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.2018.5307

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 Efficient AI Model for Financial Market Prediction Optimized by SVR

Sonalika Dash[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: ***Stock market is dynamic in nature. So prediction of the stock index accurately is a tedious task. But fortunately if the prediction will be accurate then it is easy for the traders to make decisions easily. The main challenge of the stock market is the dynamic nature which is influenced by several factors such as political, financial etc. So a robust and stable model which can give the accurate prediction is needed to be designed. In this paper we have proposed a hybrid model using Multi Layer Perceptron (MLP) optimized by Support Vector Regression Algorithm (SVR) to predict the JSPL stock data. This developed model will be evaluated and compared based on the various evaluations.***

Index Terms: Stock Market Prediction, Artificial Intelligence, Artificial Neural Network, Multi Layer Perceptron, Support Vector Regression.

I. INTRODUCTION

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] Stock market is market to trade the company's stock. Stock market has become one of the primary indicators of the economic condition of the country. Stock price prediction is very important as it used by the business people. [2] In 1991 the National Stock Exchange (NSE) of India was set up by Government of India on the recommendation of Pherwani Committee.

It was incorporated in November 1992 as a tax-paying company. Then it was recognized as a stock exchange in April 1993 under the Securities Contracts (Regulation) Act, 1956. In the past many researchers have applied different statistical and soft computing techniques to predict the movements of different stock indices like NASDAQ, Dow Jones Industrial Average (DJIA) index, Standard & Poor (S&P) 500 etc.

[3] Support vector Regression (SVR) has a characteristic that instead of minimizing the observed training error, SVR tries to reduce the generalized error bound so as to achieve generalized performance. [4] This generalization error bound is the combination of the training error and a regularization term that controls the complexity of the hypothesis space [5][6]. In this paper MLP is optimized by SVR.

II. LITERATURE REVIEW

In this paper the researcher has done the research on the Turkish Stock market. The importance of the Turkish stock market has been increased with the establishment of the Istanbul stock Exchange. ISE has a major growth in terms of trading volume, market capitalization, no. of listed corporations and foreign investments. Hence the stock markets forecasting has been the major area of interest for most of the researchers. Among them ANN is the suitable technique for handling the non-linear, highly complex, and dynamic data of stock markets.

The authors have used six different ANN's which includes Multi layer Perceptron generalized feed forward to predict the ISE market index value. Hence the results showed that the for each ANN model the highest accuracies were obtained with 1 hidden layer and also ANN models give more accurate results than the ones based on moving average. [6] The rest of the paper is as organized below. In Section III we have described the working flow of the MLP and SVR and the process for retrieving the open price and the optimized error value. In Section IV we have the experiment and result evaluation of the used models. Finally Section V provides the conclusion of the paper.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

III. METHODOLOGY

A. Multi Layer Perceptron

A multi layer Perceptron is known as a network of simple neurons called Perceptron. Multi Layer Perceptron is commonly used for regression problems. It has 3 layers known as Input Layer, Hidden Layer and Output layer. The basic idea behind it is the Perceptron computes a single output from the multiple inputs. Multi Layer Perceptron involves the Back Propagation Algorithm which will help in calculating the error between the input and the computed output. Continuous revision of error is required till the minimum error is received.[7] A common architecture of the Multi Layer Perceptron is given below figure.

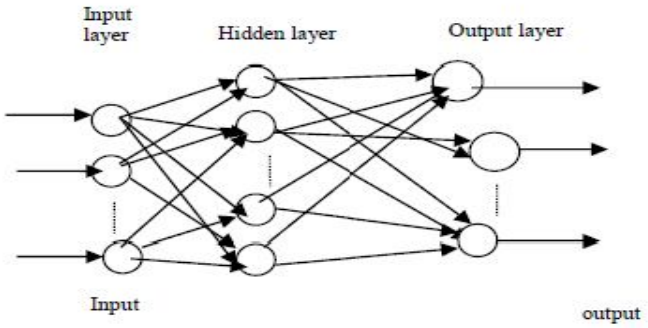


Figure 3.1: Architecture of Multi-Layer Perceptron model

We can mathematically represent the Multi Layer Perceptron as given below:

$$y = \varphi(\sum_{i=1}^{n} w_i x_i + b) = \varphi(\mathbf{w}^T \mathbf{x} + b) \quad \text{.....(3.1)}$$

where ,

w denotes the vector of weights
X denotes the vector of input
b denotes the bias and
ϑ denotes as activation function.

A typical Multi Layer Perceptron consists of set of source nodes as Input layers, more than one hidden layers and an output layer. The input signal flows through the network layer by layer. The signal flow graph for the MLP networks with one Hidden Layer as shown in Figure 2 [6]

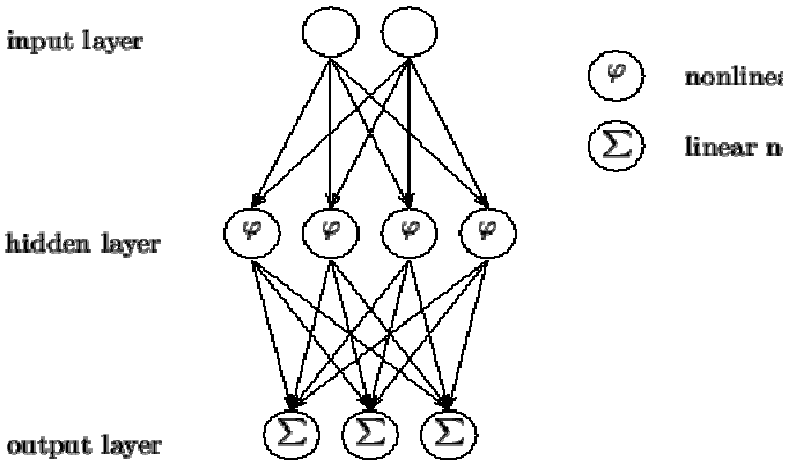


Figure 3.2: Signal-flow graph of an MLP

MLP can solve the supervised learning problem using the Back Propagation algorithm. The algorithm consists of two steps. One is the predicted outputs correspond to the given inputs are evaluated in the forward pass. And the second one is the partial derivatives of the cost function with respect to the different parameters are propagated back through the network in the backward pass.[6]

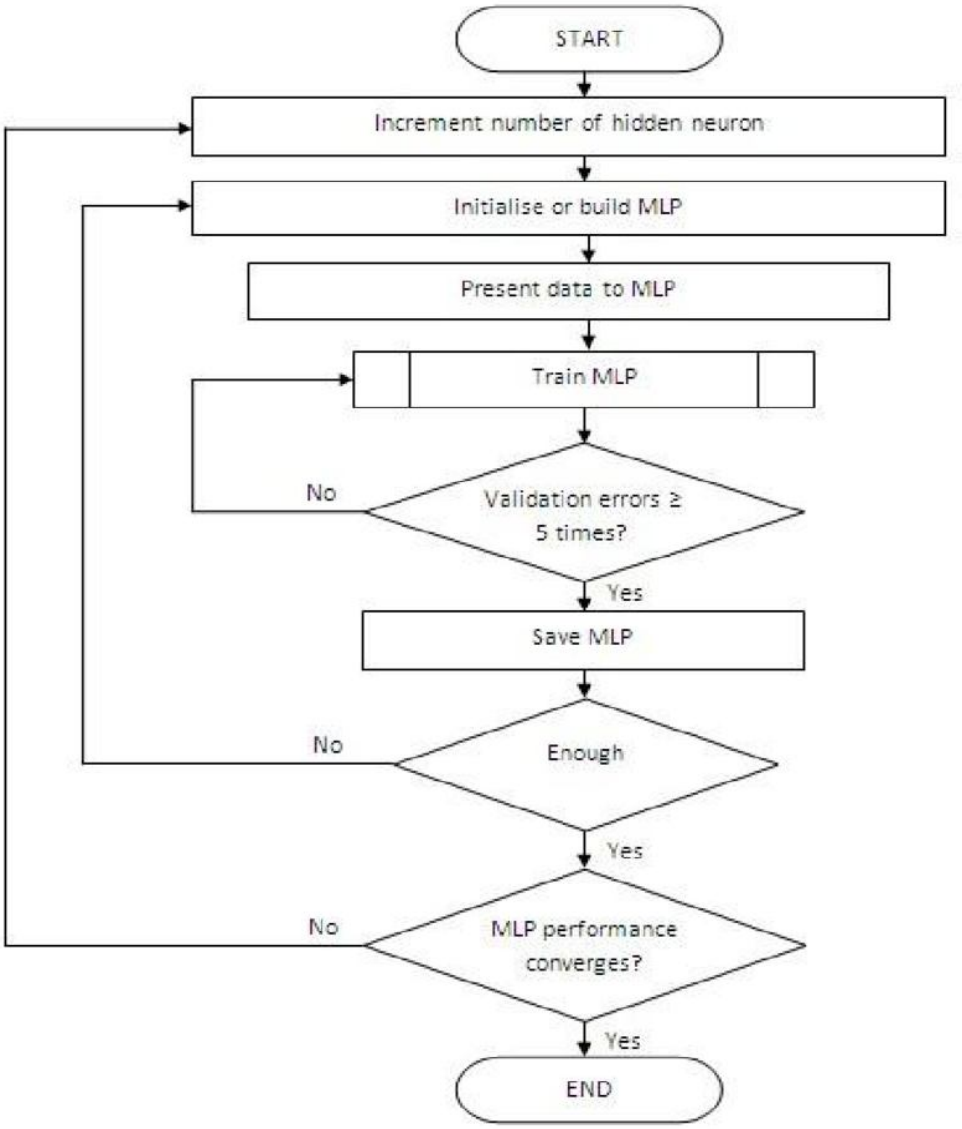


Fig 3.3: Flowchart of MLP training process

B. Support Vector Regression

Support Vector Regression principle is same as the Support Vector Machine. We can also say that Support Vector Regression is derived from Support Vector Machine. But the dependent variable in Support Vector Regression is numerical rather than categorical.[7] Support Vector Regression is technically non-parametric in nature. The SVR results never depend on the underlying dependent and independent variable as like other techniques. There are several Kernels are available. Selecting the appropriate kernel function is really an important task. Some commonly used functions are polynomial and radial basis. It can also use various kernel functions till we reach the maximum accuracy level. [8] SVR also permits for developing a non-linear model without changing the explanatory variables. Also helps in better interpretation of the model. The important benefit of SVR is it does not care about the prediction as long as the error is less than certain value, according to the principle of maximal of margin. [8] Linear SVR can be defined as a method of dividing a set of data using a linear classifier in a maximal margin of Hyper plane. Hence, the Linear SVR equation can be represented as below.[8]

$$y = \sum_{i=1}^{N} (a_i - a_i^*) \cdot \langle x_i, x \rangle + b$$ ….....(3.2)

Non linear SVR can be defined as a method of dividing the set of data in a linear form with the help of appropriate Kernel function. Hence, the Non-Linear SVR equation can be represented as below.[18]

$$y = \sum_{i=1}^{N} (a_i - a_i^*) \cdot \langle \varphi(x_i), \varphi(x) \rangle + b$$

$$y = \sum_{i=1}^{N} (a_i - a_i^*) \cdot K(x_i, x) + b$$ ….....(3.3)

C. Proposed Hybrid Prediction Model

The proposed Prediction model works on MLP as basic network and optimized by SVR to give better result. The model works on normalized time series data. The data processing is required before the model training and testing to remove unwanted data.

D. Data Preprocessing And Normalization

1) *Data Processing:* It is a very important step in many applications where data cleansing and transformation are applied to improve the results. MLP must be needed to be considered as sensitive to the form of data preprocessing. Some well-defined methods are developed for the preprocessing. Those preprocessing can be applied to the cluster and classification process. Crone et al. analyzed the different forms of data preprocessing on classification accuracy. Preprocessing the data is plays an important role in MLP. [4] Pre-processing leads to high accuracy and less computational cost. Pre-processing helps in modifying and reducing the data set size, removing a typical and noisy training samples and correcting possible erroneous training sample. [10] Data collecting methods are less controlled resulting in outliers, variety data combinations and missing values. Performing the data analysis on such type of erroneous data is really confusing. [9]
2) *Data Set Normalization:* Normalization is a process of scaling up the dataset. It's a very crucial step when the when there is a

$$Y = (Y - Y_{min}) / (Y_{max} - Y_{min})$$ .....(3.4)

collection of various datasets of different units and different scales. Normalization process scales up all the numerical data in arrange of [0,1]. [8] Various types of Normalization process are available like Min-Max Normalization, Z-score Normalization, and Decimal Scale Normalization. It helps in the predicting and forecasting. Basically there are various steps of data normalization available for the neural network. First process is the simple linear scaling process. In this process at the very least data must be scaled up to the input range of the neural network. The range is basically from -1 to 1 or 0 to 1. Assuming the values D_{min} and D_{max} respectively. The input range for the neural network is defined as I_{min} and I_{max}. Hence the equation is described below. In the proposed model stock market data is normalized using equation (3.4) can be stated as:

[Table-1 stock-market data Set of JSPL from 2012 to 2017]

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

IV. 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 JSPL as compared to other models. The error and Mean absolute percentage error are calculated using following equations. The assessment of actual and predicted values of JSPL shown in figure. In the figure both values Actual and predicted are indistinguishable for stock data of year 2012, to 2017. The figure-3.4 shows error in network for dataset of the network for JSPL for year 2012-17. The error in network almost reduces to zero as shown in figure.

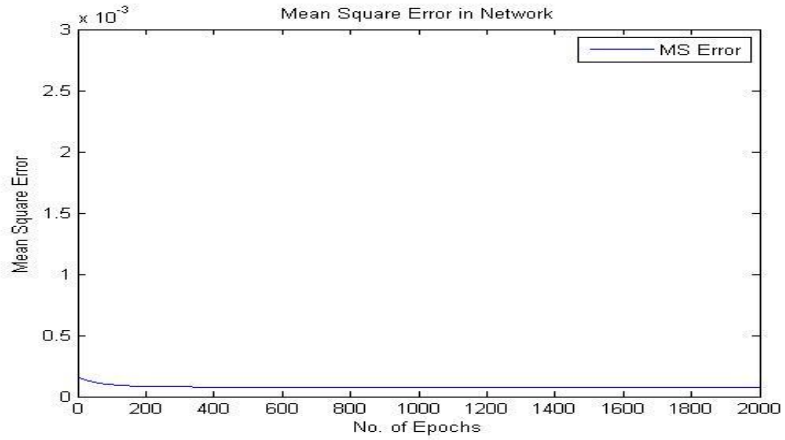


Figure-3.4: Shows the mean square error in the network

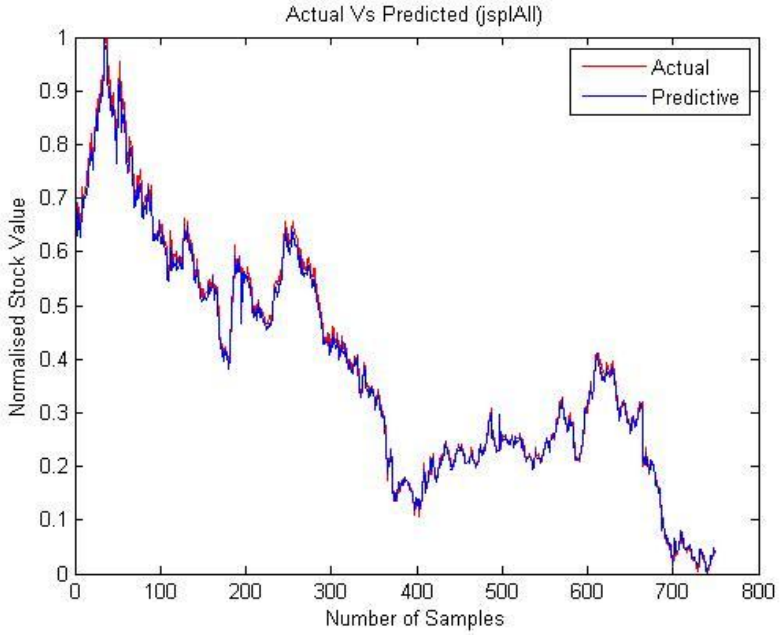


Figure: 3.5 Shows the result of the proposed Model

C. Performance evaluation

The performance for projected replica can be evaluated by help of equations (3.5) which shows mean absolute percentage error. The proposed representation is evaluated on basis of following equations. The accuracy of the proposed hybrid model is very high in comparison with MLP model so we strongly recommend that the hybrid model will be good choice to predict stick open price.

$$MAPE = \frac{1}{n}\sum_{1}^{n}\left(\frac{\beta i - \gamma i}{\gamma i}\right) \qquad (3.5)$$

Data Set	No Records	Classification Accuracy	MAPE (MLP)	MAPE (MLP-SVR)
JSPL-12	252	99.1	2.3	0.9
JSPL-13	251	98.8	2.9	1.2
JSPL-14	246	98.9	2.6	1.1
JSPL-15	251	98.8	2.9	1.2
JSPL-16	251	98.9	2.6	1.1
JSPL-17	246	98.9	2.6	1.1

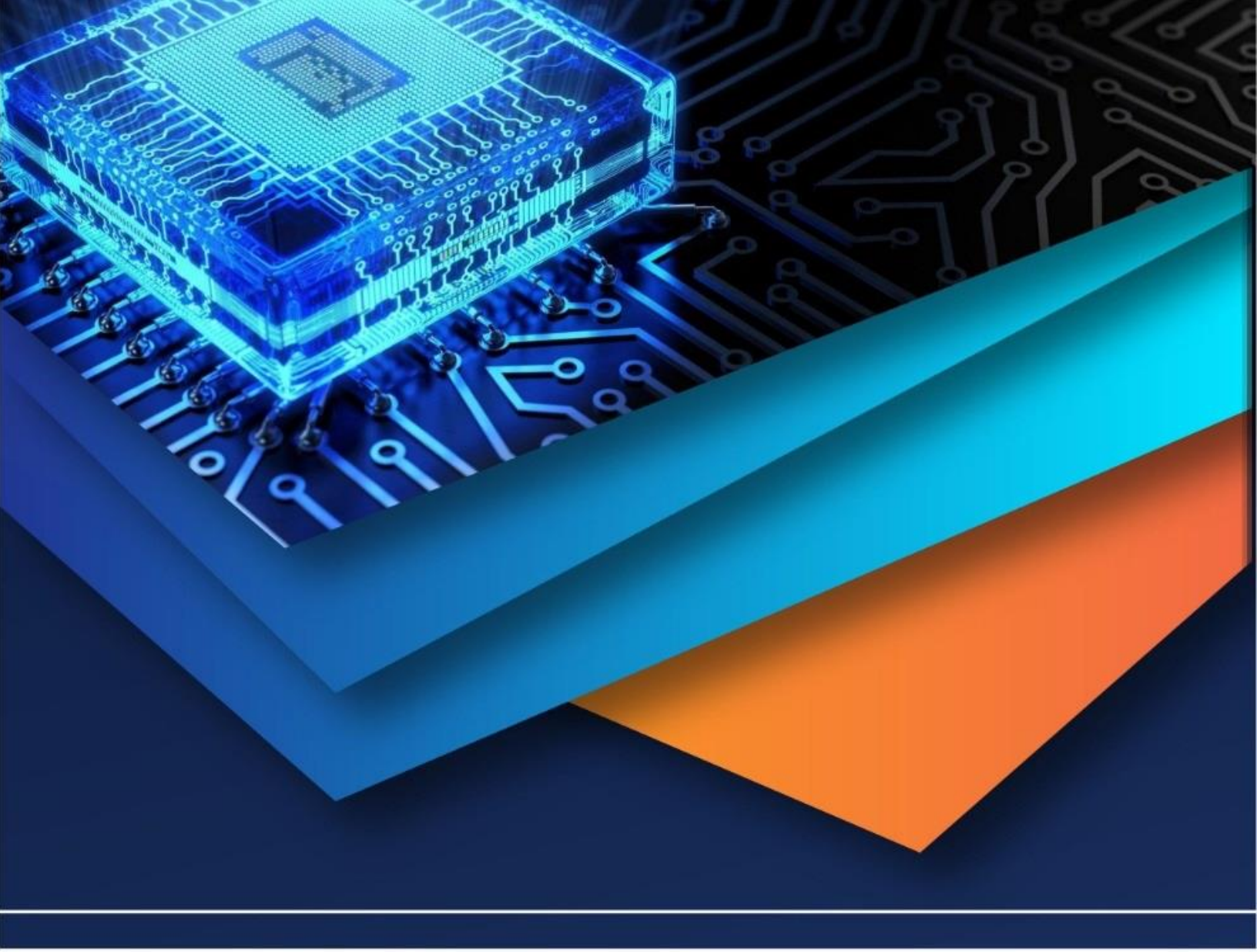
[Table-2: Represents the result obtained from the proposed model and MLP]

V. CONCLUSION

This novel neural network approach optimized by SVR is presented which is efficient sculpts for predicting the open price of stock market and tested in real dataset of JSPL. 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 1.2 for JSPL. 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] Kimberley Amadeo, What Is the Stock Market?,https://www.thebalance.com/what-is-the-stock-market-how-it-works-3305893,Updated January 10, 201
[2] Stock Markets,https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/stock-market/
[3] E. F. Fama, "The Behavior of Stock Market Prices," Business, vol. 38, pp. 34-105, 1965
[4] H.M. Huang, M. Pasquier, and C. Quek, “Financial Market Trading System With a Hierarchical Coevolutionary Fuzzy Predictive Model,” IEEE Transactions on Evolutionary Computation, vol. 13, no.1, pp. 56 - 70, 2009
[5] P. N. Refenes, A. N. Burgess, and Y. Bentz, "Neural networks in financial engineering: A study in methodology," IEEE Transactions on Neural Networks, vol. 8, no. 6, pp. 1222 - 1267, 1997.
[6] L. Cao and F. Tay, “Support Vector Machine with adaptive parameters in financial time series forecasting,” IEEE Transactions on Neural Networks, vol. 14, no. 6, pp. 1506–1518, 2003.
[7] U. Thissena, R. van. Brakela, A.P. de. Weijerb, W.J. Melssena, L.M.C. Buydensa, “Using support vector machines for time series prediction,” In: Chemometrics and Intelligent Laboratory Systems 69, pp.35– 49, 200
[8] Lijuan. Cao, “Support vector machines experts for time series forecasting,” In: Neurocomputing 51, pp. 321 – 339, 2003
[9] J. Smola. Alex, SchoLkop. Bernhard, “A tutorial on support vector regression,” In: Statistics and Computing 14, pp.199–222, 200
[10] 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