iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** V **Month of publication:** May 2018

DOI: http://doi.org/10.22214/ijraset.2018.5220

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

An Advanced Prediction Model for Stock Market Forecasting using Soft Computing

Chuinkle Sahoo[1], Ajit kumar Pasayat[2], Sumanjit Das[3]
[1]*Department of Computer Science and Engineering, Centurion University of Technology and Management, Bhubaneswar, Odisha, India.*
[2, 3]*Assistant Professor, Department of Computer Science and Engineering, Centurion University of Technology and Management, Bhubaneswar, Odisha, India.*

Abstract: ***Big Data deal with large volume of data sets which describes exponential growth and availability of data used in both structured data and unstructured data. Big data is realization of greater business intelligence and also both in all department of engineering and medical sciences which is used to solve new problems and old problems in a better way. This paper presents Hybrid Models for forecasting open price . The data mining architecture provides a data driven model which involves sharing of data and data privacy, providing knowledge how to apply big data in real time and processing data from scattered , uncertain , incomplete , complex and dynamic data . In this paper we analyze the challenging issues in the data driven model. To predict Stock market future behavior it mostly depends on past & present financial data. The aim of this paper is to study construct and evaluate investment strategies in order to predict future stock exchange. For evaluation of future stock exchange we need past sample data . Through data mining techniques these collected data are used to evaluated & gives useful knowledge through the calculations of financial factors . another strategy is artificial intelligence which used to construct perfect decision making tree . In this paper we have proposed a hybrid model using Auto Regressive Integrated Moving Average(ARIMA) and Particle Swarm Optimization (PSO) to predict the HDFC stock data.***
Keywords: Big data , data mining , time series , stock exchange , trading , forecasting and artificial intelligence

I. INTRODUCTTION

Recent days, large quantity of data is being gathered in the data warehouse. Usually there is a huge gap from the stored data to the knowledge that could be constructed from the data. This transition won't occur automatically, that's where Data Mining comes into picture. In Exploratory Data Analysis, some initial knowledge is known about the data, but Data Mining could help in a more in-depth knowledge about the data. Seeking knowledge from massive data is one of the most desired attributes of Data Mining [1]. A stock market is a public market for companies or for people to raise money. Stock market helps companies to buy or sell their shares. The price of shares depends upon the demand and supplies of shares. This process of buying and selling of shares is called trading; only the Listed Companies are allowed to carry out trading. Forecasting is the process of making projection about future performance based on existing historical time series data. An accurate forecast helps in decision making and planning for the future. Forecasting empower people to modify current variables in the present to predict future to result in favorable scenario.The selection and implementation of a proper forecasting methodology has always been an important planning and control issue for most firms and agencies. The organizational and financial stability of an organization depends on the accuracy of the forecast since such information will most likely be used to make key decisions in the areas of human resources, purchasing, marketing, planning and development of any organization and firms [4].

Stock market prediction is the process of trying to determine the future stock value of a company. The successful prediction of a stock's future price could yield significant profit. Stock price movements are governed by the theories random walk hypothesis and efficient-market hypothesis [1] [2]. The Forecasters of stock market focus on developing approaches which successfully forecast/predict stock prices using well defined trading strategies. A successful prediction model is the one which works with best accuracy having minimum input requirements and least complex model. Investors and government organizations rely on forecasting tools to guard against risks and to monitor market actuations. For researchers, these serve as a reference for studies of financial issues like pricing of financial derivatives and portfolio selection. Stock market forecasters' focus on developing approaches to successfully predict index values or stock prices, the main objective is at high profits using well defined trading strategies. The central idea to successful stock market prediction is achieving best results using minimum required input data and the least complex stock.

II. LITERATURE REVIEW

[5] Two widely used techniques in Time Series Forecasting are statistical methods and computational methods. Some popular time series forecasting methods which are well known are exponential smoothing, autoregressive model such as ARIMA and MARMA and ARCH and GARCH. The future value of a variable is assumed as a linear function of several past observation and random errors. [6]. Authors like Arshad et al utilized Box Jenkins univariate ARIMA model to predict the short run monthly price of CPO. To extract noise from of WT datas divided into two categories i.e. 80% for training set & 20% for test or validation . In WDBP we first predict the close of CRO in five days , the error increased for each day ahead .i.e. 5th day error is 0.5 greater than 1st day in terms of RMSE . To overcome from this difficulty Author forecasting in direction in change of stock price of Brazilian Oil Company Petrobar by means of ANNs . This is very suitable predictions for closing price used construction of neural models based on MLP to predict behaviour of PETR4 closing price in stock companies .This methodology validates the ANNs with data from JAN 2012 to NOV 2012 where MAPE=26.47% . Analysis of time series focuses on archiving dependency relationship between their historical data . Sequence of data specified at regular time intervals during a period is done which determines a structure & pattern & develope a model that predicts their behaviour normally . They are also known as regression models . we highlight AR , MA , ARMA statistical models used to represent stationary time series . This consists of constant mean & variance along with time . For mean stationary time series the ARIMA model emerged & it can be considered as generalization of ARMA . Here the polynomial AR model is a unit root model . Here we used dataset Crossed Validation. Here Wevlet Denoising Back Propagation neural network (WDBP) is used for prediction of month's closing price (Shanghai Composite Index) . This WDBP result compared to Convential Back Propagation Training Algorithm for MLPs . Frequency Decomposition characteristics of wevlet transform (WT) extract noise from data set . Data taken from Jan 1993 to Dec 2009. 80% data has taken for training & 20% for test . After validation WDBP presented MAPE=19.48% while MLP with Convential BackPropagation Algorithm has MAPE= 24.92% . In this paper the system used for closing price prediction is Adaptive Network based Fuzzy Inference System(ANFIS) . Objective is to predict closing price of Zagreb Stock Exchange Index Corbex (CRO) by using historical data fro beginning of Nov 2010 to end of Jan 2012 . Here the o/p we get i.e. 5th day RMSE is 0.5 greater than 1st day i.e. day by day error increases . To overcome from this problem he changed the direction to Brazilian oil companies Petrobars (PETR4) which gives more accurate result having less MAPE=26.47% . ANN with MLP architecture was employed in this work with Matlab toolbox & optimized by Iterative Process . The predicted result compared to Convential Back Propagation Algorithm . Maximum day stock prices were lower than 0.9% based on MAPE & Minimum day stock prices lower than 2.1% .This result is only possible due to the combined use of attribute selection based on correlation analysis & MLP analysis . We should select right attribute that should composed the database . Another important aspect is the chronological organization of the data base because time factor is crucial . In Levenberg – Marqdut algorithm the preparation & preliminary analysis of data can be concider effective methodology for estimating the range of variation . [3]

II. METHODOLOGY

A. Autoregressive Integrated Moving Average Model (ARIMA)

In statistics and econometrics ,and in particular in time series analysis , an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. Both of these models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting) . ARIMA models are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied one or more times to eliminate the non-stationarity .

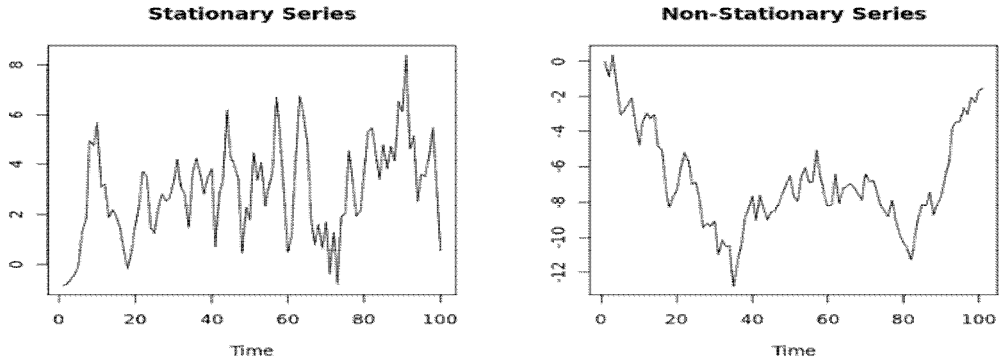
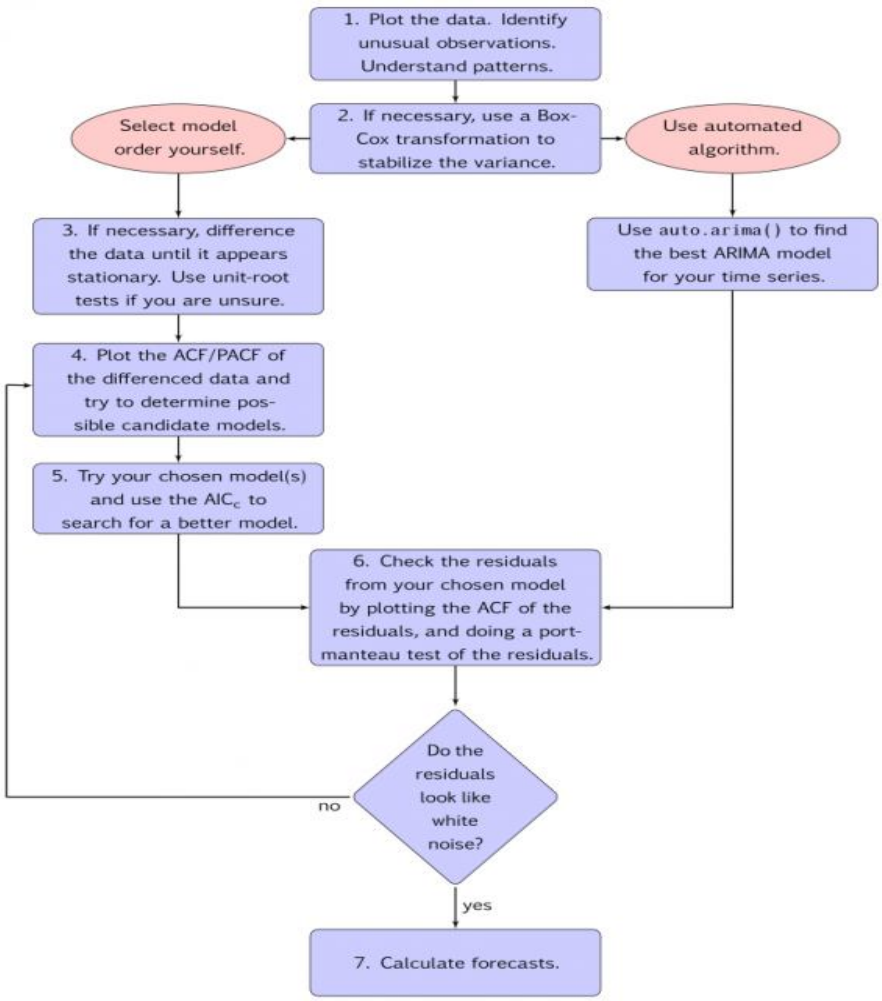


Figure-3.1: Stationary Series and Non Stationary Series

B. Working Step of ARIMA Algorithm

1) Step 1 : Identify and interpret a non-seasonal RIMA mode
2) Step 2 : Distinguish ARIMA terms from simultaneously exploring an ACF and PACF
3) Step 3 : Test that all residual autocorrelations are zero
4) Step 4 : Convert ARIMA models to infinite order MA models
5) Step 5 : Forecast with ARIMA models
6) Step 6 : Create and interpret prediction intervals for forecasts

C. ARIMA Flow Chart



D. Particle swarm optimization (PSO)

In PSO each element flies throughout the multidimensional search gap and regulates its location in each step until it reaches an optimum solution. In this technique each particle has some fixed distance from the food source and the fitness value of each particle becomes the output. Commencing the importance of fitness, the finest value is considered as particle best (Pbest). Then all particle moves in the direction of Pbest particle by changing their velocity. The velocity and location of particles in iterations are updated. From that particle best (Pbest) the global best (gbest) value is determined.

1) Functional Modified PSO

Public Parameters:

c_1& c_2: Acceleration constant

r_1& r_2: random numbers

x: vectors of n random variable.
Loc: local best value
gbest: global best value.
Pbest: Particle best.

2) Working Process

Step:1 Initialize the swarm particle in the seek space randomly.
Step:2 Compute the fitness value by using objective function and consider it as Pbest.
Step:3 update the rapidity and the location for each element. Velocity of each particle is updated by using the following equation.

$$V_{jk} = w\left(V_{jk} + c_1 \times r_1\left(Loc_{jk} - x_{jk}\right)\right) + \left(c_2 \times r_2\left(g_{best} - x_{jk}\right)\right) \quad (3.1)$$

The acceleration coefficients c_1 and c_2 are considered as constant value to influence particle velocity.
Location of each particle is updated by using the equation

$$Xjk = Xjk + Vjk \quad (3.2)$$

Step: 4 update the value of p_{best} and g_{best}.
Step: 5 stop if max iteration is accomplished or else repeat from step 2.

3) Data Validation (m-fold Cross Validation): In m-fold cross validation total dataset is split randomly into m mutually special folds of equal size, (Set_1, Set_2….Set_m). To calculate the cross validation (CV) accuracy can be represented as

Cross Validation $(Cv) = {}^{1}/_{M}\left(\sum A_j\right)$ (3.3)

Where in equation (1), A_i is fold accuracy and m is the number of folds. The entire dataset is trained and tested m number of times. We used 5-fold cross validation procedure with 5 equal numbers of folds for entire dataset.

Table 1: Stock market data of HDFC 2012 to 2017

Sl No	Stock Name	No of records.	No of features	No of input attributes	Type of Attributes
1	HDFC-12	252	12	High, Low, Close and Average Price.	Numerical
2	HDFC-13	251	12	High, Low, Close and Average Price.	Numerical
3	HDFC-14	246	12	High, Low, Close and Average Price.	Numerical
4	HDFC-15	251	12	High, Low, Close and Average Price.	Numerical
5	HDFC-16	251	12	High, Low, Close and Average Price.	Numerical
6	HDFC-17	246	12	High, Low, Close and Average Price.	Numerical

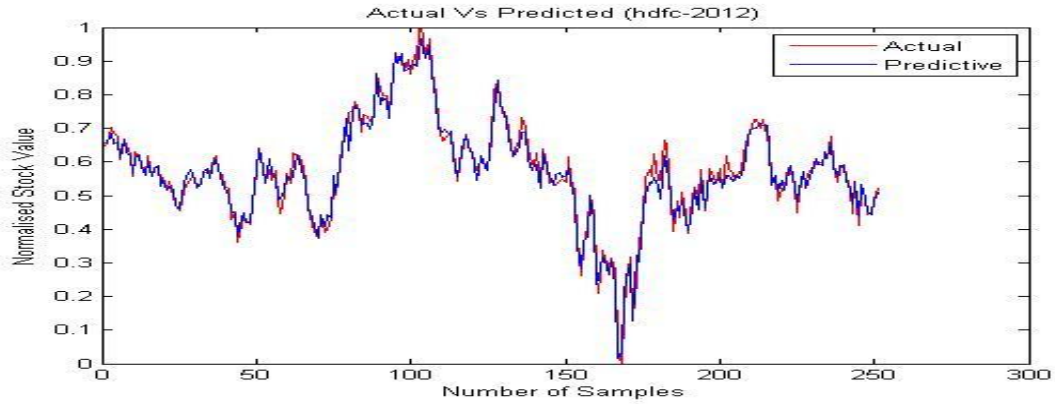
4) Working steps proposed Model:

Step 1 Collection of stock market data and variables.
Step 2 Verify the features of data collected by feature extraction.
Step 3 Normalize the data set using equation (3.6)
Step 4 Initialize the particles in sample space.
Step 5 Determine fitness value of each particle using Mean Square error as objective function.
Step 6 Training the dataset using proposed Neural Network and calculating the error
Step 7 Set the Modified PSO parameters.
Step 8 Modified PSO search of finest solution according mean Square error.
Step 9 Test the dataset after successful training of new model.

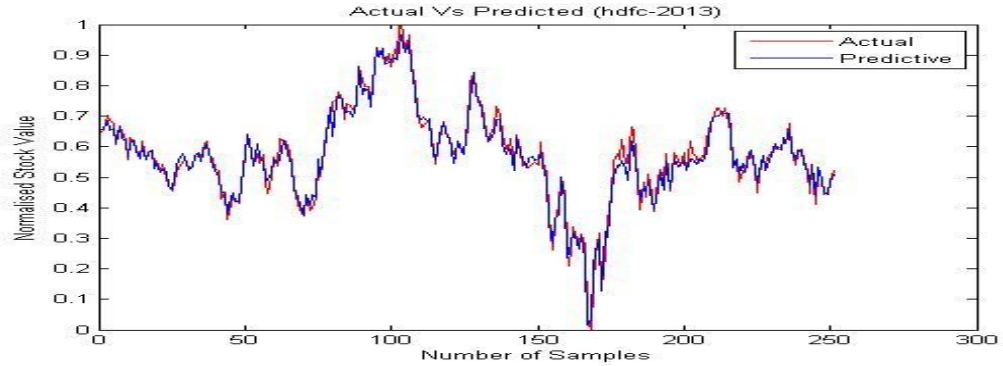
Step 10 After Successful testing the model will predict response through ARIMA objective model if the optimal solution is reached.

II. EXPERIMENTAL RESULT AND ANALYSIS

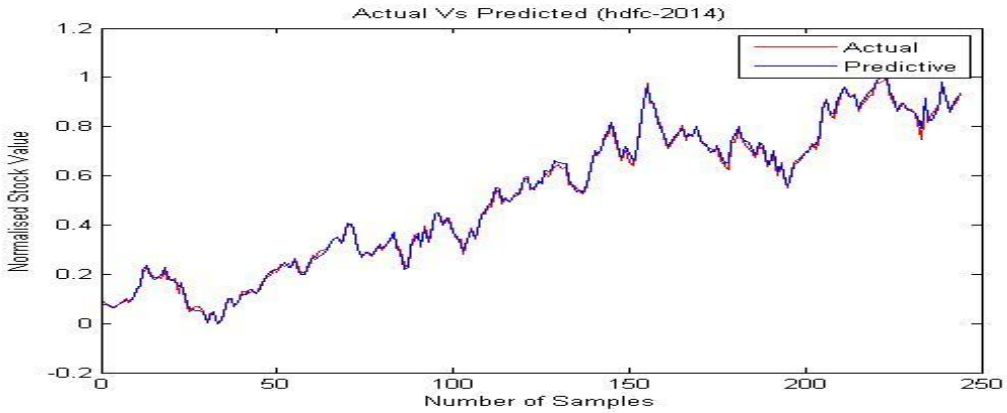
The proposed forecasting model is trained well to give reduced Mean Square Error (MSE) and reduced Mean Absolute Percentage error (MAPE) in case of stock data of HDFC as compared to other models. The error and Mean absolute percentage error are calculated using equation (3.5) & (3.8) respectively. The assessment of actual and predicted values of HDFC shown in figure . In the figure both values Actual and predicted are indistinguishable for stock data of year 2012 TO 2017. The figure-3.5 shows error in network for dataset of HDFC for the year 2012-17. The error in network almost reduces to zero as shown in figure.



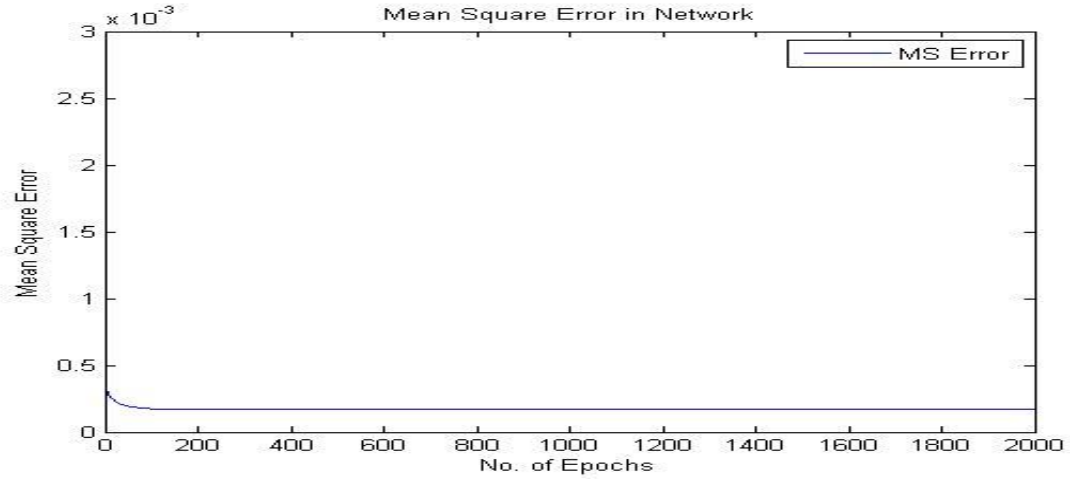
[Fig 3.2 : Actual Verses Predicted stock price of HDFC 2012]



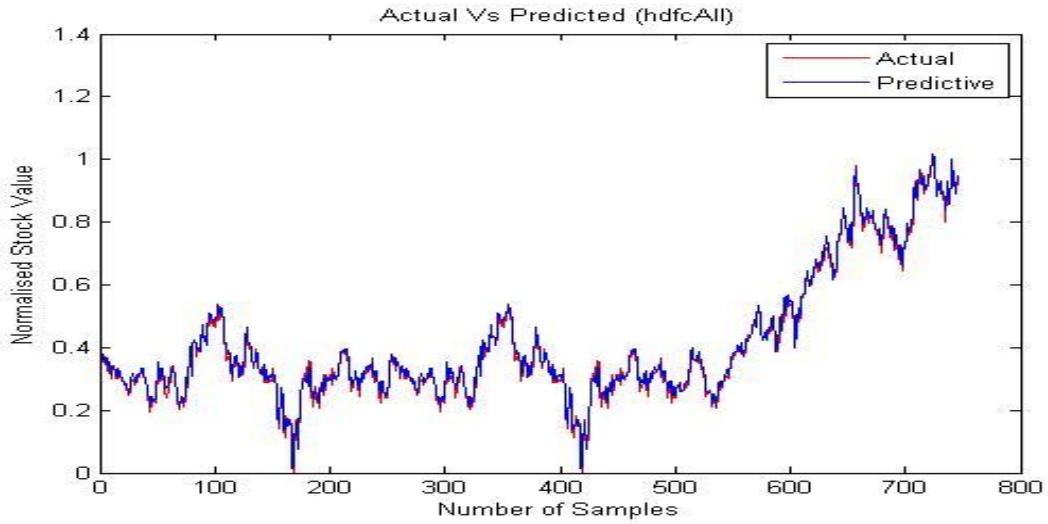
[Fig 3.3 : Actual Verses Predicted stock price of HDFC 2013]



[Fig 3.4 : Actual Verses Predicted stock price of HDFC 2014]



[Fig: 3.5Mean Square Error of Network of HDFC Entire Data set



[Fig-3.6: Actual Verses Predicted stock price of HDFC Entire dataset]

1) *Performance Evaluation:* The performance for projected replica can be evaluated by help of equations (3.8) which shows mean absolute percentage error. The proposed representation is evaluated on basis of following equations.

$$MAPE = \frac{1}{n}\sum_{1}^{n}\left(\frac{\beta i - \gamma i}{\gamma i}\right) \quad (3.8)$$

2) Comparison of Model Performance with existing approaches: The final result of actual vs predicted value for HDFC is represented in figure-3.6. The actual value represented in red lines and predicted value represented in blue lines. The actual and predicted lines are close to each other. We have compared different models of financial time series data with our proposed model. Our projected replica is comparatively enhanced than other models in terms of classification accuracy and it reduces error up to 1.1% MAPE as shown in table-2. So we strongly recommend our proposed model is best for financial time series data.

[Table-2: Comparison of performance based on MAPE]

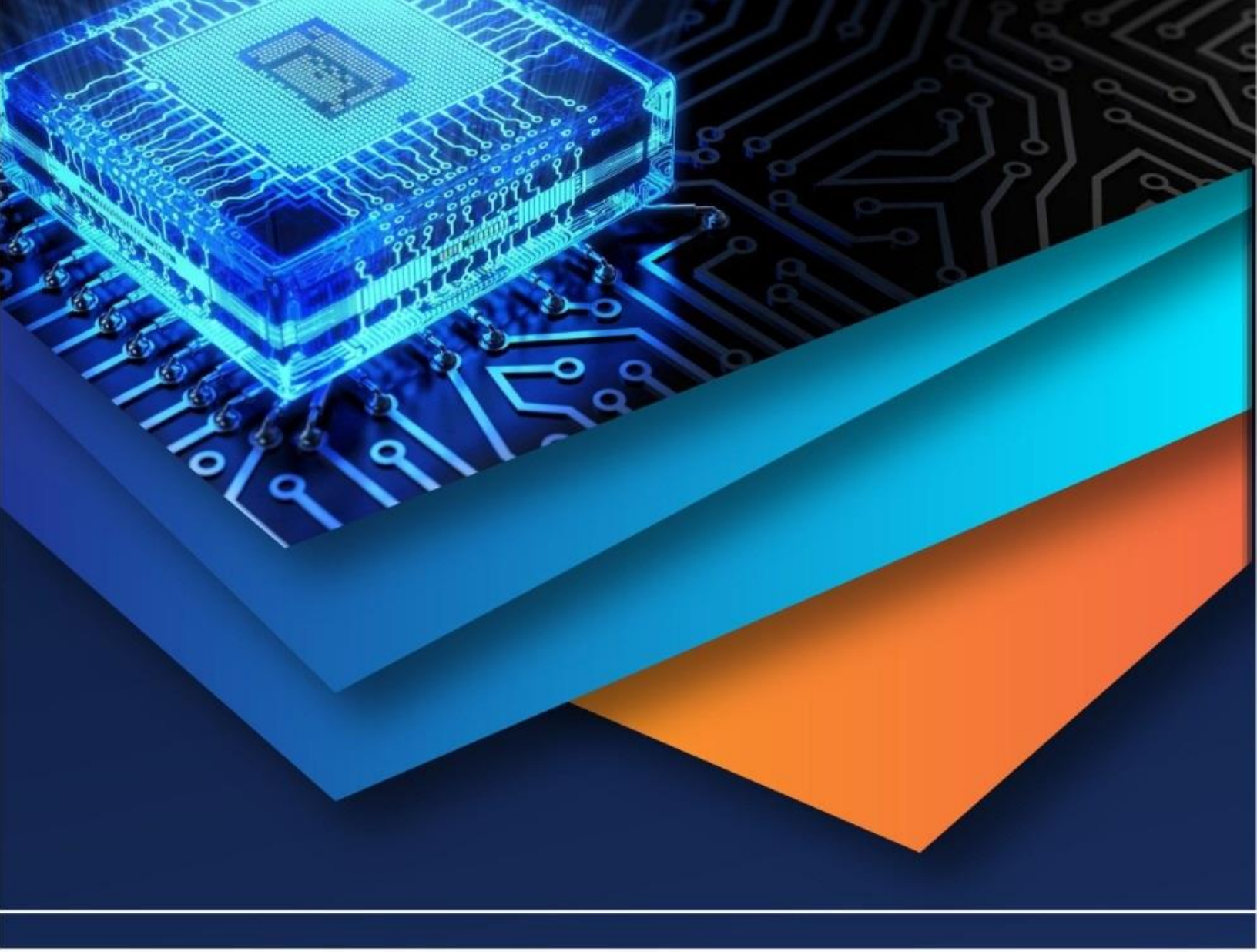
Author	Method	Classification accuracy	MAPE
[Proposed]	HAP	98.9	1.1%
Ticknor[8]	Bayesian-ANN	97.3	2.7%
Hu et al [2]	Interval Measurement	97.7	2.3%
Mukhopadhyay [3]	CMS-PSO	95.8	4.2%

V. CONCLUSION

This novel neural network approach optimized by PSO is presented which is efficient sculpts for predicting the open price of stock market and tested in real dataset of HDFC. It seems that this model is proficient in stock prediction. From the results shown in table -2, we can observe that MAPE is minimum for the proposed technique which is 1.1 for HDFC. The parameters used in this model are the optimal parameter and data input is from real world normalized data series. The network is trained, test and validated using same series of normalized data. We believe that our representation is improved than other models used for perdition of stock market open value. This experiment is executed and tested in MatLab2013a of standard processor of core2Duo 2.94 GHz; Ram 2GB and 32 bits operating system.

REFERENCES

[1] Bennett, Coleman & Co. Ltd , 2018 . Stock Market , The Economic Times [Web log message]. Retrieved from https://economictimes.indiatimes.com/definition/stock-market
[2] Gabriel Fiol-Roig, Margaret Miró-Julià, and Andreu Pere Isern-Deyà , Applying Data Mining Techniques to Stock Market Analysis , Trends in Practical Applications of Agents and Multagent System , Yves Demazeau et al. (Eds.)
[3] Leonel A. Laboissiere ,24 June 2015. Department of Electrical and Computer Engineering, University of São Paulo, Av. Trabalhador São-carlense 400, São Carlos 13566-590, SP, Brazil , Maximum and minimum stock price forecasting of Brazilian powerdistribution companies based on artificial neural networks , Applied Soft Computin
[4] Akhter Mohiuddin Rather ,11 December 2014. School of Computer and Information Sciences, University of Hyderabad, Hyderabad 500046, A.P., India , Recurrent Neural Network & a Hybrid Model for Prediction of Stock Retu
[5] Kasturi Kanchymalay1,2, N. Salim2, Anupong Sukprasert3, Ramesh Krishnan4, Ummi Raba'ah Hashim1. "Multivariate Time Series Forecasting of Crude Palm Oil Price Using Machine Learning Techniques", IOP Conf. Series: Materials Science and Engineering 226 (2017) 012117
[6] Amin Hedayati Moghaddama, Moein Hedayati Moghaddamb,∗, Morteza Esfandyaric, Stock market index prediction using artificial neural network, Journal of Economics, Finance and Administrative Science 21 (2016) 89–93



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY