network, different training-testing data set ratio to make accurate network which can further used to get accuracy in the results.

## ACKNOWLEDGEMENT

Apart from the efforts of us, the success of our research work depends largely on the encouragement and guidelines of many others. I take this opportunity to express my gratitude to the people who have been instrumental in the successful completion of this research work.

We would like to show our greatest appreciation to our research work mentor, **Mr. Virrat Devaser**. I can't say thank you enough for the tremendous support and help. I feel motivated and encouraged every time i attended his meeting. Without his encouragement and guidance this research work would not have materialized.

I'm highly grateful to **Mr. Dalwinder Singh**, Head of Department, for his thorough guidance right from day 1 to the end of research work. He actually laid the ground for conceptual understanding of research work.

**Pankaj Kumar Sood**

## DECLARATION

I hereby declare that the dissertation proposal entitled, “**Stock Price Prediction using Neural Network**” submitted for the M.Tech Degree is entirely my original work and all ideas and references have been duly acknowledged. It does not contain any work for the award of any other degree or diploma.

Date:

Investigator

Registration No

## CERTIFICATE

This is to certify that **Pankaj Kumar Sood** has completed M.Tech dissertation proposal titled **Stock Price Prediction using Neural Network** under my guidance and supervision. To the best of my knowledge, the present work is the result of her original investigation and study. No part of the dissertation proposal has ever been submitted for any other degree or diploma.

The dissertation proposal is fit for the submission and the partial fulfilment of the conditions for the award of M.Tech Computer Science & Engineering.

Date:

Signature of Advisor:

Name: Virrat Devaser

UID: 14591

## TABLE OF CONTENTS

| S. No. |  | Page No. |
|--------|--|----------|
| 1.     | INTRODUCTION.....                                      | 01       |
|        | 1.1 Description.....                                   | 04       |
|        | 1.2 Motivation.....                                    | 04       |
|        | 1.3 Neural Network.....                                | 04       |
|        | 1.3.1 Modelling Artificial Neuron.....                 | 04       |
|        | 1.3.2 Artificial Neural Networks Implementation.....   | 07       |
|        | 1.3.3 Biological Neuron.....                           | 08       |
|        | 1.3.4 Learning in Artificial Neural Network.....       | 08       |
|        | 1.3.5 Types of Learning.....                           | 08       |
|        | 1.3.6 Implementing Supervise Learning.....             | 12       |
| 2.     | 1.3.7 Learning Rule of Back Propagation Algorithm..... | 15       |
| 3.     | 1.3.8 Layer Architecture.....                          | 16       |
|        | 1.4 Neural Network Development .....                   | 16       |
|        | 1.4.1 Time frame and the target .....                  | 16       |
|        | 1.4.1 Domain Experts .....                             | 16       |
|        | 1.4.1 Data gathering .....                             | 17       |
|        | 1.4.1 Data pre-processing.....                         | 17       |
|        | 1.4.1 Data Splitting.....                              | 17       |

|    |  |    |
|----|--|----|
| 4. | 1.5 Training and Testing.....                        | 17 |
| 5. | 1.5.1 Weight update .....                            | 18 |
| 6. | 1.5.2 Eleminate Inputs .....                         | 18 |
|    | 1.5.3 Training and testing on loop.....              | 18 |
|    | 1.5.4 Deploying the Network.....                     | 19 |
|    | 1.6 Fundamental Variable and Technical Variable..... | 21 |
|    | 1.7 Basic Terms Regarding Share Market.....          | 30 |
| 2. | REVIEW OF LITERATURE.....                            | 30 |
| 3. | PRESENT WORK.....                                    | 30 |
|    | 3.1 Problem Formulation.....                         | 31 |
|    | 3.2 Objective.....                                   | 33 |
|    | 3.3.2 Data Preprocessing.....                        | 34 |
| 4. | RESULT AND DISCUSSION.....                           | 37 |
| 5. | CONCLUSION AND FUTURE SCOPE.....                     | 44 |
| 6. | REFERENCES.....                                      | 45 |
| 7. | APPENDIX.....  | 48 |



## LIST OF FIGURES

| S. No. |  | Page No. |
|--------|--|----------|
| 1.     | Fig. 1: Artificial Neural Network Modelling .....                                  | 03       |
| 2.     | Fig. 2: Artificial Neural Network Implementation.....                              | 03       |
| 3.     | Fig. 3: Human Neuron.....  | 04       |
| 4.     | Fig. 4: File Systems .....   | 05       |
| 5.     | Fig. 5: Network Architecture .....   | 06       |
| 6.     | Fig. 6: Transfer Functions.....  | 07       |
| 7.     | Fig. 7: Training Method .....  | 08       |
| 8.     | Fig. 8: Feed Forward Network .....   | 09       |
| 9.     | Fig. 9: Feed Forward Neural Network Structure .....                                | 10       |
| 10.    | Fig. 10: Prediction Process overview using NN.....                                 | 11       |
| 11.    | Fig. 11: Training Process.....   | 12       |
| 12.    | Fig. 12: Performance plot for Network.....   | 13       |
| 13.    | Fig. 13: Regression plot for SBI stock.....  | 14       |
| 14.    | Fig. 14: Plot of Training of the Network.....                                      | 16       |
| 15.    | Fig. 15: Plot of Actual vs. Predicted for Period of NOV, 2014 to<br>Jan, 2015..... | 18       |

## LIST OF TABLES

| S. No. |  | Page No. |
|--------|--|----------|
| 1.     | Table 1: Parameters of Activation Function .....             | 03       |
| 2.     | Table 2: Testing data and training data information.....     | 09       |
| 3.     | Table 3: Performance of Stock Prices.....                    | 12       |
| 4.     | Table 4: HMM Processing.....                                 | 13       |
| 5.     | Table 5: Comparison of Error.....                            | 14       |
| 6.     | Table 6: Comparison b/w TARGET and simulated price By NN.... | 16       |

# Chapter 1

## INTRODUCTION

---

The study of the stock market attracts everyone because it can change the lives of investors based on the decisions they make on what stocks to purchase. A lot of people have tried to predict the stock market, but very few among them have succeeded to predict. Basically it is impossible for everyone to predict the market for a long period of time because it is very dynamic. But with the correct data mining techniques such as neural network, linear regression, some pattern can be predicted which helps to know how the stock will act from its previous behaviour. In the past years, this thing has become very clear that neural network is very important part of analyst's class. The theory of neural network was developed in 1943. The neural network domain has become fastest growing area in computer field from the last nine to ten years.

Neural Network is basically used to recognize the pattern. It receives two things: inputs and expected output. After receiving, it searches for the relationships exist between inputs and expected output. If once the relationship has been met, then the neural network is able to give output on any kind of input with few percent of error which is normal thing. In stock market, there are two types of analysis for prediction: Fundamental Analysis and Technical Analysis. Technical Analysis totally depends upon the myth that History repeats itself and future movement can be predicted by analysing the past behaviour of securities.

The experimental challenges occur in the stock market motivates to research in this field. The most important theory Efficient Market Hypothesis states about the financial market that it is efficient and prices reflect all known information regarding a stock and that price at a very fast rate adjust according to any new information. Information includes not only current information known about a stock, but also the future information like future earnings or dividends. It attempts to explain the random walk hypothesis by stating that stock prices will be moved significantly by only new information and new information that is unknown presently and randomly occur leads to future movements in stock prices are also unknown and move randomly. Therefore, it is not possible to

perform better than the market by picking undervalued stocks, since the EMH states that there is not any kind of undervalued or overvalued stocks. The second theory Random Walk Theory related with EMH states that there exist not any trend or pattern in the future prices and these future prices are randomly departure not logically departure from the previous prices. Many researches in the past have said that the stock market cannot be predicted and it is a chaos system that appears random and non linear. The artificial intelligence technique named as neural network experts to learn these kind of chaos system. In the past many models have been propose using fundamental and technical variables to give reasonable predictions. Fundamental analysis deals with in-depth analysis of movement in the stock prices. On the other hand, technical analysis is based on charts to figure out the price movements and use this data to predict future price value. Technical analysis always tries to search the patterns in the chart. But the problem with this analysis is that it is fully subjective in nature. Apart from these two methods of stock price prediction, time series predictions are also used for the same purpose. It works by analyzing the past data to capture the pattern in the historical data and then used this data for forecasting. Time series forecasting works under two approaches. The first one is linear approach and the other one is nonlinear approach. The methods comes under linear approach are moving average and time series regression etc. The existence of the nonlinearity in prices is offered by many researchers. From the last few years, researchers have been shown more interest in the application of neural network for stock market prices forecasting due to its non linear nature. A neural network is nonlinear in nature. So, it is appropriate than any other models to predict the stock market returns. It is used to extract those patterns and trends from the historical data that are too complex to be extracted by humans. There are a lot of features present in neural network that make preferable over all other traditional models used for stock price prediction because all the traditional models lacks the ability in understanding the relationship exist among the input and the output attribute due to chaos nature of the system. The linear models have not the ability to deal with the data pattern and analyze when the underlying system is a nonlinear nature vice. Most of the non linear techniques the learning capability, therefore neural network suits well for the forecasting tasks. Stock Market has become a most important part of the economy, so any variation in market affects the economical

health of the nation. Due to its dynamic nature, everyone finds it difficult to predict. The characteristic that is similar in all stock markets common is the uncertainty that can be long term or short term. This characteristic is not desirable from the investor side but it is also not avoidable in the case of Stock Market. From the last few years, researchers has been doing to overcome the uncertainty comes in predictions. The stock price can't be exactly predicted, but from previous researches, Roughly prediction can be made about tomorrow and with the correct data mining techniques such as neural network, linear regression, some pattern can be predicted which helps to know how the stock will act from its previous behaviour. In the past years, this thing has become very clear that neural network is very important part of analyst's class. Tang and Fishwick [19] start the general introduction of developing the neural network for economic time series. Arka Ghosh [1] the designing and training of neural network is totally free from human intervention because it is completely automated process. The prediction purpose by neural network consists of two sessions: training session and testing session. To train the neural network for the prediction, we need a lot of data. In this paper, five years historical data is used to predict the prices. The most import thing is to choose predictors and predictands. Predictor is the input parameters and predictand is the target parameter. In this paper, feed forward neural network with three layers and back propagation as a learning algorithm is used. The neurons lies in input layer depend on input parameters. For instance, if numbers of input parameters are three, then number of neurons will also be three in input layer. The number of output parameters leads to neurons in output layer. One single output parameter leads to single neuron in output layer. To choose the neurons in hidden layer is itself a problem. For this, no specific concept is derived until now. It is totally like a hit and trial method. According to rule of thumb, number of neuron in the hidden layer should be twenty. In this paper, feed forward neural network and back propagation are used for the prediction and supervise training of network architecture. The learning algorithm back propagation algorithm works by looking for the minimum value of the error function in weight space with the help of gradient descent method. The weight combination with minimum error function is considered to be a best solution of the learning problem. The basically aim is to use those input parameters which affect the target value a lot.

## **1.1 Description**

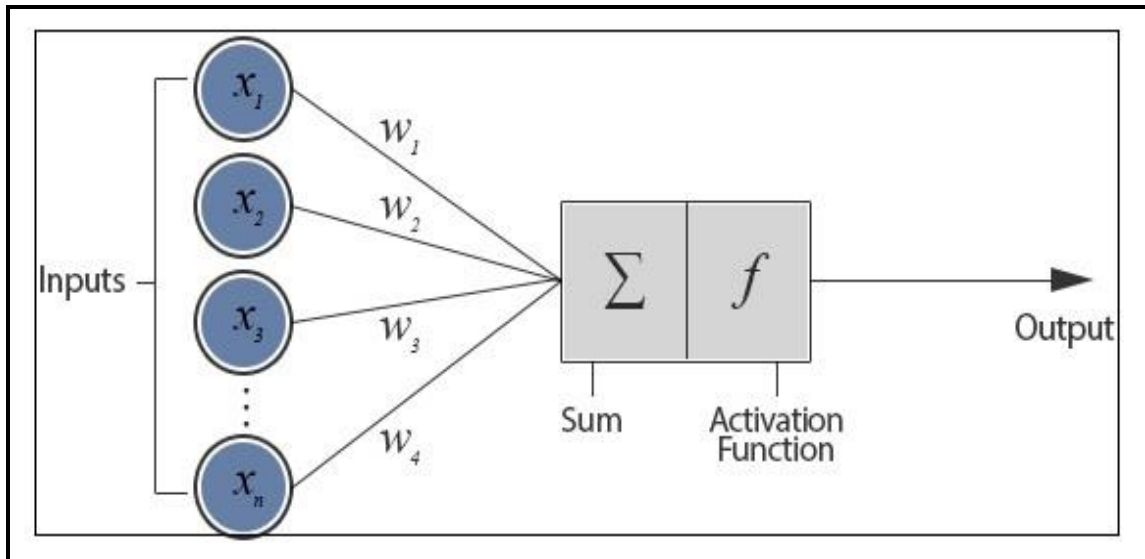
The main purpose of this Dissertation Proposal is to make a neural network that can predict the stock price with maximum accuracy for a number of sectors or companies. To get the accurate result, different-different type of neural network is used. The result of this different kind of neural network will be compared with each other and in the end among these; one network is selected for forecasting.

## **1.2 Motivation**

The basic motive of the content written in the proposal is to research or inspect how neural network used in time series predictions. Stock prices prediction also comes under time series prediction. A lot of researchers have researched on this field. If the future stock prices can be predicted for week or even a month, then many investors would be purchasing stocks to get gain. Only this fact motivates researchers to answer the question of how the stock market can be predicted. In this proposal, some different approach is used which separates it from all the previous researches. In all the previous researches, all the researchers gives open price, low price and closing price as input and try to predict the future closing price by training the network. Rapidly, all the researchers are trying to improve their prediction with the approach. But the content written in this proposal tries to do some different by varying the input parameters.

## **1.3 Neural Network**

**1.3.1 Modelling Artificial Neuron:** The neuron model in the neural network is a very simple model that is fully based and work like biological neuron. Basically the artificial neurons can also be referred as the perceptrons. The perceptron structure looks like



**Figure 1:** Artificial Neuron Modelling

In the above given diagram, the perceptron has many inputs and each individual input has some weight value assigned. The weight of the perceptron has the ability to amplify and deamplify the input signal. For instance, if the value of is 2 and weight assigned to input is 0.1 then due to the lowest value of weight assigned to input, the input signal will be decreased from 2 to 0.2. In the same way, these all weighted signals are added together and passed to the activation function. The conversion of the input into an appropriate and useful output is done by activation function. The simplest activation function is step function that will provide output 1 if the input given is higher than threshold value, otherwise zero output will be provided by the function.

For instance:

$$I_1 = 0.5, I_2 = 2.0$$

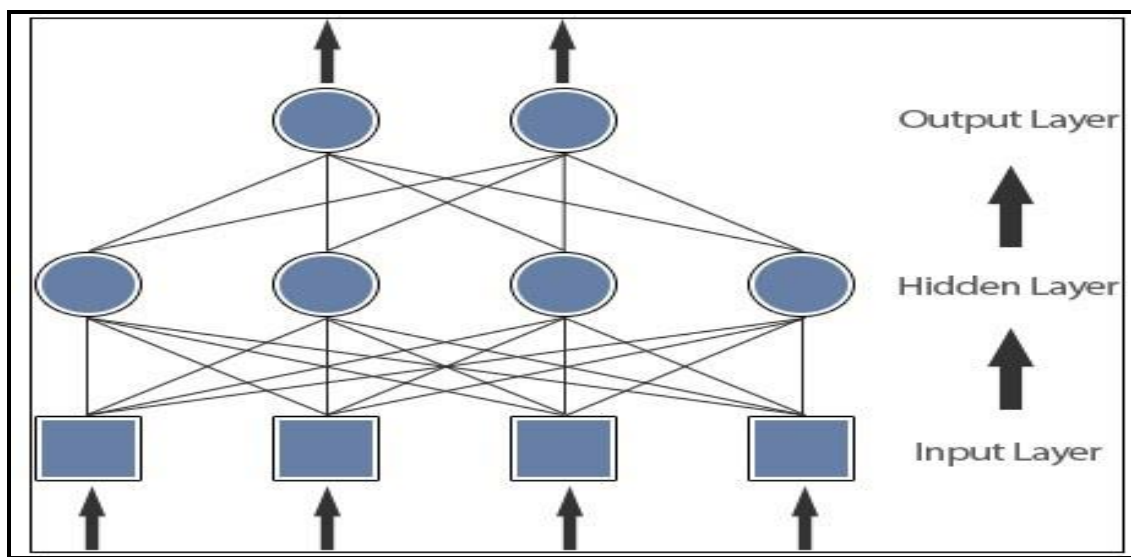
$$W_1 = 0.3, W_2 = 0.8$$

$$\text{Threshold} = 1$$

$$\text{So } I_1 * W_1 + I_2 * W_2 = (0.5 * 0.3 + 2.0 * 0.8) = 1.75$$

Now calculated value 1.75 is greater than the value of threshold, so now step function will give output as 1.

**1.3.2 Artificial Neural Networks Implementation:** In the feed forward network, each node in the hidden layer takes the input from input nodes via input layer. There is not any kind of restriction on the number of nodes per layer. Mostly, there are multiple hidden layers used to pass through the input and make them reachable to output layer. Selection of right and appropriate number of layers and nodes is important to make the network optimistic to work with given problem. In the diagram given below, it shows that how signals are passed in a single direction through the neural network layers. In the case of feedback neural network, signals can be travelled in both directions.

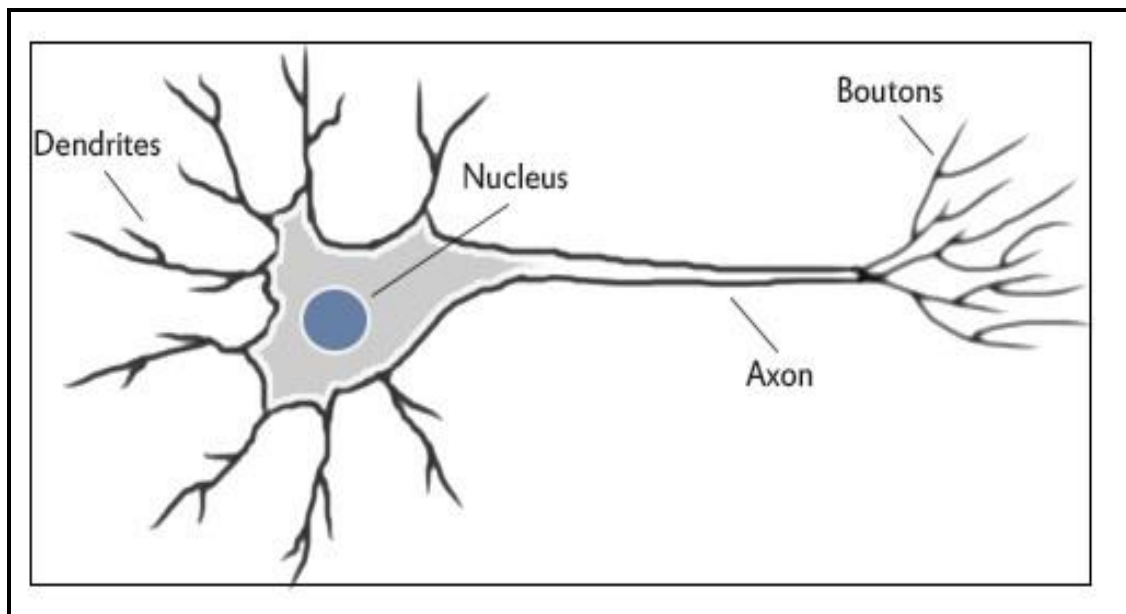


**Figure 2:** Artificial Neural Network Implementation

Neural network theory comes out from Artificial Intelligence. It is like a computer program or hardware type machine that is designed to learn in the same way like human brain learns. It can say that neural network is kind of technique to emulate the human brain. It is also a part of Data Mining. Many researchers use this for data mining purpose. The main capability of neural network is its self-learning capability. It is a nonlinear network. It matches with the brain in two ways: Knowledge is acquired by the network in the same way like human brain acquire through a learning process and the strength of inter neuron connection known as weight which is important part of neural network are used to store the knowledge or pattern.



**1.3.3 Biological Neuron:** Like human brain consists of 10 billion neurons; neural network can also consist of number of neurons. This knowledge will further help to make prediction on unknown data patterns. Weight basically treats like a connection between the nodes; weight hold all the information related the system. Due to its ability to deal with nonlinear data, it uses most in business applications. Artificial Neural Network is used in a lot of applications like Pattern Recognition, Data Classification and Stock Prices Prediction.



**Figure 3:** Human Neuron

Different kind of fields like banking, business and industries are using neural network for their tasks. The main application of neural network is forecasting. The Artificial Neural Network converts the human learning ability into computing environment with its thumb of rule and experience. Its concept begins with the functioning of human brain. Human brain has 10 billion neurons .In the same way, neural network also consists of the neurons. The basic and important part of the brain and the neural network is the neuron.

**1.3.4 Learning in Artificial Neural Network:** The learning ability is the best feature of the neural network. The working of neural network is same like human brain and it is also inspired from the human brain. Like human brain, it has also the ability to learn

from the examples and huge amount of logical data. In biological nervous system, while learning process, the neural structure of the brain tries to increase or decrease the connection strength. The changes in the strength of the connection all depends upon the brain activities. In the real life, to recall the relevant information is easier than to recall the information which has not been recalled from a long time as the connection strength of relevant information is stronger than any other irrelevant information. By adjusting the weight of the connections, learning process can be modelled by the neural network.

**1.3.5 Types of Learning:** The main objective of learning is to find the optimum weight matrices that can be applied to the network and make network able to do mapping of input set to correct output set. There are three major categories of learning which are given below:

- **Supervise Learning:** It involves input-output pairs to train the network and the weight matrix is updated based on the error that is calculated based on the difference between the actual output and target output of the network. The changes or updating made in weight values are only for making correction to neural network.
- **Unsupervised Learning:** In Unsupervised Learning, there is absence of output pair. It involves only inputs and to find the patterns in the data provided is the responsibility of neural network. Data Mining is the major application area where this kind of learning is used.
- **Reinforcement Learning:** Supervise Learning and Reinforcement Learning is almost same. Instead of providing the target value, the feedback is provided to the network. Feedback in the context of reward is given over the system well performed. The main objective of this learning is that the system can receive maximum reward by trial and error.

**1.3.6 Implementing Supervise Learning:** In this proposal, supervise learning is used to develop the model. If the network is well trained, then unseen input can also be given and mapped to correct output. A lot of learning algorithm has been using the supervise

learning. But the most popular learning algorithm comes under supervise learning is back propagation.

**1.3.7 Learning Rule of Back Propagation Algorithm:** The principle of Back Propagation Algorithm is to prevent the happening of same error again and again by identifying anything wrong and making some correction in the network.

**Working:**

- Firstly it compares the network output to the target output given in training set.
- If the target output and actual output of the network does not match, then it is the clear indication of going something wrong in the network.
- To find out the amount of error occurred in the network, it calculates the difference between target output and actual output.
- Based on the error calculated on above given step, it updates the weight of the connections.

- The output on each output node is calculated by:

$$\text{Output} = f(I_1 * W_1 + I_2 * W_2 + \dots)$$

I: input at the node

W: Strength of the connection and value of the weight

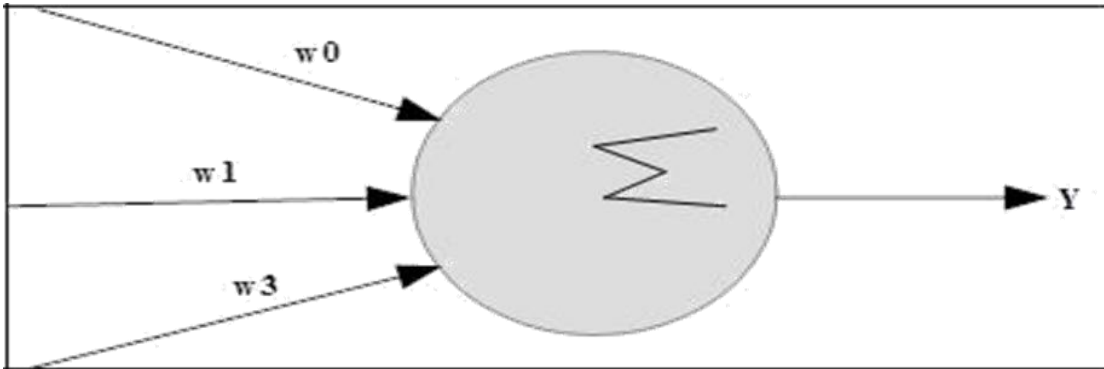
- The error on each output node is calculated by:

$$\text{Error Value} = \text{target output in training set} - \text{actual output}$$

The formula or equation used for updating the weight is

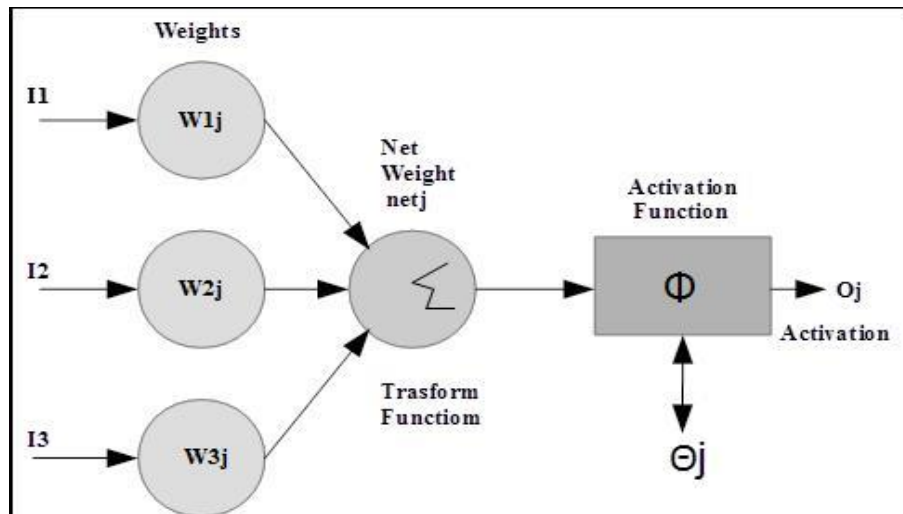
$$\text{Change in Weight} = \text{Input} * \text{Output} * \text{Learning rate}$$

The basic element of neuron model is known as node that receives input which has some associated weight from the other units. The output of the node in turn serves as the input to other nodes or units.



**Figure 4:** Neural Model

The architecture of the neural network defines the way in which these neurons are connected with each other.

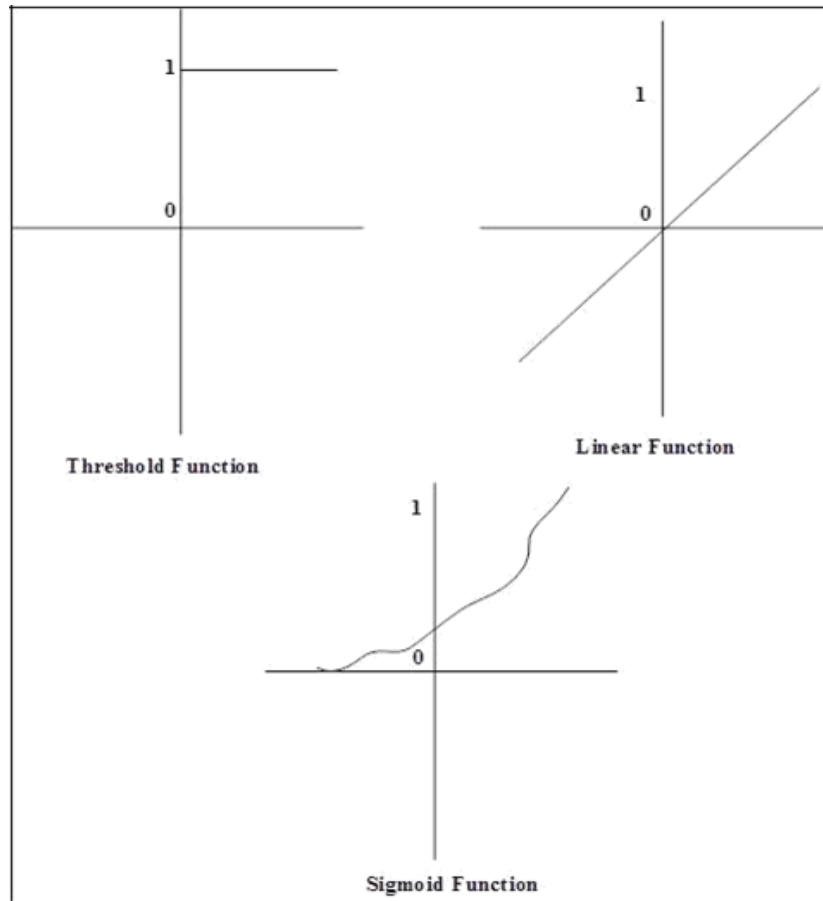


**Figure 5:** Network Architecture

The between the neurons has some value named as a weight. Each single neuron can be defined as a processing unit that takes input and after doing processing of the input with activation functions gives the output. The most common used activation functions are pure linear, the sigmoid function and the tan sigmoid function.

**Table.1** Parameters of Activation Functions

| Activation Function | Function             | Range              |
|---------------------|----------------------|--------------------|
| Pure linear         | $f(x)=x$             | $(-\infty,\infty)$ |
| Sigmoid             | $f(x)=1/1+e^{-x}$    | $[0,1]$            |
| Tan sigmoid         | $f(x)=2/1+e^{-2x}-1$ | $(0,1)$            |



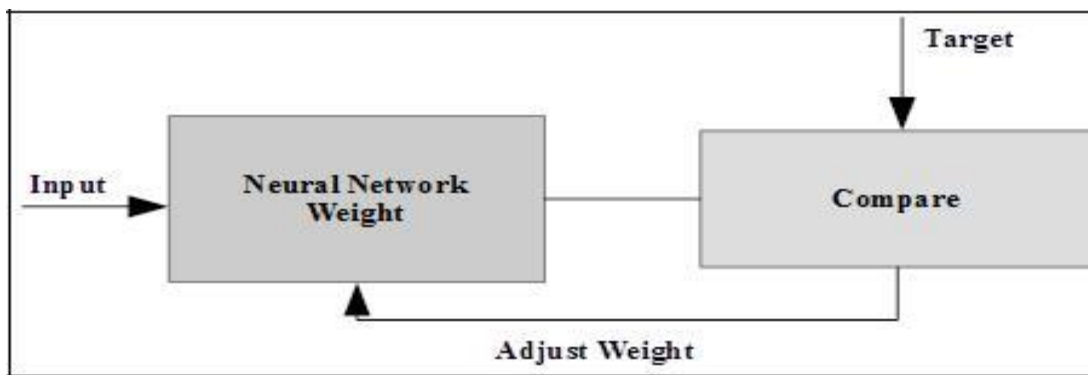
**Figure 6:** Transfer Functions

Artificial neural network can have no of layer of neuron. The connection between the layers can be done in two ways: fully connected and partially connected. There is a term named as weight which represents the connection between the nodes. With the help of predefined input and target or expected output, the value of weights can be adjusted on the behalf of difference between the actual target value and expected target value. The

learning behaviour of artificial neural network consists of basic three steps:

- First to calculate the temporary output.
- After that, compare the temporary output with the output which you desire.
- Adjusting the value of the weight based on the difference between the temporary output and the desired output.

Basically training means adjustment of weight in a neural network. The training process of neural network is shown in below diagram:



**Figure 7:** Training Method

In starting, by default the value of weight in network is within the range of  $[-1, 1]$ . After that the first sample input is presented to network and network is calculated output. The calculated output is then compared with target output. After comparing, weight is adjusted in such a way that the error or difference between calculated output and expected output is reduced. There are two types of function which is used to calculate the error: mean square error and mean absolute error.

The network can be trained with two techniques. First one is incremental training and the other one is batch training. With the incremental training, weights are adjusted again and again when any input sample has given to the network. With each input sample, the weight adjustment process will be occurred in the incremental technique. On the other hand, in the batch training, weights are adjusted only once on supplying all the input samples. The parameters used in training purpose are: learning rate, no of iterations and epochs. The value of learning rate is 0.01 by default. If the values are increased, then

there is a risk of becoming unstable network. If the values are decreased too much, then as a result the converging time of the algorithm may become very large.

### 1.3.8 Layer Architecture

Neural network can be designed by grouping the neurons into input layer, output layer and number of hidden layers. In the diagram, it is not mandatory for neuron in one layer to connect with all the neurons of next layer. The network with one neuron in one layer is not connected with all the neurons of other layer are known as sparse network. The network with neuron in one layer is connected with all the neurons of other layer are known as complete network.

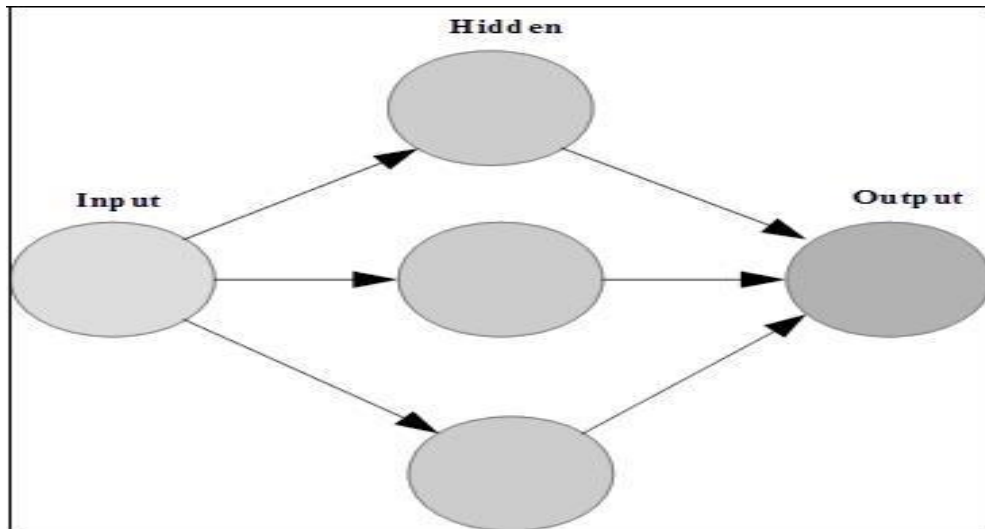


Figure 8: Feed Forward Network

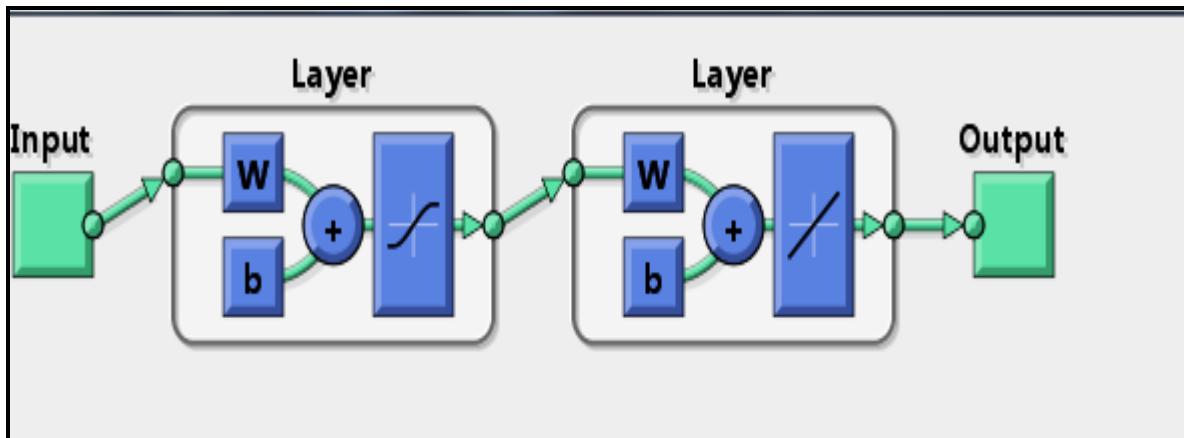


Figure 9: Feed Forward Neural Network Structure

**Input Layer:** Input Layer is the area of network in which we set all the environment parameters. This layer has no incoming connection. It gets its value from external sources. The connection going out from these layers sends the external input to the neurons of next layer.

**Hidden Layer:** Hidden layer always exist between the input layer and output layer. Network can have more than one hidden layer. Now the question arises why these layers called as hidden layers. The answer is that these layers are invisible to external a process that interacts with network. The neurons lie in the hidden layer has two connections: one is internal connection that coming from the proceeding layer and other one is external connection that outgoing to succeeding layer.

**Output Layer:** The name of this layer gives the meaning to it that which layer gives final output is known as output layer. These layer neurons have no any kind of outgoing connection. The no of neurons in this layer depends on the attribute of our target. If the desire is to forecast close price as well as change price then the network output layer will have to hold two neurons. Their values are read directly by the user.

**Number of Neurons and Layers in a Network:** This question is very import that how to choose the number of hidden layers and number of neurons in the Input Layer and number of neurons in the output layer. The number of neurons in the input data totally depends on the input data. For instance, while loading a excel sheet into the Matlab which holds the input data as well as the target data.

```
Input load = xlsread ('C:\Users\acre\Desktop\abc.xls');
```

```
Input = inputload (1: end, 2:4);
```

Here no. 2 to no. 4 as input data. So now there are total three input parameters. That's why in the network; Input layer should have three neurons.

```
target1 = inputload (1: end, 9);
```

In the same way in the output layer, the number of neurons depends on the number of features or number of columns that desire in the output. Like in the above given line of code, Column number 9 taken as an output that leads to one neuron in the output layer. In this case, if anyone wants to predict the closing price, then they have to take only one



neuron in the output layer. In the hidden layer case, one thing comes to know by experience that one hidden layer is sufficient for all kind of problems. There is very less number of situations in which performance show some improvement by adding more hidden layers. Neural Network is the processing system. There is large number of interconnected element which processes like the human brain. It is totally based on human brain geometry. There are many kind of neural network. The classification of neural network is based on its learning process. By giving a sample data as an input form and by giving the different-different values of weight factor, the network is trained. An artificial neural network is always preferred for the forecasting application because of its learning capability. The learning methods are of two types: supervised learning and unsupervised learning. Along with types of learning process, there are also types of feed forward neural network: single layer neural network and multilayer neural network. In the past, by extracting knowledge from the source like human expert and converting into programming language, forecasting was difficult. But with the neural network, automated procedure comes in form which made the learning process easy and rapid. Neural Network always learns with the help of observing the past data and expert's interviews are not mandatory. It is still able to make decision with incomplete information provides to it. The features which attract researchers to choose neural network for forecasting purpose are:

- It has the ability to learn from examples and experience.
- It always provides a practical ways which is feasible to solve any real world problem.
- Generalization plays a vital role in increasing popularity of neural network.
- It is non linear network which compatible with the real world which is also a nonlinear.

## **1.4 Neural Network Development**

There are lot of steps used in building a neural network model for forecasting.

- The first step in building a model is to decide a target. Closing price is the target value for the model given in this proposal.
- Determining the time period or frame that user wish to predict.

- Gathering the important information about the domain of problem
- Gathering the data that is needed to develop the model and try to find out the relationship for each input to target.
- Transformation of the data into appropriate format.
- Tried to reduce the dimensions of the data used as input.
- Design and make some decision related with the network architecture that includes number of input layers, hidden layers and output layers, learning parameter.
- Training and testing the network on loop.
- Try to remove the correlated inputs while doing step 9 to get the better results
- Deploying and testing the network on new and unseen data and redefining it as per desired results.

**1.4.1 Time frame and the target:** The choice of target is very first step to develop the network. The target of the model given in this proposal is to forecast the value of closing price of stock. The direction of the stock whether it will be up or down can also be predicted. But here the model is predicting the values not a direction. The decision which is very important in the development of the model is to select the time frame size. The development of the neural network to predict a short term price is harder than to develop the network for long term predictions. The noise and the variations at short term period might be the reason of it.

**1.4.2 Domain experts:** The knowledge of the domain experts is very important in the development of the effective model as they have a lot of knowledge about the factors which influence the stock market a lot. So investigation of the parameters and stock market before training can be very helpful in the development of the appropriate and effective model

**1.4.3 Data gathering:** The data used to train and test the network is received from [www.nseindia.com](http://www.nseindia.com). It is a national stock exchange website that offers all kinds of data related with the stock.

**1.4.4 Data pre-processing:** The data pre-processing is very important before the training of the network. The network accuracy for the predictions depends largely on the quality of data used. Data should be in homogeneous form. In the case of stocks data, sometimes heterogeneity exists in the data pattern due to stock split. Stock split is actually the decision made by the board of directors of the companies to increase the number of outstanding shares by issuing more shares to current shareholder. There are several reasons of stock split like psychology, stock's liquidity. It's important to convert the data into appropriate form before implementation because the heterogeneous data leads to strange graphical representation of results.

**1.4.5 Data splitting:** 70% of the data is used for training, 15% used for validation and the rest is used for testing.

## 1.5 Training and Testing

Training data is used to train the network. When a minimum error is achieved while training, then apply the same network on the testing data and observe the error. Restarting the whole process again and again with different weight matrices until the best and the less amount of error on training and testing data sets is achieved.

| Stock Symbol | Training Data |           | Testing Data |           |
|--------------|---------------|-----------|--------------|-----------|
|              | From          | To        | From         | To        |
| SBI          | 1/1/2009      | 31/9/2014 | 1/10/2014    | 29/1/2015 |

**Table 2:** Testing data and training data information

**1.5.1 Weight update:** The steps used in updating the weights are given below:

$$weight_{ij} = weight_{ij} + \text{change in } (weight_{ij})$$

$$weight_{jk} = weight_{jk} + \text{change in } (weight_{jk})$$

$$\text{chnage in weight} = \alpha * \text{input neuron}_i * \gamma$$

$\alpha$  – Learning rate

$\gamma$  – Error

The value of learning rate is usually between 0 or 1. The learning speed of the neural network is affected a lot by its value. It leads to amount of weight adjustment required for the network to give better result.

**1.5.2 Eliminate Inputs:** Eliminating the number of inputs and observing the error on the test set and training set can make network very effective.

**1.5.3 Training and testing on loop:** Repeat the training and testing process again and again to achieve better result.

**1.5.4 Deploying the network:** Testing data set is used to see the network performance. Training, testing and the designing phase can be repeated on the unsatisfactory error.

**The following given parameters are used to develop the model:**

|                            |                   |
|----------------------------|-------------------|
| Number of Inputs           | 2                 |
| Number of Hidden Layers    | 1                 |
| Number of Hidden Neurons   | 1                 |
| Back Propagation Algorithm | trainlm           |
| Performance                | Mean Square Error |

## 1.6 Fundamental Variable and Technical Variable

**Fundamental Variable:** One of the most important parameters which affect the movement of the stock market is its trading volume. Stock Market is lot of depend upon Trading volume. If the trading volume of a stock is higher, then the more the stock will be active. Trading volume is one of factor of technical analysis for price patterns. If

anyone will become able to predict the moving direction of stock trading volume in the future, then many patterns like the prices movement etc. with more accuracy and confidence can be predicted. Predicting stock prices and its movements is very difficult task due to influence of several macro-economic factors, including political events, economical conditions, investors' expectations, institutional investors' choices, and unexpected situations in other stock market and psychology of investors. Moving average method is a very good method for reduces the fluctuations in prices. When price fall below the moving average, then it has tendency to keep on falling. In the same way, when prices rise above moving average, they tend to keep on rising. In the stock market, there are two types of market: Bearish Market and Bullish Market. In the Bullish Market, the trend will have a relatively high close price within its daily range and have increasing volume. But if a stock regularly closed with a relatively low close price within its daily range with high volume, then this would be the indicator of weak security. There is very strong pressure on investors to buy when a stock closes in the upper half of a period's range and similarly there is selling pressure when a stock closes in the lower half of the period's trading range.

## 1.7 Basic Terms Regarding Share Market

- **Volume Weighted Average Price:** It is ratio of the value traded to total volume traded over particular time.
- **Average Price:** It is used to reduce the value of a range into single value. It can be calculated by
 
$$\frac{\text{Total spent}}{\text{Total Number of Shares}}$$
- **Beta:** It is used to show the movement or volatility in stock. If the value of beta is less than one, then it is the indication of less volatility in securities as comparative to market. If the value of beta is greater than one, then it is the indication of more volatility in securities as comparative to market.
- **Open Price:** It is the price of the first share traded for the current trading day.
- **Change:** It may be defined as the difference between the last closing

price and current price.

- **Debt Ratio:** Debt Ratio is the ratio of company's total debt to its total assets. If the ratio is less then it leads to less risk and strong position of equity. On the other hand, if the ratio is higher, then it leads to high risk and weak position of equity.
- **Outstanding Shares:** It may be defined as the total number of shares that exist for a company.
- **Daily Volume:** Daily Volume can be defined as the number of shares.

## Chapter 2

# LITERATURE REVIEW

---

Huang Fu-yuan in “**Forecasting Stock Price Using a Genetic Fuzzy Neural Network**” [10] the research paper is based on combination of two techniques fuzzy neural network and Genetic Algorithm to predict stock prices. Fuzzy Neural Network is the combination of fuzzy logic and artificial intelligent technique neural network. By adding rules to neural network, it works very well. The major drawback of neural network is black-box syndrome that can be avoided with the help of fuzzy logic. But the problem is that fuzzy logic has no self-learning ability. Therefore, single fuzzy logic cannot be used for forecasting purpose like a neural network. So researcher combines to overcome the problem of self- learning capacity of fuzzy logic, researcher introduced the concept of Genetic Algorithm. Genetic Algorithm is used for optimization purpose to optimize the neural network parameters like weight etc. The learning process of Genetic Fuzzy Neural Network is very different from neural network. In this newly technique, the combination of back propagation algorithm and genetic algorithm is taken to learn the network. The working of learning algorithm is given below:

- First to do the coding of Network.
- Produce the population for selection purpose.
- Then choose the best solution by crossover operation, mutation operation and optimal individuals.
- After that, do the decoding of network.
- Finally trained the network by back propagation algorithm.

They found that the performances of these two models are same for data which is used to train the network, but the difference comes in testing data. This paper concludes in the final that the forecasting accuracy of Genetic Fuzzy Neural Network is better than single Neural Network.

**Table 3: Performance of Stock Prices**

| Data Sample     | Neural Network | Genetic Fuzzy Neural Network |
|-----------------|----------------|------------------------------|
| Training Sample | 100%           | 100%                         |
| Testing Sample  | 76%            | 88%                          |

Farhad Soleimani Gharehchopogh, Tahmineh Haddadi Bonab and Seyyed Reza Khaze in “A Linear Regression Approach to Prediction of Stock Market Trading Volume: A Case Study” [8] is discussed new technique linear regression to predict stock prices. This technique predicts values only in numeric form. Basically regression is used to establish the relationship between predicting values and a target values. In this approach, they split the data sets into two parts: training data and testing data. Approximately 80% of the data is used for training purpose and the rest is used for testing, in these days, Share market has been known for research field only because of his effect on financial challenges. Many researchers are predicting next day movement of a stock in order to proof better than other techniques. Some researchers wanted to estimate the dependent variables with the help of some independent variable. For this purpose, they tried to predict stock future movement with Linear Regression. They took economic variables as an independent variable and targeted output as a dependent variable. The value of dependent variable was predicted using linear regression equation in which dependent variables are entered. Linear regression is mostly used data mining technique for predicting the future value of Stock prices based on the linear relationship it has with other variables. There are only two types of variable: one is known as an independent variable and the other one is known as dependent variable. This formula describes this relationship of a straight line:

$$y = a + bx \dots\dots\dots 1.1)$$

Here y is working as a variable x is working as an independent variable.

a and b are the coefficients of line. From that formula, the value of a can be calculated by given below formula:

$$a = y - b \dots\dots\dots 1.2$$



The formula to calculate b is given below

$$\frac{(N \sum xy - (\sum x \sum y))}{(N \sum x^2 - (\sum x)^2)}$$

Mean of (y) = current market price mean

Mean of (x) = percentage earnings ratio mean

N = Total number of occurrence of the variables

The percentage earnings ratio, current market price and the earning per Share given by the formula below:

Price Earnings Ratio = Current market price / Earning per share

Marc J. Schniederjans, Karyl B. Leggio, Quing Cao in “**A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market**” [12] study neural network on firms came under Shanghai stock exchange. He did a comparison of linear model predicting power with neural network predicting power and he analyzed from his experiment that neural network performed better in comparison to linear model. Neural Network is a flexible and non linear function that first time used in engineering field and after that it started to use for time series forecasting. It is very clear that Neural Network is used for those problems which require knowledge that is insufficient to specify but a lot of observation should present for it. The very first prediction by neural network was done using linear neural network for weather forecasting.

Md. Rafiul Hassan and Baikunth Nath in “**Stock Market Forecasting Using Hidden Markov Model: a New Approach**” [13] introduced new approach regarding stock price prediction of airline stocks. This approach is basically a model named as Hidden Markov Model (HMM) that is a like finite state Machine which has a number of fixed states. It was introduced in the beginning of the 1970's as a tool for speech recognition. This model is fully based on statistical methods. Therefore it has become increasingly popular due to its strong mathematical structure. In recent years, researchers declared HMM as a classifier or predictor for speech signal recognition, recognition of handwritten character, natural language domains etc. So it is clear that HMM is a very powerful tool for various

applications. Due to its strong application, we try to develop an HMM based tool for forecasting stock market.

While implementing the HMM for stock price prediction, the choice of the number of states and observation symbol become a very complex and difficult task for a researchers. For that instance they used left-right HMM with 4 states. In this scenario, they consider only 4 input for simplicity. They consider four input features for a stock. The prediction of next day's closing price is taken as the target price associated with the four input features. Here their observations are being continuous rather than discrete.

In order to train this model, researcher divided the dataset into two sets, one is the training set and the other one is test set. For instance, they trained an HMM with the daily stock data of South west Airlines from the period of 18 December 2002 to 29 September 2004 for predict the closing price on 30 September 2004. The trained model will produced likelihood value of -9.4594 for the stock price on 29 September 2004. With the help of this trained HMM and the past data, they discover the exact place of likelihood value -9.4544 on 01 July 2003. Now it make feel quite logical that 29 September 2004 stock behaviour follow the behaviour that of 01 July 2003. In this model, the main aim was only to predict the next day's closing price for a specific stock market share with the help of that HMM model. For training the model, they are used the past one and a half years daily data and to test the efficiency of the model, recent last three month's data were used. For a specific stock at the market, the values of the four variables open, high, low, close of that day: and with these values, researcher tried to predict next day's closing price.

|   | Open    | High    | Low     | Close   | Predicted Close |
|---|---------|---------|---------|---------|-----------------|
| <b>29 sep 2004</b>  | \$13.63 | \$13.73 | \$13.49 | \$13.62 | \$13.8          |
| <b>Matching Data Pattern</b><br><b>1<sup>st</sup> july,2003</b> | \$17.10 | \$17.20 | \$16.83 | \$17.13 |                 |
| <b>Next day Data</b><br><b>2<sup>nd</sup> july,2003</b>         |         |         |         | \$17.36 |                 |

Table.4 HMM Processing

In this way, they calculated the difference between the Closing prices on 01 July 2003 and the next day 02 July 2003. That different was  $\$17.36 - \$17.13 = \$0.23$ . Then that difference was added to the closing price on 29 September 2004 to forecast closing price for 30 September 2004. Above given shows the predicted and the actual prices of stock on 30<sup>th</sup> September 2004.

**Leaders & Followers:** This technique is based on the principle that stock of the companies can influence the stocks of other companies. In this technique, when leader increase or decrease the stock then the follower will respond its leader by increasing or decreasing its stock value some day later.

Yuling Lin, Grad. Sch. of Inf, Production & Syst, Waseda Univ, Kitakyushu, Japan, Haixiang Guo, Jinglu Hu in “**An SVM-based approach for stock market trend prediction** “ [18] proposed the support vector machine approach which consists of major two phases: feature selection and prediction model. Feature selection works by selecting best subset of index’s with the help of support vector machine filter. The indicators used to predict stock prices are evaluated based on their ranking. It is used to predict movement of stock prices with the help of financial indexes that are selected by feature selection phase. Weighted input behaves like a financial index. For the generalization, they concluded that this approach is better than others. In feature

selection, the importance level of each indicator is analyzed in such a way that which indicators can affect maximum to the movement of stock market.

Feature selection means to select subset of input variables .These variables may be technical variable or fundamental variable. The movement of stock market is classified into two categories: category 1 and category 2.In category 1, the stock prices of the next day is higher than today price. In category 1, the stock price of the next day is less than today price. To evaluate the prediction performance, they used the concept of hit ratio.

$$\text{Hit Ratio} = \frac{1}{N} \sum_{i=1}^N H_i$$

The value of hit ratio will be 1 when the actual and model output will be same. The major advantage of this approach is to reduce the noise, selection of a feature which affects the target value maximum and ranking to selected features. But this approach takes some further research. With the help of this technique, researcher can use maximum influencing factor in future in their predictive system.

Fagner Andrade de Oliveria, Luis E. Zárate, and Marcos de Azevedo Reis in “**The use of artificial neural networks in the analysis and prediction of stock prices**” [9] paper mainly focuses on the usage of neural network to predict stock closing price. Three phases are discussed in this paper to predict the closing price. In the first step, the work of obtaining the sample is done and in the second step, pre-processing of the input data is done to reduce noise and over fitting problem and at the last, the neural network configuration is done. Historical data as well as technical indicator, macroeconomics Parameters are used for forecasting closing price .In stock market, there are two types of method which are used for predicting. Fundamental Analysis and Technical Analysis are two most common models which used more in investor decision making process. Fundamental Analysis leads to study in detail about financial position of the company. When researcher want to make prediction for longer period, then they always give preference to fundamental analysis. It is used basically for long term investment and used to publish the financial report of company. It leads to the company financial position whether it's strong or not.

Technical Analysis is used to analyze the behaviour of investor. It works with mostly two terms price data and volume data. This analysis is not concern about company financial health. It is used by traders for short period investment. It is very subjective kind of process. The predictions which are based on technical analysis are not reliable. . The main focus in technical analysis is on price movement and with the help of that movement; future movement can be easily predicted. At the last, they conclude that neural network needs more data for training purpose to make able to give a better result.

Esmaeil Hadavandi, Hassan Shavandi, and Arash Ghanbari in **“Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting”**[7] paper are fully based on the usage of integrated approach to predict stock prices. They try to make an expert system to get the best accurate result with minimum amount of required input. In very first, they take the help of support vector machine to find out the most influenced factor which affects stock market. In the second state, they try to convert the raw data into clusters with the help of SOM neural network. After that, the rules are assigned to each cluster and fed into the genetic fuzzy system. The evaluation of the system is done by applying it on stocks. At the last, compare the results of this technique with the results of older techniques by the function named as MAPE. After applying on stocks, they come to know that the approach proposed by them outperforms all other older techniques.

Binoy B. Nair, N.Mohana Dharini, V.P. Mohandas in **“A Stock market trend prediction system using a Hybrid decision tree-neuro-fuzzy system”**[4] proposed the new approach Hybrid decision tree-neuro-fuzzy system that use technical analysis to extract the specific features for the forecasting purpose. After extraction, decision tree is used to select the features form the extracted features. After selecting features, they reduce the dimensions of the selected features and received the reduced data set that can further used as an input for neuro fuzzy system. The system that researcher developed is tested on 4 major stock market at international level. They found that combination of decision tree and neuro fuzzy system performed or gives result better than single

decision tree approach and single neuro fuzzy system. In actually, decision tree adds the ability of feature selection in Neuro fuzzy system that helps it results to get better accuracy.

Divyang J. Joshi, Pratik H. Bhavsarin “**Predicting Nifty 50 Movement Use of Advance Decline Ratio**” [6] This paper deals with the Advance/Training ratio that further basically deals with the trends and fully based on the daily ration of advancers and decliners. It has strong connection with sentiments as it helps to quantify stock up with respect to stock down. Volume on a given day doesn't matter for the ratio, what matters is only ratio of upward stocks to downward stocks. There are a lot of factors used to determine stock returns, but a financial ratio has been the most important factor of all stock researches. It works by comparing the number of stocks that closed higher with the no of stocks that closed lower than their previous day closing price.

$$\text{Advance/Decline ratio} = \frac{\text{Number of Advance Shares}}{\text{Number of Declined Shares}}$$

This ratio can be calculated with the above given formula for different time periods. This Period can be one day, one week or one month. Basically, Advance/Decline provides a signal about the changed direction of the market. If the ratio is greater than 1, then it will be the signal of bullish market and if the ratio is less than 1, then it will be the signal of bearish market.

S. Kheradyar, I. Ibrahim and F. Mat Nor “Stock Return Predictability with Financial Ratio” studies the roll of financial ratio in stock price prediction. According to this paper, B/M (book to market ratio) has higher ability to predict stock return as comparative to DY (dividend yield) and EY (earning yield). This paper specially focus on only three ratios that are used as stock return predictor in US and test these ratio in Malaysia market. After testing, these ratios in Malaysia market, they noticed that B/M ratio also performs well in Malaysia market than DY and EY.

Kohei Aono and Tokuo Iwaisako “Forecasting Japanese Stock Returns with Financial Ratios and Other Variables” examine the ability of P/E ratio to predict stock return in Japan. According to research given in this paper, price-earnings ratio performance is weaker than price dividend ratio to predict stock return in Japan. They also introduce the new variable interest rate to predict Japan stock market. In this paper, monthly data is used to analyse these factors.

Divyang J. Joshi, Pratik H. Bhavsar “Predicting Nifty 50 Movement Use of Advance Decline Ratio” by uses the Regression analysis to know the relationship exist between Advance/Training ratio and Nifty. They found that A/D ratio can’t be used as a better indicator by regression analysis as it shows very low value of R.

### 3.1 Problem Formulation

The earlier studies, researchers have established the impact of historical prices of the specific stock on the future prices of that stock comes under NSE. Since Banks Nifty have played a major role in Indian economy. Therefore to predict the stock prices of banking sector, it is important to include the bank nifty factors as an input to predict future price movement of the specific bank stock. The present study is carried out to know about the effect of historical prices of bank nifty and SBI on the future price movement of SBI stock prices. Further anyone can apply other technique instead of neural network for forecasting stock prices. Support Vector Regression which is famous for its generalization ability can use for stock price prediction in the future. Neural Network is the new technique that still not utilized with a proper way. In all the previous researches, stocks are predicted on the daily basis. With the help of new ideas, anyone can predict them in future on weekly and monthly basis. Also in the future, researchers can focus on other network that is very less used like Hopfield's Kohonen's on these types of applications. They can come with unique ideas like combination of neural network with new technologies like of intelligent agents. The researcher in this field improve their weakness daily, therefore it is very valuable thing to do some type of improvement by using different network architecture and techniques.

### 3.2 Objectives

The objective of this proposal is to consider the best variable to predict future prices of stocks of companies that acting under National Stock Exchange. It also concentrates on picking the best stocks of companies and sectors stocks so that designed network can give accurate prediction on them. It will inspect the use of neural networks as a forecasting tool. For the forecasting purpose, first need to train our network with different-different parameters so that accuracy can be achieved. To validate the network, apply the network on different-different stock. After that which stock will fit in the network, choose it for further research. To train the work MATLAB tool can be used. In



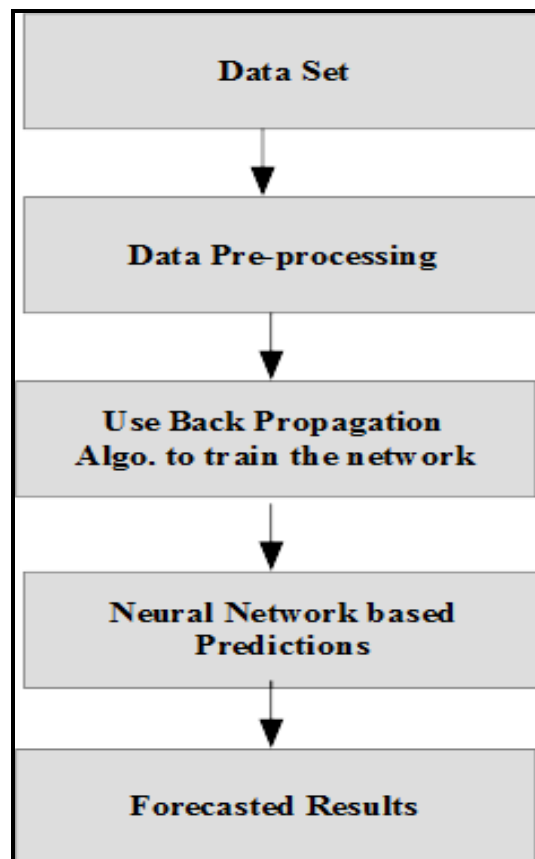
the MATLAB, there is a neural network toolbox, which can use for prediction purpose. In the neural network, to train the network, the input parameters plays very important role. Well trained network will give better accurate result. There are a lot of input parameters which reflects the future stock price movement. In this paper, the closing price of bank nifty also taken as one parameter to show that how it reflects the price movement of individual Banks that comes under it. But the challenge is to pick those variables which affect the stock indices maximum. In the input and target, a lot of relation exists: many-one, one-one. In this paper, the results of both the relationships are compared with each other to show that the reasonable accuracy cannot achieve with taking any kind of parameter related with a stock, it can achieve by taking those parameters which reflect the future closing price maximum. One parameter is also sufficient to achieve good result but the condition is that it should be highly correlated with the target value.

### **3.3 Methodology**

The main motive of work done in this report is to improve the price prediction of stock in any way with the help of artificial neural network. The methodology given in this research report leads to three layers feed forward neural network. Three layer model leads to one hidden layer, one input layer and one output layer. Back propagation algorithm is used for training purpose. The main factor in the share market is the best choice of stocks. The main source of historical stock prices collection is NSE (National Stock Exchange) website in which up to date Information of all the stock has been always available. The activation function that is used in the report is sigmoid function. The main reason to choose this function is literature survey that helps to know that it has been most widely used function. The approach using in this report has six variables as an input:

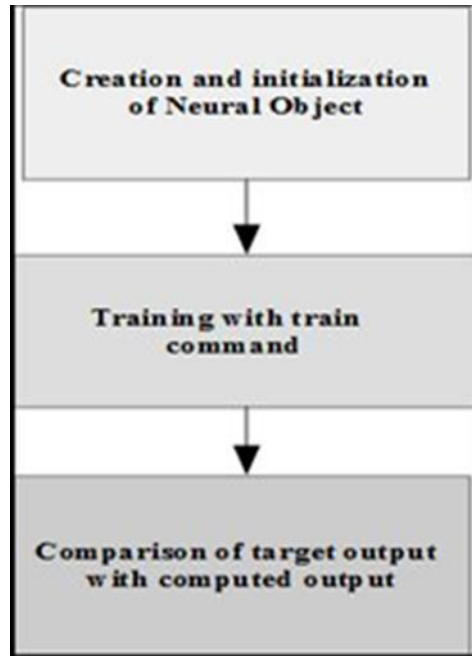
- The open price of the historical data
- The high price of the historical data
- The low price of the historical data
- The close price of the historical data

- Change in closing price (previous day close-current-day close)
- Percentage change
- Volume Weighted Average price of the historic data
- The algorithm to develop neural network is shown below:
- First step is to define the target.
- Select the input data parameters.
- Choosing the correct neural network architecture, training function.
- Select the learning function to learn the network from examples and training set.
- After that, the training and testing process is performed for every iteration.
- If user is satisfied with the network produced output, go to next step, otherwise start performing from step 3
- Finally store or record the network result.



**Figure 10:** Prediction Process overview using NN

To design the neural network, MATLAB Tool is used. In the MATLAB, there is a Neural Network Tool Box named field that provides all the needed functionality to develop neural network.



**Figure 11:** Training Process

The whole dataset used in this research hold the data of trading days from 1st January, 2009 to 29th January, 2015. The attributes involve in the dataset are previous day close prices of Nifty Index and SBI Bank. The source of historical data collection is [www.nseindia.com](http://www.nseindia.com). The research given in this paper simply tries to makes an appropriate neural network model, because the efficiency of the predictions heavily depends on the performance of the model. There are lot of critical issues while designing the network, for i.e. input variables selection, design of network architecture are considered very carefully.

**3.3.1 Selection of input variables:** The performance of the network depends upon the selection of input variable because a lot of information related with the complex linear

and non linear structures of the data lies in it. These variables play very important role in understanding the movements in the time series. While selecting the input variables, the thing that should be taken into account very seriously is that the variables should be highly predictive. While developing model for forecasting, selecting the quantity of input variables for the forecasting is the important aspect of this whole process of forecasting, because forecasting via neural network is fully based on input variables. We have considered various input structure and the performance of the network is measured on the basis of a trial and error method. The historical data is divided into known and unknown data. The known data (training set) is used to develop the neural network model and for training the same. The unknown data (testing set) is used for evaluating the network performance with the unseen and new dataset and this new dataset was not used in any training process of the network. The network performance is tested with the different-different for observing the performance of the model in different-different range of data historic periods. In this paper, the parameters used in the input structures are given below

- closing price of previous\_day of Stock
- closing price of previous\_day Bank Nifty

The next day closing price is predictand. The prediction of the future of the stock on next day ( $x+1$ ) is based on above given two predictors. The whole data set consist of 1509 days SBI stock data. In the figure, the data set is divided into input series and target series to train the network. Input series hold the historical previous close price of stock and bank nifty. Target series hold the next day closing price of stock. While training, network tries to learn the relationship between the given parameters in input series and target series and further uses the relationship to predict the future. All five input parameters are given to the network & also corresponding target is given and with the help of back propagation train, the mapping between input & output and minimization of error is learnt by the network. Error in this context is the difference between actual target value and predicted value. After getting the minimum mean square error over different epochs the appropriate and required training is done & finally the prediction system becomes ready for forecasting. In this paper, two input configurations are discussed. In

the second configuration, the previous close of bank nifty also taken as an input to show that the variation in banking stocks value also depends on the stock value of bank index.

**3.3.2 Data Pre-processing:** The network accuracy for the predictions depends largely on the quality of data used. Data should be in homogeneous form. In the case of stocks data, sometimes heterogeneity exists in the data pattern due to stock split. Stock split is actually the decision made by the board of directors of the companies to increase the number of outstanding shares by issuing more shares to current shareholder. There are several reasons of stock split like psychology, stock's liquidity. When the price of a stock goes upward, some stock investors may have the feeling that the price is too high for them to buy, or small investors may take it as an unaffordable. The advantage of splitting the stock is to bring the share price down to a level at which more investors can be attracted and it also gives existing shareholders the feeling that they suddenly have more shares than they did before, and if once the price rises, they have more stock to trade. It's important to convert the data into appropriate form before implementation because the heterogeneous data leads to strange graphical representation of results. There are various methods to analyze that the specific day closing price will increase or decrease. These methods are given below:

- **Typical Price:** It can be calculated by adding the highest price, lowest price and close price and then dividing the sum by number three.

**Algorithm:**

1. Give high, open and close price of a specific day.
  2. In the output array store the addition of values of these three parameters.
  3. Divide this output array by three.
- **Relative Strength Index:** It works by comparing the number of days in which stock finished with up and finished with down. It is calculated for a certain fixed time period. This time span may be 9 to 15 days. The ratio of average number of up days to average number of down days is added to number one. After that the result is divided with 100 and is subtracted from 100. Its range lies from 0 to 100.

If the value of result is 100 or above that, then it is the clear indication of overbought due to fall in price. When the value is below 30, then it is the clear indication of oversold.

- **Moving Average:** It has been the most useful indicator. It leads to average price for some period of time. To calculate moving average for 30 days, take sum of closing price of each of 30 days and then divided the sum by number of days. The most common interval used to calculate moving average is 20, 30, 50, 100, and 200 days. Actually mostly researcher always try to prefer longer time because it is very less affected by fluctuations occur in daily price. When any price value show some fall up to below the moving average then the stocks have a tendency to keep on having falling trend. On the opposite, when prices show some increment up to above the moving average then the stocks have a tendency to keep on having rising trend.

## Chapter 4

# RESULT AND DISCUSSIONS

---

The major use of the neural network in forecasting purpose is studied in this proposal. From the testing, one thing that come to know that to train the neural network and to make it able to give a better result, a lot of data is required at least one year data. A lot of work in future can be done on this topic by using different kind of neural network such as Feed Forward Network, Multilayer Neural Network and Recurrent Network. The result receives are showing that Neural Network is the best technique for time series forecasting processes. In this report, a non linear and the flexible technique Neural Network has been used to predict the future stock prices of selected companies comes under National Stock Exchange. The input parameters used in the research are historical open price, close price, change in close price, high price and low price of the stocks. From the results, one thing is very clear that neural network has the ability to predict stock prices with a lot of accuracy.

The dataset used in model is divided into three sets: training, validation and testing dataset. 70% of whole data is used as training to calculating the gradient and for updating the values of weights according to MSE value and biases, 15% of total data is used for validation. In the validation phase, the error is monitored on the validation set during the training process. In most of the cases, in the initial phases while training there is reduction in the error on the validation dataset. However, when the data begins over fitted by the network, then the error for the validation dataset begins to increase. At the minimum of the validation set error, the final values of network parameters like weights and biases values are attained and saved. Test set is also very important as well as validation set because the test set error is used to choose the appropriate model by comparing with each other. Using gradient descent back propagation algorithm the network is trained once up to 45 epochs. After training ANN model is tested over test dataset. The comparison of mean square error over training data set and testing data set is given below in the table:

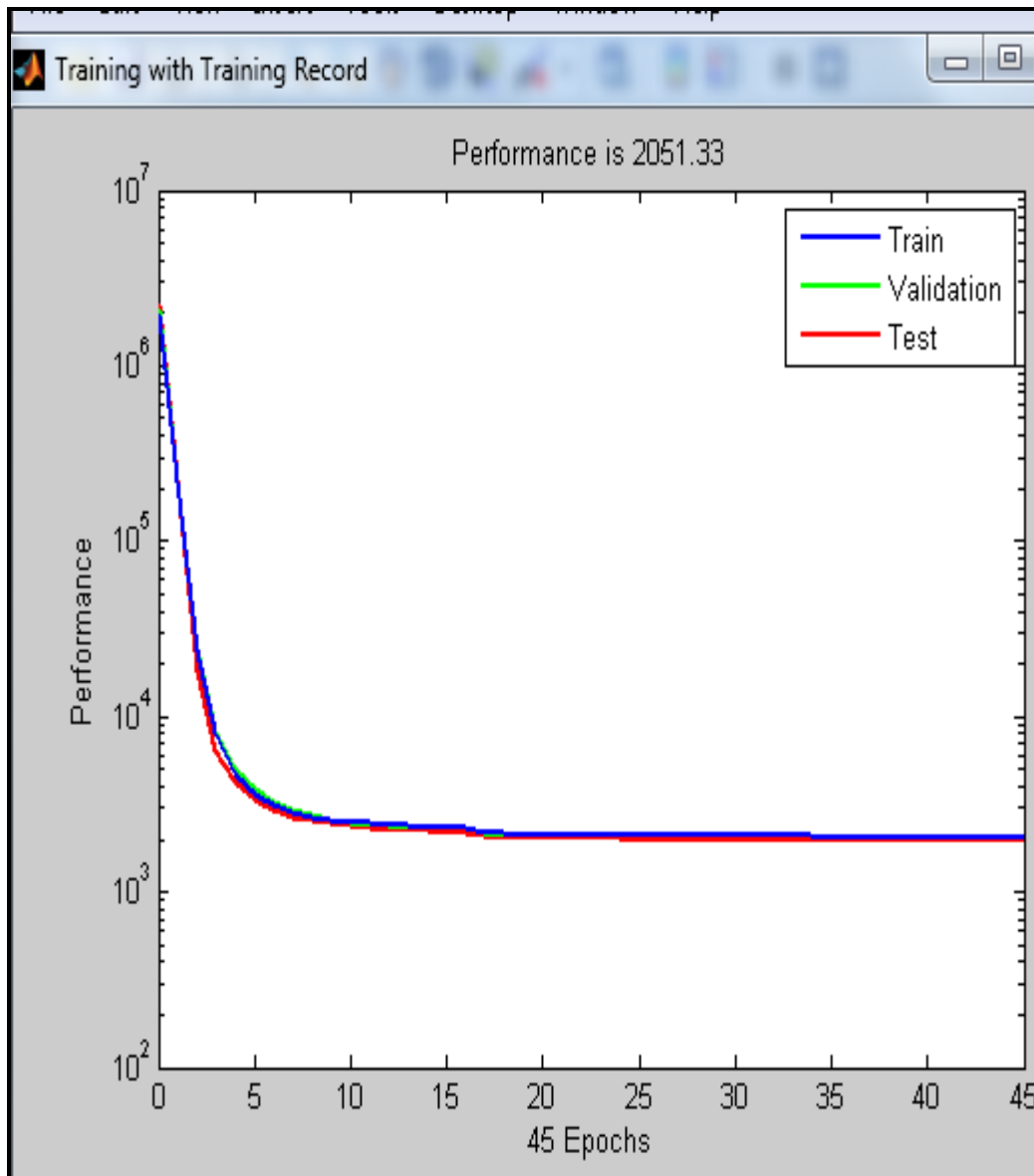
**Table 5: Comparison of Error**

| <b>NETWORK DATA</b>       | <b>1NN</b> | <b>NN2</b> |
|---------------------------|------------|------------|
| <b>USING TRAINING SET</b> | 3.90       | 4.03       |
| <b>USING TESTING SET</b>  | 13.20      | 13.34      |

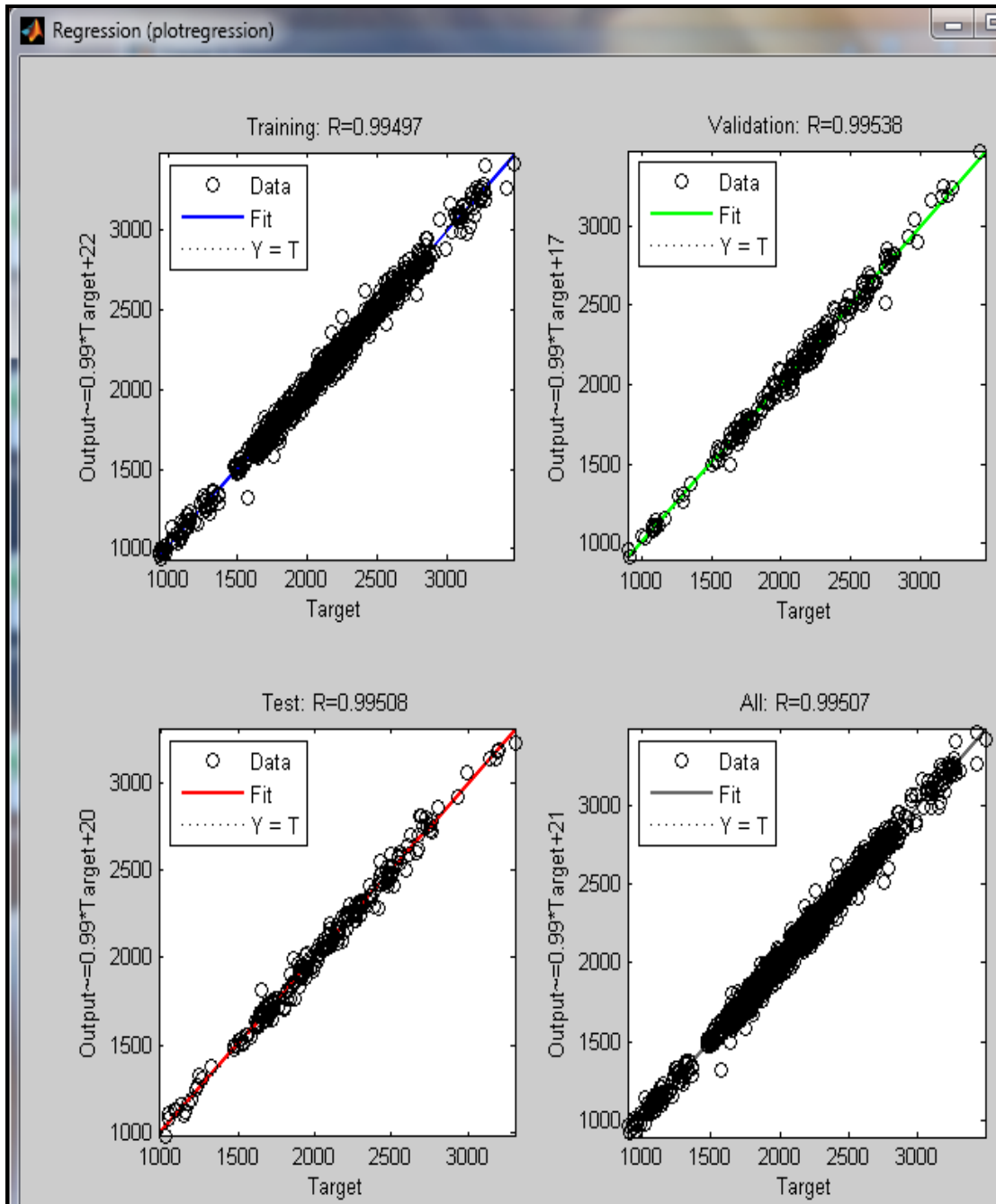
The relationship between independent and dependent variable can be described with the help of regression. In this paper, the regression graph is computed and plotted for networks over SBI data forecasting problem. To validate the network, the creation of regression plot is important as it graphically shows how much network output and the targets values are near about each other.

Perfect training leads to the network outputs and the targets should be exactly same and equal. But it is very rare in practice. The three plots given below represent the performance of training set, validation set, and testing dataset. The perfect and right results are represented by the dashed line drawn in each plot. It graphically tries to say that outputs = targets. The best fit linear regression is represented by the solid line. The R value represents the relationship between the outputs and targets. If  $R = 1$ , then this leads to the linear relationship between outputs and targets. If R is close to zero, then it represents the absence of linear relationship.





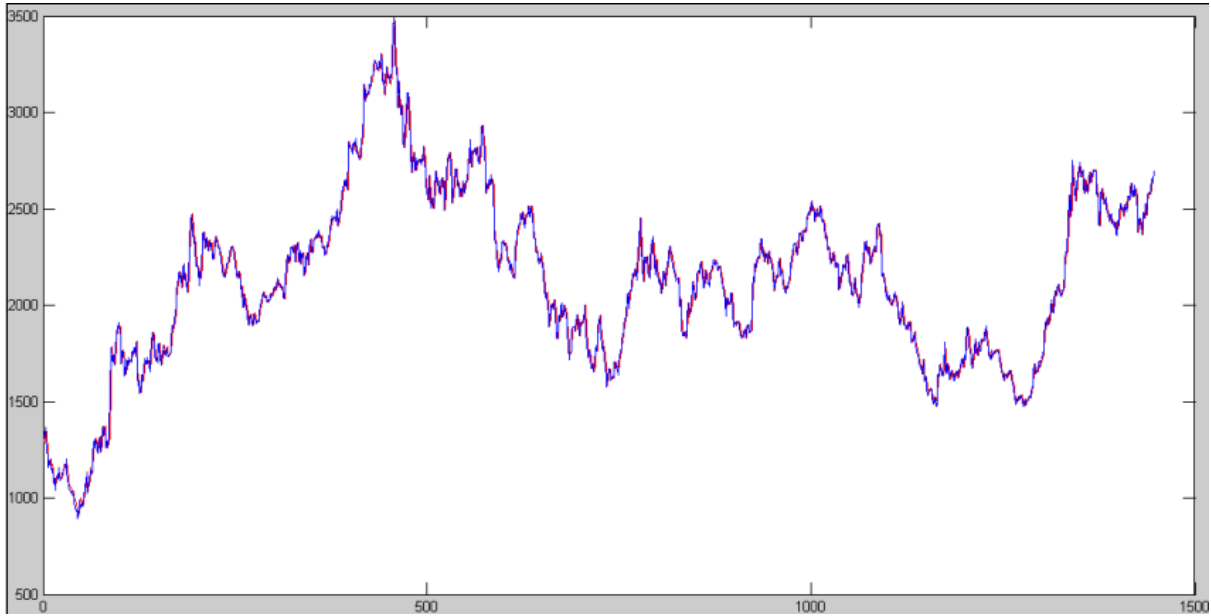
**Figure 12:** Performance plot for Network



**Figure 13:** Regression plot for SBI stock

This figure defines the performance of the network training and it is indicated from the figure that the training process has not any problem as the similarity exist in the validation and test curves. It is used to ensure the existence of over fitting

Over fitting is actually a problem that exists with the network training. When the network generalization capability is not very well, then there may be chances of over fitting problem. If the test curve had increased significantly before the validation curve increased, then it is the indication of that some over fitting might have occurred.



**Figure 14:** Plot of Training of the Network

**Performance Measurement:** The relative percentage error is used to measure the performance of the model. The error in percentage can be calculated by:

$$\frac{\text{actual value} - \text{predicted value}}{\text{actual value}} * 100$$

In the model given in the proposal, the error value is 25%. Therefore the accuracy of the network will be 75%. The accuracy of the network can be calculated by given below equation:

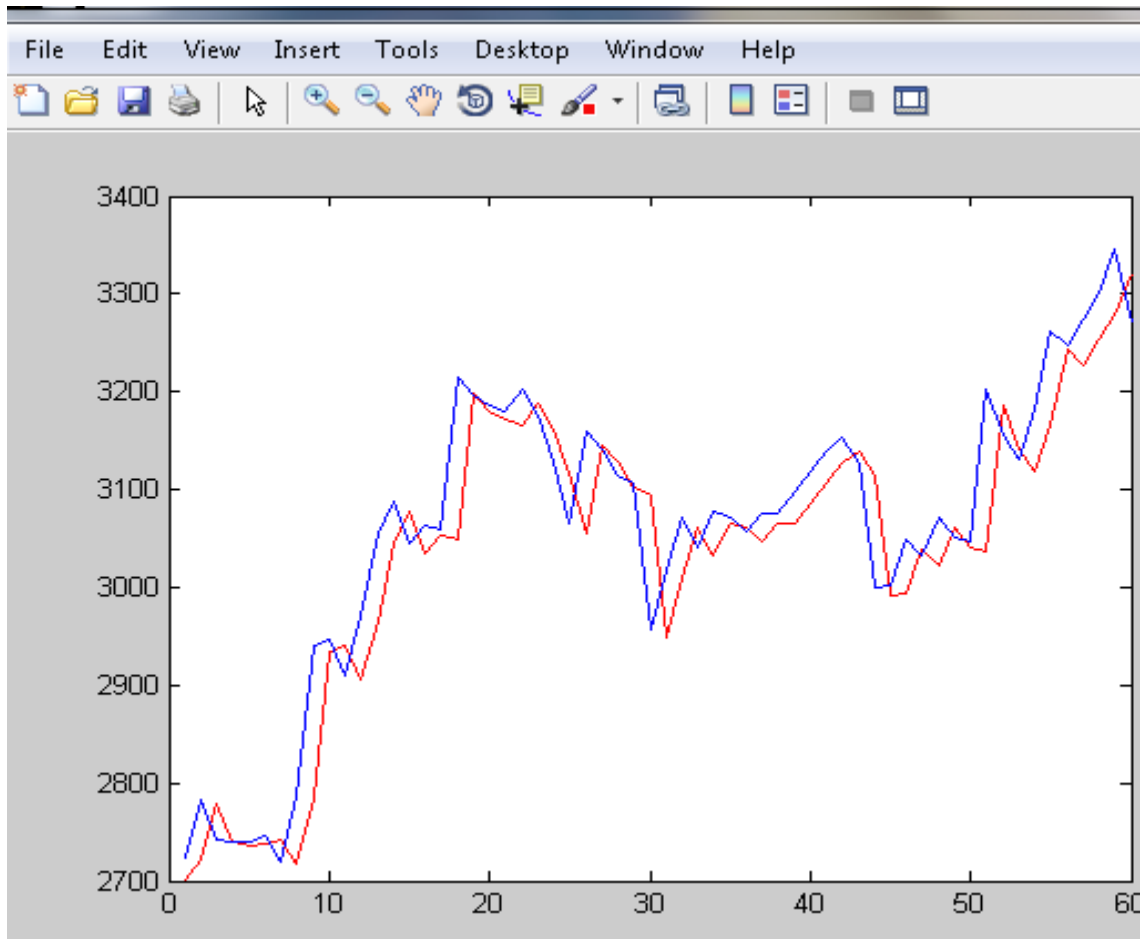
$$\text{Accuracy} = 100 - \text{Error of the Network}$$

| TARGET  | 1NN    | 2NN    | TARGET<br>3187 | 1NN<br>3189.1 | 2NN<br>3180.9 |
|---------|--------|--------|----------------|---------------|---------------|
| 2723.45 | 2700.1 | 2700.4 | 3179.5         | 3180.2        | 3172.3        |
| 2781.9  | 2720.6 | 2720.8 | 3203           | 3172.8        | 3165          |
| 2742.5  | 2778.6 | 2778.5 | 3175.5         | 3196          | 3187.4        |
| 2739.5  | 2739.5 | 2739.5 | 3129           | 3168.8        | 3161.1        |
| 2740.35 | 2736.5 | 2736.6 | 3065.5         | 3122.8        | 3116.6        |
| 2746.05 | 2737.4 | 2737.4 | 3159.5         | 3059.9        | 3055.5        |
| 2719.3  | 2743   | 2742.8 | 3140.5         | 3153          | 3146          |
| 2787.85 | 2716.5 | 2716.4 | 3113.5         | 3134.2        | 3127.8        |
| 2940.6  | 2784.5 | 2784.3 | 3105.5         | 3107.5        | 3101.9        |
| 2945.75 | 2936.1 | 2934.5 | 2956           | 3099.5        | 3094.1        |
| 2910.5  | 2941.2 | 2939.5 | 3017.5         | 2951.4        | 2949.4        |
| 2971    | 2906.3 | 2905.1 | 3071           | 3012.4        | 3009.3        |
| 3055    | 2966.3 | 2964.3 | 3041.5         | 3065.4        | 3060.8        |
| 3088.5  | 3049.5 | 3045.6 | 3077           | 3036.2        | 3032.2        |
| 3044.5  | 3082.7 | 3077.8 | 3072           | 3071.3        | 3066.3        |
| 3062.5  | 3039.1 | 3035.5 | 3057.5         | 3066.4        | 3061.6        |
| 3058.5  | 3057   | 3053   | 3076.5         | 3052          | 3047.6        |
| 3214    | 3053   | 3049.1 | 3075.5         | 3070.8        | 3065.9        |

**Table-6:** Comparison between original stock price (TARGET) & simulated price by ANN.

1NN-Simulated output using neural model with input configuration 1 (previous close price of equity)

2NN-.Simulated output using neural model with input configuration 2 (previous close price of equity and index)



**Figure 15:** Plot of Actual vs. Predicted for Period of NOV, 2014 to JAN, 2015

## Chapter 5

# CONCLUSION AND FUTURE SCOPE

---

Stock market prediction relies completely on input parameters that play a crucial role as the improper input configuration choice may be the reason of lack of accuracy in stock predictions. Stock market is influenced by different-different parameters as some parameters lead to huge impact on stock price and some have less impact and that are why the correct selection of input variables is very important. Mostly machine learning techniques are used technical analysis variables. Some other variables like fundamental variables, microeconomic indicators, news articles, etc can also be used by other techniques. But these all can utilize properly for long term predictions not for short term. In this paper, short term predictions are made. From the survey one thing can be concluded that technical variable give better prediction accuracy in the context of machine learning used as a technique to predict stocks.

There is a lot of scope and alternatives to enhance the accuracy of the predictions made by the neural network. In this paper, only historic previous closing prices of the bank equity and bank nifty are used for prediction. Other economical factors can also be utilized through adding more inputs related with bank nifty to predict the banking sector stock. By using the economical data, fundamental analysis based data; we can predict bank equity prices for longer term. So the various factor of the Bank Nifty in future can also be utilized as the input variables to check improvement in the performance and the accuracy for the stock predictions.

## Chapter 6

# REFERENCES

---

### I. Websites

1. [www.nseindia.com](http://www.nseindia.com)
2. [www.mathworks.in/help/nnet/ug/train-and-apply-multilayer-neural-networks.html](http://www.mathworks.in/help/nnet/ug/train-and-apply-multilayer-neural-networks.html)
3. [www.byclb.com/TR/Tutorials/neural\\_networks/](http://www.byclb.com/TR/Tutorials/neural_networks/)

### II. Research Papers

- [1] Arka Ghosh “**Comparative study of Financial Time Series Prediction By Artificial Neural Network with Gradient Descent Learning**” International Journal of Scientific and Engineering Research, Volume 3 Issue 1, January 2012
- [2] Alex A, Freitas. “**Data Mining and Knowledge Discovery with Evolutionary Algorithms**”. Springer 2002
- [3] Ashish Ghosh, Lakhmi C. Jain, “**Evolutionary Computation in Data Mining**”. Springer 2004
- [4] Binoy B. Nair, N.Mohana Dharini, V.P. Mohandas. “**A Stock market trend prediction system using a Hybrid decision tree-neuro-fuzzy system**” Proceedings of International Conference on Advances in Recent Technologies in Communication and Computing, 2010 IEEE
- [5] Dong Jing-rong “**Combining Stock Market Volatility Forecasts Using Bayesian Technique**” IEEE 2008

- [6] Divyang J. Joshi, Pratik H. Bhavsarin **“Predicting Nifty 50 Movement Use of Advance Decline Ratio”** Journal of Applied Research, Volume 1 Issue 3, December 2011
- [7] Esmaeil Hadavandi, Hassan Shavandi, and Arash Ghanbari. **“Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting”** Journal of Knowledge-Based Systems, Volume 23 Issue 8 ,Pages 800-808, December 2010
- [8] Farhad Soleimanian Gharehchopogh, Tahiminesh Haddadi Bonab and Seyyed Reza Khaze. **“A Linear Regression Approach to Prediction of Stock Market Tradition Volume: A Case Study”** International Journal of Managing Value and Supply Chain Vol.4, No.3, September 2013
- [9] Fagner Andrade de Oliveria, Luis E. Zárate, Marcos de Azevedo Reis and Cristiane Neri Nobre **“The use of artificial neural networks in the analysis and prediction of stock prices“** in SMC 2011
- [10] Huang Fu-Yuan **“Forecasting Stock Price Using a Genetic Fuzzy Neural Network”** Proceedings of International Conference on Computer Science and Information Technology, ICCSIT '08 , 2008 IEEE
- [11] Khalid A lkhatib, Hassan Najadat<sup>2</sup>, Ismail Hmeidi, Mohammed K. Ali Shatnawi **“Stock Price Prediction Using K-Nearest Neighbour Algorithm”** International Journal of Business, Humanities and Technology Vol. 3 No. 3; March 2013
- [12] Marc J. Schniederjans, Karyl B. Leggio, Quing Cao. **“A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market”** Journal of Computers and



- [13] Md. Rafiul Hassan and Baikunth Nath. **“Stock Market Forecasting Using Hidden Markov Model: A New Approach”** Proceedings of 5th International Conference on Intelligent Systems Design and Applications, ISDA '05 , 2005 IEEE
- [14] Robert J. Van Eyden. **“The Application of Neural Networks in the Forecasting of Share Prices”**. Finance and Technology Publishing, 1996
- [15] R. Schwaerzel and B. Rosen. **“Improving the accuracy of financial time series prediction using ensemble networks and high order statistics”**. IEEE 2003
- [16] S. Kheradyar, I. Ibrahim, and F. Mat Nor **“Stock Return Predictability with Financial Ratios”** International journal of Trade Economics and Finance, Vol. 2, No. 5, October 2011
- [17] Wei Yang; Zhejiang Univ, Hangzhou. **“Stock Price Prediction Based on Fuzzy Logic”** 2007 IEEE
- [18] Yuling LIN, Haixiang GUO and Jinglu HU. **“An SVM-based approach for stock market trend prediction”** Proceedings of 2013 International Joint Conference on Neural Network, 2013 IEEE
- [19] Z.Tang and P.A.Fishwick, **“Back propagation neural nets as models for time series forecasting”** ORSA journal on computing, vol.5, No. 4, pp 374-384, 1993

### 9.1 Abbreviations

1. **NN (Neural Network)** – Neural networks is technique used to model the human brain.
2. **P/E (Price/Earning Ratio)** – It is the ratio of current market price to earning per share.
3. **SVM (Support Vector Machine)** - It is used as a classifier to select the best input variables.
4. **BP (Back Propagation)** –It is a learning algorithm that is used to train the network.
5. **RSI (Relative Strength Index)** – It works by comparing the number of days in which stock finished with up and finished with down.
6. **MSE (Mean Square Error)** – The **mean squared error (MSE)** of an ESTIMATOR measures the average of the squares of the "errors", that is, the difference between the estimator and what is estimated.
7. **IEEE (Institute of Electrical and Electronics Engineers)** - It is a non-profit organization of technical professional association for electrical and electronics engineering.

