International Journal of Computational Intelligence Research
ISSN 0973-1873 Volume 13, Number 5 (2017), pp. 1145-1156

http://www.ripublication.com

Time Series Data Prediction Using Sliding Window Based RBF Neural Network

H.S. Hota[1] , Richa Handa[2] and A.K. Shrivas[3]

[1]*Department of CSA, Bilaspur University, C.G., India*

[2,3] *Department of Information Technology, Dr. C.V. Raman University, C.G., India*

Abstract

Time series data are data which are taken in a particular time interval, and may vary drastically during the period of observation and hence it becomes highly nonlinear. Stock index data are time series data observed daily, weekly or even monthly. Prediction of these types of data is very challenging. For accurate prediction of time series data different intelligent techniques are being used by the researchers, on the other hand, prediction of next day close price on the basis of current day price is not appropriate, instead an average of a particular range of stock data known as window may be suitable for prediction of highly nonlinear stock data. This paper explores an Artificial Neural Network (ANN) technique: Radial Basis Function Network (RBFN) for data prediction using the concept of sliding window, which produces data for current day using historical data of earlier days calculated by Weighted Moving Average (WMA). Experiments were carried out using 10-fold cross validation technique with MATLAB written code for BSE30 Index data. Result produced through RBFN were measured through MAPE, MSE, MAD and RMSE and found satisfactory.

Keywords: Weighted Moving Average (WMA), Sliding Window, Radial Basis Neural Network (RBFN), K-fold cross validation.

I. INTRODUCTION AND LITERATURE

The stock market is dynamic, non-stationary and complex in nature, the prediction of stock price index is a challenging task due to its chaotic and non linear nature. The prediction is a statement about the future and based on this prediction, investors can decide to invest or not to invest in the stock market [2]. Stock market may be influenced by many factors which cause the performance of stock market either in positive direction or in negative direction which includes political events, general economic conditions etc.

Artificial Neural Network (ANN) is a promising technique and quiet popular among the researchers due to its capability of mapping highly non linear input-output data samples unlike any statistical regression model. During last one decade researchers are focusing to develop prediction model based on neural network techniques. Authors [25,26,27] have developed many models based on Back Propagation Network (BPN) and Radial Basis Function Network (RBFN). However hybridization [27] and ensemble of various techniques are now becoming popular, on the other hand data preprocessing is one of the crucial step of stock price prediction which includes data smoothing, feature extraction and feature selection [26] etc.

Cheng Yeh et al. [6] have analyzed a new evolution approach to stock trading system to focus on evaluating the generalization capability at the model level, It clarify the issue of over-learning at the model and the system level. Z. Uykan et al.[8] have uses RBFN to determine the centers of RBFN to analysis of Input-Output clustering. They apply clustering algorithm and present the approach for investigating the relationship between clustering process of input output training samples and mean square output error in context of RBFN. Leonel A. Laboissiere et al. [5] propose a methodology that forecast the maximum and minimum stock prices. This methodology is based on calculation of distinct features to be analyzed by mean of attribute selection and actual prediction is carried out by ANN and performance is evaluated by MAE, MAPE and RMSE. Pei-Chann Chang et al. [7] have proposed a novel model by evolving partially connected neural networks (EPCNN) to predict the stock price trends using technical indicators as input, the proposed architecture of this paper provide some features different from Artificial Neural Network like random connection between neurons, more than one hidden layer and evolutionary algorithm is employed to improve the learning algorithm and training weights. Mohammad Awad et al. [10] dealt with the problem of time series prediction, the prediction is based on historical data. They provide a new efficient method of clustering of centers of RBFN. This clustering method improves performance and prediction of time series data as compared to other methods. R.J. Kuo et al. [22] proposed three stage forecasting model by integrating wavelet transform, k-means algorithm and support vector machine (SVM). The experimental results show that the forecasting algorithm with both wavelet transform and clustering has performed better. Besides, firefly algorithm-based SVR outperforms the other algorithms. However researcher have worked a lot with hybrid model but very few have used Weighted Moving Average (WMA) as data preprocessing. This paper emphasizes more on data preprocessing rather than integrated model development. Due to non-linearity of time series data historical data of previous days were considered to produce new data using WMA. A moving average (MA) is commonly used with time series data to smooth the noisy data by filtering the noise from dynamically fluctuated data. WMA smoothes the price curve [5] for better trend direction and assigns a weight factor to each value in the time series data based on its appearance. The highest weight is assigned for most recent data and a comparatively small weight is chronologically assigned to the other historical data. Time series data of 5 years of BSE 30 Index were collected from [24] and presented to RBFN after preprocessing using WMA technique. RBFN were trained and validated using popular K-fold cross validation technique [14] to strengthen the prediction model. Model was measured using well known measures and found to be satisfactory. The rest of the part of paper is organized in 4 different sections. Section 3 explains about data preprocessing, section 4 elaborate about RBFN technique used for stock price prediction, section 5 briefly explain about experimental work done using MATLAB software and at last the work has been concluded.

II.PROCESS FLOW

A process flow diagram of entire research work is shown in Figure 1 which consists various blocks representing steps during model building process for stock price prediction. As a first step, stock index data of BSE30 consisting four features open, low, high and close obtained from [24] was preprocessed using sliding window [21] with WMA technique to produce various time series data as 5WMA, 10WMA, 15WMA and 20 WMA. In the next step data were presented one by one to RBFN [20] and trained and tested using k-fold cross validation technique [22] and finally model is evaluated using different error measures like Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE) and Mean Square Error (MSE).

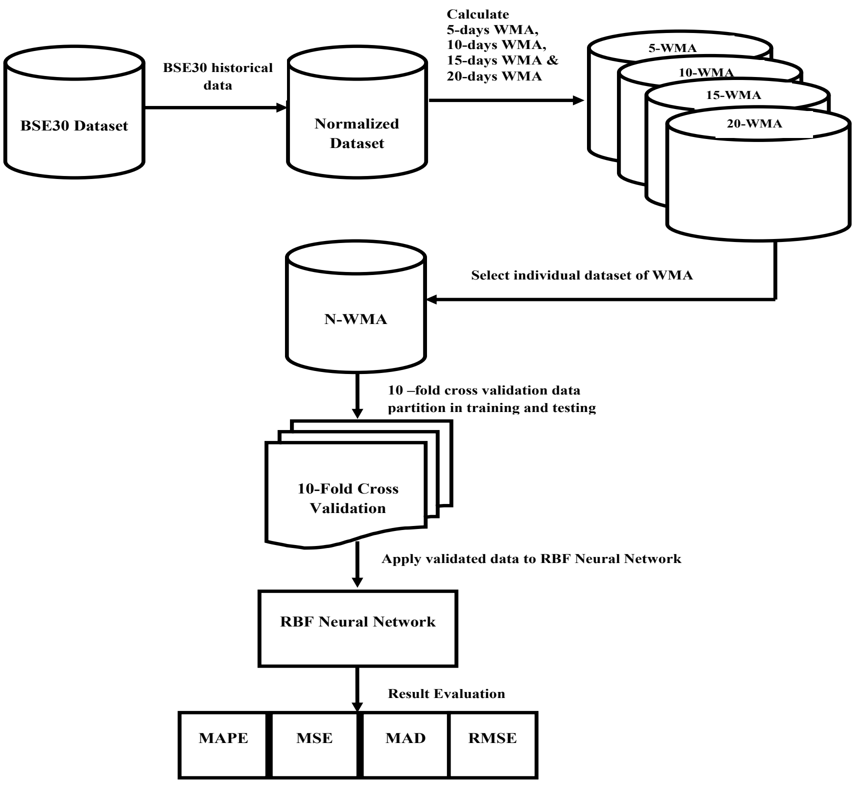


Figure 1: Process flow diagram of proposed work

III. DATA PREPROCESSING

Data preprocessing is a technique to transform the raw data into some meaningful and understandable format. When raw data is collected from specific sources then preprocessing of data is require to produce accurate prediction using neural network [2]. Normalization is widely used preprocessing technique for data smoothing. Equation 1 is used to normalize the dataset, which scales all data in range of [0 1]:

$$X_{new} = \frac{X}{X_{max}} \qquad \ldots(1)$$

Where x is daily observation of time series data obtained from [24] comprises of opening, highest, lowest and closing price of the day of 5 years from December 2010 to November 2016 and X_{max} is highest value of observation of a particular feature while X_{new} is obtained normalize observation.

A. Sliding Window

Sliding Window is a temporary approximation over the actual value of the time series data [3]. The size of the window and segment increases until we reached the less error approximation [9]. After selecting the first segment, the next segment is selected [11] from the end of the first segment. The process is repeated until all time series data are segmented. The process of sliding window is shown in Figure 2 with window size=5. Sliding window accumulate the historical time series data [21] to predict next day close price of stock. Figure 2 shows process of sliding window with window size=5. Each number (1, 2, 3…..10) represents daily observation of time series data of day 1, 2, 3….10 respectively. Initially window has covered from 1 to 5 which represents that 5 days historical data are being used for prediction of next day close price, then window slides right side by one day to cover another 5 days (from 2 to 6) observations to predict next day close price . The process will be continued till time series data of a particular time period considered for experimental purpose is exhausted, in this manner we have retrieved 1166 observations from total of 1171 observations and the result will be a new time series data calculated through WMA technique with window size 5, same process is applied for window size 10, 15 and 20.

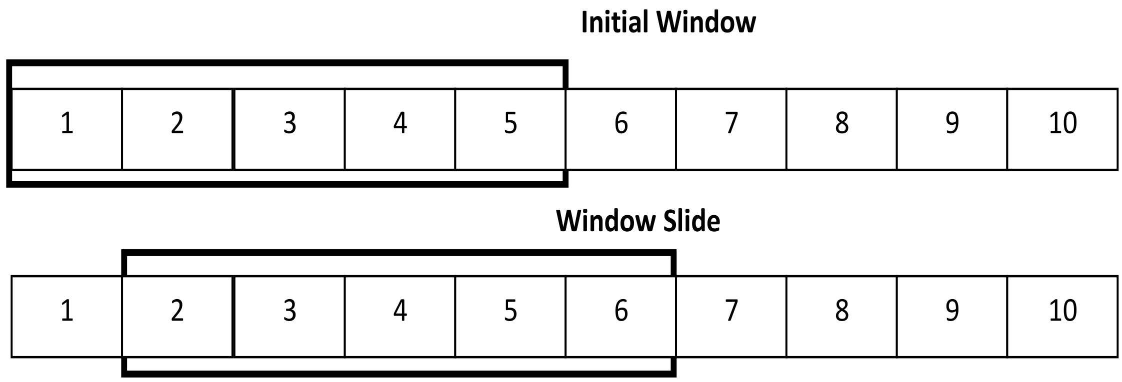


Figure 2: Process of sliding window.

B. Weighted Moving Average (WMA)

Moving average (MA) is a technique to smooth out the time series data by removing noise from it by calculating the average price over a specific time period, the time period can be 10 days, 10 weeks, 10 months or any time period chosen by the investor, there are three types of Moving Averages can be used: Simple Moving Average (SMA), Weighted Moving Average (WMA) and Exponential Moving Average (EMA). In this paper we have used WMA technique for data preprocessing. WMA takes the average of several periods of data using weights [5]. It is usually gives more importance on some periods then others. The most recent data gets the highest weight and each value of data set gets the smaller weight as we count backward direction in time series data [16]. WMA is calculated by taking each daily observation of time series data over a time period and multiply them by certain position in data series, ones the positions of time period has been counted then they have summed together and divided by the summation of number of time period as given in equation 2.

$$F_t = \frac{\sum_{i=1}^{n} W_i A_{t-1}}{\sum_{i=1}^{n} W_i} \qquad \ldots(2)$$

Where F_t= Prediction for coming period, W_i= the weight to be given to the actual occurrence for the period t-i. A_i is the actual occurrence for the period t-i and n is the total number of periods in prediction.

C. K-fold cross validation

Time series data obtained as per sliding window technique of WMA are 5WMA, 10WMA, 15WMA and 20WMA were dynamically partitioned in training and testing data sets using k-fold cross validation [15]. Cross validation is the method that is better than static method of partitioning the data. Static data partition with fixed percentage of training and testing data may bias ANN and may have problem of network paralysis. Also training and testing data sets may or may not contain non-linear data patterns of time series data. On the other hand dynamic partitioning of data as training and testing changes the fold dynamically. In k-fold cross validation the data set is divided into k subsets. Each time, one of the k subset is used as the test set and the other k-1 subsets are put together to form a training set. Then the average error across all k trials is computed and in this manner each fold takes part in training and testing both. The advantage k-fold cross validation is that size of each test set is independently chosen. A process of k-fold cross validation is shown in Figure 3.

Figure 3: Process of k-fold cross validation

IV. RADIAL BASIS FUNCTION NETWORK (RBFN)

Radial basis function network is an artificial neural network [17, 23] that uses radial basis functions as activation functions [18, 12]. The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters. Radial basis function networks have many uses, including function approximation, time series prediction [1,4], classification, and system control. Radial Basis Function Network (RBFN) is characterized by transfer function in the hidden layer [13] which has radial symmetry with respect to center. RBFN provides the possibility of learning the weights efficiently [10] without local minima problem .RBFN is a three layer architecture, first layer is called Input layer where source node is given, second layer is called hidden layer in which each neuron computes its output using radial basis function and this output is send to the third layer called output layer [19]. The architecture of RBFN is shown in Figure 4.

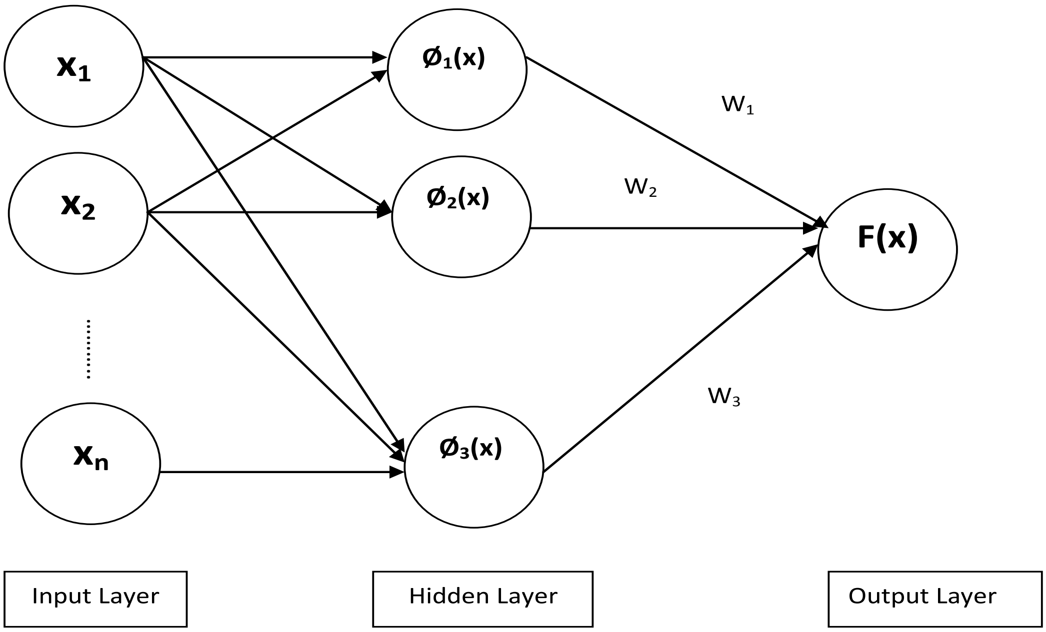


Figure 4: A simple architecture of RBFN

The most general formula for radial basis function is:

$$F(x, \emptyset, w) = \sum_{i-1}^{m} \emptyset_i(x).w_i \qquad \ldots(3)$$

Ø $\emptyset$ is the Gaussian function that is used in radial basis function and w_i is the associated weight for every Radial Basis Function. Ø can be calculated by following expression in Equation 4:

$$\emptyset(x, c, r) = exp\left(\frac{\|x-c\|}{r}\right) \qquad \ldots(4)$$

Where c is center point of function ø and r is its radius and x is the input vector.

V. RESULT ANALYSIS

Experimental work for data preprocessing and prediction [24] through RBFN is done by writing MATLAB program under window 7 environment. A predefined formula of RBFN as newrb() shown below in equation 5 is used to add neurons to the hidden layer and to simulate the work of a radial basis network until it meets the specified mean squared error goal.

net = newrb (P, T, goal, spread, MN, DF) ...(5)

Above function takes matrices of input and target vectors P and T respectively, and parameters goal with value 0.001 and spread as radius of the RBFN. The large value of spread smoother the function approximation and too small value of spread means many neurons are required to fit approximation function, simulation result of above function returns RBF network as net as shown in Figure 5 with randomly selected 200 neurons at hidden layer.

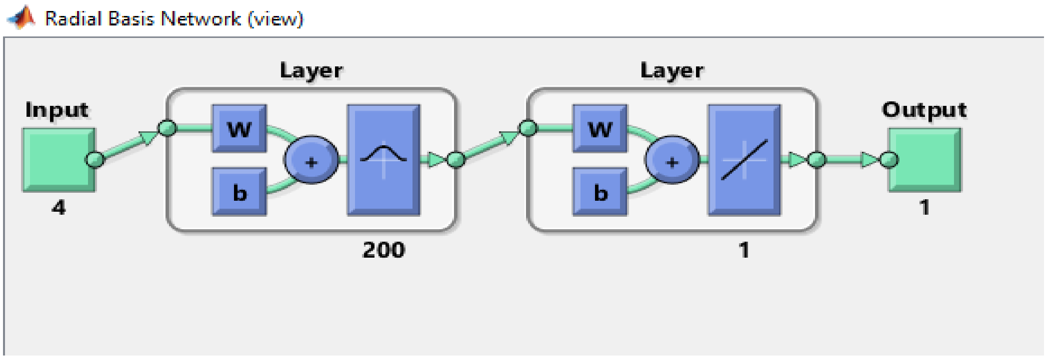


Figure 5: MATLAB generated view of RBFN through newrb() function of MATLAB.

The preprocessed data obtained through process, explained in section III were used to train and validate the RBFN model as per 10-fold cross validation technique explained in section III C. The predicted next day close price from RBFN were compared with actual next day close price in term of error measures RMSE, MAD, MAPE and MSE.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(Y_{a,i} - Y_{p,i})^2} \qquad \ldots(6)$$

$$\mathrm{MAD} = \frac{\sum_{i=0}^{n} |Y_{a,i} - Y_{p,i}|}{n} \qquad \ldots(7)$$

$$\mathrm{MAPE} = \frac{\sum_{i=0}^{n} |Y_{a,i} - Y_{p,i}|}{n} \qquad \ldots(8)$$

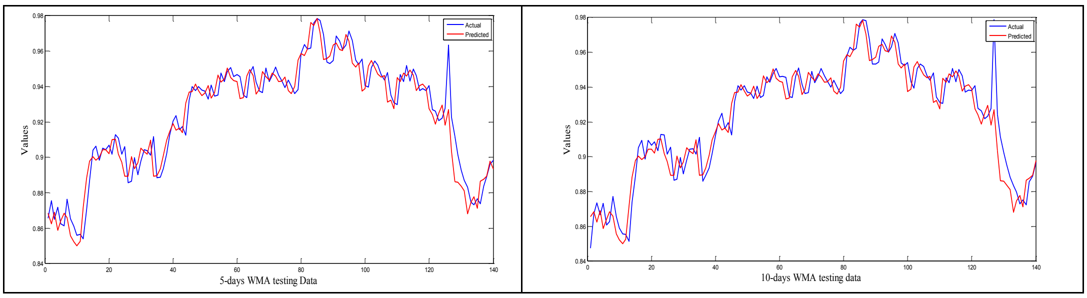
$$\mathrm{MSE} = \frac{\sum_{i=0}^{n} |Y_{a,i} - Y_{p,i}|^2}{n} \qquad \ldots(9)$$

Where Y_a=Actual observation, Y_p= Predicted observation and n=Total number of observations.

The simulated result through MATLAB code is shown in Table 1 which clearly reflects the impact of data preprocessing using sliding window based WMA. Table show the result in terms of error measures calculated using equation 6, 7, 8 and 9, however MAPE is more capable to reflect performance of any predictive model in more explanative way as compare to other error measures. If other measures are not consistent, then MAPE is considering as standard of measure. If we analyze data of above table, then MAPE is being decreased while increasing window size, say for example MAPE=0.8069 for 5WMA while it is 0.8062, 0.8054, 0.8046 for 10WMA, 15WMA and 20WMA respectively at training stage. The increased size of WMA does not able to reduce MAPE more but there were slight changes in MAPE, on the other hand at testing stage MAPE is being increased while window size is increased. MSE is being decreased for both training and testing data for WMA= 5, 10, 15, 20 while window size is being increased. However MAD is almost stationary at training stage with 0.0059 and at testing stage with 0.0065. On the other hand Figure 6 shows a comparison in between actual and predicted next day close price for WMA=5, 10, 15, 20 respectively at testing stage. Comparative graph shown in figure 6 proves positive prediction trends of RBFN.

Table 1: Comparative error measures: MAPE, MSE, MAD and RMSE for WMA=5,10,15 and 20.

WMA	MAPE		MSE		MAD		RMSE	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
5 WMA	0.8069	0.8776	5.99E-05	9.09E-05	0.0059	0.0065	0.0077	0.0091
10WMA	0.8062	0.8851	5.97E-05	8.67E-05	0.0059	0.0065	0.0059	0.0067
15WMA	0.8054	0.8834	5.96E-05	8.24E-05	0.0059	0.0065	0.0077	0.0088
20WMA	0.8046	0.8877	5.96E-05	8.18E-05	0.0059	0.0066	0.0077	0.0088



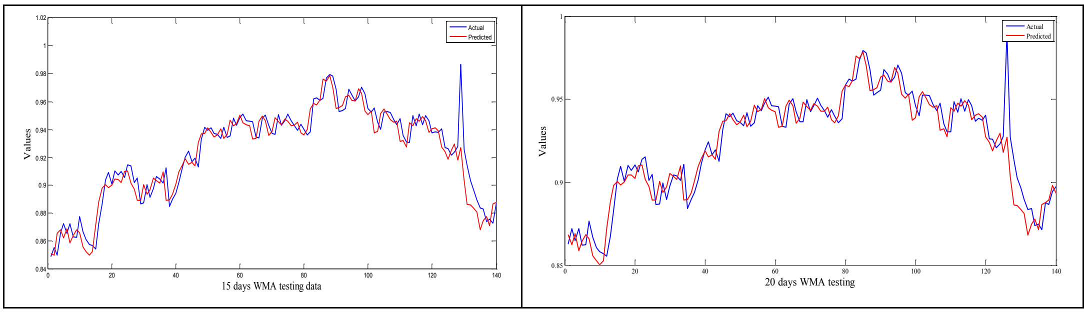


Figure 6: Comparative graph of actual and predicted value during testing for WMA=5 (Upper left) , 10 (Upper right), 15 (Lower left), 20(Lower right).

VI. CONCLUSION

Due to non-linear behavior of stock data, it is very difficult to predict it using conventional techniques. This paper present a model for next day close prediction of time series data based on the concept of sliding window and WMA as data preprocessing, 10-fold cross validation was used to train RBFN model with preprocessed data for accurate prediction. The performance of the model is measured through different accuracy measures like MAPE, MAD, MSE and RMSE. The empirical result show that results at testing stage are as per expectation but same is not true in case of testing data. In future new prediction models can be introduced with hybrid or ensemble techniques, also new features will be extracted and will be applied on models.

REFERENCES

[1] Usman,O.,L., & Alaba,O.,B. (2014). Predicting Electricity Consumption Using Radial Basis Function (RBF) Network. International Journal of Computer Science and Artificial Intelligence, 4(2), 54-62.

[2] Majhi, R.,Panda,G.,& Sahoo,G. (2009). Development and performance evaluation of FLANN based model for forecasting of stock market. Expert sytem with Applications, 36, 6800-6808.

[3] Yahmed,Y.B., Bakar.A.a., RazakHamdan,A., Ahmed, A., & Abdullah, S.M.S.(2015). Adaptive sliding window algorithm for weather data segmentation. Journal of Theoretical and Applied Information Technology, 80(2), 322-333.

[4] Yu,Y., Zhu,Y., Li,S., & Wan,D.(2014). Time Series Outlier Detection Based on Sliding Window Prediction. Mathematical problems in Engineering, 2014, 1-14.

[5] Leonel, A. L., Ricardo A.S. F., & Guilherme G. L. (2015). Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks. Applied Soft Computing, 35, 66-74.

[6] Yeh, C., Lien,C. & Tsai, Y.(2011) . Evaluation approach to stock trading system using evolutionary computation. Expert Systems with Applications, 38,794-803.

[7] Chang, P., Wang, D. & Zhou, C. (1985). Fuzzy identification of systems and its application to modeling and control. IEEE transactions on systems, man, and Cybernetics, 15(1), 116-032.

[8] Uykan, Z. ,Guzelis, C. & Celebi, M.E. (2000). Analysis of Input-Output clustering for determining centers of RBFN. IEEE Transaction of neural network, 11(4), 851-858.

[9] Mozaffari, L.,Moxaffari, A. & Azad, N.L.(2015). Vehical speed prediction via a sliding-window time series analysis and an evolutionary least learning machine: A case study on San Francisco urban roads.. Engineering science and technology, an international journal, 18,150-162.

[10] Awad,M., Pomares, H., Rojas,I. , Salameh, O. & Hamdon, M. (2009). Prediction of time series using RBF neural networks: A new approach of clustering. The international Arab journal of information technology, 6(2),138-143.

[11] Vafaeipour, M., Rahbari, O., Rosen, M.A., Fazelpour, F. & Ansarirad, P. (2014). Application of sliding window technique for prediction of wind velocity time series. International journal of Energy and environmental engeering (springer), 5,105-111.

[12] Majhi, B., Rout, M. & Baghel, V. (2014). On the development and performance evaluation of a multi objective GA-based RBF adaptive model for the prediction of stock indices. Journal of king saud university computer and information sciences, 26, 319-331.

[13] Mohammadi, R., Ghomi, S.M.T.F. & Zeinali, F. (2014). A new hybrid evolutionary based RBF networks method for forecasting time series: A case study of forecasting emergency supply demand time series. Engineering Applications of Artificial Intelligence, 26, 204-214.

[14] Wong, T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognition, 31, 1-8.

[15] Jiang, P., & Chen, J. (2016). Displacement prediction of landslide based on generalized regression neural networks with K-fold cross-validation. Neurocomputing, 198(c), 40-47.

[16] Lucas, A. & Zhang, X. (2016). Score-driven exponentially weighted moving averages and Value-at-Risk forecasting. International Journal of Forecasting, 32, 293-302.

[17] Haviluddin, & Tahyudin, I. (2015). Time Series Prediction Using Radial Basis Function Neural Network. International Journal of Electrical and Computer Engineering (IJECE), 5(4), 31-37.

[18] Niu, H. & Wang,J. (2013). Financial time series prediction by a random data-time effective RBF neural network .Soft Comput.

[19] Frank, R.J., Davey, N. & Hunt, S.P. (2001). Time series prediction and neural networks. Journal of Intelligent and Robotic Systems, 31,91-103.

[20] Haviluddin & Jawahir, A. (2015). Comparing of ARIMA and RBFNN for short-term forecasting. International Journal of Advances in Intelligent Informatics, 1(1), 15-22.

[21] Yu,Y., Zhu,Y., Li, S. & Wan, D.(2014). Time Series Outlier Detection Based on Sliding Window Prediction. Hindawi Publishing Corporation, 2014.

[22] Kuo, R.J. & Li,P.S. (2016). Taiwanese export trade forecasting using firefly algorithm based K-means algorithm and SVR with wavelet transform. Computers & Industrial Engineering, 99,153-161.

[23] Feng,H.M. & Chou, H.S. (2011). Evolutional RBFNs prediction systems generation in the applications of financial time series data. Expert Systems with Applications, 38, 8285-8292.

[24] BSESN Historical prices | S&P BSE SENSEX Stock - Yahoo Finance. (n.d.). Retrieved April 28, 2017, from https://in.finance.yahoo.com/q/hp?s=%5EBSESN.

[25] Sharma, D.K., Sharma, H.P. & Hota, H.S. (2015). Future Value Prediction of US Stock Market Using ARIMA and RBFN. International Research Journal of Finance and Economics (IRJFE), 134, 136-145.

[26] Handa, R., Hota, H.S., & Tandan, S.R. (2015). Stock Market Prediction with various technical indicators using Neural Network techniques. International Journal for research in Applied Science and Engineering Technology (IJRASET), 3(4) , 604-608

[27] Sharma, H., Sharma, D.K., & Hota, H.S. (2016). A hybrid Neuro-Fuzzy Model for Foreign Exchange Rate Prediction. Academy of Accounting and Financial Studies Journal, 20 (3), 1-13.