

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

Shishir Kumar Sharma

Research Scholar, Kalinga University Nawa Raipur(C.G.) India

Abstract

Stock exchange prediction is that the method of determining further values of a company's stocks and other financial values. Market lacks a efficient software where the correct suggestions of obtainable stocks and therefore the proper investment analysis are presented in an efficient way. The investors should be guided and encouraged to take a position within the stocks in an exceedingly coherent way. the event of a vibrant application for analyzing and predicting securities market prices is also a basic tool aimed toward increasing the speed of investor's interest in available markets[14]. stock exchange price data is generated in huge volume and it changes every second. securities market could be a challenging and complicated system where people can either gain money or lost their saving. during this work, I made an effort for Prediction of stock exchange index and it's very challenging task. the most objective of investors is to take a position the money and make profit. My task is to predicts the right foresting with minimum error. during this research work i've got 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 S&P BSE AUTO data. The ANN gives the higher prediction with very less Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). we've 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.

KEYWORDS : Stock Market, Support Vector Machine(SVM), Artificial Neural Network(ANN), Mean Absolute Error(MAE) and Mean Absolute Percentage Error(MAPE)

1. INTRODUCTION

Trading in shares is business in many economies. supported the knowledge on their websites, Stockbrokers don't seem to possess any intelligent tool which is able to help them advise clients on which stocks are suitable for any buy or sale trade. Stock price prediction is extremely important because it's employed by most of the business people also as folk. People will either gain money or lost their saving available market activity. an honest prediction of future behaviour of market would be extremely valuable in various areas. Now days, accurate prediction of economic data for N-days ahead forecasting is extremely challenging task thanks to very noisy and nonlinear nature of information time to time. during this research work i've got used data processing technique to predict the model correctly, thanks to the normal model isn't giving satisfactory result.

1.1 EXISTING SYSTEM

Earlier the the stock market prediction was done by some complex mathematics by manual methods which were not much accurate and it consumes a lot of time as well.

1.2 PROPOSED SYSTEM

In the proposed system we are using data processing techniques to predict the sock prices. These techniques are fast and far accurate as compared to previous ones.

In this research work i've got used data processing techniques like CART, CHAID, ANN and SVM to analysis and prediction of S&P BSE AUTO data. There are various researchers who are developed the 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 within the context of two ensemble learning: Bagging and Boosting (Least Square Boost: LSBoost) for two 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 exchange 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 employed for candlesticks patterns which provides 0.3% of MAPE. B. Weng (2017) [8] have suggested disparate data sources to come up with a prediction model together with a comparison of various machine learning methods. R. Handa et al. (2015)[9] have suggested ANN based module. they need used 10 years of historical daily Indian stock data were used for the experimental purpose where 7 features out of 16 are selected for exchange 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 are selected for the prediction and achieved 2.614% and 2.627% of MAPE just in case of error back propagation network (EBPN) and radial basis function network (RBFN) respectively. S. Shrivastava et al. (2017) [11] have suggested various processing based predictive models to predict the stock in financial domain.

The remainder of this paper is organized as follows : In Section 2, we discuss the methods used for prediction and materials. Then in section 3 results and discussion and in section 4 conclusion and future scope of this research work.

2. METHOD AND MATERIALS

2.1 Data Mining

There is an infinite amount of information available within the data Industry. This data is of no use until it's converted into useful information. It's necessary to research this huge amount of data and extract useful information from it. Extraction of knowledge isn't the only real process we wish to perform; processing also involves other processes like Data Cleaning, Data Integration, Data Transformation, processing, Pattern

Evaluation and Data Presentation. Once these processes are over, we would be able to use this information in many applications like Fraud Detection, research, Production Control, Science Exploration, etc.

Listed below are the various fields of market where processing is used:

Customer Profiling - processing helps determine what quite of individuals buy what quite products.

Identifying Customer Requirements - processing helps in identifying the best products for various customers. It uses prediction to look out the factors which can attract new customers.

Cross research - processing performs Association/correlations between product sales.

Target Marketing - processing helps to go looking out clusters of model customers who share equivalent characteristics like interests, spending habits, income, etc.

Determining Customer purchasing pattern - processing helps in determining customer purchasing pattern.

Providing Summary Information - processing provides us various multidimensional summary reports. Corporate Analysis and Risk Management processing is used within the subsequent fields of the company Sector: Finance

Planning and Asset Evaluation - It involves income analysis and prediction, contingent claim analysis to judge assets.

Resource Planning - It involves summarizing and comparing the resources and spending.

Competition - It involves monitoring competitors and market directions.

Fraud Detection - processing is additionally utilized within the fields of MasterCard services and telecommunication to detect frauds. In fraud telephone calls, it helps to go looking out the destination of the choice, duration of the choice, time of the day or week, etc. It also analyzes the patterns that deviate from expected norms.

Tools and techniques play a very important role in every domain of research work. during this research work I even have used for processing based predictive models like CART,CHAID, ANN and SVM for exchange Index prediction.

2.2 Decision Tree

The basic idea of a decision tree [1] is to separate our data recursively into subsets so as that every subset contains more or less homogeneous states of our target variable (predictable attribute). At each split within the tree, all input attributes are evaluated for his or her impact on the predictable attribute. When this recursive process is completed, a

choice tree is formed. during this research work we've used CART and CHAID are decision tree technique that's used for classification and prediction.

2.3 Artificial Neural Network (ANN)

Artificial Neural networks (Giudici, P., et al., 2009) [4] are often used for predictive processing. They were originally developed within the sector of machine learning to undertake to imitate the neurophysiology of the human brain through the combination of straightforward computational elements (neurons) in a very highly interconnected system. during this research work we've focused on the ANN with different learning rate and 1 hidden layer with 20 neurons. Learning rate updates the burden at the time of learning and wont to improve the performance of predictive model.

2.4 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 are often either a classification function (used to categorize the input data) or a regression function (used to estimation of the specified output). For classification, nonlinear kernel functions are often wont to transform the input file to a high dimensional feature space during which the input file 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 supported the worth of the linear combination of features. They are also said to belong to "kernel methods".

2.5 Data Set

I have collected S&P BSE AUTO data set from <https://www.bseindia.com/indices/IndexArchiveData.html> [5] to analysis of data and predict the stock. The dataset is from Jan 2015 to Feb 2020 and contains 1241 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, i've got used data processing based predictive models like CART, CHAID, SVM and ANN for predicting stock exchange Index price with window7 environment. i've got used two error measures namely Mean absolute Error (MAE) and Mean absolute Percentage Error (MAPE) to live the performance of this predictive model. the most objective of this research work is to attenuate these two error measures. i've got also compared the performance of assorted predictive models using MAE and MAPE error measures. Table 1 shows that training and testing error measures of those 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 assorted predictive models respectively.

The ANN predictive model is given better performance with none hidden layer, so we've got to test the performance of ANN model with different learning rate and 1 hidden layer. i've got 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

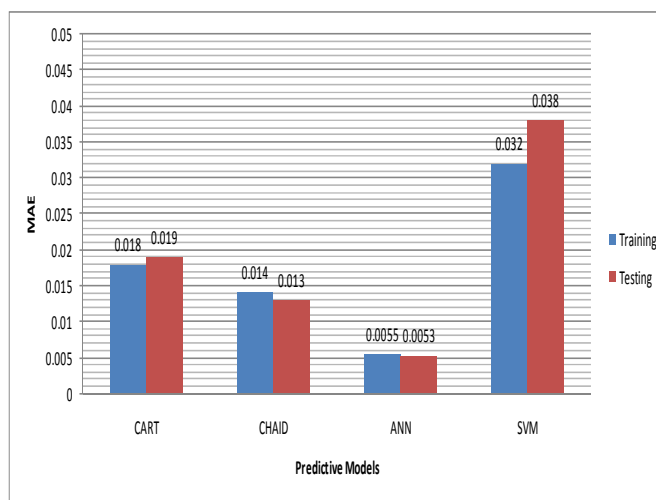


Figure 1(a) : MAE of various predictive models

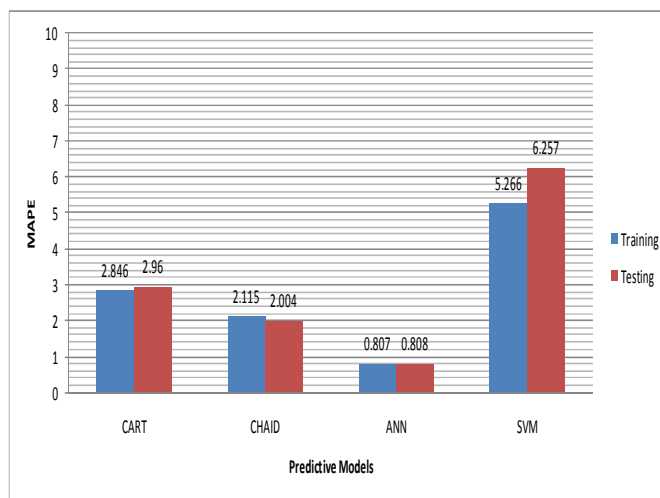


Figure 1(b) : MAPE of various predictive models

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

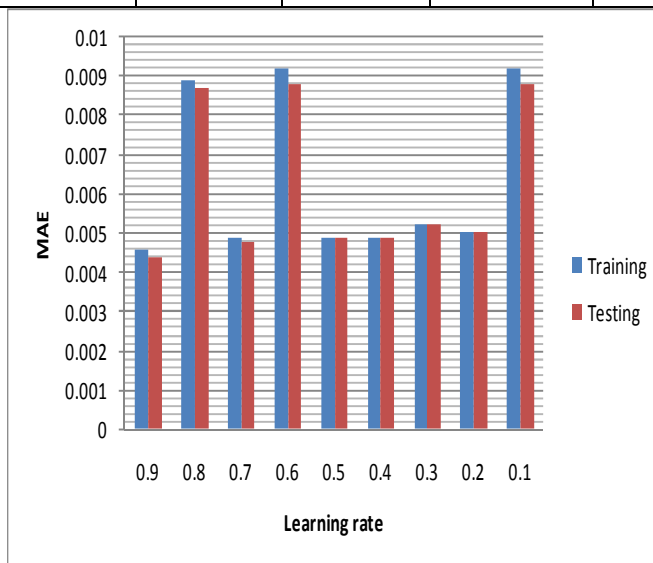


Figure 2 (a): MAE of ANN predictive model

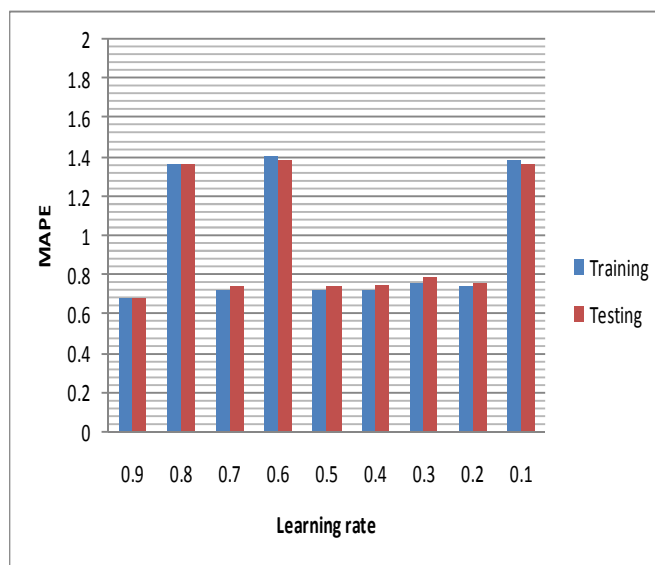


Figure 2 (b): MAPE of ANN predictive model

4. CONCLUSION AND FUTURE WORK

The noisy and nonlinear behaviour of costs in financial markets make the prediction in those market a difficult task. An improved prediction is often gained by having better variables. The higher prediction is also helpful for investors to speculate currency and take profit of their invested currency. During this research work we've got used data processing based predictive techniques and compared the performance in terms of MAE and MAPE error measures with S&P BSE AUTO data. We've suggested ANN gives better prediction for securities 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 I will try to propose new integrated predictive model for better stock market index prediction and also analyzed and validate our new integrated predictive model with new financial data set like YAHOO FINANCE, BSE 500 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/~mllearn/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]. <https://www.bseindia.com/indices/IndexArchiveData> (Browsing date : Jan 2019)

- [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. Shrivastava, "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.
- [12]. A. Nayak, M.M. Monahari Pai and R.M. Pai "Prediction Models for Indian Stock Market", ELSEVIER, 2016 pp. 441-449.
- [13]. M. Vijh, D. Chandola, V.A. Tikkiwal and A. Kumar "Stock Closing Price Prediction Using Machine Techniques", *International Conference on Computational Intelligence and Data Science (ICCIDS)*, 2019 pp. 599-606.
- [14]. M.A. Siddiqui, M.A. Afroz, S.N. Khan and M. Irshad "Stock Price Prediction using AI", *International Journal of Merging Technology and Advance Research in Computing(IJMTARC)*, Vol. VIII, 2020.