International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 05 | May-2018 www.irjet.net p-ISSN: 2395-0072

# A Robust Predictive Model for Stock Market Index Prediction using Data Mining Technique

**A. K. Shrivas[1], Shishir Kumar Sharma[2]**

*[1,2]Dr. C. V. Raman University, Bilaspur, India.*

---

***Abstract***-*Prediction of stock market index is very challenging task for investors to invest the currency. The main objective of investors are invest the currency and make profit. This research work explores to develop the robust predictive model which predicts the correct foresting with minimum error. In this research work we have used machine learning techniques like Classification and Regression Technique (CART), CHAID, Artificial Neural Network (ANN) and Support Vector Machine (SVM) for analysis and prediction of BSE SENSEX data. The ANN gives the better prediction with very less Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). We have also extend the experimental work and analyzed the ANN predictive model with different learning rate and achieved less error measures like MAE =0.0044 and MAPE= 0.676 with 0.9 learning rate and 1 hidden layer.*

## 1. INTRODUCTION

Now days, accurate prediction of financial data for N-days ahead forecasting is very challenging task due to very noisy and nonlinear nature of data time to time. In this research work we have used data mining technique to predict the model correctly, because of the traditional model is not giving satisfactory result. In this research work we have used data mining techniques like CART, CHAID, ANN and SVM to analysis and prediction of BSE SENSEX data. There are various researchers have developed the robust predictive model and suggested to different predictive model for N-days ahead forecasting. D. K. Sharma et al. (2017) [6] have compared regression technique with ensemble regression techniques in the context of two ensemble learning: Bagging and Boosting (Least Square Boost: LSBoost) for two Foreign Exchange (FX) data namely INR/USD and INR/EUR. The comparative results show that regression ensemble with LSBoost is performing better than others with MAPE =0.6338. S. A. Hussein et al. (2015) [7] have suggested an efficient model for accurate stock market prediction, like Artificial Neural Networks (ANN). Multi-Layer Perceptron (MLP) is used and trained with Kullback Leibler Divergence (KLD) learning algorithm and Radial Basis Function Neural Network (RBFNN) trained with Localized Generalization Error (L-GEM) is used for candlesticks patterns which gives 0.3% of MAPE. B. Weng (2017) [8] have suggested disparate data sources to generate a prediction model along with a comparison of different machine learning methods. R. Handa et al. (2015)[9] have suggested ANN based module. They have used 10 years of historical daily Indian stock data were used for the experimental purpose where 7 features out of 16 are selected for stock market prediction. The MAPE found 5.48 with these seven features. H. S. Hota et al. (2016) [10] have presented hybridization of ANN and wavelet transform techniques for stock prediction model. Feature extracting and selection methods have used where seven features out of sixteen features have been selected for the prediction and achieved 2.614% and 2.627% of MAPE in case of error back propagation network (EBPN) and radial basis function network (RBFN) respectively. S. Shrivastava et al. (2017) [11] have suggested various data mining based predictive models to predict the stock in financial domain.

## 2. METHOD AND MATERIALS

Tools and techniques are very important role in every domain of research work. In this research work we have used for data mining based predictive models like CART,CHAID, ANN and SVM for stock market Index prediction.

### 2.1. Decision Tree

The basic idea of a decision tree [1] is to split our data recursively into subsets so that each subset contains more or less homogeneous states of our target variable (predictable attribute). At each split in the tree, all input attributes are evaluated for their impact on the predictable attribute. When this recursive process is completed, a decision tre is formed. In this research work we have used CART and CHAID are decision tree technique that is used for classification and prediction.

### 2.2 Artificial Neural Network (ANN)

Artificial Neural networks (Giudici, P., et al., 2009) [4] can be used for predictive data mining. They were originally developed in the field of machine learning to try to imitate the neurophysiology of the human brain through the combination of simple computational elements (neurons) in a highly interconnected system. In this research work we have focused on the ANN with different learning rate and 1 hidden layer with 20 neurons. Learning rate updates the weight at the time of learning and used to improve the performance of predictive model.

### 2.3 Support Vector Machine (SVM)

Support vector machines (SVMs) [3] are supervised learning methods that generate input-output mapping functions from a set of labelled training data. The mapping function can be either a classification function (used to categorize the input data) or a regression function (used to estimation of the desired output). For classification, nonlinear kernel functions are often used to transform the input data to a high dimensional feature space in which the input data becomes more separable (i.e., linearly separable) compared to the original input space. SVMs belong to a family of generalized linear models which achieves a classification or regression decision based on the value of the linear combination of features. They are also said to belong to "kernel methods".

### 2.4 Data Set

We have collected BSE SENSEX data set from http://www.bseindia.com/indices/IndexArchiveData.aspx) [5] to analysis of data and predict the stock. The dataset is from 2 Jan 2012 to Feb 2018 and contains1240 instances with 4 features namely open, high, low and close and 1 class level that is next-day-close with different continuous value.

## 3. RESULT AND DISCUSSION

In this research work we have used data mining based predictive models like CART, CHAID, ANN and SVM for predicting stock market Index price in window7 environment. We have used two error measures namely Mean absolute Error (MAE) and Mean absolute Percentage Error (MAPE) to check the performance of predictive model. The main motive of this research work is to minimize these two error measures. We have also compared the performance of various predictive models using MAE and MAPE error measures. Table 1 shows that training and testing error measures of these four predictive models where ANN gives better performance in terms of less MAE and MAPE error measures in without hidden layer neurons. The suggested ANN model gives 0.0053 and 0.808 MAE and MAPE at testing stage respectively. Figure 1(a) and Figure 1 (b) shows that MAE and MAPE of different predictive models respectively.

The ANN predictive model is given better performance without any hidden layer, so we have to check the performance of ANN model with different learning rate and 1 hidden layer. We have checked the performance of ANN model with learning rate from 0.1 to 0.9 and 1 hidden layer with 20 neurons. Table 2 shows that error measure of ANN with different learning rate and 1 hidden layer with 20 neurons where ANN gives better result at training and testing stages. The suggested ANN model gives error measures MAE = **0.0046 and MAPE= 0.0044.** Figure 2(a) and Figure 2 (b) shows that MAE and MAPE of ANN predictive models respectively with different learning rate.

Table 1: Performance Measure of Various Predictive Models

| Predictive Model | MAE | | MAPE | |
|---|---|---|---|---|
| | Training | Testing | Training | Testing |
| **CART** | 0.018 | 0.019 | 2.846 | 2.960 |
| **CHAID** | 0.014 | 0.013 | 2.115 | 2.004 |
| **ANN** | **0.0055** | **0.0053** | **0.807** | **0.808** |
| **SVM** | 0.032 | 0.038 | 5.266 | 6.257 |

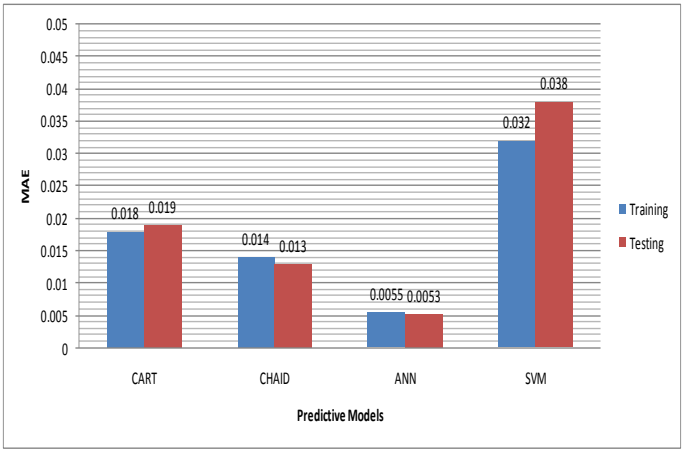


Figure 1(a) : MAE of different predictive models

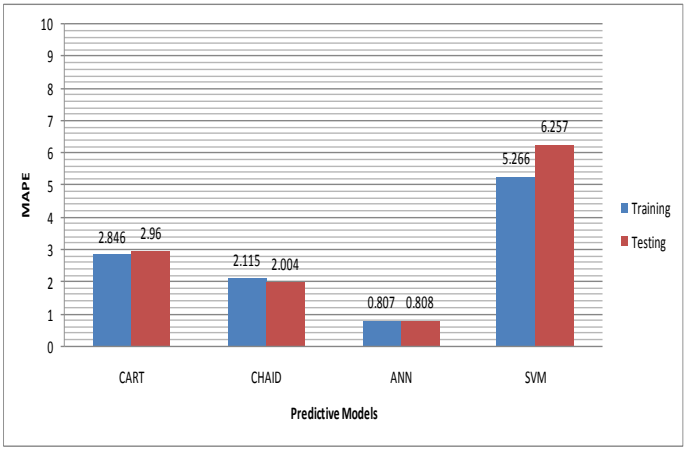


Figure 1(b) : MAPE of different predictive models

Table 2: Error measures of ANN with different learning rate

| Learning Rate | MAE | | MAPE | |
|---|---|---|---|---|
| | Training | Testing | Training | Testing |
| 0.9 | **0.0046** | **0.0044** | **0.682** | **0.676** |
| 0.8 | 0.0089 | 0.0087 | 1.363 | 1.362 |
| 0.7 | 0.0049 | 0.0048 | 0.717 | 0.742 |
| 0.6 | 0.0092 | 0.0088 | 1.405 | 1.384 |
| 0.5 | 0.0049 | 0.0049 | 0.718 | 0.742 |
| 0.4 | 0.0049 | 0.0049 | 0.721 | 0.752 |
| 0.3 | 0.0052 | 0.0052 | 0.759 | 0.787 |
| 0.2 | 0.0050 | 0.0050 | 0.736 | 0.757 |
| 0.1 | 0.0092 | 0.0088 | 1.387 | 1.363 |

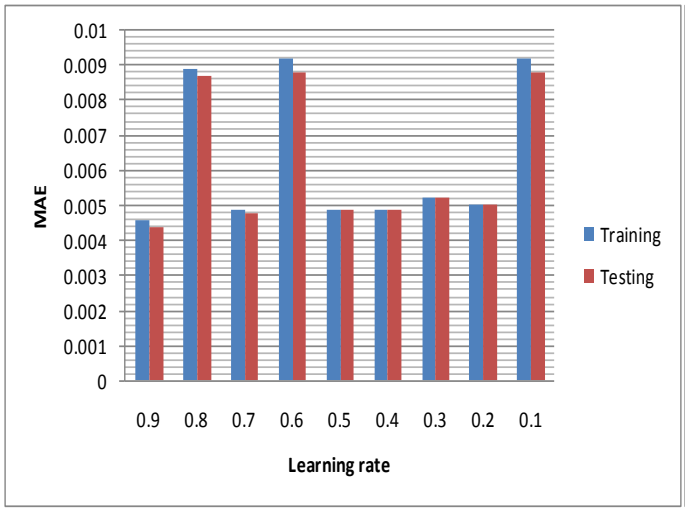


Figure 2 (a): MAE of ANN predictive model

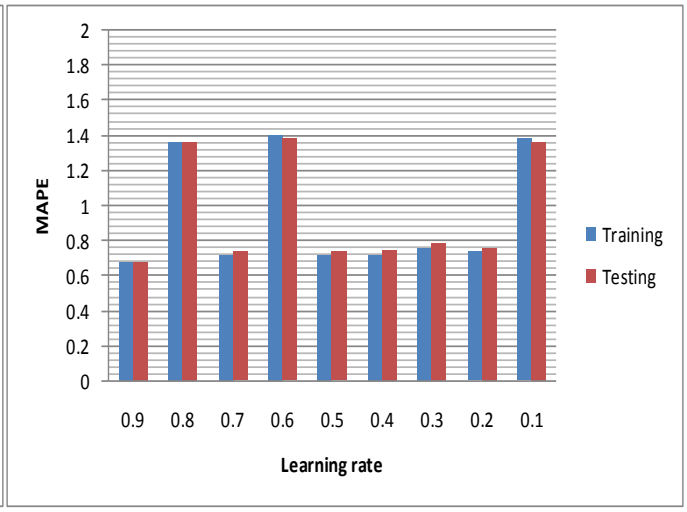


Figure 2 (b): MAPE of ANN predictive model

## 4. CONCLUSION AND FUTURE WORK

Stock market index prediction is very essential and challenging task for every investors to invest currency in market. The better prediction may be helpful for investors to invest currency and take profit of their invested currency. In this research work we have used data mining based predictive techniques and compared the performance in terms of MAE and MAPE error measures with BSE SENSEX data. We have suggested ANN gives better prediction for stock market index with less MAE and MAPE error measures. The ANN gives MAE =0.0044 and MAPE =0.676 at testing stages with learning rate 0.9.

In future we will proposed new integrated predictive model for better stock market index prediction and also analyzed and validated the our new integrated predictive model with new financial data set like YAHOO, BSE 100 and others.

## REFERENCES

[1]. A. K. Pujari, "Data Mining Techniques", Universities Press (India) Private Limited, 4th ed., 2001.

[2]. UCI Repository of Machine Learning Databases, University of California at Irvine, Department of Computer Science. Available:http://www.ics.uci.edu/~mlearn/databases/ (Browsing date: Jan. 2018).

[3]. D. L. Olson and D. Delen, "Advanced Data Mining Techniques", USA, Springer Publishing, 2008

[4]. P. Giudici, Applied Data Mining for Business and Industry, 2nd edition, John Wiley & Sons, 2009.

[5]. http://www.bseindia.com/indices/IndexArchiveData.aspx) (Browsing date : Jan 2018)

[6]. D.K. Sharma H.S. Hota and R. Handa, "Prediction Of Foreign Exchange Rate Using Regression Techniques" , Review of Business and Technology Research, Vol. 14(1), 2017,pp. 29-33.

[7]. A.S. Hussein, I.M. Hamed, and M.F. Tolba, "An Efficient System for Stock Market Prediction", Springer International Publishing Switzerland, DOI: 10.1007/978-3-319-11310-4_76, 2015.

[8]. B. Weng, "Application of Machine Learning Techniques for Stock Market Prediction" , A dissertation Submitted to the Graduate Faculty of Auburn University, 2017.

[9]. R. Handa H.S. Hota and S.R. Tandan, "Stock Market Prediction with Various Technical Indicators Using Neural Network Techniques", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol. 3(6), 2015,pp. 604-608.

[10]. H.S. Hota and R. Handa, "Hybridization of Wavelet Transform and ANN Techniques as Stock Prediction Model", International Journal of Innovations in Engineering and Technology (IJIET) Vol. 7(2), 2016, pp.227-232.

[11]. S. Shrivastava and A. K.Shrivas, "Performance Analysis of Data Mining Techniques for Stock Price Forecasting", International Journal of Engineering Research and Technology (IJERT) Vol. 6(5), 2017, pp. 861-864.