RESEARCH REVIEW JOURNALS
Volume-04
Issue-05
May-2019

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com[UGC Listed Journal]

An Analytical Study on Model of Stock Index prediction Based on Data Mining

[1]Y Rohita and [2]Dr.V.S.S.S.Balaram

[1]*Research Scholar, Sri SatyaSai University of Technology*
[2]*Research Guide, Sri SatyaSai University of Technology*

ARTICLE DETAILS

Article History
Published Online: 25 May 2019

Keywords
Data Mining, Stock Index

ABSTRACT

Numerous individuals consider stock market prediction as gambling. Anyway it is conceivable to create helpful patterns by the investigation of stock prices. Data mining systems can be connected on over a wide span of time budgetary data to create examples and decision making algorithms. Prediction of stock market list is testing errand investors to contribute the money. The primary targets of investors are contributing the money and make profit. The point of this paper is to build up the robust predictive model which predicts the right foresting with minimum error. The instruments utilized in this investigation are machine learning techniques like Classification and Regression Technique (CART), CHAID, Artificial Neural Network (ANN) and Support Vector Machine (SVM) for examination and prediction of BSE SENSEX data.. Anticipating stock return is a vital financial subject that has pulled in analysts' consideration for a long time The ANN gives the better prediction with extremely less Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). We have additionally expand the experimental work and broke down the ANN predictive model with various learning rate and accomplished less error estimates like MAE =0.0044 and MAPE= 0.676 with 0.9 learning rate and 1 hidden layer This think about attempts to help the speculators in the stock market to choose the better planning for purchasing or selling stocks dependent on the information separated from the prices of such stocks.

1. Introduction

These days, with the advancement in innovation we have help in each field. We Humans are constantly keen on recognizing what will occur straightaway. Same runs with the field of stock market, People for the most part need to put their cash in stock market and anticipate exceptional yields in limited capacity to focus time. Be that as it may, there is no settled example with which the market goes here and there. Different parameters are to be considered, through which the market carries on nowadays stock prices are influenced because of numerous reasons like organization related news, political occasions, natural disasters and so forth.

1.1 Data mining

Data mining as a procedure of finding helpful patterns, relationships has its own specialty in financial modeling. So also to other computational strategies pretty much every data mining strategy and system has been utilized in financial modeling. An inadequate rundown incorporates an assortment of direct and non-straight models, multi-layer neural systems, k-implies and hierarchical clustering; k-nearest neighbors, decision tree investigation, regression (calculated regression; general multiple regression), ARIMA, essential part examination, and Bayesian learning

1.2 Stock market

Stock market is where purchasing and selling of stocks/shares happens. This trading should be possible physically on trading floor or on the web. At the point when an investor purchases stocks of a specific organization he turns into a section proprietor of that organization as per the quantity of offers held by him. To make most extreme profit, right speculation ought to be set aside a few minutes. As indicated by the ideal opportunity for which stocks are held, stock trading can be named position trading, swing trading and day trading. A stock exchange is a place or association by which stock traders (individuals and organizations) can trade stocks. Organizations might need to get their stock recorded on a stock exchange. Different stocks might be traded "over the counter", that is, through a merchant. A substantial organization will more often than not have its stock recorded on numerous exchanges over the world. Exchanges may likewise cover different kinds of security, for example, settled interest securities or interest derivatives.

1.3 Specifics of Data Mining In Finance

Specifics of data mining in finance are originating from the need to: forecast multidimensional time arrangement with abnormal state of commotion; oblige explicit proficiency criteria (e.g., the limit of trading profit) notwithstanding prediction exactness, for example, R2; make composed multi-resolution forecast (minutes, days, weeks, months, and years); join a flood of content flags as information data for forecasting models (e.g., Enron case, September 11 and others); have the capacity to clarify the forecast and the forecasting model ("discovery" models have constrained interest and future for huge speculation decisions); have the capacity to profit by unobtrusive patterns with a short life time; and consolidate the effect of market players on market regularities.

The current effective market hypothesis/theory debilitates endeavor to discover long haul stable trading rules/regularities with critical profit. This hypothesis depends on the possibility that if such regularities exist they would be discovered and utilized by most of the market players. This would make governs less profitable and eventfully pointless or even damaging

1.4 Models of Data Mining and Practice in Finance

Prediction errands in finance normally are presented in one of two structures: (1) straight prediction of the market numeric characteristic, e.g., stock return or exchange rate, and (2) the prediction whether the market characteristic will increment or decline. Having as a main priority that we have to consider the trading cost and hugeness of the trading return in the second case we have to forecast whether the market characteristic will increment or decline no not exactly some limit. Along these lines, the contrast between data mining techniques for (1) or (2) can be more subtle, in light of the fact that (2) may require some sort of numeric forecast.

- **Portfolio management and neural networks**- The neural system most normally utilized by financial foundations is a multilayer perceptron (MLP) with a single hidden layer of nodes for time arrangement prediction. The pinnacle of research exercises in finance dependent on neural systems was in mid 1990s that covered MLP and intermittent NN. Other neural systems utilized in prediction are time defer systems, Elman systems, Jordan systems, GMDH, milti-intermittent systems
- **Interpretable trading rules and relational data mining -**The rationale of portfolio management dependent on discovering interpretable trading rules is equivalent to for neural systems with the substitution of NN for rule discovering procedures. Contingent upon the rule discovering methods delivered rules can be very unique. Underneath we present classifications of rules that can be discovered.

2. Significance of Data mining for Stock price predictions

A few works are done to foresee stock price developments dependent on outlining authentic qualities. A significant number of them neglected to deliver promising outcomes since it can't suit the genuine development. The data mining strategies can address such issues the shrouded patterns can be discovered by applying data mining pattern coordinating procedures. It fills in as the primer stage for social affair business data to appraise future needs. A considerable lot of the regular techniques neglected to convey in this issue. Data mining systems viably addresses the difficulties in which the customary techniques neglected to convey promising outcomes. Some of them are

- Stock price prediction
- Pattern generation with historical data
- Effective usage of business capital

3. Review of literature

Anthony J. T. Lee, Ming-Chih Lin, Rung-Tai Kao (2010) -proposed HRK (Hierarchical agglomerative and recursive KMeans clustering) which predicts transient stock developments with reference to financial reports. This technique contains three stages. First the financial reports are changed over to highlight vector and HRK strategy is connected to isolate them into clusters. In second step K-Means clustering technique parcels each cluster into sub clusters with the goal that each sub cluster has a place with same class. Third for each sub cluster the centroid is picked as agent include vector. These vectors are utilized for stock price development prediction.

Kranti M. Jaybhay (2012) -This paper proposes a novel technique for the prediction of stock market shutting price. Numerous analysts have contributed here of disorderly forecast in their ways. Data mining systems can be utilized more in financial markets to settle on subjective decisions for investors. Central and specialized examinations are the conventional methodologies up until this point. ANN is a prominent method to recognize obscure and concealed patterns in data is utilized for offer market prediction. A multilayered feed-forward neural system is worked by utilizing blend of data and printed mining. The Neural Network is prepared on the stock statements and extricated key expressions utilizing the Back engendering Algorithm which is utilized to foresee share market shutting price. This paper is an endeavor to decide if the BSE market news in mix with the recorded statements can productively help in the estimation of the BSE closing index for a given trading day.

Prasanna (2012) -proposed a technique to appraise genuine estimation of stock price utilizing hybrid and half Mcniven approach. The predictions are produced for three classes. Reasonable esteemed, Overvalued and underestimated stocks. This encourages the investors to choose great stocks which are reasonable esteemed hence expanding the profit for investors. This strategy yields preferred outcomes over 3D subspace clustering technique in chose cases.

Li, S., Wang et al. (2013) -connected assessment investigation by utilizing twitter data for predicting achievement rate of motion picture. For this reason films were named Flop, Hit, and Average. The tweets from 2009 to 2013 length were separated, and each tweet was delegated positive, negative, unbiased and superfluous. Lingpipe sentiment analyzer was utilized to test the sentiments, and result demonstrated that movie prediction exactness of the created framework was 64.4 % superior to conventional framework

S. Prasanna (2016) -The utilization of AI procedures for stock price prediction prompts voluminous development of wealth of investors with the approach of innovation. A few prediction and estimations are coming up for practically all parts of the market. Especially any sort of stock price prediction isn't at all conceivable without over the top data control which should be possible viably just through data mining. The deliberate factual control of data should be possible viably just by applying reasonable business knowledge and AI procedures. Especially Indian stock market is running in progressively complex situation and requirements over the top data mining. A few works in regards to stock price prediction is improved the situation global just as Indian market with the appearance of data mining methods. In this survey we endeavor to portray some imperative works improved the situation stock price prediction utilizing data mining procedures.

VivekKanade (2017) –The Stock market process is loaded with vulnerability and it's influenced by numerous variables. Henceforth the Stock market prediction is one of the essential efforts in business and finance. There are two kinds of investigation feasible for prediction, principal and specialized. In this paper both key and specialized investigation are considered. Crucial examination is finished utilizing online life data by applying sentiment investigation process. Online networking data has high effect today than any time in recent memory, it can supportive in predicting the pattern of the stock market and Technical examination is finished utilizing recorded data of stock prices by applying machine learning algorithms. The strategy includes gathering news and furthermore gathers web-based social networking data and extricating sentiments communicated by person. At that point the connection between's the sentiments and stock qualities are broke down. The learned model would then be able to be utilized to make future predictions about stock esteem. It very well may be demonstrated that this strategy can foresee the stock execution and sentiment and social data are likewise firmly corresponded with ongoing news.

4. Objectives of the study

1. To conceptualize the term Data mining and stock market prediction.
2. To investigate the data mining models and practices in finance, and also investigate the specifics of data mining in finance.
3. To determine the significance of data mining in stock price prediction in market.
4. To analyze the predictive models like CART, CHAID, ANN and SVM for stock market Index prediction.
5. To analyze the error measure of ANN with different learning rate.

5. Research methodology

Tools and techniques are imperative job in each space of research work. In this examination work we have utilized for data mining based predictive models like CART, CHAID, ANN and SVM for stock market Index prediction.

• Decision Tree

The fundamental thought of a decision tree is to split our data recursively into subsets with the goal that every subset contains pretty much homogeneous conditions of our objective variable (unsurprising trait). At each split in the tree, all info properties are assessed for their effect on the anticipated trait. At the point when this recursive procedure is finished, a decision tree is shaped. In this exploration work we have utilized CART and CHAID are decision tree technique that is utilized for arrangement and prediction.

• Artificial Neural Network (ANN)

Counterfeit Neural systems can be utilized for predictive data mining. They were initially created in the field of machine learning to endeavor to emulate the neurophysiology of the human brain through the blend of basic computational components (neurons) in a profoundly interconnected framework. In this examination work we have concentrated on the ANN with various learning rate and 1 concealed layer with 20 neurons. Learning rate refreshes the weight at the season of learning and used to enhance the execution of predictive model.

• Support Vector Machine (SVM)

Support vector machines (SVMs) are directed learning strategies that create input-output mapping capacities from a lot of labeled training data. The mapping capacity can be either an arrangement work (used to sort the input data) or a regression work (used to estimation of the ideal output). For arrangement, nonlinear bit capacities are frequently used to change the input data to a high dimensional element space in which the input data turns out to be increasingly divisible (i.e., linearly detachable) contrasted with the first input space. SVMs have a place with a group of summed up linear models which accomplishes an order or regression decision dependent on the estimation of the linear combination of highlights. They are likewise said to have a place with "kernel strategies".

• Data Set

We have gathered BSE SENSEX data set to investigation of data and predict the stock. The dataset is from 2 Jan 2012 to Feb 2018 and contains1240 cases with 4 includes in particular open, high, low and close and 1 class dimension that is following day-close with various continuous value.

6. Result and discussion

In this examination work we have utilized data mining based predictive models like CART, CHAID, ANN and SVM for Table 1: Performance Measure of Various Predictive Models predicting stock market Index price in window7 condition.

Table 1Performance Measure of Various Predictive Models predicting stock market Index price in window7 environment

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

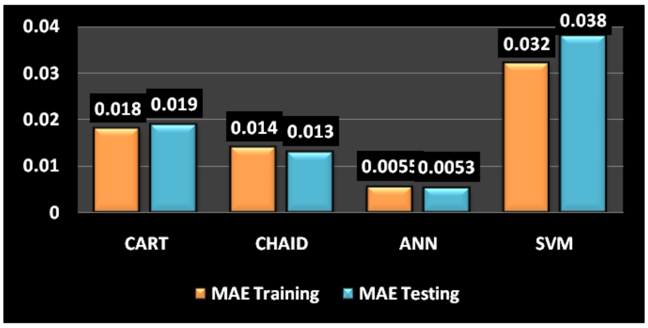

Figure 1 MAE of different predictive models

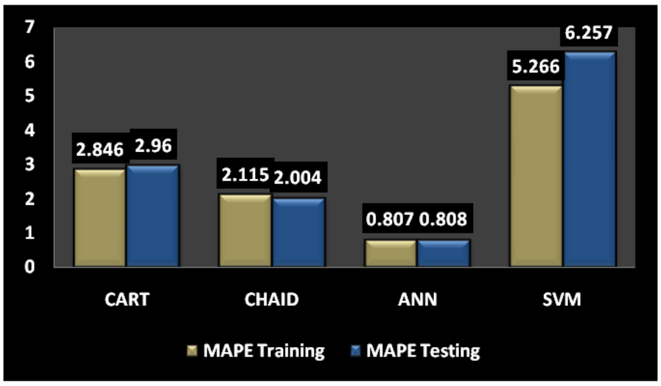

Figure 2 MAPE of different predictive models

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

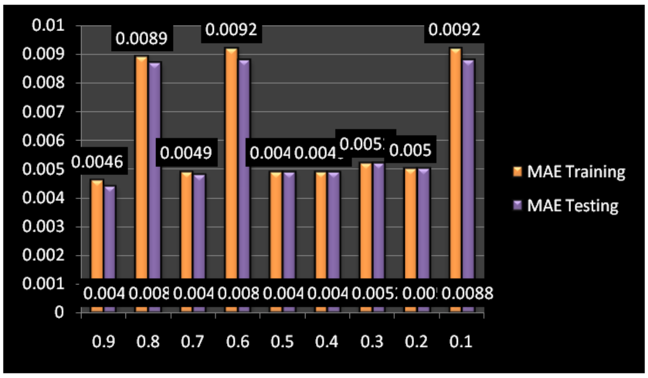

Figure 3 MAE of ANN predictive model

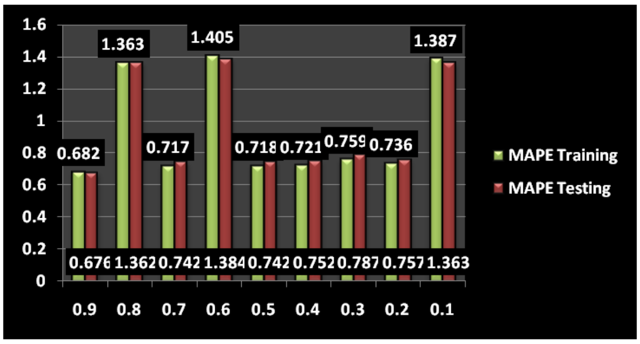

Figure 4 MAPE of ANN predictive models

We have utilized two error estimates in particular Mean total Error (MAE) and Mean total Percentage Error (MAPE) to check the execution of predictive model. The principle intention of this exploration work is to limit these two error measures. We have likewise looked at the execution of different predictive models utilizing MAE and MAPE error measures. Table 1 demonstrates that training and testing error proportions of these four predictive models where ANN gives better execution as far as less MAE and MAPE error measures in without hidden layer neurons. The recommended ANN model gives 0.0053 and 0.808 MAE and MAPE at testing stage separately.

Figure 1 and Figure 2 demonstrates that MAE and MAPE of various predictive models individually.The ANN predictive model is given better performance with no hidden layer, so we need to check the performance of ANN model with various 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 demonstrates that error proportion of ANN with various learning rate and 1 hidden layer with 20 neurons where ANN gives better outcome at training and testing stages. The proposed ANN model gives error measures MAE = 0.0046 and MAPE= 0.0044. Figure 3 and Figure 4 demonstrates that MAE and MAPE of ANN predictive models individually with various learning rate.

7. Conclusion

With respect to the future work, there is still huge space for testing and enhancing the proposed model by assessing the model over the entire organizations recorded in the stock market. Likewise, the assessment of a bigger accumulation of learning techniques, for example, neural systems, hereditary algorithms, and affiliation rules can speak to a rich territory for future examination. At long last, rethinking the components influencing the conduct of the stock markets, for example, trading volume, news and financial reports which may affect stock price can be another rich field for future studying. Stock market index prediction is basic and testing errand for each investor to put money in market. The better prediction may be helpful for investors to invest currency and take profit of their invested cash. In this exploration work we have utilized data mining based predictive techniques and analyzed the performance as far as MAE and MAPE error measures with BSE SENSEX data. We have recommended 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 coordinated predictive model for better stock market index prediction and furthermore broke down and approved the our new incorporated predictive model with new financial data set like YAHOO, BSE 100 and others.

References

[1] Anthony J. T. Lee, Ming-Chih Lin, Rung-Tai Kao (2010) - " An Effective Clustering Approach To Stock Market Prediction" Department of Information Management, National Taiwan University, Taipei, Taiwan

[2] Kranti M. Jaybhay (2012) –"Stock Market Prediction Model by Combining Numeric and News Textual Mining", *International Journal of Computer Applications (0975 – 8887) Volume 57– No.19, November 2012*

[3] S. Prasanna, D. Ezhilmaran,(2012) -"Estimation of True Stock Value: A Hybrid McNiven Approach with Predictive Data mining Concepts", EJSR Vol110_Issue1_No. 7 2012.

[4] Wang, X., &Luo, X (2013) -, Sentimental Space Based Analysis of User Personalized Sentiments, IEEE 9th International Conference on Semantics, Knowledge and Grids, October 2013, pp. 151-156.

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

[6] S. Prasanna (2016) –" A Survey of Stock Price Prediction & Estimation Using Data Mining Techniques", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 6 (2016) pp 4097-4099

[7] 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.

[8] .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.

[9] VivekKanade (2017) –" Stock Market Prediction: Using Historical Data Analysis", IJARCSSE, Volume 7, Issue 1, January 2017 ISSN: 2277 128X
[10] 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