# Stock Investment Returns Prediction: A Comparison of ANN and ARIMA

Shraddha Painjane
*Department of Computer Engineering*
*Shree L.R. Tiwari college of engineering*
*Mira Road, Mumbai, India*
shraddhapainjane@gmail.com

Prof. Sachin Bojewar
*Department of Computer Engineering*
*Vidyalankar Institute of technology*
*Mumbai, India.*

Prof. Deepali Patil
*Department of Computer Engineering*
*Shree L.R. Tiwari college of engineering*
*Mira Road, Mumbai, India*

***Abstract*** - Every Investors dream is to obtain prior knowledge of the market's ups and downs before it actually happens. It is very difficult to forecast it accurately and consistently but it has now become possible to create market forecast with algorithms. Estimation and prediction of stock price performance is very complex and difficult. Multiple analytical techniques are available and can be used to improve investors forecasting ability and enhance investment decision making.The growing trend of utilizing advanced algorithms is a solution for predictions and to gage market movements.

***Index Terms*** - *Machine learning, prediction, supervised learning, stock market.*

*****

## I. INTRODUCTION

Data mining perfectly sets on the theory that the past or historic data holds the important memory for prediction of future direction.The prediction of stock markets is considered as a toughest task of financial time series prediction. This technology helps the investors to discover hidden patterns from the past, historical data to enhance the predictive capability for investment decisions[6]. There is a rapid progress in digital data acquisition which has led to fast growing content of data that are stored in databases or data warehouses.

Stock market is an essential part of the finance and economy of a country. Stock Market has high profit and high risk features due to which its prediction must be highly accurate, but the main issue about such data are that these are highly complex, random, nonlinear function and can only be learnt by a data mining method such as neural networks and some other to recognize future market trend. The continuous highs and lows in the stock market makes it very difficult to directly predict what will be the price upcoming in the market next day. Hence without prior knowledge about the market, it is not easy for investors as well as the new comers in stock market to invest for particular stock. Due to complex and unpredictable movements in stock market prices, it becomes difficult for investors to invest without knowing if it will be a profit or loss.

Stock trend prediction has been an exciting topic for researchers from different fields. Machine learning is considered as a best algorithm for prediction of financial market. The large data volume makes it difficult to extract them manually hence powerful tools are required to extract valuable information hiding behind the data. Financialinvestors who invest in stock market are usually not aware of the behaviour of stock market. Due to this, these stock traders face a lot of problems while buying or selling any stock as they do not know or have any idea which would gain them more profit. Although the investors have the idea that they have to deal with daily news of stock market and vast amount of information, but extracting valuable knowledge from such a vast data exceeds human capabilities. Hence data mining techniques using appropriate algorithms makes it easier to extract useful knowledge from resources. Machine learning is the process that is used to learn the large dataset and avail useful knowledge from it.

Machine Learning was basically organized to satisfy three primary focuses: Task oriented studies, theoretical analysis and cognitive simulation[5]. In task oriented studies, there is development and analysis of learning system so that the performance can be improved in predetermined set of tasks. In theoretical Analysis, there is a theoretical exploration of space of possible learning algorithms that are not dependent on application domain. In cognitive simulation,investigation and computer simulation of learning process is done.Machine learning can be classified along many different dimensions.Firstlyon the basis of strategies that can be used in learning. secondly on the basis of the representation of knowledge by the learner.and lastly in terms of application domain in which knowledge is acquired for performance system.

Supervised Learning demands the learning of mapping between a set of input variables x and an output variable y and then this mapping is applied to predict the outputs for unseen

data. It is the most important learning technique in machine learning.The goal of supervised learning is to build a concise model of the distribution of class labels in terms of predictor features. The resulting classifier is then used to assign class labels to the testing instances where the values of the predictor features are known, but the value of the class label is unknown. Supervised learning can be generalised in several ways: Semi supervised learning-In this the desired output value s are provided to training data as a subset[7]. Active Learning-This makes query to a human user. Structured prediction-In this the standard methods are extended when a desired output value is a complex object. Learning to rank-In this too the standard methods are extended when the input is a set of objects and the output that is expected is a ranking of those objects.

The proposed system uses the historical stock data of past 2 to 3 years. This historical data contains high, low, close price, open price, percent change etc. These Parameters will be used to calculate the further prediction. The user will want to know the returns he will get after he invest a particular amount in a particular stock. Hence this will be predicted on the basis of historical stock data by applying prediction algorithms like Neural network and ARIMA. The proposed system will predict the returns by applying both the algorithms and will give the result that which algorithm suits best for the stock prediction investment returns. This proposed system will not only tell which algorithm gives accurate prediction but will also give the system which predicts investment returns with minimized error.The objective of this system is to predict the investment returns for a particular years using prediction algorithms. Prediction of stock market is very complex and difficult to estimate hence need to use analytical methods to enhance forecasting ability and improve decision making in stock market. The main objective of the project is to calculate the investment returns using the historical stock data which contains the parameters such as highs, lows, opening price etc and apply prediction algorithms like neural network and ARIMA and give the result that which algorithm predicts best with minimized error in return.

## II ARTIFICIAL NEURAL NETWORK

Artificial neural network has been an option for last several decades and are successfully used in the applications of different domain such as finance, medicine etc.

"A neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes" – DARPA Neural Network Study (1988)

"Artificial neural systems, or neural networks, are physical cellular systems which can acquire, store, and utilize experiential knowledge." – Zurada (1992)

Artificial neurons are nothing but the ones that are inspired from biological neuronal structure. As the biological neuron works, Artificial neuron works in a similar manner. In Feed Forward network, neurons are arranged in the form of layers. The neurons in a layered structure gets input from the previous layer and feed their output to the next layer. In this type, connections to the neurons on the same layer or the layer before is not permitted. [1]The last layer is known as output layer. The layer which intakes the input is known as the input layer and the layers between the input layer and output layer are known as hidden layers. Input layer consist of input neurons which transmits only input provided to their outputs.

Network consisting only a layer of input nodes and a single layer of neurons are called single layer network.[1] Network consisting of a layer of input node and one or more hidden layers are called multilayer networks. The networks in which connections to the same neurons or the connections to the neurons before are allowed are called recurrent networks. The artificial neuron given in this figure has *N* input, denoted as u1,u2....each line connecting these inputs to the neuron is assigned a weight which are denoted as w1,w2…respectively. Weights in the artificial model corresponds to the synaptic connections in biological neurons. The threshold in artificial neurons is represented by Θ.

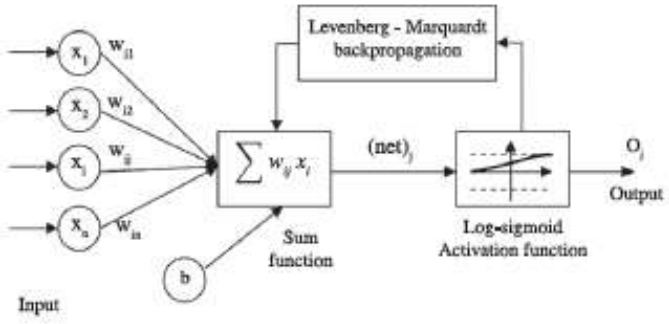


**Figure 1.** Architecture of applied neural network[36].

The inputs and the weights are real values. A negative value for a weight indicates an **inhibitory** connection while a positive value indicates an **excitatory** one. Although in biological neurons, has a negative value, it may be assigned a positive value in artificial neuron models. Sometimes, the threshold is combined for simplicity into the summation part by assuming an imaginary input $u_0$ =+1 and a connection weight $w_0$ = **θ .**

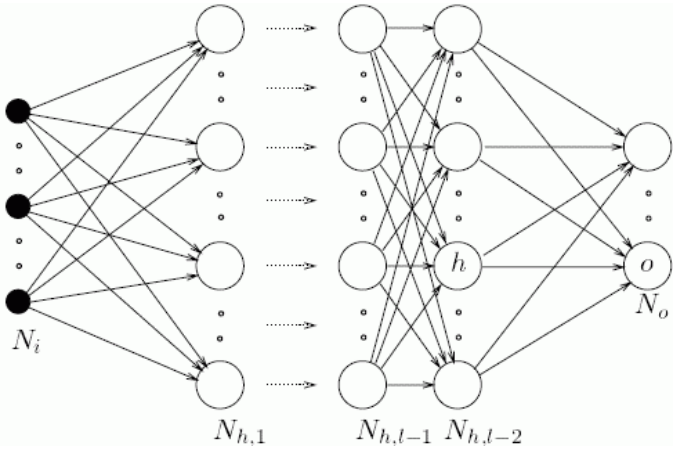


**Figure 2. Recurrent Neural Network**

The artificial neural network that we train for the prediction of stock data has an arbitrary number of hidden layers, and arbitrary number of hidden nodes in each layer, which is decided by the user in the run-time.

The error for convergence is calculated as the rms error between the target values and the actual outputs. This same error is used to report the performance of the algorithm on the test set. Hence after learning the artificial neural network one can say that it is an effective tool for stock market prediction and can be used on real world dataset.

## III AUTOREGRESSIVE INTEGRATED MOVING AVERAGE

Generally, it is found that prediction can be done using two perspectives: statistical and artificial intelligence techniques.ARIMA model uses statistical perspective.ARIMA models are said to be more robust and efficient when it comes to be applied on short term prediction in financial time series forecasting.Box and Jenkins in 1970 introduced the ARIMA model[3]. It is referred as Box-Jenkins methodology which composes set of activities for identifying, estimating and diagnosing ARIMA models with time series data.

The ARIMA transaction estimates and evaluates equally spaced univariate time series data, intervention data and transfer function data using Auto Regressive Moving Average (ARMA) or Autoregressive Integrated Moving-Average (ARIMA) model[4]. A value is predicted in a response time series by an ARIMA model as a linear combination of its own values, current and past values of other time series, and past errors. A comprehensive set of tools is provided to ARIMA procedures for identification, parameter estimation and forecasting of univariate time series model and it offers great flexibility that can be analyzed.In ARIMA model, the future value of a variable is a linear combination of past values and past errors

The steps in building ARIMA predictive model consist of model identification, parameter estimation and diagnostic checking.

ARIMA has four major steps in model building: Identification, Estimation, Diagnostics & Forecast. Using these four steps first tentative model parameters are identified through graphs ACF and PACF and then coefficient are determined and a likely model is identified, then the next step is to validate the model and at the end use simple statistics and confidence intervals to determine the validity of the forecast and track model performance. ARIMA model uses the historic data and decomposes it into Autoregressive (AR) which indicates weighted moving average over past observations, Integrated (I) which indicates Moving Average and linear trends or polynomial trend which indicates weighted moving average over past errors. Therefore, it has three model parameters AR(p), I(d) and MA(q) which when combines into ARIMA(p, d, q) model where

p = order of autocorrelation
d = order of integration (differencing)
q = order of moving averages

A non-seasonal stationary time-series can be modelled as a combination of the past values and the errors which can be denoted as ARIMA (p, d, q).

## CONCLUSION

Investment return price after certain years of investment which the user will give as input. This price will calculated using neural network and ARIMA. It will consider all the historical stock market data on which expected result will be provided to the user.

## REFERENCES

[1] Zhang, Guoqiang, B. Eddy Patuwo, and Michael Y. Hu. "Forecasting with artificial neural networks:: The state of the art." *International journal of forecasting* 14.1 (1998): 35-62.

[2] Sharma, Kiran, AnkitNaik, and Purushottam Patel. "Study of Artificial Neural Network." *Int. J. of Advanced Research Trends in Engineering and Technology (IJARTET)* 2.4 (2015): 46-48.

[3] Pai, Ping-Feng, and Chih-Sheng Lin. "A hybrid ARIMA and support vector machines model in stock price forecasting." *Omega* 33.6 (2005): 497-505.

[4] Yu, Yanhua, et al. "Network traffic prediction and result analysis based on seasonal ARIMA and correlation coefficient." *Intelligent System Design and Engineering Application (ISDEA), 2010 International Conference on*. Vol. 1. IEEE, 2010.

[5] Dash, Manoranjan, and Huan Liu. "Feature selection for classification." *Intelligent data analysis* 1.1-4 (1997): 131-156.

[6] Han, Jiawei, Jian Pei, and MichelineKamber. *Data mining: concepts and techniques*. Elsevier, 2011.

[7] Chapelle, Olivier, Bernhard Scholkopf, and Alexander Zien. "Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]." *IEEE Transactions on Neural Networks* 20.3 (2009): 542-542.