Comparison of Regression, ANNs and SVMs methods for Prediction of the Indian Stock Market

Vivekananda Journal of Research
January - June 2021, Vol. 10, Issue 1, 125-146
ISSN 2319-8702(Print)
ISSN 2456-7574(Online)
Peer Reviewed Refereed Journal

http://www.vips.edu/vjr.php

Deepanshu Verma*
Subathra V**

Abstract

Focusing on the Indian context, this paper draws attention towards predicting the movements in stock markets. It locates and maps the future position of the index, which is portrayed as a benchmark of the entire economy. For the analysis, NSE was tested for 5 years with techniques of econometrics & soft computing to efficiently predict the index and safeguard the funds invested by the investors.

For this study, a number of tests were applied and data has been analysed. This study found that SVMs has the best prediction and outperform in minimising the error while the ANN's model sets back in front of SVMs and Regression.

Index terms – *Stock Market, Regression, Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Nifty, India*

INTRODUCTION

In developing countries like India, where markets are growing and the economy is among the fastest growing economies of the world. The returns are above anticipated but losses can also hamper the confidence of the budding investors, in the economy where 65% of the population is below the age of 35 years. The importance and vitality of the prediction model then becomes very high when markets are immensely volatile. India is a liberalised, privatised and globalised economy, affected by international markets and exposed to big international events like Trade war between US-China, Middle East disturbances, etc.

* Management Graduate,School of Business, Public Policy & Social Entrepreneurship (SBPPSE), Ambedkar University Delhi (AUD), Email:- dpv.567@gmail.com
** Research Scholar, Faculty of Management Studies (FMS), University of Delhi
Email:- subathravnkt@gmail.com

While internal factors like Budget, Monsoon, Election polls, leading Companies performance affects the markets and indices. Due to these unpredictable events, the volatility and fluctuations become very high and affects the investor's sentiments. In such a scenario, the importance of forecasting tools which is beyond the boundaries of mere psychological or behavioural aspects rises more. The confidence level and right discretion has to reach out to investors regarding expected movement of the market. Such tools may not only empower the investors but also restrict the loss of investor's capital and nation's wealth when dealing with international exposure.

There had been better and improved forecasting models, which somehow applies and uses these techniques as primary tools like Autoregressive Integrated Moving Average (ARIMA) and Generalised Autoregressive Conditioned Heteroskedasticity (GARCH) model. However, with the development of Information Technology, there have been soft computing and machine learning tools like Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) which have reached a significant accuracy level in the western world but still have not been applied on Indian markets [2].

According to the paper, we have distinguished our tools into two categories:

- **Econometric Model** — These are statistical tools like Regression, ARIMA and GARCH. These models need to be guided with certain prerequisite assumptions like linearity of the model and stationery of the time specified financial data.
- **Soft Computing Model** — These techniques consist of Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Fuzzy Logic (FL) and many more; which feeds data into it and simulates the biological process.

SVMs and ANNs have efficiently performed and predicted the Stock market [1], when applied on the Istanbul Stock Exchange.

ANNs have successfully made it possible to predict the financial markets in the western world with satisfying all the diversity elements [2][3][5]. ANNs are refined and advanced versions of regression analysis where preposition of constraints assumptions do not restrict the results. For example, prediction of stock market movement was experimented in [4]. Authors have tested two models ANNs and SVMs and presented validated results. Another type of ANN, radial basis function (RBF) was used to predict the Shanghai Stock Exchange (SSE) in China [11]. In [12] ANNs were trained to predict the American and Singaporean stock exchanges like DJIA, NASDAQ and STI index.

This paper entails comparison between traditional source of prediction i.e., Regression model and soft computing techniques like ANNs and SVMs for the prediction of Nifty. The flow of the research paper is as follow:

Section III gives us a brief idea about the National Stock Exchange (NSE). Section IV gives us information about the research methodology adopted for the paper. Section V talks about evaluation of the criterion and Section VI briefs about the results found out after evaluation. Section VII concludes and shows the path for future work and References thereon.

LITERATURE REVIEW

2.1 Identification of variables

This section summarises the relevant literature and discusses the various attributes which affects the stock market globally.

Alaa F. Sheta and Sara Elsir M. Ahmed (2015) experimented the use of Artificial Neural Networks (ANNs) and Support Vector Machines (SVM) to establish prediction models for the S&P 500 stock index. Authors also get to know how conventional models like multiple linear regressions (MLR) behave in this case. He has used the Regression model, the Artificial Neural Network model and the Support Vector Machine model for forecasting the S&P 500 stock index. SVM outperformed the MLP (ANNs) and MLR (Regression) models in both training and testing cases. SVMs have many advantages such as using various kernels, which allows the algorithm to suit many classification problems. SVM are more likely to avoid the problem of falling into local minimum. 26 potential financial and economic variables.

Mingyue Qiu and Yu Song (2016) analysed the Nikkei 225 stocks data from 1993 to 2013. In this paper, the author had selected 18 input variables to run within the neural network system and inserted in 1, 10, 20, 100…hidden neural networks to pick out the data with output variable and present the justification/validation of the macroeconomics factors in predicting the market behaviour. The accuracy of the data set and methodology was MSE at 0.0044, which suggests that it is a better estimate to predict and conduct the study.

Diebold and Yilmaz (2008) study the link between macroeconomic fundamentals and stock market volatilities for a panel of 40 countries, using Granger causality. Real GDP, real personal consumption expenditures and consumer price inflation are used as macroeconomic indicators. The study concludes that there exists a clear link between macroeconomic fundamentals and stock market volatilities, with volatile fundamentals translating into volatile stock markets.

Mukhopadhyay and Sarkar (2003) analyse the Indian stock market returns for the pre-and-post liberalisation period, to study the impact of macroeconomic factors, and find a strong causal effect of output, inflation, money supply, and foreign direct investment on stock returns, especially post liberalisation.

2.2 Adoption of different methodology for the study

Hakob Grigoryan (2015) achieved results using this technique compared with the ones found from some ARIMA models. Author used the mean square error (MSE) measure to evaluate the performances of these two models (ARIMA & ANN). The comparative analysis assumes that the given model can be successfully applied to predict the financial instructions.

Amin Hedayati Moghaddam (2015) studied NASDAQ for a period of 70 days (20th Jan to 7th March 2015), where the author has distributed data into two input datasets of four prior days and 9 prior days. Then, he applied different techniques like ANN, MLP, FIS and other quantitative tools to forecast the market movement by laying down different hidden neuron structures into different sets and validating the results. In conclusion, the conclusion author has claimed satisfactory results achievement for stock market movement prediction through these tools.

Y. Kara (2011) investigates the Istanbul Stock Exchange National 1000 Index using ANNs and SVMs using 10 macro variables to predict the market. Research data consists of 10 years of financial series from 1997-2007. The model produced 75.74% accuracy for ANNs and 71.52% accuracy for SVMs testing. The paper also shed light on the importance of technical indicators for direction and movement. ISE crisis is held responsible for decrease in the accuracy of the models.

Kwon and Shin (1999) investigate whether economic activities can explain the movement of the stock market in Korea using Engle-Granger co- integration and the Granger-causality tests from the Vector Error Correction Model (VECM). The study reveals that the Korean stock market is co integrated with macroeconomic variables, whereas Granger causality identifies that the Korean stock index is not a leading indicator for economic variables.

Shangkun Deng (2011) examines the relationship between sentimental analysis and stock price movement. ROC, MACD, BIAS and Multiple Kernel Learning are applied where technical and news factors merge to evaluate model performance. Sony, Panasonic and SHARP are Japanese companies whose price movement is traced on the US market with one-week (7 days) data set. This paper suggests more detailed deep text mining for further study.

Schwert (1990) analyses monthly data for the U.S. over the period 1857 to 1987, and finds strong evidence of the impact of volatility in financial asset returns on macroeconomic variable volatility. There are two-way agents between stock market fluctuation and money growth volatility, and between industrial production volatility and stock market volatility.

Wai Shen and Xiaopen Guo (2011) understand the non-linearity of the stock market data and try to forecast Shanghai Stock Exchange. They used ANN (RBF) and Artificial Fish Swarm Algorithm (AFSA) and different optimisation techniques like K-Mean clustering, Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). They also compared the results with ARIMA, BP and SVM. RBF outperforms among all the other models and this paper suggests room for further improvement and development in the AFSA model. This paper also draws scope for adding non-quantitative factors like investor's psychology to increase accuracy.

2.3 Selection for Training & Testing case

Xiaotian Zhu and Hong Wang (2008) investigates the relationship between Stock return and trading volume where three Stock Indices are chosen – DJIA, NYSE and STI using Artificial Neural networks. By establishing the relationship this paper tries to forecast performance under different periods of forecasting outlook. It divides data into two parts – 90% for training and 10% for testing ANNs. It finds only trading volume to be an inefficient tool for forecasting purpose and demand addition of further variable to enhance the accuracy.

Tripathi (2011) examines the relationship between Indian Stock market and macroeconomic variables (91-days Treasury bill is used as proxy for short-term interest rate, wholesale price index for inflation, S&P 500 Equity Index of U.S. for international market index, and exchange rates) during the period January 2005 to February 2011. The results of the Granger causality suggest that there exists bidirectional causality between inflation rate and stock market, exchange rate, stock market, interest rate, stock market, international stock market, BSE volume, exchange rate, and BSE volume.

2.4 Adoption of Financial Data Time-Series

The financial time-series taken for the study is the 5-year period from 1st April 2014 to 31st March 2019. This period had all the economic stability with respect to Indian economy and no major international event took place like war. During the considered period, India had a stable or majority government, which was elected for the next 5 years without any coalition. All these factors have made 2014-19 an appropriate period to study and formulate prediction models under normal and certain conditions.

In 2014, a newly formed government was formed which had a clear mandate or supreme majority in the House of Representatives – Lok Sabha. It was a political shift after 10 years of functioning of UPA I and UPA II. A new leader was elected and NDA came into power with approximately 2/3rd of lion's share.

2.5 Components for prediction purpose

There have been 26 variables indicated based on [1] (Alaa, 2015), where variables were considered for the prediction use. The data has been inscribed and collected on a weekly *basis* in order to map movement in a concise manner. Weekly data provides access to locate events and attribute causes more appropriately within a collection of days. Especially a country like India where festivals and working holidays are very frequent and non-symmetric due to cultural significance it becomes very difficult to trace the relationship between Indian and International markets. There are *261 weeks* from the period of 5 years. These weeks are further segmented to assess and analyse data.

What is Nifty?

India has the oldest stock market in Asia, i.e., Bombay Stock Exchange (BSE) but after the introduction of National Stock Exchange (NSE), which is a completely digitised stock exchange and the number of trade settlement surpassed the BSE by 1997.

The lower concentration on one-sector makes Nifty a suitable index for experiment purpose as it diversifies and consists of broader exposure with association of larger number of constituents.

Nifty or Nifty50 (National Stock Exchange Fifty) started from 3 November 1995 with a base point of 1000. It is the flagship index for NSE, which consists of 50 large cap diversified stocks from different sectors and portrays the outlook of the economy. The benchmark value is measured in concurrent.

NIFTY 50 is measured using Free Float (FF) Market Capitalization weighted method, wherein the level of index reflects the free float market capitalization of all stocks in Index. It is operated and managed by NSE Indices Limited. In order to quantify Nifty, we have to calculate the sum of current market price of all the companies and divide it with the base market capital and base index value.

$$\textbf{\textit{Index level}} = \frac{\sum \textbf{\textit{CMPi}}}{\textbf{\textit{Base Market Capital}} \times \textbf{\textit{Base Index Value}}}$$

CMP implies current market price of the stocks and Base Market Capital is the total market capitalization of each scrip in the benchmark during the base period.

E1. Index Elements

A. Nifty (Weekly Opening)

The Nifty weekly opening is a measure that connects the slack of the weekend. Nifty closing is not exactly the opening for the index. The index starts from 9-9:15am when it starts accepting the contracts deal and officially opens from 9:30am for normal settlements and contacts. These 15 minutes window fills the gap and reflects the variation from last week is closing. The level of variation in the opening and closing defines the urgency and takes into account the investors sentiment for the initial span.

B. Trading Volume

Trading volume *(million transactions)* is the quantity of shares transacted on the stock exchange during a specified point of time. Trading volume has a frictional degree of relationship with volatility and equity trading volume (Jeff Fleming, 2011). It is an important variable to trace the fluctuations.

E2. Nifty Companies

Nifty is a weighted index, which constitutes a portion of companies based on the market capitalization. Therefore, the major companies are the primary sources for fluctuations of the index. These companies constitute approximately 39.9% of the index- HDFC Bank (10.9%), Reliance Industries (10.2%), HDFC Ltd. (7.5%), ICICI Bank (6.11%) and Infosys Ltd (5. 13%). Tracking the movement of these companies may denote the direction for the index forecast.

E3. Global Currencies

International trade and relations play a vital role in stabilising and depreciation or appreciation of the home currency. Indian Rupee (INR) has been tracked with respect the biggest trade partner of India, which are USA, UK and UAE. Therefore, in this regard we have taken these 3 currencies – Dollar ($), Pound (£) and Dirham based on the report of Directorate General of Foreign Trade India.

E4. Commodities

India imports 84% of crude oil from the rest of the world and a leading importer of gold as well. So, these two commodity factors are chosen to track and record the impact on the index. Crude Oil and Gold prices lead to affect the currency movement in respect to inflow/outflow.

E5. Financial Market

Bond markets are the substitute markets for the index investment where there fixed income securities guaranteed with lock-in period of 5-years or specific time duration. The interest rate fluctuations and during crisis period, these markets provide an investment platform. Hence, the relationship and impact of financial markets and stock markets had to be considered to draw conclusive results.

E6. Indian Economy

The drivers of economic outlook and attraction for foreign investors as well as key fundamental factors like level of inflation, GDP growth rate forecast and Industrial production play supportive and sentimental roles to depict the true and right image of the economic conditions in a nation.

E7. Global Indices

The relative movement and impact of international events can be traced on the global indices and they do affect the Indian Stock Exchange. Denoting the time when the UK's Brexit deal took place Indian markets were also fluctuating and relative impact could be noted.

E8. Banking

The notifications and regular notices by the central bank of India, i.e., Reserve Bank of India play a very significant role in determining the liquidity of the economy and provides layout for the monetary, which is issued quarterly by the apex institution. Key rates and rations include Repo Rate, Cash Reserve Ratio (CRR), Statutory Liquidity ratio (SLR) and Foreign Exchange Reserve with the government information to deal with the demand & supply parity.

E9. Outlier – Demonetisation

The data set also considered an outlier event of the Indian economy on 8th November 2016, when the Prime Minister of India banned the currency denomination of Rs.500 and Rs.1000 to curb the flow of counterfeiting currency, cut off funding for terrorist groups and formalisation of the economy.

This step withdrew Rs.15.31 lakh crores from the economy as denomination of Rs.500 and Rs.1000 as per Finance Minister Report in the parliament.

RESEARCH GAP AND CONTRIBUTION

Predominantly, there have been studies which incorporated prediction tools technical models like 52 weeks high/low, Moving Averages (MA), Candle-stick charts, Relative strength index (RSI) Moving averages converging diverging (MACD), etc. but all these could not accurately predict and significant confidence level had not been achieved.

There had been paper published globally in context of accurately forecast the market conditions in different time period but the technological tools had not been that regressive and capable of absorbing all that information in an efficient manner which could produce satisfactory results.

Particularly, in Indian context there had not been study based on multiple macro-economic factors but they were highly niche and concentrated on niche components which could not enhance the model accuracy. This paper attempts to cover maximum number of effective variables and achieve a better level of precision to claim higher amount of accuracy.

RESEARCH METHODOLOGY

A. Regression Analysis

Regression analysis is an econometric package used to draw inference between two or more than two variables called dependent and independent variable, wherein dependent variable is constituted to be influenced by the independent variable. This technique has been efficiently used for mapping and prediction of different scenarios [2]. It is important to draw relationships and quantify the range of impact of independence on dependent variables. It is crucial to identify how stock index fluctuates over the period.

Multiple Linear Regression (MLR)

The simple linear model can be elaborated as the equation (1) given below:

$$y = a_1x_1 + a_2x_2 + \cdots a_nx_j \quad (1)$$

This equation is capable for terms up to x_j independent variable. In this case, we need Least Square (LS) approximation to generate the efficient values for the standards, $a_1, \ldots. a_j$, which is given in (2)

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{a}_1x_1 - \hat{a}_2x_{2-\ldots}\hat{a}_nx_n)^2 \quad (2)$$

In order to calculate the optimal value of the standards $\hat{a}_1, \ldots, \hat{a}_n$ using differentiation for the functions

$$\frac{\partial L}{\partial \hat{a}_1} = \frac{\partial L}{\partial \hat{a}_2} = \cdots = \frac{\partial L}{\partial \hat{a}_j} = 0 \quad (3)$$

By answering the above equation, we can generate the optimal value and apply the multiple regression problems. This technique is efficient on a larger sample and weak on smaller sample.

B. Artificial Neural Networks

Artificial Neural Networks (ANNs) are mathematical models, which were inspired by the human brain. It also functions on the footnotes of the biological process where every input assigned with a bias and all inputs collated with biases to form an output.

Neural networks can be applied for making prediction, classification and developing pattern identification [2]. Due to the collation of all input sets in hidden layer, it has the ability to perform on incomplete data sets and uncertainty.

In Feedforward, Multilayer Perceptron (MLP) contains neurons, which are ordered in layers. Each of these neurons contains two types of functions – Summation function and Activation function. The summation equation is written in (4) and Activation equation written in (5).

Data sets are trained using Training case through input layers, which are bridged with hidden layers. Hidden layers weigh the information and assigns weights. At last, all hidden layers derive the values and produces output through sigmoid function. A training algorithm like Back Propagation can be used to adjust neural networks and minimise the error between actual and the predicted value [3].

$$S = \sum_{i=0}^{n} w_i x_i \tag{4}$$

$$\phi(S) = \frac{1}{1+e^{-S}} \tag{5}$$

There must be assignment of tasks into the system to normalise the data set like number of neurons, hidden layer, epochs, adopted learning algorithm.

C. Support Vector Machines

Support Vector Machines are advanced supervised learning models for forecasting and scheduling the data set. The primary job of SVM is to map data into higher dimensional areas in lieu of separating and classification of data [7]. Data mapping is used by locating the predestined kernel function. Data separation task is done by finding the optimal hyperplane, separating two classes.

Learning Process in SVM

Training data in the SVM can be explained by an example suppose a data set $\{x_i, y_i\}_{i=1,\ldots,n}$ where the input vector $x_i \in R^d$ and the predicted $y_i \epsilon R$. The model objective in SVM is to obtain the linear decision function as given in (6).

$$f(x) \leq w, \phi_i(x) > +b \tag{6}$$

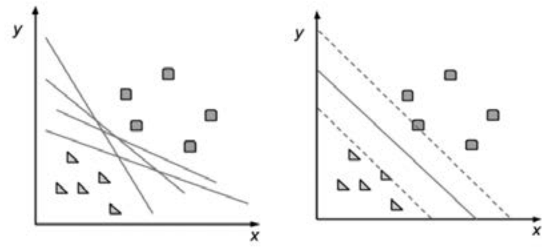
In this equation, *w* is the vector weight whereas *b* is constant. This equation can be further elaborated in the function below, which is formed to minimise the regularized risk function.

$$R(C) = \frac{C}{n} \sum_{i=1}^{n} L_\varepsilon(f(x_i), y_i) + \frac{1}{2} \parallel w \parallel^2 \tag{7}$$

Where $L_\varepsilon(f(x_i), y_i)$ is the concise version of ϵ- intensive loss equation and denoted by the following equation:

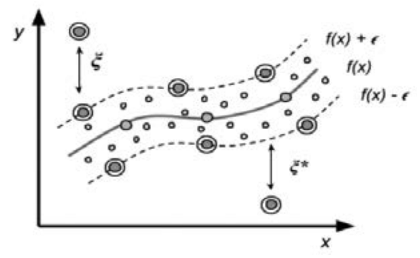
$$L_\varepsilon(f(x), y) = \begin{cases} \{|f(x) - y| - \epsilon & |f(x) - y| \geq \epsilon \\ 0 & otherwise \end{cases} \tag{8}$$

To quantify the level of misclassification to achieve a satisfactory level of error, so we use ξ_i and ξ_i^* (refer figure 2)



(1)

Figure 1: *Optimal hyperplane in Support Vector Machine*



(2)

Figure 2: *Optimal hyperplane with slack variable*

$$Min. R(w, \xi_i^*) = \frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{9}$$

In the equation (9), C is the regularized constant greater than zero. C constitutes a restriction for prediction difference, i.e., greater than ϵ. ξ_i and ξ_i^* are slack variables that represent the error between the predicted and actual values of ϵ. The prime function of SVM is to minimise the slack variables size and weight square (w^2).

This equation (9) can be transformed by means of Lagrarian multipliers to a quadratic equation:

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x) + b \tag{10}$$

In equation (10) α_i and α_i^* are Lagrarian multipliers. Equation (11) is subject to the following limitation:

$$\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0$$

$$0 \leq \alpha_i \leq C \qquad i = 1, \ldots\ldots, n \tag{11}$$

$$0 \leq \alpha_i^* \leq C \qquad i = 1, \ldots\ldots, n$$

K(.) is the kernel function and its components of two vectors (x_i, x_j) in the space $\phi(x_i)$ and $\phi(x_j)$ which assures Mercer's condition. Hence,

$$K(x_i, x_j) = \phi(x_i).\phi(x_j) \tag{12}$$

Based on the literature, SVMs have majorly multiple merits over traditional methods like better output with high dimensional area and can produce good results even on small set of data using above kernel function & tools.

EVALUATION CRITERIA

There are certain parameters to apply to test and standardise the results of different prediction models. These criterions are tested to quantify how well they analyse and predict the market conditions. They include the following tools: Mean Absolute Error (MAE), Root Mean Standard Error and coefficient of correlation, given below (14) (15) and (16).

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |(y_i - \hat{y}_i)| \quad (13)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (14)$$

$$R = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{y})(\underline{y}_i - \overline{\hat{y}})}{\sum_{i=1}^{n}(y_i - \overline{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \overline{\hat{y}})}} \quad (15)$$

In the proposed model, *y* is the actual index point, $\hat{y}$ is the predicted value, whereas *n* is the total number of weeks.

ANALYSIS AND RESULTS

A. INDEX DATA SET

For this study, there has been identification of 26 crucial features, which affects the Nifty in India. These factors consist of 9 broad areas from Index, companies, currencies, commodities market, financial markets, economy, global indices, banking notifications and an outlier — Demonetization. All these features are transcribed into various variables $\{x_1, \ldots . x_{26}\}$ for numeric calculations and modelling of the data. The financial data of 5 years from 1st April 2014 - 31st March 2019 is broken into 261 weeks' duration to access data space and larger quantity of data for accurate prediction model.

In this set of 261 weeks, 183 weeks (70%) will be treated as *Training set* to feed into the data and rest 78 weeks (30%) is considered as *Training set* to verify the prediction accuracy and locating residuals. The objective of larger proportion for training set is due to feeding of larger data set for supervised learning and model accuracy purpose.

B. MULTIPLE REGRESSION MODEL

The regression model applies the given equation for the analysis:

$$y = a_0 + \sum_{i=1}^{26} a_i x_i \quad (16)$$

The values of the features shall be equated using least square estimation to generate optimal results. The model results are represented using chart, which are given below.

Figure (3a) and (4a) are the set segmented data set for the prediction purpose. The training data has been fitted to the model and testing in the next phase with the real or actual data.

The difference in the training case is showcased in Figure (3b), which depicted the mismatch in fitting the data. Figure (4b), depicts the error of difference in the actual and predicted data set.

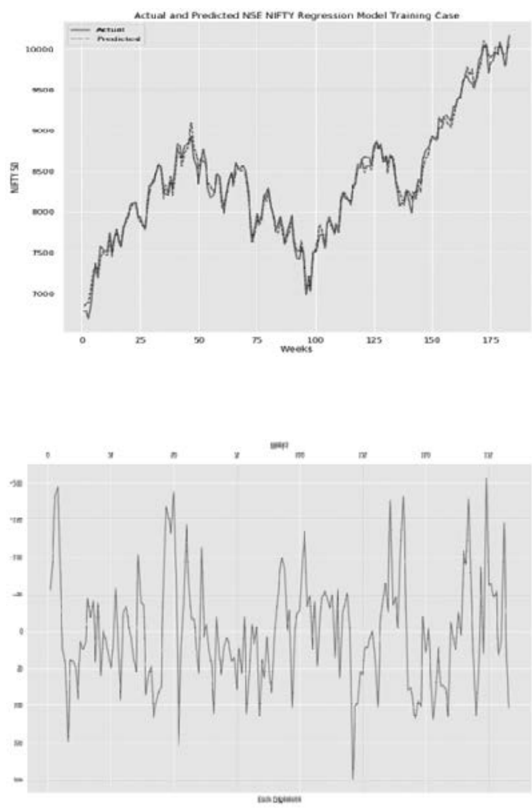
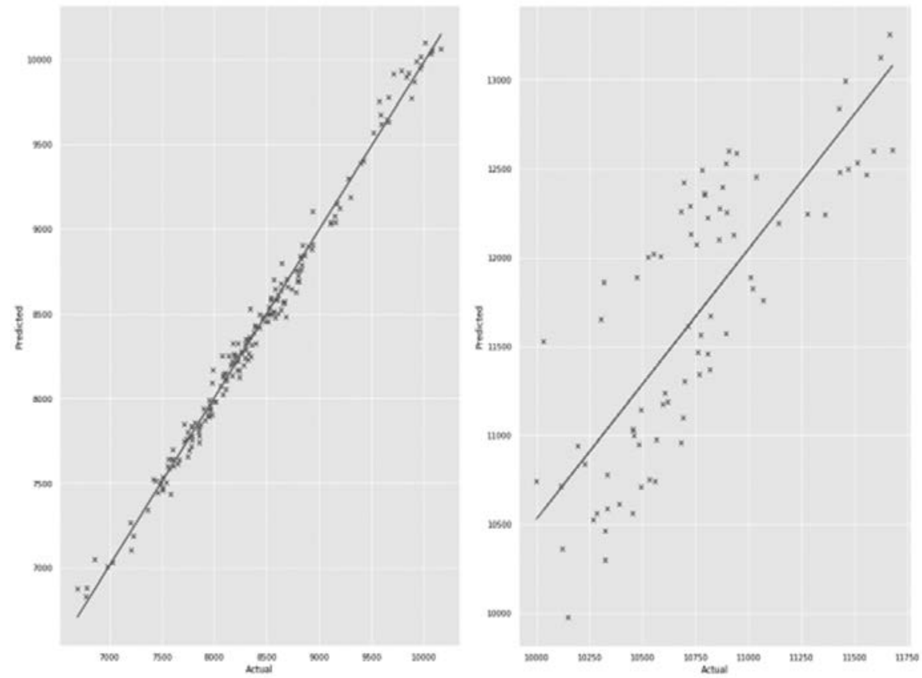


Figure 3: *(a, b) (L-R) Regression: Actual and Predicted Nifty Index values in Training Case*

Figure 4: *(a, b) (L-R) Regression: Actual and Predicted Nifty Index values in Testing Case*



R for Training Data — 0.9948 *R for Testing Data — 0.8043*

Figure 5: (a, b) (L-R) Regression Scatter Plot

TABLE 1: A REGRESSION MODEL WITH INDEPENDENT VARIABLES: $\boldsymbol{x1}, \ldots\ldots, \boldsymbol{x26}$

$$\begin{aligned}\hat{y} = (5330.27) + (&-533.84 * x_1 + 269.62 * x_2 + 77.45 * x_3 + 1351.88 * x_4 + 1658.89 * x_5 \\ &+ 1726.96 * x_6 + 975.77 * x_7 + 740.01 * x_8 - 2117.05 * x_9 + 2305.67 \\ &* x_{10} - 199.33 x_{11} + 224.36 * x_{12} + 382.67 * x_{13} + 61.08 * x_{14} - 8.54 * x_{15} \\ &+ 928.50 * x_{16} + 8.36 * x_{17} + 488.24 * x_{18} - 30.27 x_{19} + 153.38 * x_{20} \\ &+ 291.87 * x_{21} + 342.27 * x_{22} + 211.32 * x_{23} + 38.22 * x_{24} + -408.74 * x_{25} \\ &+ 0 * x_{26})\end{aligned}$$

TABLE 2: LIST OF VARIABLES WITH p-VALUE

	Variables	Field	Constant	SE	Beta	t value	p value	R
	Constant		5,330.27					
x1	Nifty Open	Index		0.09	0.40	4.63	-	-533.84
x2	Trading Volume			-	-0.04	- 1.95	0.06	269.62
x3	ICIBK	Companies		1.08	0.10	2.68	0.01	77.45
x4	RELI			0.40	0.09	1.08	0.29	1,351.88
x5	HDFCB			1.07	0.19	1.01	0.32	1,658.89
x6	HDFC			0.47	0.13	1.10	0.28	1,726.92
x7	INFY			0.61	0.11	2.67	0.01	975.77
x8	USD/INR	Currency		_	0.36	0.10	0.92	740.01
x9	AED/INR			98.70	-0.13	- 1.89	0.07	-2,117.05
x10	GBP/INR			11.36	-0.05	- 0.91	0.37	2,305.67
x11	Oil	Commodities markets		0.05	-0.01	- 0.15	0.88	-199.33
x12	Gold			0.53	-0.01	- 0.41	0.69	224.36
x13	GILT	Financial Market		135.77	0.13	0.42	0.68	382.67
x14	Corp Bond (%)			180.80	0.03	0.35	0.72	61.08
x15	GDP	Economy		4,075.03	0.05	1.63	0.11	-8.54
x16	IIP			886.87	0.01	0.67	0.51	928.50
x17	WPI			1,727.54	0.01	0.18	0.86	8.36
x18	FTSE	Global Indices		0.15	0.10	1.84	0.07	488.24
X19	Nikkei 225			0.04	0.01	0.16	0.88	-30.27
X20	NYSE			0.14	-0.01	- 0.10	0.92	153.38
x21	SSE			0.07	-0.05	- 1.82	0.08	291.87
x22	SLR	Banking		6,963.26	-0.01	- 0.17	0.87	342.27
x23	Repo Rate			7,636.93	-0.01	- 0.31	0.76	211.32
x24	CRR			_	-0.01	- 0.17	0.87	38.22
x25	Foreign Exchange Reserve (Rs. Cr)			0.00	0.01	0.10	0.92	-408.74
x26	Demonetisation	Outlier		-246.00	199.81	- 0.10	- 1.23	0.22

C. DESIGNED ANN MODEL

He proposed the structure of the MLP Network consists of three layers (input, hidden layer and output layer). The model has 26 nodes, which are connected to one hidden layer, which produced single node for the output layer. The Back Propagation (BP) algorithm is used to train the data set. Table IV is giving all the technical details applied for testing.

TABLE 3: THE SETTINGS FOR MLP

Epochs	5000
Number of Hidden Layer	1
Number of neurons in hidden layer	1000
Optimizer	Adam
Loss of data	MSE

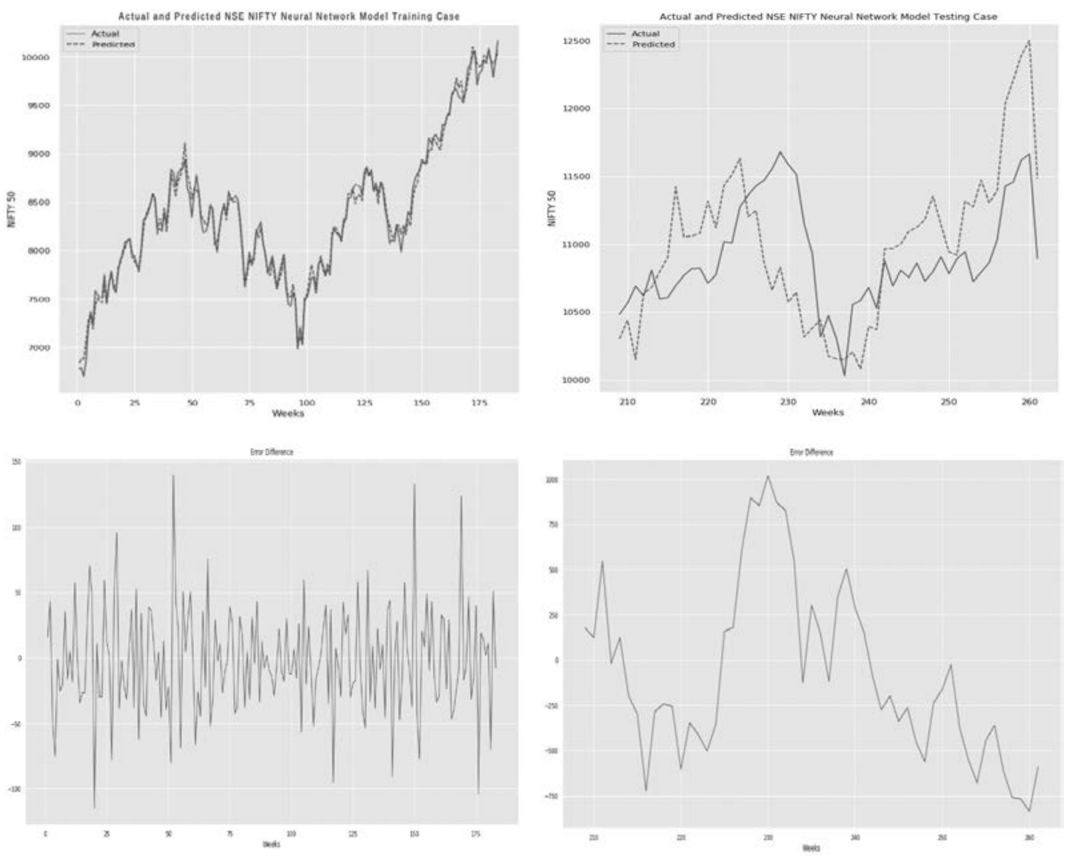
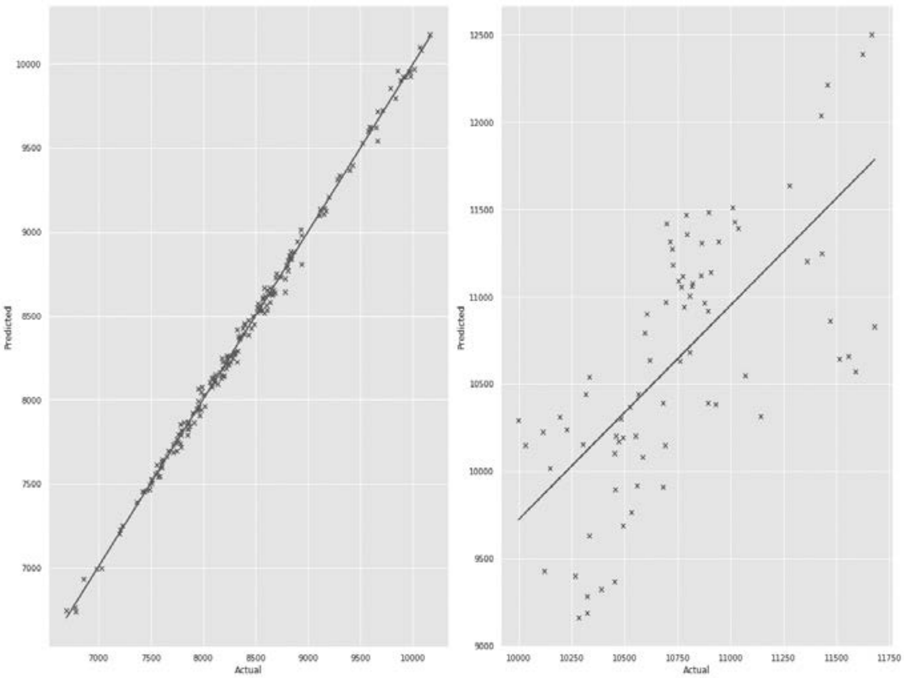


Figure 6: *(a, b) (L-R) ANN - Actual and Predicted Nifty Index values in Training Case*

Figure 7: *(a, b) (L-R) ANN - Actual and Predicted Nifty Index values in Testing Case*



R for Training Data – 0.9906

R for Testing Data – 0.7692

Figure 8: *(a, b) (L-R) ANN — Scatter Plot*

D. DESIGNED SVM MODEL

SVM is trained with RBF kernel to develop the Nifty index model. RBF function has the potential to plot non-linear function to ease implementation [7]. The values of *C (Width of Support Vector)* and have high affect the accuracy of the model. The top-of-the-line results was achieved with *C* = and = *0.01*. Figure 9 and 10 shows the training case and testing case with error of difference. Figure 11 shows the scatter plot with the coefficient of correlation *(R)*.

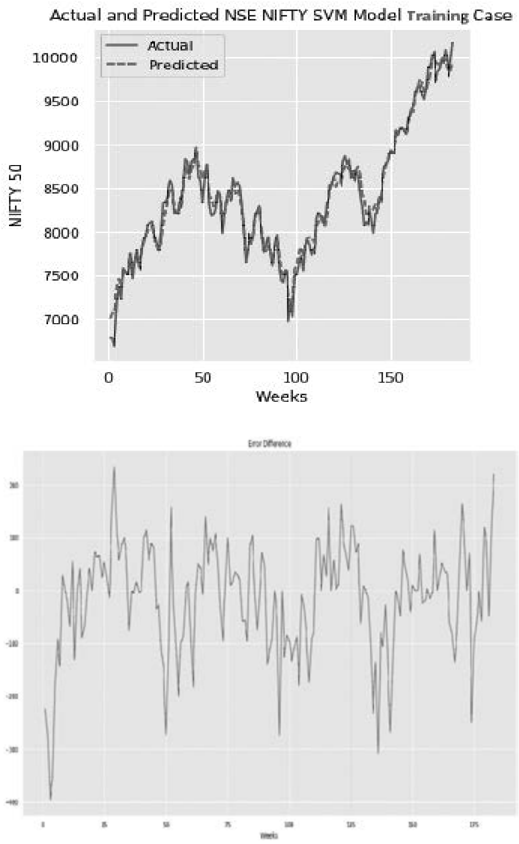


Figure 9: (a, b) (L-R) ANN - Actual and Predicted Nifty Index values in Training Case

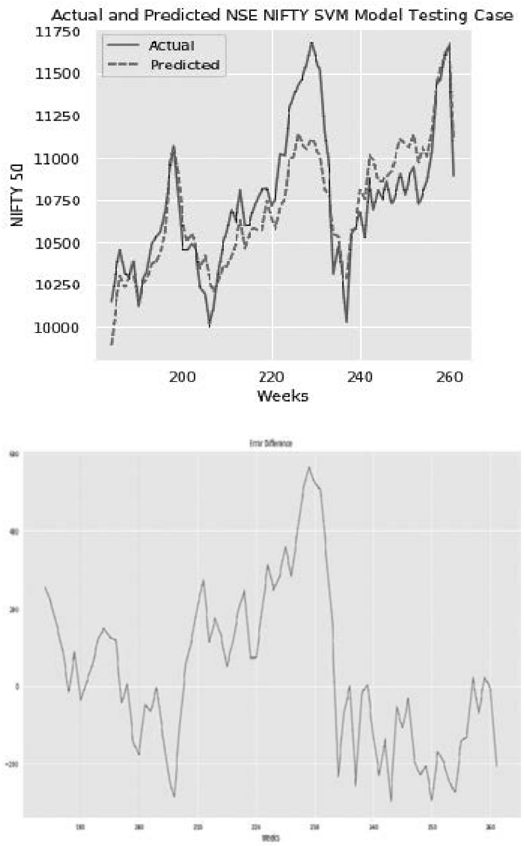
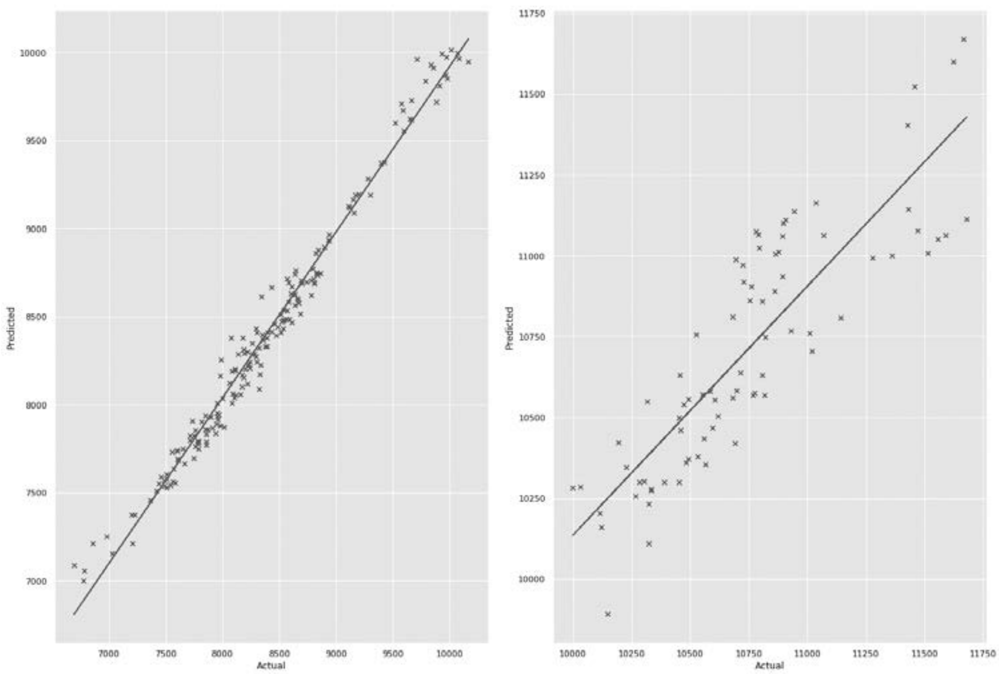


Figure 10: (a, b) (L-R) ANN - Actual and Predicted Nifty Index values in Testing Case



R *for Training Data — 0.9906* *R for Testing Data — 0.8627*

Figure 11: *(a, b) (L-R) SVM — Scatter Plot*

RESULTS & OUTCOME

The computed evaluation criterion of the regression, ANNs (MLP) and SVMs (RBF) model for training and testing cases are displayed in Table IV. On the basis of these outcomes, it can be concluded that SVM model outperformed between MLP and MLR class of models. SVM has the advantage of least loss of estimation between predicted and actual values. The second-best alternative is MLR, which outperformed MLP model in both accuracy and data loss with minimum value of 168.29.

TABLE 4: EVALUATION SCHEME FOR THE MODELS

Criteria	Regression Model		ANNs		SVMs	
	Training	Testing	Training	Testing	Training	Testing
1. Coefficient of Correlation (R)	0.995	0.804	0.991	0.769	0.991	0.863
2. Coefficient of Correlation Square	0.990	-5.497	0.979	-11.980	0.973	0.736

3. Mean Absolute Error (MAE)	58.900	921.090	78.690	1,113.260	78.690	168.290
4. Root Mean Squared Error (RMSE)	74.829	1,047.080	106.450	1,479.930	106.450	211.130
5. Relative Absolute Error (RAE)	0.319	1.700	0.369	2.028	0.369	0.727

CONCLUSION AND FUTURE WORK

In this paper, we scrutinized and analysed data set on the different prediction models – Regression, ANNs and SVMs for the prediction of Nifty. Using these models, we produced Training and Testing results. Based on these results it is observed that these 26 features of economic and market movement significantly affect the market movement for the specified during the study.

The data was bundled on a weekly basis to mitigate any event loss. Between the generated models, SVM models outperformed in predicting the stock market movement with minimised loss of data function. All the results are verified using the evaluation criterion set.

For future work, there should be more focus on new and developing soft computing and machine learning techniques to forecast and build a more robust model that emphasise on increasing the accuracy even more.

REFERENCES

Alaa F. Sheta (2015). A Comparison between Regression, Artificial Neural Networks and Support Vector Machines for Predicting Stock Market Index - (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No.7.

Mingyue Qiu (2016). Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market -Chaos, Solitons and Fractals 85: 1–7.

Hakob Grigoryan (2015). An Artificial Neural Network for Data Forecasting Purposes", Informatica Economic vol. 19.

Amin Hedayati Moghaddam (2016). The Stock market index prediction using artificial neural network: Journal of Economics, Finance and Administrative Science Volume 21, Issue 41, December pp. 89-93.

Y. Kara (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul stock exchange, Expert Syst. Appl., vol. 38.

Diebold, F., & Yilmaz, K. (2008). Macroeconomic Volatility and Stock Market Volatility, Worldwide. NBER Working Paper No. 14269 (N)

Kwon, C., & Shin, T. (1999). Co-integration and causality between macroeconomic variables and stock market returns. Global Finance Journal, 10 (1), 71-81.

Shangkun DENG, (2011). Combining Technical Analysis with Sentiment Analysis for Stock Price Prediction: Ninth IEEE International Conference on Dependable, Autonomic and Secure Computing, pp. 5-8.

Mukhopadhyay, D., & Sarkar, N. (2003). Stock return and macroeconomic fundamentals in model-specification framework: Evidence from Indian stock market. Indian Statistical Institute Economic Research Unit, ERU 2003-05 Discussion Paper, 1-28.

Schwert, G. (1990). Stock Returns and Real Activity: A Century of Evidence. Journal of Finance, 45, 1237-1257.

W. Shen (2008). Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm," vol. 24, no. 3, pp. 378–38.

X. Zhu (2008). Predicting stock index increments by neural networks: The role of trading volume under different horizons. Expert System Application, vol. 34, no. 4, pp. 3043–3054.

Tripathi, N. (2011). Causal Relationship between Macro-Economic Indicators and Stock Market in India. Asian Journal of Finance and accounting, 3 (1)

Reserve Bank of India (2019). QUARTERLY ESTIMATES OF GROSS VALUE ADDED At BASIC PRICE. Retrieved Jan 2020. https://dbie.rbi.org.in/DBIE/dbie.rbi?site=home

Department of Commerce (2019). System on Foreign Trade Performance Analysis (FTPA) Version 3.0. Retrieved Jan 2020. https://tradestat.commerce.gov.in/ftpa/rgncntq.asp