International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.03, April 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Prediction of Stock Value Using NIFTY 50 Market Index and RBF-Kernel Based Support Vector Regressor

Sambit Hore[1], Raj Vipani[2], Dr. Prithwiraj Das[3] and Dr. Saibal Dutta[4]

[1]*Department of Electrical Engineering, Heritage Institue of Technology, Kolkata(India)*

[2]*Department of Electrical Engineering, Heritage Institue of Technology, Kolkata(India)*

[3]*Department of Electrical Engineering,GCETT, Berhampore, W.B., Kolkata(India)*

[4]*Department of Electrical Engineering, Heritage Institue of Technology, Kolkata(India)*

ABSTRACT

Support Machine Regression (SVR), a category of the Support Vector Machine (SVM) is a machine learning method gaining increased popularity for its wide range of applications from pattern recognition to regression estimation with a generalized performance. In our work, we have employed a Support Vector Regression algorithm for the prediction of market day closing value of companies listed under the National Stock Exchange of India's NIFTY 50 market index with the intraday opening value of the index. Owing to the highly non-linear and noisy nature of financial stock market time series, it is important to employ a SVR model as other conventional models are not adequate. We have used a SVR based on a Radial Basis Function (RBF) Kernel and the proposed model had been simulated on the MATLAB platform. All results reported have been averaged over ten tests and the error in forecasting has been obtained below 1%.

Keywords—Financial Forecasting; Support Vector Reggresor; RBF- Kernel.

I. INTRODUCTION

It is of utmost importance to accurately forecast the value of stocks so that correct investments can be made with the minimum amount of risk percentage. The financial time series data is highly non-linear, noisy and non-stationary in nature [1 & 10]. This calls for the usage of machine learning and data mining techniques for forecasting the trends of stocks in the market. Over the past two decades, a number of neural network algorithms and models have been developed which have an advantage over traditional models used for describing the dynamics of financial time series data [2, 3 & 9]. Neural Networks have a better accuracy at forecasting market trends due to its high tolerance towards noisy, incorrect and corrupted data [13].

The Support Vector Machine developed by Vapnik & co-workers is a supervised machine learning algorithm that can be applied for the purpose of regression [4, 5 & 14]. Support Vector Regressor (SVR) is a category of the Support Vector Machine, which is a machine learning approach to perform non-linear regression analysis. Based on the Statistical Learning Theory, the Support Vector Machine used as a regressor has the objective of creating a hyper-plane that passes through as many data points as possible [15]. This hyper-plane is called the Optimal Separating Hyper-plane.

In this work, we have suggested a predictive model that uses a Radial Basis Function (RBF) kernel based Support Vector Regressor for predicting the intraday closing value of companies listed under the National Stock Exchange of India's NIFTY 50 stock market index from the opening value of this index. The subsequent sections describe the data used and the algorithm proposed in this study. We have also briefly described the Support Vector Machine as a regressor in section IV. The results of comprehensive testing have been described in section V.

II. DATA DESCRIPTION

In this paper, we have used the publicly available [16] historical data of the National Stock Exchange (NSE) of India's NIFTY 50 stock market index and 8 companies from different sectors of the Indian economy listed under this stock market index. Table 1 lists the companies that have been used to perform this study. The NIFTY 50 index, a free float market capitalization weighted index, is calculated by the weights assigned in accordance to the size of the companies listed under this stock market index. Thus, the NIFTY 50 index gives the overall health of the companies listed under this index. In this work, the trading day opening value of NIFTY 50 has been used to forecast the closing value of the same day of 8 companies which are a part of the said index. Analysis has been done for 30, 50, 100 and 200 trading days. Starting from September 1, 2016, the data of the NIFTY 50 index and 8 companies under the index has been obtained for 201 trading days i.e. till June 23, 2017. The market trend of the NIFTY 50 index and 2 companies for the said period has been shown in Fig.1.

Name of the Company	***NSE Symbol***
Bajaj Auto Ltd.	BAJAJ-AUTO
Tata Consultancy Services Ltd.	TCS
Wipro	WIPRO
Larsen & Turbo Ltd.	LT
Bosch Ltd.	BOSCHLTD
Coal India Ltd.	COALINDIA
Cipla Ltd.	CIPLA
Bharti Airtel Ltd.	BHARTIARTL

Table 1: List of companies used in this study

Fig.1: Market Trend of NIFTY 50, Tata Consultancy Services Ltd. and Larsen and Turbo Ltd. for the period 01/09/2016 to 23/06/2017

III. PROPOSED METHODOLOGY

Fig.2 shows the flowchart of the algorithm proposed in this work. We have used the opening value of NIFTY 50 stock market index to forecast the closing value of a particular trading day of companies listed under this index.

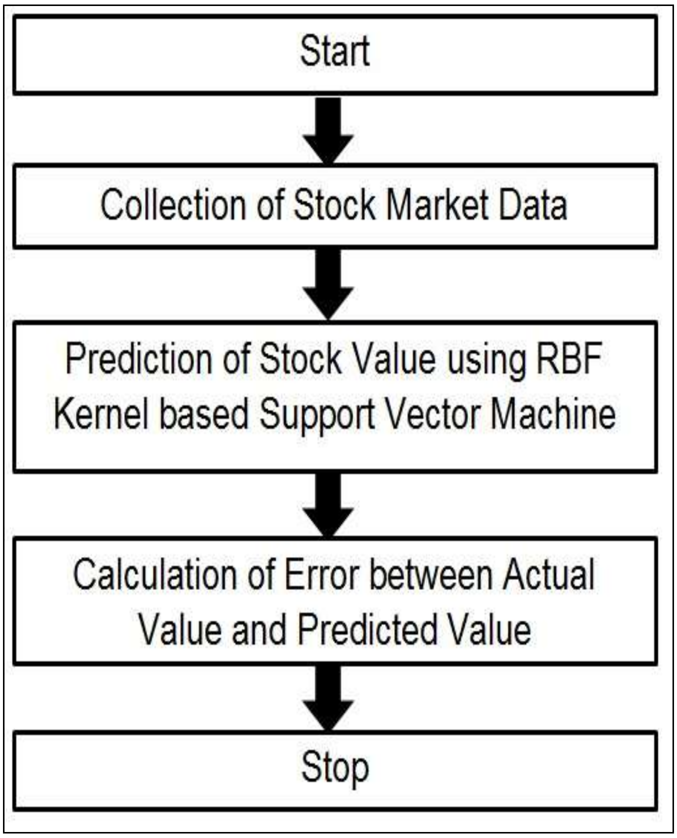

Fig.2: Flowchart of the proposed methodology

IV. SUPPORT VECTOR REGRESSOR

Support Vector Machine is a discriminative model of a linear classifier with the objective of classifying a set of data points using the characteristics of the object. It is a supervised machine learning algorithm that can also be used for regression, data mining and other machine learning tasks [12]. Support Vector Regression (SVR) is an extension of the Support Vector Machine which uses a loss function to minimize the generalization error boundary. This loss function is called Epsilon Insensitive.

In linear regression, a regression function is developed with the help of a paired input-output set and parameters determined from the set. This function is used to predict an output value from of an input value. The input value is the independent variable whereas the output value is the dependent variable. Similarly, in case of SVR, this output value is an approximate value of the input data given by the Optimal Separating Hyperplane (OSH) with an error bound. The OSH is represented by the function:

$$f(x) = <\omega, x> + b \; with \; \omega \in x, b \in R$$

(1)

where, x is the feature vector of the data points, ω is the weight vector which is learned from a set of training data points, < > is the dot product in the vector space x and b is called the Bias which alters the position of the decision boundary [6].

For a given data set $\{(x_1,y_1),(x_2,y_2), \ldots, (x_n,y_n)\}$, in e-SVR(Epsilon- Support Vector Regression) we find the function f(x) such that it has a maximum deviation of ε from the measured target y_i corresponding to the input x_i. In this study, the target variable is the market closing value of the listed companies whereas the input is intraday opening value of NIFTY 50 market index. ε is also called the margin of tolerance. Fig.3 shows linear regression analysis with epsilon insensitive band.

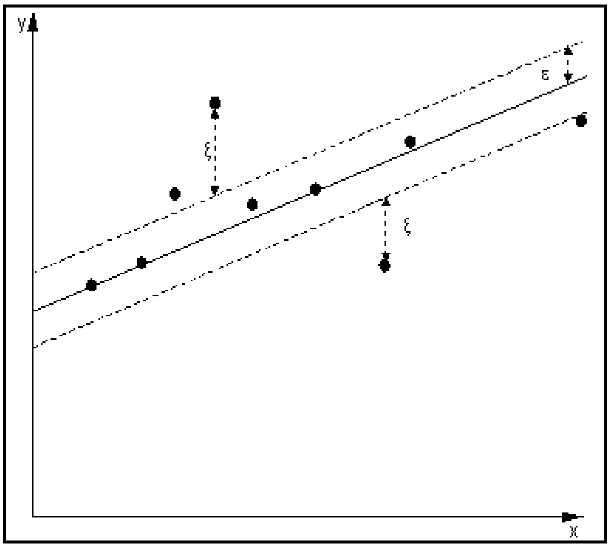

Fig.3: Linear Regression in one dimension vector space with epsilon insensitive band

For better results, it is important to make the function f(x) as flat as possible. The flatness of f(x) gives a small value of the weight vector ω, which can be achieved by minimizing the Euclidean Norm i.e. $||\omega||^2$ by solving it as a convex optimization problem by requiring

$$minimize\ \frac{1}{2}||\omega||^2$$

$$subject\ to\ \begin{cases} y_i - < \omega, x_i > -b \le \varepsilon \\ < \omega, x_i > +b - y_i \le \varepsilon \end{cases}$$

(2)

In equation (2) it is assumed that the function f(x) exists and approximates all pairs (x_i, y_i) with an error margin not exceeding ε , in other words the optimization is possible. But this might not always be the case and does not provide a good generalization model. To deal with these infeasible constraints we introduce slack variables (ξ_i, ξ_i^*) in the optimization problem. Hence (2) can be written as [8]:

$$minimize\ \frac{1}{2}||\omega||^2 + C\sum_{i=1}^{\ell}(\xi_i + \xi_i^*)$$

$$subject\ to\ \begin{cases} y_i - < \omega, x_i > -b \le \varepsilon + \xi_i \\ < \omega, x_i > +b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \quad \ge 0 \end{cases}$$

(3)

The constant C > 0 finds the balance between the two optimization requirements i.e. between the flatness of f(x) and the amount to which the deviations larger than ε can be considered [14]. Equation (3) can be easily solved in its dual formation with the help of Lagrange Multipliers [11]. Using the primal function and its constraints, the Lagrange function is developed as:

$$\begin{aligned} L := \frac{1}{2}||\omega||^2 &+ C\sum_{i=1}^{\ell}(\xi_i + \xi_i^*) \\ &- \sum_{i=1}^{\ell}\alpha_i(\varepsilon + \xi_i - y_i + < \omega, x_i > +b) \\ &- \sum_{i=1}^{\ell}\alpha_i^*(\varepsilon + \xi_i^* + y_i - < \omega, x_i > -b) - \sum_{i=1}^{\ell}(\eta_i\xi_i + \eta_i^*\xi_i^* \end{aligned}$$

(4)

Equation (4) is then optimized using saddle point condition of the function. The final conditions are then obtained as:

$$maximize \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{\ell}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) < x_i, x_j > \\ -\varepsilon\sum_{i=1}^{\ell}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{\ell}y_i(\alpha_i - \alpha_i^*) \end{cases}$$

$$subject\ to \begin{cases} \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \quad \in [0, C] \end{cases}$$

(5)

Using the conditions in Eq. (5), we get

$$\omega = \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)x_i \; and \; therefore \; f(x) = \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*) < x_i, x > +b$$

(6)

This is called the Support Vector expansion i.e. we can describe ω as a linear combination of the training patterns x_i which entails that the complexity of the functions representation by Support Vectors is independent of the dimensionality of the input space and depends only the number of Support Vectors.

Using the Karush-Kuhn-Tucker (KKT) condition which states that at the optimal solution the product between the dual variables and constraints has to vanish; b in Equation (1) is found as:

$$b = \begin{cases} y_i - < \omega, x_i > - \varepsilon \; for \; \alpha_i \in (0, C) \\ y_i - < \omega, x_i > + \varepsilon \; for \; \alpha_i^* \in (0, C) \end{cases}$$

(7)

In SVR in order to make the algorithm valid for nonlinear inseparable sets we preprocess the training patterns x_i by a map $\varphi: X \rightarrow F$ into some feature space F of higher dimensionality. This technique is called the kernel trick. In this study, we have used the Radial Basis Function (RBF) Kernel to achieve the said transformation.

V. RESULTS AND DISCUSSIONS

In our work, a Radial Basis Function (RBF) kernel based Least Squares Support Vector Regressor is used to perform the regression analysis. The kernel parameters 'C' and 'σ' have been found using the grid search algorithm [8]. The RBF kernel is mathematically expressed as:

$$K(x, x') = e^{-\gamma \,|\, X - X' \,|^2}$$

(8)

The trading day closing value of 8 companies listed in Table1 has been forecasted from the trading day opening value of the NIFTY 50 stock market index. The Support Vector Regressor is trained using the 30 trading days data of the market opening value of NIFTY 50 and the market closing value of the companies included in this study. The market closing value is then forecasted from the market opening value of the NIFTY 50 stock index of the 31st trading day from the starting date of the period of this study. The average of ten time forecasted value is taken to calculate the error as:

$$Error\,(\%) = \frac{|A.C.V - P.C.V|}{A.C.V} \times 100$$

(9)

Where, A.C.V is the actual closing value and P.C.V is the predicted closing value. The calculated errors are listed in Table 2.

Similarly, the errors between the actual closing value and predicted closing value for the 51^{st}, 101^{st} and 201^{st} market day is calculated by training the Support Vector Regressor for 50, 100 and 200 trading days respectively. These errors are listed in Table 3, Table 4 and Table 5 respectively.

International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.03, April 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

The results from a RBF kernel based Support Vector Regressor is compared with Support Vector Regressors having a linear kernel function and a Multi-Layer Perceptron kernel function. The errors calculated are shown in Table 6.

Name of Company	**Actual Closing Value (□)**	**Predicted Closing Value (□)**	**Error (%)**
Bajaj Auto Ltd.	2787.95	2814.86	0.96
Tata Consultancy	2395.10	2398.40	0.13
Wipro	495.00	492.26	0.55
Larsen & Turbo	1491.05	1478.80	0.82
Bosch Ltd.	22376.20	22205.80	0.76
Coal India Ltd.	312.55	315.62	0.98
Cipla Ltd.	596.35	591.01	0.89
Bharti Airtel Ltd.	306.95	304.02	0.95

Table 2: Prediction of market closing value on October 19, 2016 with NIFTY 50 opening at 8697.50

Name of Company	**Actual Closing Value (□)**	**Predicted Closing Value (□)**	**Error (%)**
Bajaj Auto Ltd.	2549.80	2565.40	0.61
Tata Consultancy Services Ltd.	2137.20	2148.40	0.52
Wipro	438.40	436.36	0.46
Larsen & Turbo Ltd.	1371.00	1375.60	0.34
Bosch Ltd.	19226.80	19231.00	0.02
Coal India Ltd.	306.85	308.45	0.52
Cipla Ltd.	550.95	545.69	0.95
Bharti Airtel Ltd.	297.15	298.32	0.39

Table 3: Prediction of market closing value on November 17, 2016 with NIFTY 50 opening at 8105.10

Name of Company	Actual Closing Value (□)	Predicted Closing Value (□)	Error (%)
Bajaj Auto Ltd.	2854.10	2828.70	0.89
Tata Consultancy Services Ltd.	2357.80	2380.30	0.95
Wipro	465.55	463.90	0.35
Larsen & Turbo Ltd.	1439.90	1443.70	0.27
Bosch Ltd.	22254.90	22426.00	0.77
Coal India Ltd.	317.75	320.00	0.58
Cipla Ltd.	582.25	582.00	0.05
Bharti Airtel Ltd.	323.75	325.70	0.63

Table 4: Prediction of market closing value on January 27, 2017 with NIFTY 50 opening at 8610.50

Name of Company	Actual Closing Value (□)	Predicted Closing Value (□)	Error (%)
Bajaj Auto Ltd.	2824.30	2848.70	0.86
Tata Consultancy Services Ltd.	2361.70	2352.20	0.39
Wipro	256.85	258.10	0.49
Larsen & Turbo Ltd.	1722.70	1738.00	0.88
Bosch Ltd.	23247.50	23165.40	0.35
Coal India Ltd.	244.90	246.87	0.80
Cipla Ltd.	534.65	534.28	0.06
Bharti Airtel Ltd.	366.20	367.15	0.25

Table 5: Prediction of market closing value on June 23, 2017 with NIFTY 50 opening at 9643.25

Name of the company	Kernel Function		
	Radial Basis Function	Linear Function	Multi-Layer Perceptron
Bajaj Auto Ltd.	0.86	0.87	0.74
Tata Consultancy Ltd.	0.39	0.53	1.01
Wipro	0.49	0.86	0.93
Larsen & Turbo Ltd.	0.88	0.89	1.02
Bosch Ltd.	0.35	0.37	0.81
Coal India Ltd.	0.80	0.91	0.94
Cipla Ltd.	0.06	0.53	0.23
Bharti Airtel Ltd.	0.25	0.40	0.38

Table 6: Error comparison of predicted results by changing the Kernel Functions for June 23, 2017 with IFTY 50 opening at 9643.25.

VI. CONCLUSIONS

In this work, we have proposed a method to predict the closing value of companies listed under the National Stock Exchange of India's NIFTY 50 market index with the help of the opening value of the said index. The proposed method has been developed on the MATLAB platform and the mean error between the predicted value and actual value for 8 companies listed in table 1 of this study has been found to be less than 1%. This study can be extended to include all 50 companies listed under this index.

REFERENCES

JOURNALS:

[1] Y. Abu, and A. Atiya Mostafa, "Introduction to financial forecasting*", Applied Intelligence, 6 (3): 205–213." (1996).*

[2] W. Cheng, Lorry Wagner, and Chien-Hua Lin, *"Forecasting the 30-year US treasury bond with a system of neural networks", NeuroVe$t Journal 1.2 (1996).*

[3] Kaastra, Iebeling, and Milton S. Boyd, "Forecasting futures trading volume using neural networks", *Journal of Futures Markets 15.8 (1995): 953-970.*

[4] C. Cortes and Vladimir Vapnik, "Support vector machine", *Machine learning 20.3 (1995): 273-297.*

[5] Drucker. Harris, et al., "Support vector regression machines", *Advances in neural information processing systems. 1997.*

[6] Chapelle. Olivier and Vladimir Vapnik, "Model selection for support vector machines", *Advances in neural information processing systems. 2000, 230-236.*

[7] A.J. Smola, and Bernhard Schölkopf, "A tutorial on support vector regression", *Statistics and computing 14.3 (2004): 199-222.*

[8] Hsu, Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin, "A practical guide to support vector classification." (2003): 1-16.8.

CONFERENCES:

[9] Naeini, Mahdi Pakdaman, Hamidreza Taremian, and Homa Baradaran Hashemi, "Stock market value prediction using neural networks", *Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on. IEEE, 2010.*

BOOKS:

[10] J.W.Halls: Adaptive selection of US stocks with neural nets. Trading on the edge: neural, genetic, and fuzzy systems for chaotic financial markets, *New York: Wiley (1994): 45-65.*

[11] R. Fletcher: Practical methods of optimization, *John Wiley & Sons, 2013.*

[12] L. Wang: Support vector machines: theory and applications, *Vol. 177. Springer Science & Business Media, 2005, May 11.*

[13] S. Haykin: Neural Networks: A Comprehensive Foundation, *Prentice Hall, New Jersey, 2nd edition*

[14] V. Vapnik: The Nature of Statistical Learning Theory, *Springer, New York,1995, ISBN 0-387-94559-8.*

TECHNICAL REPORT:

[15] N.Ancona, "Classification Properties of Support Vector Machines for Regression", *Technical Report, RIIESI/CNR-Nr. 02/99.*

WEBSITE:

http://www.nseindia.com