2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)
International Institute of Information Technology (I²IT), Pune

A COMPARATIVE STUDY ON TECHNIQUES USED FOR PREDICTION OF STOCK MARKET

Lakshmi Tharun Ponnam
Department of Computer Science and Engineering
V R Siddhartha Engineering College
Vijayawada, India
lakshmitharunponnam@gmail.com*

V. Srinivasa Rao
HOD, Department of Computer Science and Engineering
V R Siddhartha Engineering College
Vijayawada, India
hodcse@vrsiddhartha.ac.in

K. Srinivas
Department of Computer Science and Engineering
V R Siddhartha Engineering College
Vijayawada, India
vrdrks@vrsiddhartha.ac.in

Vamsi Raavi
Department of Computer Science and Engineering
V R Siddhartha Engineering College
Vijayawada, India
raavivamsi305@gmail.com

Abstract— **The most significant issue in the recent times in finance is finding the systematic ways to abridge and envision the stock market data. Stock market analysis gives useful information to Entrepreneurs, Individuals and Institutions about the etiquette of the market helping with speculation decisions. Many prediction models have been developed during the last decade. As a comparative study, we will be analyzing the ARIMA model, Neural Nets, Holt winters and the TSLM linear model. These models are outlined to assist the investors, bankers and capitalists to find out the concealed patterns that are present in the historical data. For the analysis purpose, we will be using five different companies' data procured through the Yahoo Finance. As the prediction of stock market is always considered as a provocative job and the precision of the predictions done has become an important part a comparative study is even done while removing the seasonal trends in the data too. We have observed that the Holt Winters model while removing the seasonal trends outperformed when compared to the rest models with a higher prediction accuracy.**

Keywords— Stock Market; ARIMA model; Neural Nets; Holt Winters; Time Series Linear model; Seasonal Trends;

I. INTRODUCTION

Investing in stock market is never easy as the movement of Stock prices is highly mobilized, turbulent and dynamic process. Analysis is the only way by which we can make this process of investment easy. Financial Markets can be analyzed mainly in two ways one way is Technical analysis and the next is the Fundamental analysis. Technical analysis focuses and looks at the variations in the price of the stock. Whereas the Fundamental analysis of the stock market tries to analyze the Stock by looking at the cardinal components like the companies news articles, opinion of the Analysts, Reviewers etc. Fundamental analysis doesn't seems to be trustworthy because the decision that are made by this way may not have any kind of scientific reasons. So there by we will be more focusing on the Technical Analysis of the Stock market.

Prediction is always considered to be an interesting area in the field of Research especially in the domain of technical analysis and making researchers to engage in this domain and to ameliorate the prevailing predictive models. The most important cause for this is Entrepreneurs, Investors, Individuals, and Capitalists are entitled to make speculations and their potential to the scheme and to plan the productive scheme regarding their everyday and future attempts. As the prediction of the stock market is always considered to be an arduous task due to the complex nature of stock market Stock price [1, 2, 3]. The strong wish of the investors is to have a method to forecast which could assure maximum profits from the forecast and minimizing the losses.

Different techniques has been introduced by the researchers in order to predict the future value of the stock. As a part of comparative study we will be analyzing the different techniques from different domains including ARIMA model, Neural Nets, Holt winters and TSLM Linear model. We will be analyzing these models mainly because of their popularity and the prediction accuracy. Of these models the first one ARIMA (Autoregressive integrated moving average) model which belongs to the domain of statistical model. Predominantly it is stated that prediction can be performed in two ways includes Artificial intelligence techniques and statistical Techniques [2]. ARIMA model is known for its resilience, systematism and efficiency in the prediction of stock markets in short time when compared to the artificial neural network techniques [4, 5, 6]. ARIMA models are been used in large scale in the fields of Finance, Monetary and economic fields. Other statistics models are regression method, exponential smoothing and GARCH (Generalized Auto Regressive Conditional Heteroscedasticity).

The next model Holt winters is generally used when the data that we are working on exhibits tendency to shift and seasonality [7]. The two important Holt winter models are the Additive model and the Multiplicative model. Generally Additive model is used for the data that exhibits tendency to shift and the periodic variations in the data are constant. Multiplicative models are used for the data which exhibits periodic differences that are altering correspondingly.

The next method that we would be studying is the most popular and widely used Artificial Neural networks. There have been an increase in the attentiveness of the researchers

978-1-5090-2080-5/16/$31.00 ©2016 IEEE

and been implemented in various domains including Prediction, Physical and medical sciences, Finance, Engineering and wide range of problems. An Artificial Neural network is just nothing but an integrated system of a large number of elementary processing segments which work in collaterally and whose primary task is intended by the structure of the network, Weights or sturdiness assigned to the connections and the process that is performed by the elementary nodes [8]. The predictions that are obtained using the Artificial Neural Networks are almost precise except in the case that if there is a large change in the real data as it is very difficult to predict the changes in the stock exactly[9].

The next linear model in contrast with a nonlinear model like Artificial Neural networks that we would like to discuss is the TSLM (Time Series Linear Model). This TSLM is used to fit the linear models in to the seasonal elements as well as the tendency components [10]. TSLM is generally considered as a huge wrapper which allows the forecast of the Seasonal as well as the tendency components which are considered to glide and hail from the data. Forecasting is made simple and effective using the Time series linear model.

In order to make this comparative study a suitable programming language is necessary. In the field of analysis of time series R has created itself an alternative for many Programmers, Researchers and institutes. R being an open source language provides standard computational environment with lots of packages that are inbuilt assisting the programmers and researchers in Numerical and statistical computing techniques. In the field of Data sciences, Data mining R has just created an evolutionary environment where researchers can work in parallel and bring out some useful work. As a part of our comparative study we will be using R because of its robustness, Versatility, ability to perform object oriented techniques etc. In this paper we will be studying about various approaches that are used in forecasting stock market while comparing their accuracy of those approaches with and without removing the seasonal trends present in the data.

II. EXISTING WORK

Right now for many researchers and Institutions stock market is one of the important area to do their research owing to the fact because of its adverse effect on various fields of Economics. Many of them used various techniques in order to predict the stock market like Neural Networks, Artificial Intelligence, Genetic Algorithms and Time series analysis. Still the zeal to work in this domain and many of them are carrying their work to develop new predicting models and algorithms that works efficiently and having better accuracy than the previous ones. Hence there is a great dire need to do a comparative study in this field.

Many techniques has been introduced in the recent times and one of the popular technique that has been used for many prediction models is the ARIMA (Autoregressive integrated moving average) model. ARIMA model is being castoff as a point of reference model to juxtapose the prediction precision. [11] They have developed a model to predict the future value of the stock using the Artificial Neural Networks that are edified using the characteristics that are drawn out by utilizing the perusal of the ARIMA model. After this characteristics extraction then the unprocessed data is scrutinized for the autocorrelation and limited autocorrelation are used for deriving the functional graphs which are then used for the nonlinear analysis of the data.

In the recent times the prominence for the Artificial Neural Networks has got a huge prominence among the researchers because of its ability to be used in different domains. [12] Has devised a model using the Artificial Neural Networks for the companies that are listed in the Shanghai Stock Market. In this paper the capabilities of two weights initializing techniques and two training algorithms are juxtaposed with each other. In this comparison as per the results obtained it is clear that stock market prediction is a feasible with the both initializing technique as well as the algorithmic model. The performance and the accuracy of the predictions can be increased using the back propagation learning algorithm and with the beeline regression weights initializer.

Another method that is being used by the researchers and academicians is the Holt winters method. In order to do a prediction using the Holt winters model we need to estimate the present underlined movement of the stock prices as well as the trend present that we anticipate to occur in the present month and the oncoming month. After this estimation Holt winters model approximates the movement of the stock, trend and seasonal effects present in the data [12]. It is done by taking the weighted averages of the estimation done previously and the value that is suggested by the new estimation. An exponential smoothing is also done for the stock movement, Trend and Seasonal analysis that falls connecting zero and one.[12] Results seems to be promising stating that Holt winter is one of the best forecasting method for analyzing the stock data.

Stock market forecasting being an arduous task selecting a suitable model for prediction is never made an easy task. Of all the methods or techniques that are proposed till now Time series linear model is one of the promising model for estimating the future model of the stock prices. [14] Is generally built on assembling the occurrences as data to a function that minimizes the loss. The most possible associated area is Systematic aggregate learning [15], in this the empirical aim is to assimilate multitudinous tasks at the same time rather than one at a time.

From the preceding findings we have founded that different techniques have been introduced by the researchers and of all this each technique has their own advantage and a disadvantage. As a part of our comparative study we will be working on different techniques including the Time series linear model, Holt winters model, ARIMA Model and the neural nets. It is founded in many of the papers that researchers are trying to compare their models that they have designed with the statistical methods. This comparative study may provide an inducement to the researchers in finding out and designing new prediction methods for the analysis of the stock market which is even considered as an arduous task. We will be more focused on the Technical Analysis of the Stock market.

III. DATA SET

For any kind of research to carry on Data set plays a very crucial role. For our comparative study we will be utilizing the data set that is available through the Yahoo Finance. Yahoo finance through an open project source made the availability the historical dataset. The main intention of this open source project is to make the researchers can use those data sets and bring some use full work which could be useful for the Investors, Capitalists etc.

The stock market data of different companies that are registered in different Stock exchanges is available through the Yahoo finance in a six column format consisting of Open, High, Low, Close, Volume and Adjusted Volume. For our comparative study we will be working on only one column that is the Closing value. Because information or the trend present in the data can be collected using the Closing value. The data will be available in three periodic types including daily which is used for daily analysis, weekly for weekly analysis and monthly for monthly analysis.

As a part of our comparative study we will be using the monthly data set since we will be doing the monthly forecast of four prominent companies naming Bank of America, Phillips, Boeing and Siemens. Monthly data of these companies from January 2000 to April 2015 is taken for our comparative study.

IV. METHODOLOGY

Stock markets being dynamic and volatile in nature it's an arduous task to predict the future value of these stocks. Our comparative study is focused on evaluating the accuracy of the available prediction models that are used by the Researchers, Academicians and capitalists. We would be evaluating the accuracy of ARIMA model, Holt winters, artificial neural nets and Time series linear model. Hence for a proper evaluation a suitable programming language is also necessary. For this evaluation purpose we will be choosing the R Programming language because R is one of the most complete statistical analysis tool available for manipulating and managing the data. Many developers across the world are working and contributing towards introducing multiple repositories in various domains like Data mining, spatial analysis and econometrics.

The methodology is very simple that first we have to obtain our data set from the yahoo finance which is made available for free and after getting the data from the yahoo finance then we will be converting the data that we have obtained from the yahoo finance and later we will convert the data that we have obtained from yahoo finance into time series data. After this the seasonal trends that are present in the data are removed in order to remove any aggressive values that are present in the data and smoothen the curve. Then the prediction model is applied on the time series data which doesn't have any seasonal trends in the data.

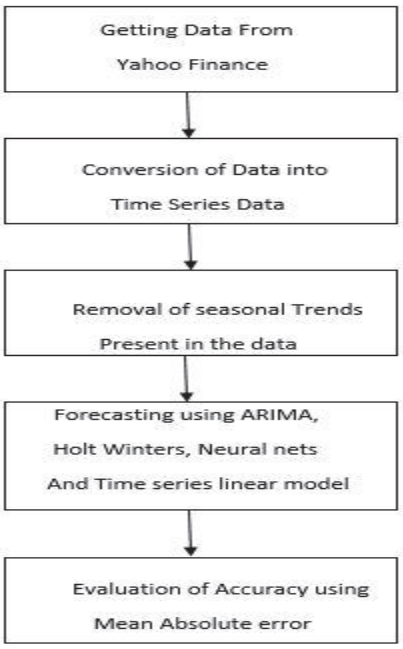

Fig. 1. Architecture of the Stock market Prediction Model

A. Conversion of Data into Time series Data

Conversion of data in to time series data involves mapping the average monthly closing value of a particular company with the month and year. So that a matrix will be generated where the rows represent the month and the columns represent the year. The R function $ts()$ helps in doing this conversion which is defined as follows.

$$timeseries = ts(rev(st\$Close), start = C(2000,1), frequency = 12)$$

Where st represents the data that we have obtained from the yahoo finance, *Close* represents closing values that are present in the st . As we are taking the data from January 2000 it is represented using *start* and the *frequency* is set as 12 since we have 12 months in a year. Let us consider a company AMG (Affiliated Managers Group Inc.) listed in the New York stock exchange and plotting the time series data we get the following result.

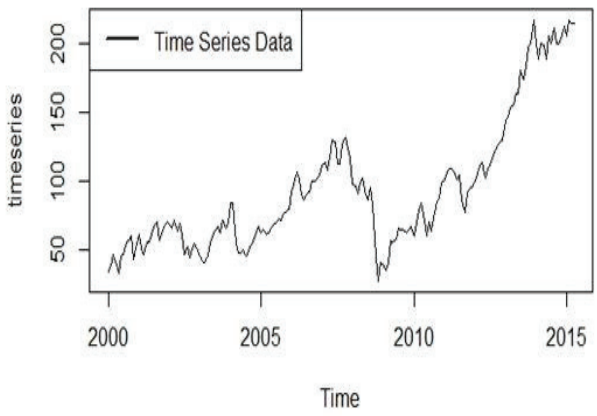

Fig. 2. Plot of Time Series data for the AMG Company

B. Removal of Seasonal Trends present in the Data

Removal of seasonal trends in the data helps to have a better forecast. There may be some unwanted trends present in the data which may have shown some adverse effects in the forecast. As a part of our comparative study we will analyze the accuracy of the prediction models by the removal of the seasonal trends. As we have seen in the plot in Fig. 1. of the time series data the graph is uneven with sudden peaks and down falls. To have a better forecast we have to estimate the overall trend present in the data to do this we have to remove the seasonal trends present in the data and smoothen the curve. In R we have an in built function named the $stl()$ which removes the seasonal trends present in the data.

$$StlStock = stl(timeseries, S.Window = "Periodic")$$

Where *timeseries* represents the time series data and the *"Periodic"* is used to represent in order to remove the periodic seasonal trends present in the data. Plotting the *StlStock* we have four components in Fig. 3. the first represents the actual time series data and the second represents the seasonal trends present in the data and the third component represents the data after the removal of seasonal trends finally the fourth component represents the reminder between the Actual data and the data after the removal of Seasonal trends.

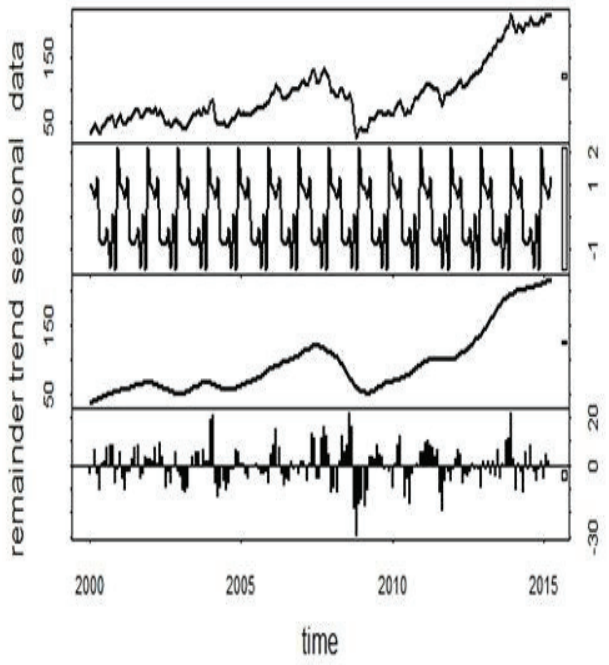

Fig. 3. Plot of Time Series data after removing the seasonal Trends for the AMG Company

C. Forecasting the Stock market

After the removal of the seasonal trends present in the data then the next step is forecasting. As per our comparative study we will forecast the stock market with the removal of the Seasonal trends present in the data. As a part of our comparative study we will be analyzing the ARIMA Model, Neural nets, Holt winters and the time series linear model. Implementing these models is made easy using R because of the packages and the prebuilt libraries that were made available by the researchers around the world. In order to implement these models we will be using the following packages they are forecast and fpp. First the data is given as input to the models and after making pre calculations these pre calculations is then given for the forecast function which predicts the future value of the stock.

1) ARIMA Model

In R generally ARIMA Model can be implemented in two ways one way is the autofit which generally initializes the parameters based up on the data and the next is the manual ARIMA in which we have to initialize all the parameters which is very difficult [16] since we have to identify the input for each and every parameter individually. For our comparative study we will be using the autofit ARIMA model and it can be implemented in R Using the following function.

$$autofit = auto.arima(StlStock)$$
$$Predictedvalues = forecast(autofit, h = 39)$$

Where $auto.arima()$ is the function used for pre calculations and forecast is the function used to forecast the future values using the pre calculations made by the $auto.arima()$ function. Plotting the forecasted values are depicted in the following graph Fig. 4.

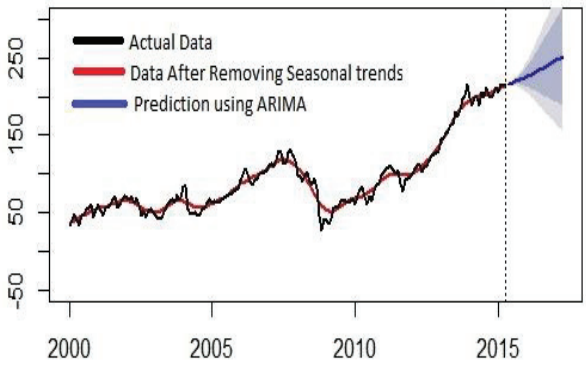

Fig. 4. Plot of Forecasting Using ARIMA

2) Holt Winters

Holt winters is known for its accuracy and the outperformed many other models in the domain of Prediction. In R generally Holt winters can be implemented directly using the prebuilt function $HoltWinters()$. The data after the removal of the seasonal trends is taken as input for this function and in return the Holt Winters makes the pre calculations that are necessary for the forecasting purpose. All the parameters that are necessary for the forecasting purpose will be initialized automatically based up on the data given to the function.

$$HWStockr = HoltWinters(StlStock)$$
$$Predictedvalues = (HWStockr, h = 39)$$

Where *StlStock* denotes that Stock data after the removal of seasonal trends present in the data and the *h* value indicates the number of values that is the number of months that we need the prediction is denoted by *h* value. The pre calculations that are done is given as input to the forecast function. Plotting the predicted values we get the following graph in Fig. 5.

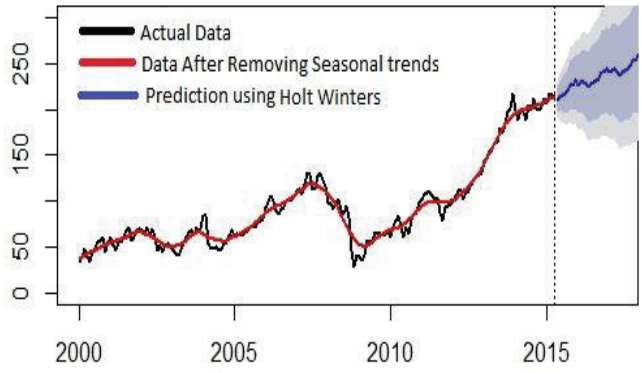

Fig. 5. Plot of Forecasting Using Holt Winters.

3) Artificial Neural Nets

In R implementation of this neural nets is made very easy. R programming supports a number of prebuilt models which trains the model automatically and even selects the number of layers that are necessary for a proper prediction based up on the data provided as input. The function $nnetar()$ which creates a model and trains by itself. Which is later used for prediction purposes.

$$NETfitr = nnetar(StlStock)$$
$$\text{Pr}edictedvalues = forecast(NETfitr, h = 39)$$

Where $nnetar()$ is the function that we use for creating the neural net model as well as training the model. *StlStock* represents the data after removal of the seasonal trends present in the data. The value of *h* indicates the number of predictions or the number of months for which we need the prediction is determined. The pre calculations that are done by the $nnetar()$ function will be used by the forecast function to make the predictions. Plotting the predicted values we got the following graph in Fig. 6.

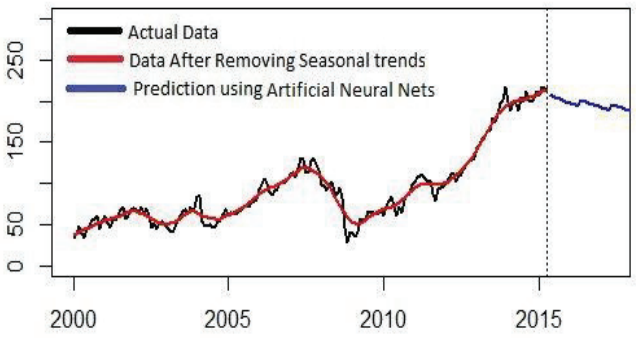

Fig. 6. Plot of Forecasting Using Artificial Neural Nets

4) Time Series Linear Model

Time series linear model is one of the stochastic way of implementing a predictive model. In a time series linear model first an ideal linear model is created and in that linear model the data is embedded into it so that the linear model reflects the properties of the actual data. Creation of this linear model is made simple using R programming using simple inbuilt function $tslm()$. The main advantage of this Time series linear model is that in embedding the actual data to the ideal linear model we can include both seasonal trends as well as the traditional trends present in the data.

$$Fit = tslm(StlStock \sim trend + season)$$
$$\text{Pr}edictedvalues = forecast(Fit, h = 39)$$

Where $tslm()$ is the function to create the ideal linear model and embeds the data *StlStock* which have seasonal trends removed. And the value *h* denotes the number of predicted or the number of months that we need to predict are determined. The $tslm()$ function does all the pre calculations that are necessary for the prediction which is used as input for the forecast function for the prediction. The forecasted values are plotted and the graph is determined in the Fig. 7.

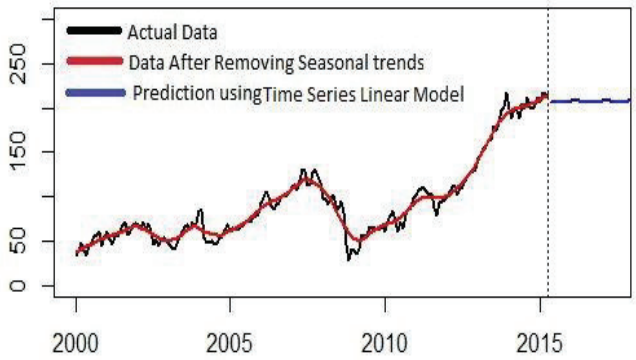

Fig. 7. Plot of Forecasting Using Time Series Linear Model

V. EVALUATION METRICS

Precision is one of the important measure that is used in order to evaluate the accuracy of the Predictions that are generated by this predictive models. Generally the Stock market data that for the particular set of companies that we have taken as a part of our comparative study is available in the Yahoo Finance that we are utilizing is taken and it is divided into two sets and one of the set is termed as R_{train} which is used to train the Algorithm and used to learn and the next set is termed as the test set R_{test} which is used to evaluate the accuracy of predictions generated. One of the important technique that is used to analyse and measure the accuracy and precision of the Predictions generated is the Mean absolute Method termed as MAE in Acronym. Mena absolute error which is termed as MAE is defined as the measure of deviation or divergence of the predicted Stock prices through the predictive models from the original Stock Prices. It is calculated as the mean or average of the absolute errors that

are calculated and it can be defined as in the following manner:

$$MAE(f) = \frac{1}{|R_{test}|} \sum_{p_s \in R_{test}} (f(s) - p_s)$$

Where R_{test} represents the training set and p_s represents the predicted stock value, and $f(s)$ represents the actual Stock value in the test set that we have taken. A lower Mean Absolute Error value indicates that the Predictions that are generated by the predictive model are accurate. So generally smaller mean absolute errors are generally recommended. In order to evaluate the best prediction model we will run the predictive models that we have designed on the five companies that are listed and calculate the Mean absolute error.

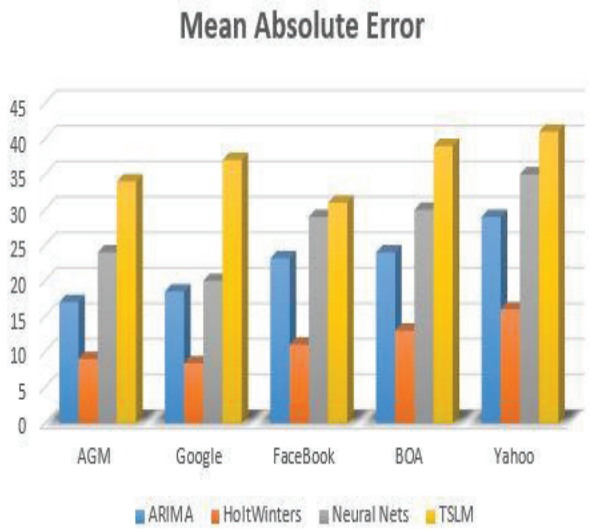

Fig. 8. Plot for the Mean Absolute Errors for the Five Companies

From the Fig. 8. It is cleared that the Holt winters model out performed with the least mean absolute error when compared to the rest of the models. When evaluating the Mean absolute error of all the five companies that we have taken Holt winters reported the least Mean absolute error than the rest.

VI. CONCLUSION

Predictions of the Stock market prices that are generated using these stochastic techniques including the ARIMA Model, Holt Winters, Artificial Neural Nets and TSLM Linear model are easy to implement, Reliable and Justifiable. As per our comparative study we have identified that the Holt winters has better accuracy when compared to the other predictive models in terms of the Mean absolute error. After this Holt winters ARIMA Model seems to be a promising model. The performance of this techniques may be effected by the data sparsity and hence there is a great chance of conducting research in this area. As the need for this predictive models especially in the domain of Stock market is increasing drastically new technologies and methods are needed to increase their performance.

In the present paper we have made a comparative study on different techniques that are used in the prediction of stock market and evaluated their accuracy of predictions using the Mean Absolute error. Our results hold the promise of using Holt winters which has a low mean absolute error when compared to the rest of the models.

References

[1] P. Pai and C. Lin, "A hybrid ARIMA and support vector machines model in stock price prediction", Omega vol.33 pp. 497-505, 2005

[2] J.J. Wang, J.Z. Wang, Z.G. Zhang and S.P. Guo, "Stock index forecasting based on a hybrid model", Omega vol.40 pp.758-766, 2012.

[3] L.Y. Wei, "A hybrid model based on ANFIS and adaptive expectation genetic algorithm to forecast TAIEX", Economic Modelling vol. 33 pp. 893-899, 2013.

[4] L.C. Kyungjoo, Y. Sehwan and J. John, "Neural Network Model vs. SARIMA Model In Forecasting Korean Stock Price Index (KOSPI), Issues in Information System, vol. 8 no. 2, pp. 372-378, 2007.

[5] N. Merh, V.P. Saxena, and K.R. Pardasani, "A Comparison Between Hybrid Approaches of ANN and ARIMA For Indian Stock Trend Forecasting", Journal of Business Intelligence, vol. 3, no.2, pp. 23-43, 2010.

[6] J. Sterba and Hilovska, "The Implementation of Hybrid ARIMA Neural Network Prediction Model for Aggregate Water Consumption Prediction", Aplimat- Journal of Applied Mathematics, vol.3, no.3, pp.123-131, 2010.

[7] P. S. Kalekar, "Time series Forecasting using Holt-Winters Exponential Smoothing," Kanwal Rekhi School of Information Technology, Tech. Rep., 2004.

[8] Abhishek Kar, "Stock Prediction using Artificial Neural Networks", Dept. of Computer Science and Engineering, IIT Kanpur.

[9] A. Victor Devadoss, T. Antony Alphonnse Ligori "Forecasting of Stock Prices Using Multi-Layer Perceptron " Volume: 02, December 2013, Pages: 440-449 International Journal of Computing Algorithm

[10] http://www.inside-r.org/packages/cran/forecast/docs/tslm

[11] Jung-Hua Wang, Jia-Yann Leu "Stock market trend prediction using ARIMA-based neural networks Neural Networks", 1996, IEEE International Conference on Volume 4, Issue, 3-6 Jun 1996 Page(s):2160 – 2165 vol.4.

[12] Abhyankar, A., Copeland, L. S., & Wong, "Uncovering nonlinear structure in real-time stock market indexes", 1997 The S&P 500, the DAX, the Nikkei 225, and the FTSE-100. Journal of Business & Economic Statistics, 15, 1–14.

[13] Kuo-Cheng, T., Ojoung, K., Tjung, "Time series and neural network forecast of daily stock prices", Investment Management and Financial Innovations 9, 32–54 (2012).

[14] G. Ristanoski, W. Liu, and J. Bailey, "Time series forecasting using distribution enhanced linear regression", In PAKDD (1), pages 484--495, 2013.

[15] Theodoros Evgeniou, Massimiliano Pontil, "Regularized multi–task learning", Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 109–117, 2004.

[16] http://www.inside-r.org/packages/cran/forecast/docs/auto.arima