International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Stock Market Prediction Using Machine Learning Techniques

A. Vani[1], K. Naga Vihari[2], Shaik Nafeez Umar[3], M. Bhupathi Naidu[4], N. Ramachandra[5]

[1]Research Scholar, Department of Statistics, S. V. University, Tirupati, A.P, India
[2]Guest Faculty, Department of Statistics, Sri Venkateswara University, Tirupati, A.P, India
[3]Assistant Professor, Department of Statistics and Computer Applications, S.V.Agricultural College, Tirupati, A.P, India
[4]Professor and Registrar, Sri Venkateswara University, Tirupati, A.P, India
[5]Department of Statistics, Sri Venkateswara University, Tirupati, A.P, India

ABSTRACT

Generally, the stock market index oscillating over the time periodand influencingmany factors. Different stock market indices determined from various mixture of stock may share similar trend in certain. The purpose of this research to predict market price index of Bombay Stock Exchange (BSE) on day wise closing using machine learning technique. The machine learning techniques ANN (Artificial Neural Network)including feedforward with lagged values of the as input, and Support Vector Machine (SVM) are compared. The price index was found to be most relevant and influenced the market performance. The results showed that performance of BSE index can be predicted with machine learningapproach. The machine learning approach shown ANN model better performance than the SVM model. The Mean Absolute Percentage Error (MAPE) of ANN is 0.5790 and the Root Mean Square Error (RMSE) is 472.12.

Keywords: Bombay Stock Exchange, ANN, SVM and Prediction

INTRODUCTION

Stock market analytics plays a vital role in business segmentation to predict indices values which is helpful for decision makers who are investing the stock market. Various advanced statistical time series models available for predicting the stock market indices likeAutoregressive Conditionally Heteroscedastic (ARCH), GeneralizedAutoregressive Conditionally Heteroscedastic (GARCH), Vector Auto Regressive Model (VAR), Auto Regressive Auto Regressive Integrated Moving Average (ARIMA), Machine learning model and deep learning models.

Data scientists was build statistical and machine learning flows to access the data, construct the data for supervised and unsupervised modeling, fit various types of analytical models and checking the performance of models to solve the upcoming business problems. Generally, all analytical models are trained and tested data sets based on requirements of the client. There is some issues on selecting the variables on which is depends dependent variable.Many researchers developed and estimated a mixture of models based on Artificial Neural Network (ANN) and Hybrid Neural Network (HNN), Support Vec-

tor Machine (SVM) emphasized, few of the literature on that models.

Gjerde.O., et al. (1999) examined the relationship between stock returns and macroeconomic variables in Norway. The results shows a positive relationship between oil price and stock returns and real economic activity and stock returns.

Malkiel B.G (2003) research shows that investors strongly focus on growth stocks while making investment decisions and are much less concerned about value stocks. This is partly due to the fact that the analysts rarely monitor stocks which have performed poorly in the recent past (value). It isargued that such lack of interest in value stocks leads their prices to depreciate to a value far below their true value, thereby giving investors the chance to benefit from the mispricing errors.

Firmansyah (2006) fluctuations in stock price are indicated by volatility to statistical measure of price fluctuation over a given period.

Duca.G (2007) employs Granger causality test to examine direction of causality among stock prices and GDP in developed market economies. But the result points out a unidirectional causality which runs from stock prices to DGP and that no causality was found in the reverse direction in the developed economies market.

Neha Saini (2014) examined and compared the forecasting ability of Autoregressive Moving Average (ARMA) and Stochastic Volatility models applied in the context of Indian stock market using daily values of Sensex from Bombay Stock Exchange (BSE). The results of the study confirmed that the volatility forecasting capabilities of both the models.

Mohandass (2013) attempted to study the best fit volatility model using Bombay stock exchange daily sectoral indices for the period of January, 2001 to June, 2012. The findings concluded that the non-linear model is fit to model the volatility of the return series and recommended GARCH (1,1) model is the best one.

PraveenKulshreshtha (2011) investigated BSE SENSEX, BSE 100, BSE 200, BSE 500, CNX NIFTY, CNX 100, CNX 200 and CNX 500 by employing ARCH/GARCH time series models to examine the volatility in the Indian financial market during 2000-14. The study concluded that extreme volatility during the crisis period has affected the volatility in the Indian financial market for a long duration.

Srinivasan P., et al. (2010) attempted to forecast the volatility (conditional variance) of the SENSEX Index returns using daily data, covering a period from 1st January 1996 to 29th January 2010. The result showed that the symmetric GARCH model do perform better in forecasting conditional variance of the SENSEX Index return rather than the asymmetric GARCH models.

Kamijo ., *et al.* (1993) was developed in, which combinations of a neural network with an expert system. The neural network was used to predict future prices and generate trading signals.

METHDOLOGY

Daily market indices data from June 2021 to June 2023, Bombay Stock Exchange (BSE) closing price were applied for this study and predict the upcoming trend using Machine Learning Techniques.

Preprocessing Data-Normalization of the data

The consistency of a neural network model has huge extent and depends on the quality of the data used. The min-max normalization of the data, rescaling is the simplest process and consists in rescaling the of features to scale the range between (0,1) or (-1, 1). The closing price of the index were used scaling

function. The function used the maximum and the minimum values of the price index data series, the function was:

$X=\frac{x-\min(x)}{\max(x)-\min(x)}$

Normality test - Jarque-Bera (JB)

The Jarque-Bera normality test one of the most commonly used in time series data. The test statistics always non negative.

$JB=\frac{n}{6}(S^2+\frac{1}{4}(K-3)$, where n is number of observations or degrees of freedom, S is the skewness and K is the Kurtosis.

Machine Learning Approach:

I. Artificial Neural Network (ANN)

The ANN is a biological brain inspired tool in which a huge number of neurons are strongly inter connected in order to solve complex problems like wise multilayer system.

A neural network has always been compared to human nervous system. Information in passed through interconnected units analogous to information passage through neurons in humans, sometime called multilayer system.There are many learning rules available in ANN. In this paper learning rules can be used in conjunction with back propagation error method. This error is explained back propagated to all the units such that the error at each unit us proportional to the contribution of that unit towards total error at the output unit. Figure 1 explains the basic structure of an easy neural network model for better understanding.

For the NNAR model, the Box-Cox transformation was applied before predicting the model and The fitted is denoted as NNAR (n,p) model, where k is number of hidden nodes, analogous to an AR(p) model with nonlinear function.

The model as and Box-Cox transformation with λ=0.05,

$y_i=f(y_{t-i})+\epsilon_t$

$y_i=(y_{t-i})+\epsilon_t$, where $y_{t-1}=(y_{t-1},y_{t-2},\cdots,y_{t-8})$

$y_{t-1}=(y_{t-1},y_{t-2},\cdots,y_{t-8})$ is a vector containing lagged values of the series, and f is a neural network with 4 hidden nodes in a single layer.The error series $\{\epsilon_t\}\{\epsilon_t\}$ is assumed to be homoscedastic

We can simulate future sample paths of this model iteratively, by randomly generating a value for $\epsilon_t\epsilon_t$, either from a normal distribution. The relationship between the output (y_t) and the inputs (y_{t-1} , y_{t-2} , …, y_{t-p}) can be mathematically represented as follows:

$$y_t = f\left(\sum_{j=0}^{q} w_j \; g\left(\sum_{i=0}^{p} w_{ij}\, y_{t-i}\right)\right)$$

Where, w_j(j= 0,1,2,…,q) and w_{ij}(i=0,1,2…,p, j=0,1,2…,q) model parameters often called the connection weights, p number of input nodes q number of hidden nodes, g and f activation function at hidden and output layer respectively. Activation function defines the relationship between inputs and outputs of a network in terms of degree of the non-linearity.

II. Support Vector Machine (SVM)

Forecasting is a critical aspect of data analysis, in particularly in agricultural and weather predictions. In recent years, Support Vector Machine (SVM) has powerful tool for time series forecasting due to handling curve nature relationship and multivariate data. It is a tool for pattern classification and

regression algorithm for data point of view. One popular method for time series forecasting is Support Vector Regression (SVR), which leverages the power of Support Vector Machines (SVMs) for regression analysis.

Generally, Support Vector Machine (SVM) is also supervised learning algorithm used for cluster, classification and regression Problems.

For this type of SVM the error function is:

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i + C\sum_{i=1}^{N}\xi^*{}_i$$

Which we minimize subject to:

$$w^T\phi(x_i) + b - y_i \leq \varepsilon + \xi^*{}_i$$
$$y_i - w^T\phi(x_i) - b_i \leq \varepsilon + \xi_i$$
$$\xi_i, \xi^*{}_i \geq 0, i = 1, \ldots, N$$

Performance of the Model

a. Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE),It is also measure of accuracy for fitted models in estimation.

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

b. Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) of the residual (prediction errors) are indicates the average magnitude of the difference between the values measures by the two devices

RMSE=$\sqrt{\sum_{i=1}^{n}\frac{(P-O)^2}{n}}$, where 'P' is predicted values is 'O'observed values and 'n' is sample size.

RESULTS AND DISCUSSION

Stock market indices (June, 2022 to June 2024) of the Bombay Stock Exchange data have been used for this study. The stock market indices shows increasingly over the time period (Fig1)

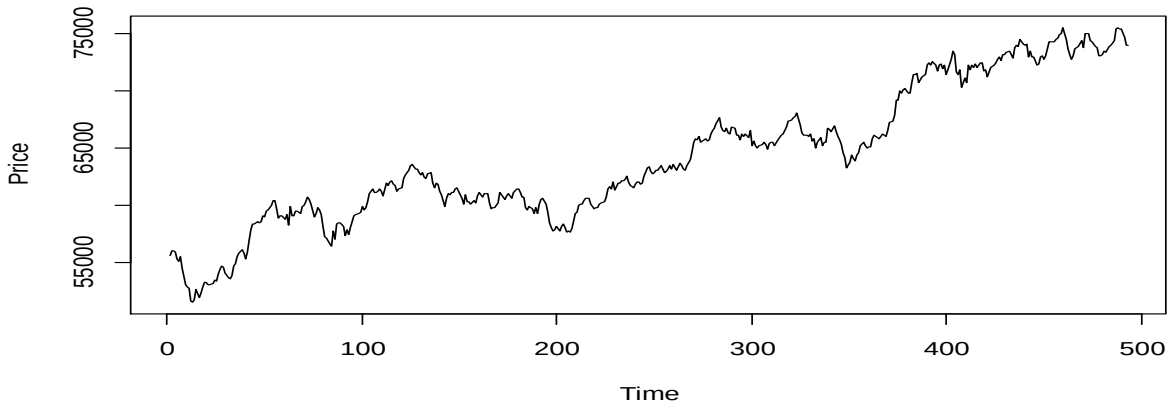

Fig 1: Market indices trend

Table 1: Normality test for Market indices

Normality Test	t-test	p-value
Jarque-Bera test	18.341	0.00

The Jarque-Beratest deployed for testing of normality of stock market indices, from the output, the p-value <0.05, The value indicates that the null hypothesis should be rejected and the data does not follow a normal distribution. The market indices showed not normally distributed.In this case applied data transformation techniques for stock market indices.

Table 2: Model performance

Model	MAPE	RMSE
ANN	0.5790	474.12
SVM	0.5843	475.79

As the above table shows, in both models, ANN model shows very good performed with SVM model. The MAPE an RMSE measures very low of ANN model than the SVM model.The Feed-forward neural networks with a single hidden layer and lagged inputs for forecasting univariate time series. The ANN forecasted values are very close to the actual observations. The residuals of stock market indices are constant mean and variances, of residuals showed a linear trend to confirm that the residuals are normally distributed

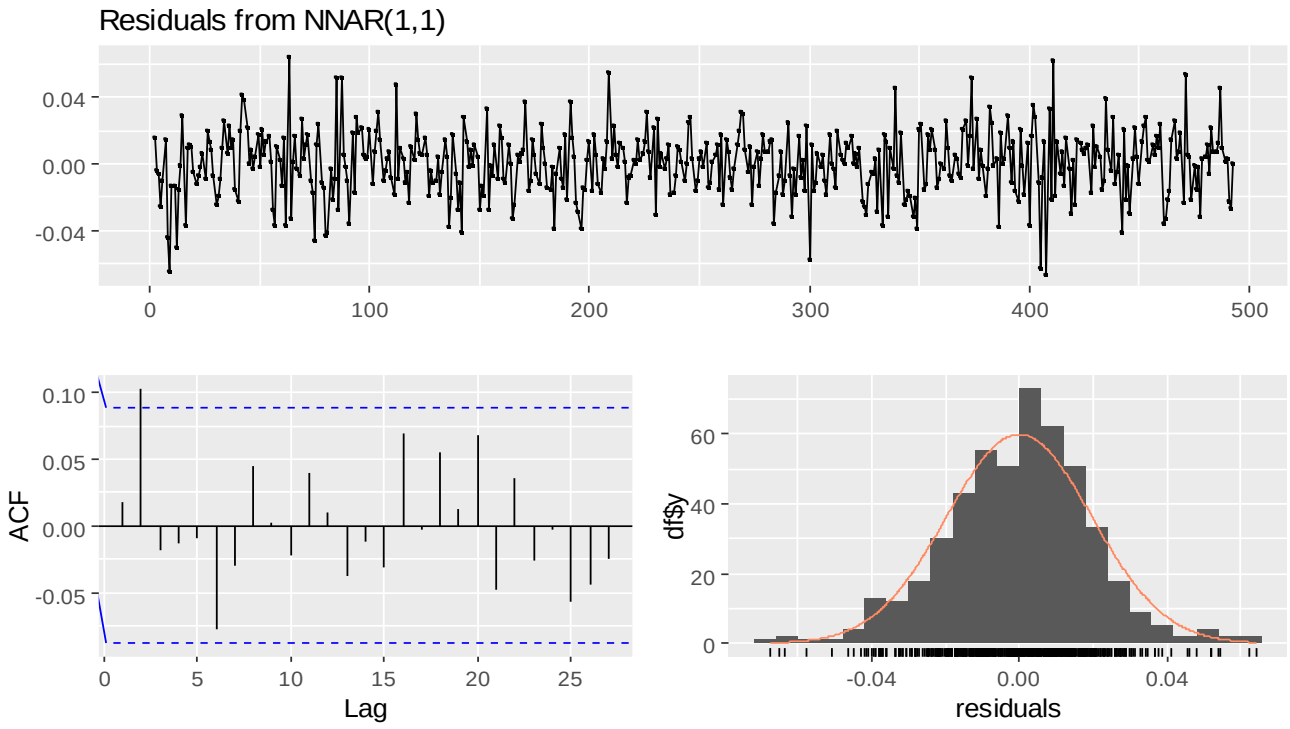

Fig2: Residuals of Stock Market indices-ANN model

The residual graphs (Fig: 2) show that the mean of the residuals is close to zero and constant variance, there is no significant correlation in the residuals series.The variation of the residuals stays

much the same across the historical data, and therefore the residual variance can be treated as constant. The Auto Correlation Function (ACF) of the residuals shows good for prediction.The histogram suggests that the residuals may be followed normally distributed, and the predicted quit good for prediction of stock market indices.

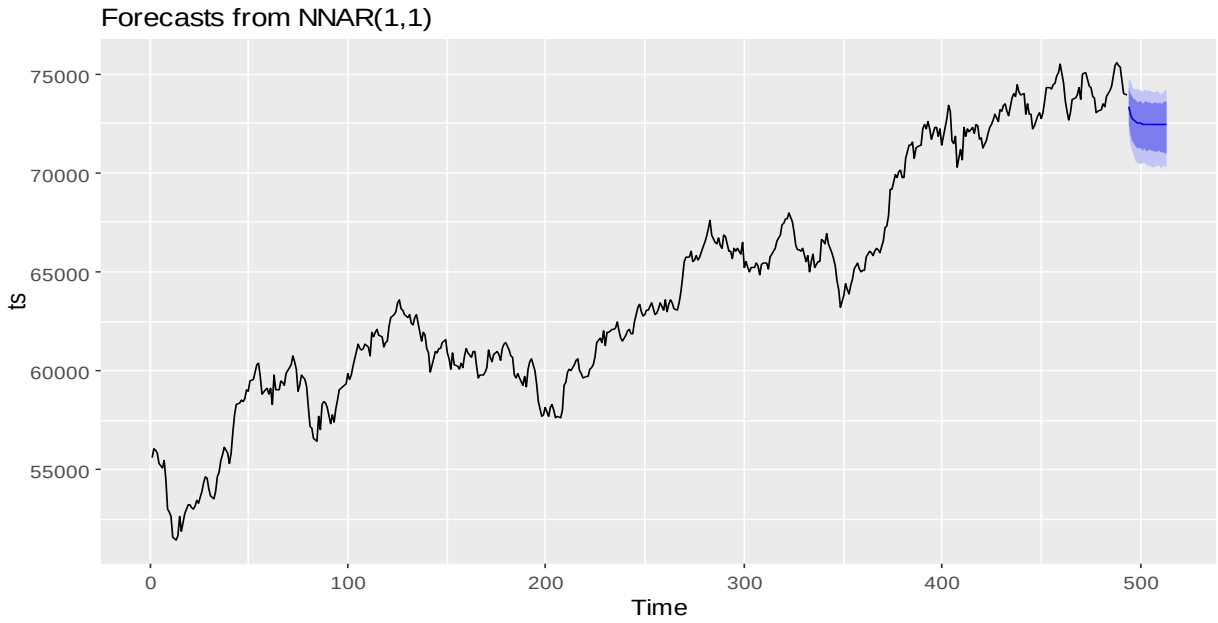

Fig3: Residuals of Stock Market indices-ANN model

CONCLUSIONS

As per the study, the stock market indices are oscillating over the time period, based on many factors. In this paper, two machine learning model were used and compared with performance level at 0.01 levels, the ANN model was good and performed well. The errors in the forecasting were very low in ANN model rather than SVM model. The ANN model useful for predicting upcoming stock market indices.

REFRENCES

1. Duca G. (2007). "The relationship between the stock market and the economy: Experience from international financial markets", *The Bank of Valletta Review*, No. 36.
2. Fama E. F. (1965). "Random walks in stock market prices", *Financial Analyst Journal*, Vol. 21, No. 5, pp. 55-59, doi. 10.2469/faj.v51.n1.1861.
3. Firmansyah (2006). *AnalisisVolatilitasHarga Kopi International*, Jakarta: PT. Usahawan.
4. Gjerde O. and Saettem F. (1999). "Casual relations among stock returns and macroeconomic variables in a small, open economy", *Journal of International Financial Markets Institutions and Money*, Vol. 9, pp. 61-74.
5. Malkiel B.G. (2003). "The efficient market hypothesis and its critics", *Journal of Economic Perspectives*, Vol. 17, No. 1, pp. 59-82.
6. Mohandass (2013). "Modeling volatility of BSE sectoral indices", *International Journal of Marketing, Financial Services & Management Research*, Vol. 2, No. 3.
7. Neha Saini (2014). "Forecasting volatility in Indian stock market using state space models", *Journal of Statistical and Econometric Methods*, Vol. 3, No. 1, pp. 115-136.

8. Srinivasan P. and Ibrahim P. (2010). "Forecasting stock market volatility of Bse-30 index using Garch models", pp. 47-60, available online at:https://journals.sagepub.com/doi/10.1177/097324701000600304.
9. Kamijo, K., &Tanigawa, T. (1993). Stock price pattern recognition: A recurrent neural network approach, NeuralNetworks in Finance and Investing, 357–370.