Journal of Theoretical and Applied Information Technology
15th August 2019. Vol.97. No 15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

STOCK MARKET PRICE PREDICTION SYSTEM USING NEURAL NETWORKS AND GENETIC ALGORITHM

[1]NADIA ABDUL JAWAD, [2]Dr.MHD BASSAM KURDY

[1]Research Scholar, Department of Web Sciences, Syrian Virtual University, Syria

[2]Professor, Department of Artificial Intelligence, Syrian Virtual University, Syria

E-mail: [1]nadia_84696@svuonline.org, [2]t_bkurdy@svuonline.org

ABSTRACT

Stock market forecasting of price/index has always been an important financial subject. Knowing the close price/index based on previous information is useful for investors who need to buy or sell the stock. Most of the applications are focused on building systems with less error and more accuracy. Most traders have used technical analysis tools to predict future stock market movements. Popular methods to find dynamic relationship between input and target output were artificial neural networks that proved to be effective recently. The evolutionary algorithms improve performance in predicting financial market results. This study uses the following: ten technical indicators as inputs, genetic algorithm (GA) to select significant features, backpropagation neural (BPN) to predict future stock price based on features of the previous day and self-organizing map (SOM) to reduce data size. Also, this paper compares three fusion hybrid prediction models, SOM-GA-BPN, SOM-BPN and GA-BPN, with a single model, which is the BPN. Three indices (S&P 500, IBM and NASDAQ) are used in order to evaluate the performance of the proposed hybrid methods. We compare these models with other models such as Support Vector Regression (SVR), Artificial Neural Network (ANN), Random Forest (RF), SVR–ANN, SVR–RF and SVR–SVR fusion prediction models using evaluation measures. The comparison proves the effectiveness and accuracy of the proposed technical indicators and methods.

Keywords: *Stock Market Price Forecast, Genetic Algorithm, Back Propagation Neural Network, Self-Organizing Map, Feature Selection.*

1. INTRODUCTION AND LITERATURE REVIEW

A major challenge facing investors is to predict stock market price/index in a short time and to decrease the difference between predicted and real price, both in financial and commodity markets. Accurately forecasting price movements is necessary for making investment decisions (buy or sell) that increases profits on invested money [1]. Available huge historical data have always helped to forecast future values of stock price because history repeats itself. Stock market data is characterized by being complex, nonlinear, constantly changing, mysterious and chaotic in nature [2]. So the relationship between present information and future stock market price is dynamic and not clear. Many applications have been made to find this relationship, in addition to the significant information that stock market price is based on it. The methods used by investors to forecast and anticipate the future trade can be grouped into two main categories: Fundamental analysis: which focuses on the external economic factors and relevant new events. It is based on the study of supply and demand that makes prices to move higher, lower, or stay the same. Technical analysis: which is the study of market movement, primarily through the use of charts and analysis data patterns, for the purpose of forecasting future price trends. The difference between the two analysis types is that the fundamental analysis studies the cause of market action, while the technical type studies the effect [3].

Technical Analysis has been widely used by traders as a tool for predicting the future behavior of the stock prices. Technical indicators are a fundamental part of technical analysis; many researchers have been proposing many indicators as inputs to predict stock market index such as [4], [5], [6].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In recent years, many artificial intelligence techniques and hybrid intelligent systems have solved the limitations of traditional and statistical methods in nonlinear and time variant problems of finance data [7]. One of the artificial intelligence technologies is artificial neural networks (ANN), which is a supervised, self-learning technique [8], [9], that can find the smooth approximation between input and output information. It can also recognize similar or new patterns even if they weren't in the training data set. Large training data help ANN to discover difficult relationships and show better results. However, the widely used method is back propagation neural network (BPN), which is effective in forecasting stock market price because its multilayer model. It is based on the principle of spreading the error internally by gradient decent technique to lower the network error [10]. The performance of BPN is affected by selecting the hidden layers with neurons, the transfer function between layers and the chosen parameters. BPN has drawbacks, it falls easily into local minimum and has slow convergence, [11] proposes a genetic algorithm to train the BP network weights, in order to improve the speed of convergence and to overcome overfitting problem. Another type of artificial neural networks is the self-organizing map (SOM), which attempts to divide the data based on similar properties into several groups or clusters. Therefore, it is a clustering and reduction of dimension technique [12]. The self-organizing map converts from a higher dimensional input space to a lower dimensional map space and then forms a semantic output map. It is an unsupervised neural network; it doesn't need external help to group data samples in regions [13]. SOM has attracted many researchers in recent years and has been successfully used in the field of text and data clustering [14], [15]. Researchers reduced the large data of IBM, MSFT, S&P 500 and NASDAQ to data with less dimension using SOM and they noticed that it helps in extracting fuzzy rules easily [5].

[16] Feature selection is a dimension reduction technique which aims at selecting a small subset of relevant features for improving the performance of the proposed model, decreasing data size, lowering computational complexity, and knowing the features which affect data. Features are categorized into three types: relevant features; which must be selected from original features, irrelevant features; which affect target result adversely, and redundant features; that don't yield good results. [17] Feature selection methods are divided into two groups: filter and wrapped models. Filter models are based on general characterizations of training data, they also separate feature selection from classifier learning. Wrapper model is a feature selection which is based on the accuracy of the classifier model. In addition, many approaches of selection features have been proposed recently [18], [19], [20].

Genetic algorithms have been used properly for selecting optimal features because it is based on searching for the optimal solution among candidate solutions [21], [22], [23]. Researches have proved that the proposed hybrid model GA-SVM is better than the single model, support vector machine (SVM) [24]. Using genetic algorithm is useful for choosing the best parameters C and Ϭ for SVM classifier. The genetic algorithm has proved its importance as a feature selection approach in finance data [25]. In [4] they propose a hybrid model that uses the genetic algorithm to select features and parameter optimization for the SVM classifier.

2. RELATED WORK

Machine learning techniques have recently received a lot of attention in order to anticipate financial markets prices. The main objective of current researches is to improve and develop a predictive system of future financial market prices with higher accuracy using machine learning methods.

Gonzalez et al. [4], they used Ensemble system based on genetic algorithm. It consisted of 10 SVM classifiers, each classifier has its own inputs and parameters, their outputs are combined by Majority Voting. Also they used technical indicators as inputs. They used the genetic algorithm to select features (best inputs) and parameter optimization for each SVM classifier in ensemble system to predict stock market price of Sao Paulo Stock Exchange index. The experimental results showed that the proposed model is more accurate than other methods like bagging, AdaBoost, SVM, and Random Forest. However, it took longer time.

KURDY and HUSSAIN [5] used self organizing map, to reduce the large data of IBM, MSFT, S&P 500 and NASDAQ to data with less dimensions, then they extracted the fuzzy rules from both the support vector machine model and the relevance vector machine model. They also used four technical indicators (MACD, RSI, Bollinger Band and Stochastic Oscillator) as inputs. They compared the results between two proposed models SOM-SVM-FIS and SOM-RVM-FIS and found that the SOM-RVM-FIS is better than the SOM-SVM-FIS

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

model because it doesn't need to get the optimal values for C and sigma before building the model. They noticed that SOM helps in extracting and analyzing fuzzy rules easily and improves execution time of proposed models.

Patel et al. [6] proposed hybrid forecasting models to predict (t + n)th day closing price of both S&P BSE Sensex and CNX Nifty indexes. They used technical indicators as inputs and Support vector regression to convert technical indicators data, from (t)th day to (t + n)th day, to predict (t + n)th day's closing price using the following models: Support vector regression, Artificial Neural Networks and Random Forest. They compared SVR-SVR, SVR-ANN and SVR-RF with SVR, ANN and RF. They noticed that the accuracy of two-stage models has increased more than the single models, when number of predicted points increased.

Another work was presented by Yizhen et al. [11] proposed a forecasting model by using the genetic algorithm in the first stage to get optimal weights for backpropagation network, then they used back propagation neural network, after that they trained and tested the network using optimal weights of the best fit chromosome, to predict close price of Shanghai index along ten days. They used the GA in order to improve the speed of convergence and to overcome overfitting problem because the BPN has two drawbacks, it falls easily into local minimum and has slow convergence. Also they used BP single model to compare it with the GA-BP proposed model results. The authors proved that experimental results of the GA-BP have achieved the satisfactory accuracy of Shanghai composite index prediction.

Jena and Padhy [24] proposed hybrid forecasting model GA-SVM. They used genetic algorithm to choose the best parameters C and б for SVM classifier. They also proved that the proposed hybrid model GA-SVM is better than the single model of support vector machine (SVM) in predicting the prices of the Indian stock market indexes. The GA-SVM model improved the prediction accuracy using evaluation measures DS, MAE and NMSE.

This paper proposes a hybrid intelligence forecasting model by integrating SOM for reducing data size, GA for selecting features and BPN for close price forecasting. It also compares SOM-GA-BPN, GA-BPN, SOM-BPN and BPN models using evaluation measures (MSE, MAPE, MAE, rRMSE, ACCURACY). In order to evaluate the performance of the proposed hybrid models, the indices S&P 500, NASDAQ and the company IBM are used as illustrative examples. Ten technical indicators and volumes along t days, are used as inputs to predict the (t+1) days closing price. The proposed study has proved its effectiveness by comparing the following proposed models: SVR-ANN, SVR-SVR, SVR-Random Forest, SVR, ANN, and Random Forest prediction models [6].

The remainder of this paper is organized as follows. In Sect. 3, BP neural network, SOM clustering, and GA are discussed. The hybrid forecasting model is presented in Sect. 4. Dataset and performance criteria are presented in Sect. 5. Section 6 evaluates the models for the S&P 500, NASDAQ, IBM and S&P BSE Sensex indexes. Finally, Sect. 7 concludes the paper.

3. METHODOLOGY

3.1 Back Propagation Neural Network

The back propagation neural network was proposed in 1986 by Rumelhart, Hinton and Williams for solving several problems because its function is to train and adjust the weights to reach less error [26]. It is a multilayer and supervised network. Choosing the number of hidden layers, the number of neurons in them and the activation function is very important and depends on the proposed problem [10]. It includes four steps: forward-propagation of the training sample data, calculation the difference, back-propagation of the error and the weights adjustment. Every neuron in the input layer, which holds information from training data, passes through at least one hidden layer after multiplication by weights, then the hidden layers process the weighted information using activation functions. Finally, the hidden layers pass the processed information to each output layer, this is the forward-propagation process. The error is calculated by the difference between the output value - which results from the output layers - and the target value. Backpropagation algorithm depends on gradient decent method that works by modifying the weights and parameters of each layer through the back-propagation process. This process starts from the output layers, then passes through the hidden layers and ends with the input layers. This method minimizes the error and these steps repeat iteratively until the network error reaches acceptable low value [8], [11]. The BPN's steps shown in the following (Figure 1).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

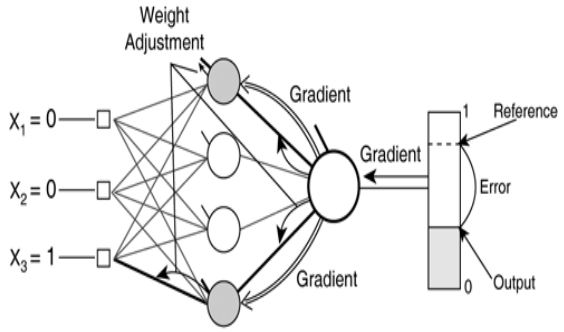


Figure 1: Shows BPN's steps.

3.2 Self-Organizing Map Clustering

A self-organizing map is a type of artificial neural networks and it was introduced in the 1980s [5]. It works as a clustering technique and converts large data to several clusters with lower dimension. It is unsupervised because there aren't input-output pairs. Also, it is a competitive learning technique. It consists only from the input layer and output layer without hidden layer. The SOM structure is shown in (Figure 2). The Input layer is an input vector from the input data set. Whereas the output layer is the matrix of nodes which are called neurons, each neuron has a weight vector which its dimension is the same as the dimension of the input vector.

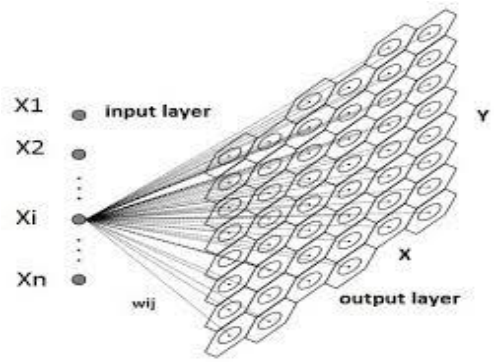


Figure 2: Show the SOM structure.

The main idea is to calculate the degree of similarity of training data vectors to form a topology map with two dimensions in the output layer. Firstly, all weights are initialized to random values. Then, Euclidean distance is computed between the input vector and all the other weight vectors to find the closest neuron (the winner neuron), which is called the best matching unit (BMU). Also the weights of the winner neuron and the neighboring neurons- which are neurons that are close to it - are updated until the input vector is reached. However, if a similar input vector is found, the winner neuron will be continually activated and thus it will learn more. This process repeats iteratively until we reach predefined training steps, runs or cycles [5], [12], [13].

3.3 Genetic Algorithm

The genetic algorithm, which was invented by John Holland in the 1960s [27], it is the heuristic search and optimization technique that mimics the process of natural evolution. GA is important for searching among possibilities of huge data to find the optimal solution. The genetic algorithm is used in many tasks such as optimization of neural networks, medical data and glass identification [22], [28]. It consists of the following five steps: Initialization, Evaluation, Selection, Crossover and Mutation. GA uses a search space that is called population. It forms a set of chromosomes which could be decoded to form a phenotype according to a proposed problem. It initializes the population randomly and then evaluates each chromosome using a fitness function which determines the quality of the chromosome in the population for the optimization task, this is the selection process. There are two operators, crossover and mutation, which can be applied to individuals for reproducing new individuals. These processes repeat until it reaches a predefined number of generations or a defined minimum or maximum value of the fitness function [4], [23].

4. THE HYBRID FORECASTING MODEL

This paper proposes two stages to predict stock market price, it uses the original data and integrates GA-BPN then it compares it with the single stage BPN. We also propose a three-stage hybrid forecasting stock market index model by reducing the size of the original data, and this is done by integrating SOM-GA-BPN. Finally, we compare it with the two stages SOM-BPN. The reprocessing data and the overall architecture of the proposed hybrid predicting model SOM-GA-BPN shown in (Figure 3).

For building the proposed forecasting model, we first select the technical indicators as input predictors, because they offer a different way for predicting the future price. We can get these indicators from the price data after applying a formula to them. These data may be high, low, open, close or any combination of them over a period of time [29]. There are many types of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

technical indicators: Trend Indicators; from this type we choose Exponential Moving Average (EMA) and Moving Average Convergence/Divergence oscillator (MACD). Momentum Indicators; from this type we select relative strength index (RSI), commodity channel index (CCI) and stochastic oscillator (SO). Volatility Indicators; from which we select Bollinger Bands (BB). Volume Indicators; from which we choose Chaikin Oscillator and On-Balance Volume (OBV) [30]. We have noticed that the period in their calculation should be short to better predict the close price. We have explained the calculations of the selected technical indicators in (Table 1) [31]. We have noticed that the data of OBV, Chaikin Oscillator and volume, are large values. However, to reduce the prediction error, these values are scaled into the range of [-1,1] by computing:

$$NewX = -1 + \frac{2(X-minX)}{(maxX-minX)} \quad (1)$$

In this study, we use the volume and ten technical indicators; the technical indicators and volume values - along t days -are used to predict the t+1day close price.

For building the models, we first determine a set of parameters. In BP model, the first step is selecting the number of input, output layer neurons, and choosing 11 neurons in the input layer and one neuron - which is the predicted close price - for the output layer. As a result, we use one hidden layer and number neurons of the hidden layer shown in the following equation:

$$num_{neurons} = \frac{2*(N+1)}{3} \quad (2)$$

where N number neurons of the input layer [10]. We choose 8 neurons in the hidden layer. The second step is to select the activation/transfer function, which gives the best evaluation (the minimum mean square error). In this study, the purelin function is used in both hidden and output layer for all experiments. The following (Figure 4) shows the linear purelin function.

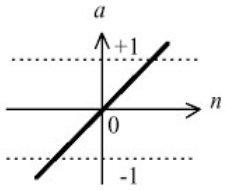


Figure 4: Shows the linear purelin function.

For building GA model, we need five steps:

1. Initial population: individuals are coded as binary chromosomes. Each chromosome represents a subset of features. As in [4], [25], we use the symbol '1' to denote the existence of a feature and '0' to its absence. The first generation is generated randomly.

2. Fitness function: it evaluates the performance of the chromosome. It plays a major role in transferring the best/higher fit chromosomes to the next generations. In this study, the fitness function is:

$$Fitness\ Chromo = \left(\frac{1}{Error}\right)\left(\frac{11-numFeatures}{11}\right) \quad (3)$$

Where Error is the mean squared error (MSE) of the BP classifier model and numFeatures are the selected features in the current chromosome. We use this equation to select the smallest number of features with less error [22].

3. Selection: it is a process that comes after the evaluation of chromosome by the fitness function for selecting the higher one from the current population, and then moving them to the next generation. In this study, we use the roulette wheel method.

4. Crossover: we use the two-point crossover method. The (Figure 5) Shows the Two-point crossover. Two parent individuals are selected from the current population. Then, two random points are generated. Then the selected part is exchanged between two parents to form two child individuals. The probability of crossover, that we use, is (0.7).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

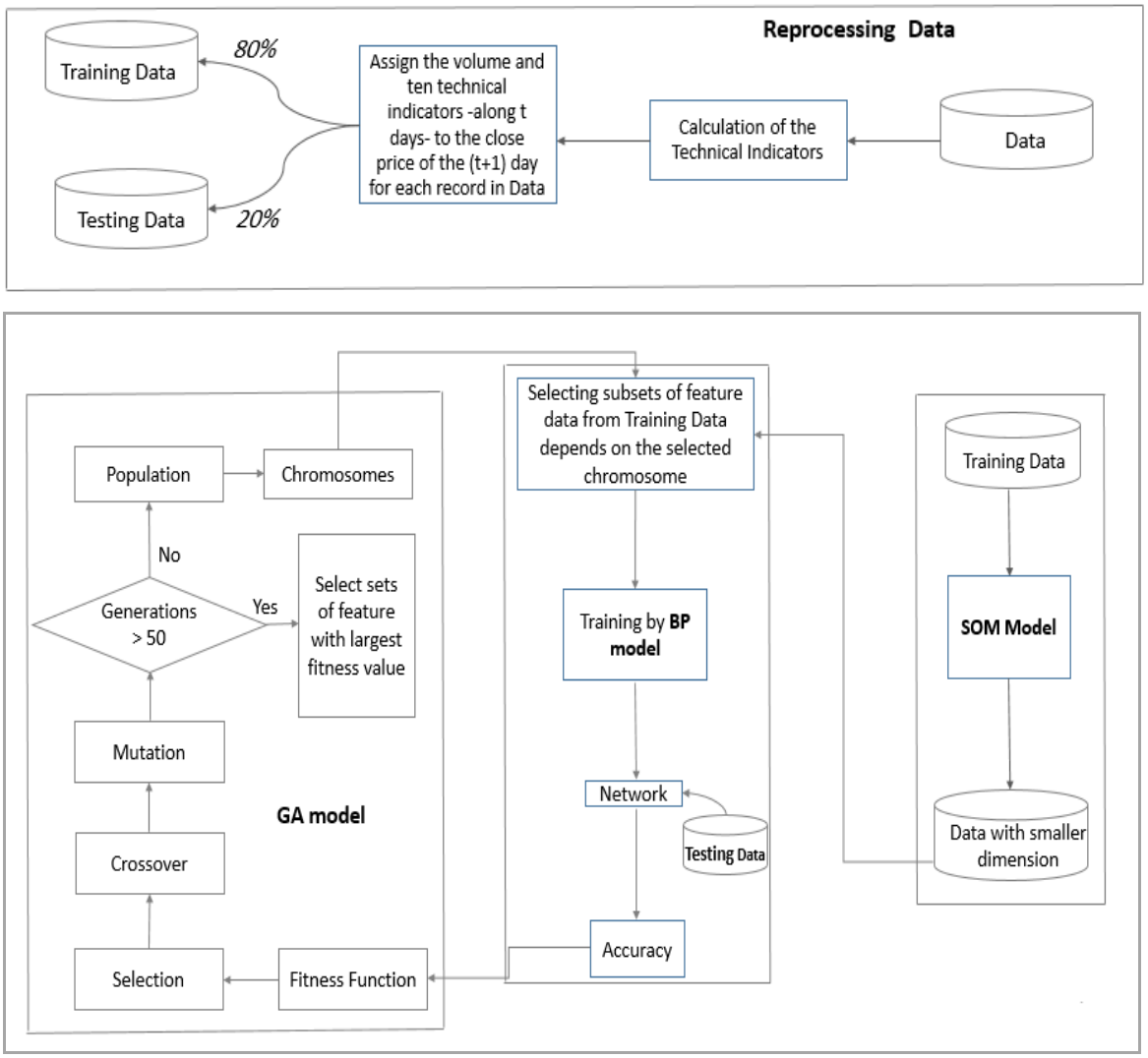


Figure 3: The proposed hybrid predicting model.

Table 1: The selected technical indicators & their formulas

Indicator	formulas
EMA	EMA = {Close − EMA (previous day)} multiplier + EMA (previous day). Where multiplier: $\frac{2}{\text{Time periods} + 1} = \frac{2}{9+1} = \frac{2}{10}$ and EMA (previous day) = $\frac{\sum_{i=1}^{n} \text{close index on day } i}{n}$, just for the first value
MACD	MACD = ((5 − day EMA) − (10 − day EMA))
Bollinger Band	Middle Band = 10 − day simple moving average (SMA) Upper Band = 10 − day SMA + (10 − day standard deviation of price . 2) Lower Band = 10 − day SMA − (10 − day standard deviation of price . 2) Where SMA = $\frac{\sum_{i=1}^{n} \text{close index on day } i}{10}$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

RSI	$RSI = 100 - \frac{100}{1 + RS}$ Where $RS = \frac{\text{Average of the past } F \text{ days up closes}}{\text{Average of the past } F \text{ days down closes}}$
SO	$SO = \frac{\text{current close} - \text{lowest low}}{\text{highest high} - \text{lowest low}} 100$ Where highest high & lowest low of the past 5 days
CCI	$CCI = \frac{(\text{Typical price}) - (10\text{-day SMA of typical})}{0.015 \text{ Mean deviation}}$ $\text{Typical price} = \frac{\text{close} + \text{high} + \text{low}}{3}$ There are four steps to calculate the Mean Deviation: First, subtract the most recent 10-period average of the typical price from each period's typical price. Second, take the absolute values of these numbers. Third, sum the absolute values. Fourth, divide by the total number of periods (10).
OBV	If the closing price is above the prior close price, then: Current OBV = Previous OBV + Current Volume If the closing price is below the prior close price, then: Current OBV = Previous OBV – Current Volume If the closing prices equals the prior close price, then: Current OBV = Previous OBV (no change)
Chaikin Oscillator	$\text{Money Flow Multiplier} = \frac{[(\text{Close} - \text{Low}) - (\text{High} - \text{Close})]}{(\text{High} - \text{Low})}$ Money Flow Volume = Money Flow Multiplier Volume for the Period ADL = Previous ADL + Current Period's Money Flow Volume Chaikin Oscillator = (3 – day EMA of ADL) – (10 – day EMA of ADL)

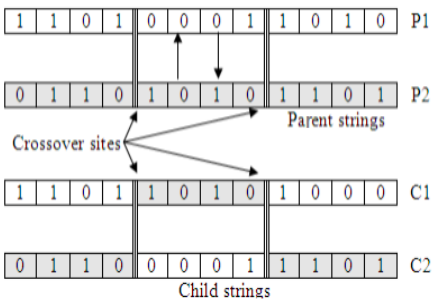


Figure 5: Shows the Two-point crossover.

5. Mutation: it is used to add new features on chromosomes in the population. It changes one bit from one to zero, or zero to one in the selected chromosome after choosing a single bit randomly. The probability of mutation, that we use, is (0.1).

we choose the number of chromosomes in population to be 100 individuals and runs/generations to be 50 in all experiments.

To build the SOM model, the essential key is to select the number of neurons of matrix output.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Selecting the size of matrix output depends on the outcome which is better or equal to the result of the original data evaluation.

5. DATASETS AND PERFORMANCE CRITERIA

For evaluating the performance of the proposed hybrid models, the daily indices of the S&P 500, NASDAQ and datasets of IBM are used in this study. And we used [32] for obtaining the datasets of financial indices and companies. The two indices S&P 500 and NASDAQ are gathered from 1/18/1993 to 12/29/2017. The dataset of company IBM is gathered from 1/16/1962 to 12/29/2017. In the first two indices, there are 6286 data points in datasets in total. The first 5029 data points (80 % of the total sample points) are used as the training samples, while the remaining 1257 data points (20 % of the total sample points) are used as the testing samples. For IBM, there are 14087 data points datasets in total. The first 11270 data points (80 % of the total sample points) are used as training samples, while the remaining 2817 data points (20 % of the total sample points) are used as the testing samples.

The prediction performance is evaluated using the following performance measures: Mean Absolute Percentage Error (MAPE), Mean Absolute Error(MAE), relative Root Mean Squared Error (rRMSE), Mean Squared Error (MSE) and Accuracy. Formulas of these evaluation measures are shown in equations (4)- (8).

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|At-Ft|}{|At|}\,100 \quad (4)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\frac{|At-Ft|}{|At|} \quad (5)$$

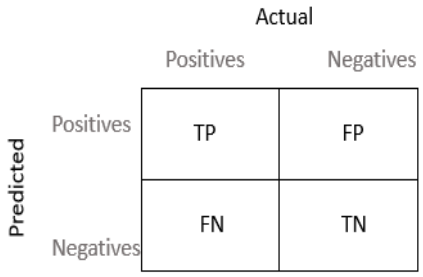
$$rRMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{At-Ft}{At}\right)^2} \quad (6)$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(At-Ft)^2 \quad (7)$$

where At is the actual value of stock price, Ft is the predicted value of stock price and n is the number of predicted points.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}\,100 \quad (8)$$

Where:



The smaller the values of MSE, RMSE, MAE and MAPE, the closer the predicted time series values are to that of the actual value. They can be used to evaluate the prediction error.

6. RESULTS AND DISCUSSION

The proposed models are implemented using Matlab R2018 on a PC with the following specifications (CPU: Intel core i5, System: Windows 10 Ultimate 64-bit, RAM: 8GB).

6.1 Evaluating the Models for the S&P 500 Index

The results in (Table 2) explain the difference between the results of the BPN and GA-BPN forecasting models. The values are convergent. The results of the BPN model are based on all features (technical indicators & volume data). Investors or experts will calculate all the features values to predict the close price value of the next day and that needs more data and time for accounting indicators' values. Although the results of the GA-BPN model are less than the BPN model, it selects two features as shown in (Table 3). It needs less time and data storage. The actual values of the S&P500 index and the prediction results of the proposed GA-BPN model from January 4, 2013 to December 29, 2017 shown in (Figure 6). Here, there is no reduction of the data dimension. We used all 5029 data points for training with the proposed approaches.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: The results of the proposed models for the S&P 500 index using error measures

Prediction Models	Error Measures				
	MAPE (%)	MAE	rRMSE	MSE	Accuracy (%)
BPN	0.5681	0.0057	0.0081	254.99	96.49
GA-BPN	0.584	0.0058	0.0083	265.02	96.49

Table 3: The resulting chromosome using the GA-BPN model for the S&P 500 index

EMA	MACD	Middle BB	Upper BB	Lower BB	RSI	SO	CCI	OBV	Chaikin	Volume
1	0	0	0	0	0	1	0	0	0	0

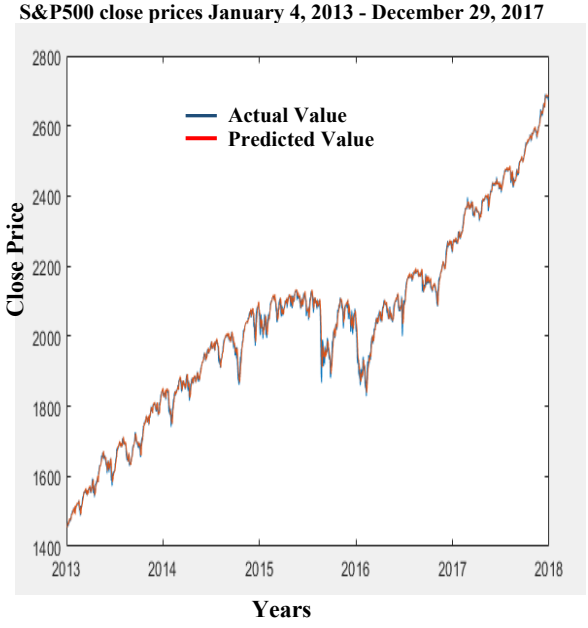


Figure 6: Prediction with S&P500 results using GA-BPN model

Reducing the data size by using the SOM model is difficult and takes more time to get fewer records of data, which gives more or equal results than the original data. we used the centers of clusters as new training data (instead of native training data). We selected the numbers of clusters 100, 121, 144 and 169 as mentioned in (Table 4). The results show a comparison between data matrices based on the results MSE and output the Chromosome of the GA-BPN model. We noticed that the closer results to (Table 3) are the results 121 and 144 number of the clusters, but we used less MSE value, whose size is 11*11 of the output matrix (121 number of clusters). We used the data with 121 data points for training by implementing the BPN model. The results of the proposed models (SOM-BPN and SOM-GA-BPN) using the evaluation measures as shown in (Table 5).

Table 4: Shows the results clustering of SOM model for S&P 500 index using the GA-BPN model

GA-BPN model	Number of clusters			
	100	121	144	169
MSE	325.03	257.9	269.35	310.66
Selected Features	EMA MACD Middle BB RSI	EMA Middle BB SO	EMA Middle BB SO	EMA Middle BB

Table 5: Shows the results of the proposed Models for S&P 500 index after clustering

Prediction Models	Error Measures				
	MAPE (%)	MAE	rRMSE	MSE	Accuracy (%)
SOM-BPN	0.5861	0.0059	0.0083	266.91	96.26
SOM-GA-BPN	0.5706	0.0057	0.0082	257.9	96.02

We noticed that two models' results are convergent. We used genetic algorithm to select small subset of features that improve the performance and need less time and data size, that's why SOM-GA-BPN model's results are better than the another model. The actual values of the S&P500 index and the prediction results of the proposed SOM-GA-BPN model from January

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4, 2013 to December 29, 2017 shown in (Figure 7).

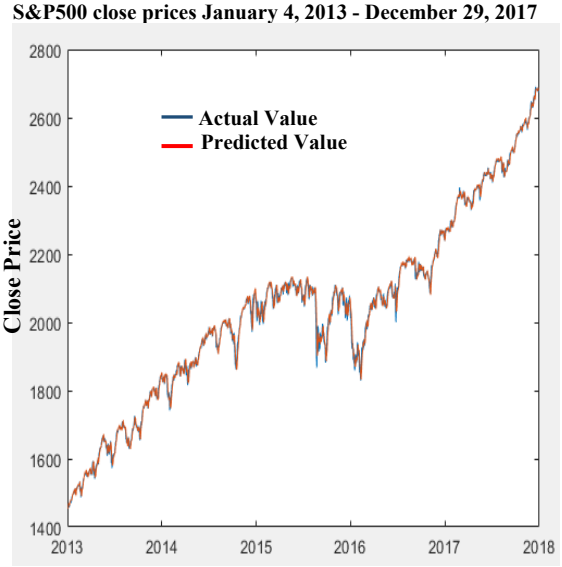


Figure 7: Prediction with S&P500 results using SOM-GA-BPN model

6.2 Evaluating the Models for the NASDAQ Index and the Company IBM

we notice that the results in (Table 6) for both indexes are convergent. The proposed GA-BPN model is used to select less subset of features. Although the results of BPN model is better, but we use all features to show evaluation measures values. We searched the less number of features that the index will depend on them to predict accurately and the GA-BPN achieves that. The resulting features of the NASDAQ and IBM indices using GA-BPN and SOM-GA_BPN models can be seen in (Table 7). The resulting number of clusters using the SOM model for NASDAQ and IBM indices shown in (Table 8). The results in (Table 9) were also similar to the previous results in (Table 6), but they took more time and they were more complex than the GA-BPN model. The actual values of the NASDAQ index and the prediction results of the proposed GA-BPN model from January 4, 2013 to December 29, 2017 shown in (Figure 8). The actual values of the IBM index and the prediction results of the proposed GA-BPN model from October 23, 2006 to December 29, 2017 shown in (Figure 9).

Table 6: The results of the proposed Models for the NASDAQ and IBM indexes using error measures

Index	Prediction Models	Error Measures				
		MAPE (%)	MAE	rRMSE	MSE	Accuracy(%)
NASDAQ	BPN	0.7206	0.0072	0.0098	2131.2	97.93
	GA-BPN	0.7166	0.0072	0.0099	2139.4	97.53
IBM	BPN	1.1053	0.0111	0.0158	4.90	98.23
	GA-BPN	1.1382	0.0114	0.0163	5.19	98.51

Table 7: The resulting features of the NASDAQ and IBM indices using GA-BPN and SOM-GA_BPN models

Index	Features of SOM-GA-BPN	Features of GA-BPN
NASDAQ	EMA, Middle BB, SO	EMA, SO
IBM	EMA, Middle BB, SO	EMA, Middle BB

Table 8: The results clustering of the SOM model for the NASDAQ and IBM indices

Index	Number of clusters
NASDAQ	144
IBM	144

Table 9: The results of the proposed Models for the NASDAQ and IBM indexes after clustering

Index	Prediction Models	Error Measures				
		MAPE (%)	MAE	rRMSE	MSE	Accuracy (%)
NASDAQ	SOM-BPN	0.7435	0.0074	0.0102	2303	97.53
	SOM-GA-BPN	0.7283	0.0073	0.0101	2242.4	97.37
IBM	SOM-BPN	1.2246	0.0122	0.0172	5.79	97.66
	SOM-GA-BPN	1.1433	0.0114	0.0164	5.25	98.51

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

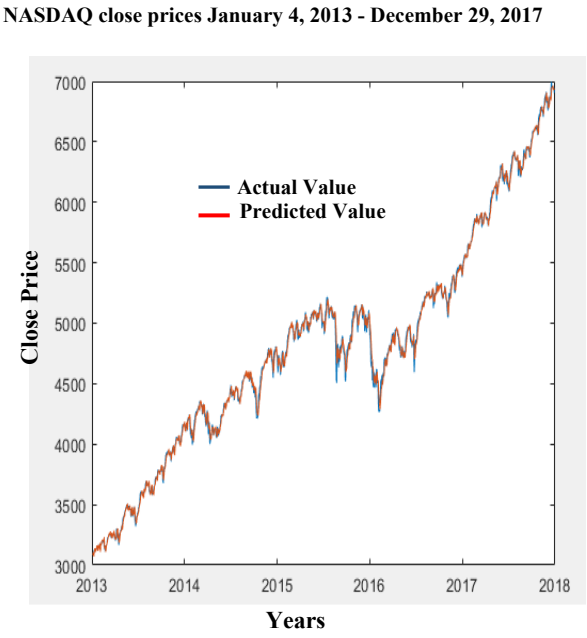


Figure 8: Prediction with NASDAQ results using GA-BPN model

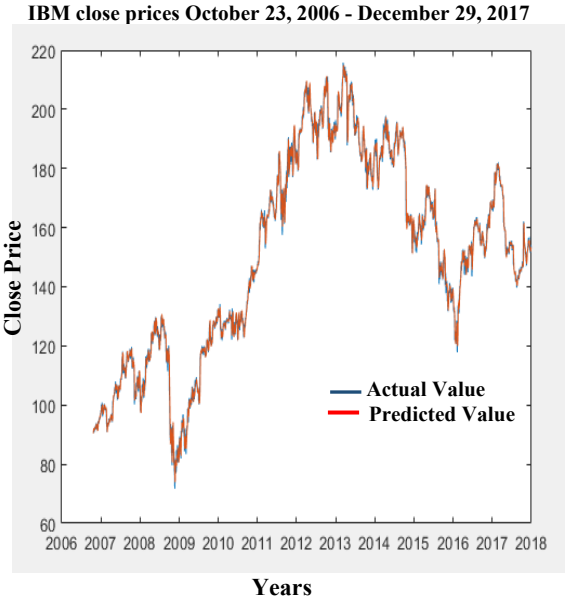


Figure 9: Prediction with IBM results using GA-BPN model

6.3 Evaluating the Models for the S&P BSE Sensex Index

We compared the proposed methods (BPN and GA-BPN) with the models shown in (Table 10). In the BPN model, the number of hidden layer neurons are 8 neurons, we use 1000 epochs and a sigmoid log is used as the transfer function of the neurons of the hidden layer, whereas the neuron in the output layer uses linear transfer function (purelin). We used the same training and testing data from Jan 2003 to Dec 2012 of stock market index S&P BSE Sensex for prediction performance of 1-Day Ahead of Time [6]. We used EMA, MACD, Middle BB, Upper BB, Lower BB, RSI, SO, CCI, OBV, Chaikin Oscillator and Volume as inputs. We noticed both BPN and GA-BPN results are the best. The genetic algorithm selected three features, which are Middle BB, Upper BB and CCI, with less difference between actual and predicted close price than other models.

Table 10: The results of the proposed models and other models for the S&P BSE Sensex index

Prediction Models	Error Measures			
	MAPE (%)	MAE	rRMSE	MSE
ANN	1.78	313.92	2.31	166090.16
SVR-ANN	1.55	272.71	1.96	118395.09
SVR	0.98	172.47	1.25	47558.47
SVR-SVR	1.48	260.05	1.89	108137.61
Random Forest	1.25	221.91	1.60	81098.60
SVR-Random Forest	1.23	216.02	1.55	73483.60
BPN	0.98	0.0098	0.0124	46504
GA-BPN	0.93	0.0093	0.0119	43391

7. CONCLUSION

This article presents some artificial intelligence techniques that form hybrid forecasting models to predict the close price of stock market. We proposed the following models: SOM to reduce data dimension, GA to select important variables and BPN to find relationship between input and output. We used financial datasets of S&P 500, NASDAQ and IBM to evaluate combined methods with error measures. Experimental results proved the effectiveness of GA-BPN model based on comparison with other models. We used ten technical indicators in this study as inputs. However, further improvement would be to collect more variables or new technical indicators to increase the prediction performance of the proposed

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

models. Future works may aim to predict (t + n) th day's closing price/value where n is two or more, also to use genetic algorithm for the optimization of BPN parameters and weights, and thus improving the performance.

REFERENCES:

[1] YASER S. ABU-MOSTAFA and AMIR E ATIYA," Introduction to financial forecasting", *Applied Intelligence,* vol. 6, no. 3, 1996, pp. 205-213.

[2] Zhang Yudong and Wu Lenan, "Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network", *Expert systems with applications,* vol. 36, no. 5, 2009, pp. 8849-8854.

[3] John J. Murphy, "Technical analysis of the financial markets: A comprehensive guide to trading methods and applications", *New York Institute of Finance*, 1999, p.586.

[4] Rafael Thomazi Gonzalez, Carlos Alberto Padilha and Dante Augusto Couto, "Ensemble system based on genetic algorithm for stock market forecasting", *In 2015 IEEE Congress on Evolutionary Computation (CEC)*, 2015, pp. 3102-3108.

[5] Mohamad Bassam KURDY and Ali AL HUSSAIN, "Fuzzy Classifier Based on SOM Approach for Stock Market Price Prediction", *Syrian Virtual University*, syria, 2016, p. 48.

[6] Jigar Patel, Sahil Shah, Priyank Thakkar and K. Kotecha, "Predicting stock market index using fusion of machine learning techniques", *Expert Systems with Applications,* vol. 42, no. 4, 2015, pp. 2162-2172.

[7] Arash Bahrammirzaee, "A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems", *Neural Computing and Applications,* vol. 19, no. 8, 2010, pp. 1165-1195.

[8] G. Sundar and K. Satyanarayana, "SMPBPM: Stock Market Prediction by Back Propagation Model", *Elsevier, Advances in Engineering and Technology*, 2015, pp. 42-50.

[9] Suraiya Jabin, "Stock market prediction using feed-forward artificial neural network", *International Journal of Computer Applications,* vol. 99, no. 9, 2014, pp. 4-8.

[10] G. Sundar and K. Satyanarayana, "Back Propagation: A Prediction Approach for Stock Market Based on Hidden Layer Identification", *International Journal of Innovative Research in Computer and Communication Engineering,* vol. 4, no. 12, 2016, pp. 21601-21607.

[11] LI Yizhen , Zeng Wenhua, Lin ling, Wu jun and Lu Gang, "The forecasting of Shanghai index trend based on genetic algorithm and back propagation Artificial neural network algorithm", *In 2011 6th International Conference on Computer Science & Education (ICCSE)* , 2011, pp. 420-424.

[12] André Skupin and Pragya Agarwal, "Self-organising maps: Applications in geographic information science", *John Wiley & Sons*, 2008.

[13] Juha Vesanto and Esa Alhoniemi, "Clustering of the self-organizing map", *IEEE Transactions on neural networks,* vol. 11, no. 3, 2000, pp. 586-600.

[14] Chih-Ming Hsu, "A hybrid procedure for stock price prediction by integrating self-organizing map and genetic programming", *Expert Systems with Applications,* vol. 38, no. 11, 2011, pp. 14026-14036.

[15] Yuan-Chao Liu, Ming Liu and Xiao-Long Wang, "Application of self-organizing maps in text clustering: a review", *In Applications of Self-Organizing Maps*. InTech, Rijeka, Croatia, 2012.

[16] Jiliang Tang, Salem Alelyani and Huan Liu, "Feature selection for classification: A review", *Data classification: algorithms and applications,* 2014, p. 37.

[17] M. Dash and H. Liu , "Feature Selection for Classification", *Intelligent Data Analysis,* vol. 1, no. 1-4, 1997, pp. 131-156.

[18] Shima Kamyab and Mahdi Eftekhari, "Feature selection using multimodal optimization techniques", *Neurocomputing,* vol. 171, 2016, pp. 586-597.

[19] Tapas Bhadra and Sanghamitra Bandyopadhyay, "Unsupervised feature selection using an improved version of differential evolution", *Expert Systems with Applications,* vol. 42, no. 8, 2015, pp. 4042-4053.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[20] Parham Moradi and Mozhgan Gholampour, "A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy", *Applied Soft Computing*, vol. 43, 2016, pp. 117-130.

[21] Swati N.Moon and Dr. Narendra Bawane, "Optimal feature selection by genetic algorithm for classification using neural network", *International Research Journal of Engineering and Technology (IRJET)*, vol. 2, no. 5, 2015, pp. 582-586.

[22] Te-Sheng Li, "Feature selection for classification by using a GA-based neural network approach", *Journal of the Chinese Institute of Industrial Engineers* , vol. 23, no. 1, 2006, pp. 55-64.

[23] Anne M. P. Canuto and Diego S. C. Nascimento, "A genetic-based approach to features selection for ensembles using a hybrid and adaptive fitness function" *In The 2012 international joint conference on neural networks (IJCNN)* , 2012 , pp. 1-8.

[24] Om Prakash Jena and Dr. Sudarsan Padhy, "Application of GA with SVM for Stock Price Prediction in Financial Market", *International Journal of Science and Research (IJSR)*, vol. 3, no. 10, 2014, pp. 498-503.

[25] Yanan Mao, Zuoquan Zhang and Dingyuan Fan, "Hybrid feature selection based on improved genetic algorithm for stock prediction", *In 2016 6th International Conference on Digital Home (ICDH)* , 2016, pp. 215-220.

[26] Laurene Fausett, "Fundamentals of neural networks: architectures, algorithms, and applications", *Prentice-Hall, Englewood Cliffs*, 1994, p. 476.

[27] Melanie Mitchell, "An introduction to genetic algorithms", *Massachusetts Institute of Technology*, MIT press, London, 1998, p. 162.

[28] Wei Shen and Mia n Xing, "Stock index forecast with back propagation neural network optimized by genetic algorithm", *In 2009 Second International Conference on Information and Computing Science*, vol. 2, 2009, pp. 376-379.

[29] "STOCKCHARTS" [Online]. Available: https://stockcharts.com/school/doku.php?id=chart_school:technical_indicators:introduction_to_technical_indicators_and_oscillators. [Accessed: 13-Apr-2019]

[30] "QUASTIC MONEY MAKERS" [Online]. Available: https://quastic.com/trading-basics/technical-indicators-categories-and-types/. [Accessed: 13-Apr-2019]

[31] "STOCKCHARTS" [Online]. Available: https://stockcharts.com/school/doku.php?id=chart_school:technical_indicators. [Accessed: 13-Apr-2019]

[32] "YAHOO FINANCE" [Online]. Available: https://finance.yahoo.com. [Accessed: 13-Apr-2019]