# Forcasting Stock Market Index: Comparison of Neural Network and Random Forest

V. Munaiah[1], J. Prabhakara Naik[2], R.V.S.S. Nagabhushana Rao[3], K.Vasu[4], G. Mokesh Rayalu[5]

[1]*Lecturer, Department of Statistics, PVKN Govt College, Chittoor*
[2]*Lecturer in Statistics, SSBN Degree & PG College (Autonomous), Anantapur*
[3]*Assistant professor, Department of Statistics, Vikrama Simhapuri University, Nellore*
[4]*Assistant Professor, Vidya Educational Institutions, Yadamari, Chittoor*
[5]*Assistant Professor, Department of Mathematics, School of Advanced sciences ,VIT, Vellore*

*Abstract: This review expected to create, access, analyze the execution of artificial neural system (ANN), and Random Forest (RF) models in anticipating and mapping SOC stocks in the Eastern Mau Forest Reserve, Kenya. Helper information, including soil inspecting, climatic, topographic, and remotely-detected information were utilized for model alignment. The adjusted models were connected to make expectation maps of SOC stocks that were approved utilizing autonomous testing information. The outcomes demonstrated that the models overestimated SOC stocks. The two expectation maps comprehensively portrayed comparative spatial examples of SOC stocks, with an expanding angle of SOC stocks from east to west. The most elevated stocks were on the backwoods overwhelmed western and north-western parts, while the least stocks were on the cropland-commanded eastern part. The most essential variable for clarifying the watched spatial examples of SOC stocks was add up to nitrogen fixation. In view of the nearby execution of RF and ANN models, we suggested that both models ought to be adjusted, and after that the best outcome connected for spatial expectation of target soil properties in different settings.*

## I.    INTRODUCTION

Securities exchange expectation is a critical field of entusiasm among financial specialists. Exact expectations imply that one conceivably can predict occasions or patterns accordingly making ventures more gainful. Making such forecasts is troublesome because of the perplexing way of the market which is impacted by a wide assortment of elements. A few strategies exist and innovative techniques are one principle branch that has turned out to be progressively prominent as of late. These sorts of strategies use algorithmic models keeping in mind the end goal to make the forecasts. The issue can be viewed as a period arrangement forecast issue which is the undertaking of assessing an esteem construct exclusively with respect to past qualities in a period arrangement. Two cases of mechanical strategies from machine learning (ML) that have been utilized are manufactured neural systems (ANN) and irregular woodlands (RF) [1, 2, 3, 4, 5]. An ANN is a model that is enlivened by the human sensory system and a RF is an accumulation of choice trees. Both these models are reasonable for applications like stock expectation as they can display complex structures, for example, nonlinear examples. These techniques make expectations by breaking down existing information, assembling a model to mirror the hidden way of the information and afterward utilizes this model to sum up on beforehand concealed information.

There exist a considerable amount of research on ANNs inside the field of stock expectation which has yielded some fascinating outcomes [1]. RFs have additionally been looked into inside the territory, however have by and large not pulled in an indistinguishable measure of consideration from ANNs. In any case, in Caruana and Niculescu-Mizil's [6] think about where they did vast scale exact examinations on an arrangement of managed learning calculations; the general execution of the RF model was positioned higher than the ANN display. In this manner an intriguing inquiry rises with respect to how ANN and RF models contrast when connected with stock expectation. The objective of this review is to think about the execution between streamlined usage of ANNs and RFs, individually.

## II.    LITERATURE REVIEW

A.  *Forecasting Stock Index Movement:* A Comparison Of Support Vector Machines And Random Forest
There exists unfathomable research articles which anticipate money markets too evaluating of stock file monetary instruments yet a large portion of the proposed models concentrate on the exact estimating of the levels (i.e. esteem) of the basic stock record. There is an absence of studies looking at the consistency of the heading/indication of stock record development. Given the thought that an

expectation with little figure mistake does not really convert into capital pick up, this review is an endeavour to foresee the course of S&P CNX NIFTY Market Index of the National Stock Exchange, one of the quickest developing money related trades in creating Asian nations. [1] Arbitrary timberland and Support Vector Machines (SVM) are certain sort of machine learning strategy, and are promising apparatuses for the forecast of money related time arrangement. The tried grouping models, which foresee heading, incorporate direct discriminant investigation, logit, counterfeit neural system, arbitrary woodland and SVM. Exact experimentation proposes that the SVM beats the other grouping strategies regarding foreseeing the bearing of the share trading system development and irregular woods technique outflanks neural system, discriminant investigation and logit demonstrate utilized as a part of this review.[2]

B. *Prediction Of Stock Market Index Movement By Ten Data Mining Techniques:*

Capacity to foresee heading of stock/record value precisely is vital for market merchants or speculators to expand their benefits. Information mining procedures have been effectively appeared to produce high anticipating precision of stock value development. These days, in stead of a solitary strategy, merchants need to utilize different anticipating procedures to increase various signs and more data about the fate of the business sectors. In this paper, ten distinct systems of information mining are talked about and connected to foresee value development of Hang Seng record of Hong Kong securities exchange. The methodologies incorporate Linear discriminant investigation (LDA), Quadratic discriminant examination (QDA), K-closest neighbor grouping, Naïve Bayes in light of part estimation, Logit demonstrate, Tree based arrangement, neural system, Bayesian order with Gaussian process, Support vector machine (SVM) and Least squares bolster vector machine (LS-SVM). Trial comes about demonstrate that the SVM and LS-SVM create unrivaled prescient exhibitions among alternate models. In particular, SVM is superior to LS-SVM for in-test expectation yet LS-SVM is, thus, superior to the SVM for the out-of-test estimates in term of hit rate and mistake rate criteria. [3]

C. *Economic Prediction Using Neural Networks: The Case Of Ibm Daily Stock Returns:*

This paper reports a few after effects of an on-going venture utilizing neural system displaying and learning methods to look for and decipher nonlinear regularities in resource value developments. We concentrate here working on this issue of IBM normal stock day by day returns. Dealing with the remarkable elements of financial information highlights the part to be played by measurable deduction and obliges adjustments to standard learning procedures which may demonstrate helpful in different settings. [4]

D. *Stock Market Prediction System With Modular Neural Networks:*

This paper examines a purchasing and offering timing forecast framework for stocks on the Tokyo Stock Exchange and investigation of intemal portrayal. It depends on measured neural systems. We built up various learning calculations and expectation strategies for the TOPIX(Tokyo Stock Exchange Prices Indexes) forecast framework. The forecast framework accomplished exact expectations and the reproduction on stocks tradmg demonstrated a fabulous benefit. The expectation framework was created by Fujitsu and Nikko Securities. [5]

E. *Predicting Stock Market Index Using Fusion Of Machine Learning Techniques:*

The paper concentrates on the errand of foreseeing future estimations of securities exchange list. Two lists to be specific CNX Nifty and S&P Bombay Stock Exchange (BSE) Sensex from Indian securities exchanges are chosen for exploratory assessment. Investigations depend on 10 years of recorded information of these two files. The expectations are made for 1–10, 15 and 30 days ahead of time. The paper proposes two phase combination approach including Support Vector Regression (SVR) in the main stage. The second phase of the combination approach utilizes Artificial Neural Network (ANN), Random Forest (RF) and SVR coming about into SVR–ANN, SVR–RF and SVR–SVR combination forecast models. The forecast execution of these mixture models is contrasted and the single stage situations where ANN, RF and SVR are utilized without any assistance. Ten specialized pointers are chosen as the contributions to each of the forecast models. [6]

III. PROPOSED SYSTEM

The motivation behind this review is to anticipate the bearings of every day change of the S&P CNX NIFTY Index. Bearing is a clear cut variable to demonstrate the development heading of S&P CNX NIFTY Index whenever t. They are arranged as "0" or "1" in the examination information. "0" implies that the following day's record is lower than today's file, and "1" implies that the following day's file is higher than today' s file.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887*
*Volume 5 Issue XII December 2017- Available at www.ijraset.com*

The examination information utilized as a part of this review is specialized markers and the bearing of progress in the day by day shutting costs for the S&P CNX NIFTY Index. The arrangement ranges from first March 2005 to first May 2015. The information is partitioned into two periods-the principal time frame is utilized for model estimation and is delegated in-test, while the second time frame is saved for out-of-test determining and assessment. The division adds up to around 25 for every penny being held for out-of-test purposes. Since this review end eavor to conjecture the bearing of day by day value change in the stock value file, specialized markers are utilized as information factors.

## IV. PREDICTION MODELS

### A. Neural networks

Measurable strategies and neural systems are generally utilized for time arrangement expectation. Exact outcomes have demonstrated that Neural Networks beat straight relapse extraordinarily on account of more mind boggling conduct of ward factors like nonlinear, dynamic and disorderly practices. Neural systems are dependable for displaying nonlinear, dynamic market forecasts. Neural Network makes not very many suspicions instead of typicality suppositions regularly found in factual techniques. Neural system can perform forecast subsequent to taking in the basic connection between the info factors and yields. From an analyst's perspective, neural systems are similar to nonparametric, nonlinear relapse models.
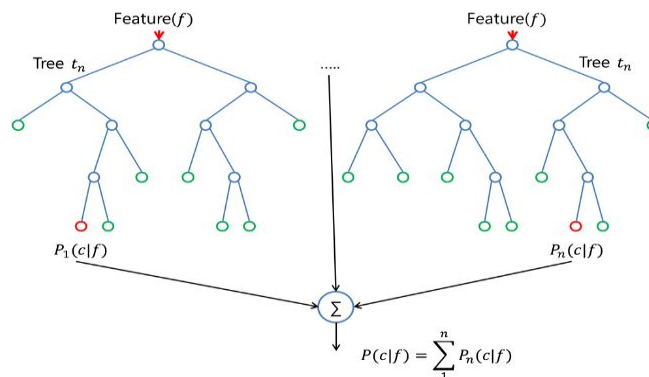
The fundamental design comprises of three sorts of neuron layers: input, covered up, and yield layers. In the forward systems, the flag stream is from contribution to yield units, which is named as sustain forward system. There are a few other neural system structures (Elman organize, versatile reverberation hypothesis maps, focused systems, and so forth.), contingent upon the properties and prerequisite of the application.
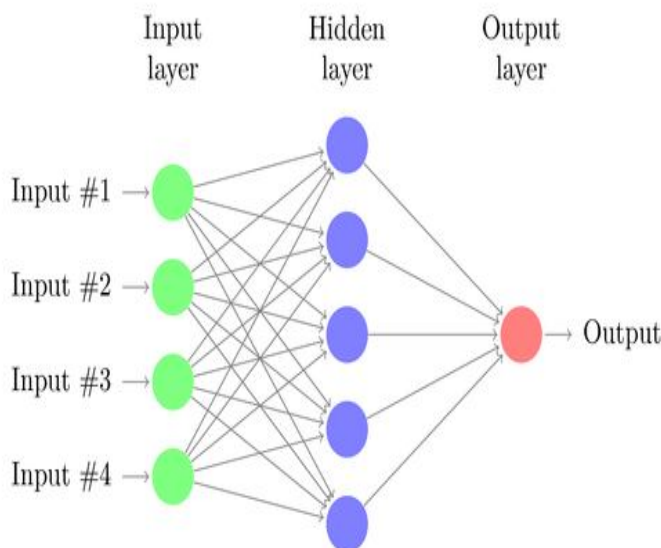


Variances of stock costs and stock records are another case of a perplexing, multidimensional, yet in a few conditions at any rate halfway deterministic marvel. Neural systems are being utilized by numerous specialized examiners to make forecasts about stock costs in light of a substantial number of components, for example, past execution of different stocks and different financial markers.

### B. Random Forests

Choice Trees give an instinctive method for characterizing any order (or relapse). These are utilized to speak to restrictive chain of command of a given framework. A choice is touched base at, by navigating down the choice tree. Choices trees can be comprehensively grouped into order and relapse trees. Various calculations exist for making a choice tree like CART (Classification And Regression Trees), Random Forests, C4.5, ID3. Truck and Random Forests are utilized for the most part utilized for relapse and C4.5 and ID3 for characterization. Arbitrary Forests – utilizes troupe learning i.e. different models utilized for better prescient execution. Technique joins stowing (clarified further as step 3 of calculation) and arbitrary determination of components to build gathering of choice trees with controlled variety.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887*
*Volume 5 Issue XII December 2017- Available at www.ijraset.com*

## C. Numerical Measures Of Errors

Numerical measures of blunders assume a vital part in demonstrative investigation of time arrangement examination. The estimate exactness can be recognized and looked at by utilizing these numerical measures of mistakes as it were. The measures are most normally reliant on the leftover which is characterized as the contrast amongst genuine and anticipated esteem. The accompanying are a portion of the vital measures of the numerical measures of blunders. Estimating models should be assessed from alternate points of view. These incorporate deciding what amount guage values go astray from real values; regardless of whether the model used to gauge is helpful or not; and the quality of direct connection amongst reliant and autonomous factors.

Guage values got from various determining models and strategies may contrast. The blunder residuals demonstrate how dangerous the gauge model or technique is. The model is tried by taking distinction between the genuine esteem and estimate esteem. The littler the distinction, the better the model is. A few criteria can be utilized to analyze distinctive estimating models. As indicated by (zeng, et, al 1998), a few techniques for assessing estimating models are as per the following:

1) Root Mean Square Error(RMSE)
2) Mean Absolute Error (MAE)
3) Mean Absolute Percentage Error (MAPE)
4) Mean square error(MSE)

## D. Mean Squared Error (MSE)

The mean squared error is useful to understand how close the predicted value is from its original value. It relay the concepts of bias, precision, and accuracy in statistical forecasting. The average of the square of the difference between the desired response and the actual system output is known as Mean Squared Error. The smaller the Mean Squared Error, the closer the forecast to the data. It is obtained from

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^{n} e_t^2$$

Where E= Residual measure or difference between predicted and original values
N= Total number of observations in the data.

## E. Root Mean Square Error (RMSE):

The RMSE measures variance of the error. In other words the difference between forecast and corresponding observed values are squared and then averaged over the sample. Finally, the square root of the average is taken. Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. This means the RMSE is most useful when we need to identify the large residuals in the data.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} e_t^2}$$

*F. Mean Absolute Percentage Error (MAPE)*

MAPE is considered as a measure of accuracy in the time series data and it is always expressed in percentage. The absolute difference between actual and predicted value is summed and averaged to minimize the effect of bulk data. Then it is converted to percentage in its final form. The expression for MAPE is given by

$$MAPE = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{e_t}{y_t} \right|$$

*G. Mean Absolute Error (MAE)*

In statistics the MAE is used to measure forecasts or predictions is how much close to the eventual outcomes.MAE is given by

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |e_t|$$

The mean absolute error used the same scale as the data being measured. this is known as a scale-dependent accuracy measure and therefore cannot be used to make comparisions between series using different scales.

The mean absolute error is a common measure of forecast error in time series analysis,where the terms "mean absolute deviation" is sometimes used in confusion with the more standard definition of mean absolute deviation. The same confusion exists more generally.

*H. Research Data*

Ten years of information of aggregate two stock value files (CNX Nifty, S&P BSE Sensex) and two stocks (Reliance Industries, Infosys Ltd.) from Jan 2003 to Dec 2012 is utilized as a part of this review. Every one of the information is acquired from <http://www.nseindia.com/> and <http://www.bseindia.com/> sites. These information shapes our whole informational index. Rate savvy increment and abatement instances of every year in the whole informational index This review utilizes 20% of the whole information as the parameter determination information. This information is utilized to decide outline parameters of indicator models. Parameter determination informational collection is developed by taking equivalent extent of information from each of the ten years. The extent of rate insightful increment and lessening cases in every year is likewise kept up. This testing technique empowers parameter setting informational collection to be better illustrative of the whole informational collection. This parameter choice information is additionally partitioned into preparing and hold-out set. Each of the set comprises of 10% of the whole information. Portrays the quantity of increment and abatement cases for parameter choice informational index. These measurements is for S&P BSE Sensex. Comparative information examination is accomplished for CNX Nifty, Reliance Industries and Infosys Ltd. Ideal parameters for indicator models are acquired by methods for examinations on parameter choice information. From that point forward, for contrasting ANN, SVM, arbitrary backwoods and innocent Bayes, correlation informational index is concocted. This informational index includes whole ten years of information. It is likewise isolated in preparing (half of the whole information) and wait (half of the whole information) set. Points of interest of this informational index for S&P BSE SENSEX. These exploratory settings are same as in Kara, Acar Boyacioglu, and Baykan (2011). There are some specialized markers through which one can anticipate the future development of stocks. Here in this review, add up to ten specialized pointers as utilized in Kara et al. (2011) are utilized. Two methodologies for the portrayal of the info information are utilized in this review.

## V. RESULT

| Results | Neural Networks | Random Forest |
|---|---|---|
| Root Mean Square Error (RMSE) | 3577.2449 | 2486.6361 |
| Mean Absolute | 14.01459 | 79.92351 |

| | | |
|---|---|---|
| Error (MAE) | | |
| Mean Absolute Percentage Error (MAPE) | 0.000871 | 0.005225 |
| Mean square error (MSE) | 214.7160 | 11349.5836 |

## VI. CONCLUSION

This review utilized neural system and arbitrary backwoods to anticipate the day by day development of course of S&P CNX NIFTY Index and contrasted the outcomes and that of conventional discriminant and log it show and manufactured methods like neural system. The exploratory outcomes demonstrated that neural system beat irregular woodland, neural system and other customary models utilized as a part of this review. The better execution of neural systems over alternate models is expected than the reason that neural systems actualize the auxiliary hazard minimization guideline which limits an upper bound of the speculation mistake as opposed to limits the preparation blunder. This in the end prompts preferable speculation over the neural system and irregular woodland which executes the observational hazard minimization guideline. Albeit neural system was marginally superior to the irregular timberland, the two models can further be assessed for various monetary time arrangement, for example, trade rates, and distinctive securities exchange record, to discover a definitive outcome. The machine learning techniques like bolster vector machines and irregular backwoods will help brokers, borrowers, FII and so forth, to settle on better speculation choice. Monetary forecaster, merchants, and dealers, can utilize distinctive exchanging approaches in light of the machine learning methods, which can prompt monetary profit. Be that as it may, every strategy has its own qualities and shortcomings. Future research should likewise be possible by consolidating models by incorporating neural system with other grouping models. The shortcoming of one strategy can be adjusted by the qualities of another by accomplishing a methodical impact. There is additionally extension to survey the heading of securities exchange record, considering the arrangement of potential macroeconomic information factors, for example, loan costs, purchaser value list, mechanical creation and so forth. Along these lines to finish up, neural system is a valuable apparatus for financial experts and professionals managing the estimating securities exchange swapping scale.

## REFERENCES

[1] Kumar, Manish, and M. Thenmozhi. "Forecasting stock index movement: A comparison of support vector machines and random forest." (2006).

[2] Manish, Kumar, and M. Thenmozhi. "Forecasting stock index movement: A comparison of support vector machines and random forest." Proceedings of ninth Indian institute of capital markets conference, Mumbai, India. 2005.

[3] Ou, Phichhang, and Hengshan Wang. "Prediction of stock market index movement by ten data mining techniques." Modern Applied Science 3.12 (2009): 28.

[4] White, Halbert. "Economic prediction using neural networks: The case of IBM daily stock returns." (1988): 451-458.

[5] Kimoto, Takashi, et al. "Stock market prediction system with modular neural networks." Neural Networks, 1990., 1990 IJCNN International Joint Conference on. IEEE, 1990.

[6] Patel, Jigar, et al. "Predicting stock market index using fusion of machine learning techniques." Expert Systems with Applications 42.4 (2015): 2162-2172.

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)