

Stock Direction Forecasting Techniques: An Empirical Study Combining Machine Learning System with Market Indicators in the Indian Context

Deepika Chandwani
MBA (Financial Analyst)
Indore, India

Manminder Singh Saluja, Ph. D
Assistant Professor (Senior Scale)
IIPS, DAVV, Indore, India

ABSTRACT

Stock price movement prediction has been one of the most challenging issues in finance since the time immemorial. Many researchers in past have carried out extensive studies with the intention of investigating the approaches that uncover the hidden information in stock market data. As a result of which, Artificial Intelligence and data mining techniques have come to the forefront because of their ability to map non-linear data. The study encapsulates market indicators with AI techniques to generate useful extracts to improve decisions under conditions of uncertainty. Three approaches (fundamental model, technical indicators model and hybrid model) have been tested using the standalone and integrated machine learning algorithms viz. SVM, ANN, GA-SVM, and GA-ANN and the results of all the three approaches have been compared in the four above mentioned methods. The core objective of this paper is to identify an approach from the above mentioned algorithms that best predicts the Indian stocks price movement. It is observed from the results that the use of GA significantly increases the accuracy of ANN and that the use of technical analysis with SVM and ANN is well suited for Indian stocks and can help investors and traders maximize their quarterly profits.

Keywords

Support Vector Machine; Artificial Neural Network; Genetic Algorithm; Financial Ratios; Technical Indicators

1. INTRODUCTION

Stock markets have now become a sought after means of amassing abundant wealth not only for institutional giants but for small traders as well. Countries around the world have experienced exponential growth in terms of stock market capitalization and an increasing number of investors are now crossing national boundaries to take advantage of undervalued stocks and countries to book profits, thereby, connecting and influencing the national economies across the globe. This in turn is making “financial forecasting” all the more imperative as the profitability of investing and trading in the stock market to a large extent depends on the predictability.

Two main schools of thoughts that have been primarily used by researchers’ to forecast stock prices are Technical Indicators and Fundamental Ratios, but both the methods fail to produce satisfactory results for financial time series data is characterized by noise, non-stationary, chaos and high degree of uncertainty. Supplementing these approaches with newly surfaced methods called “machine learning algorithms” is widely gaining popularity because of the ability of Artificial Intelligence techniques to map non – linear data.

Artificial Neural Network is till date the most widely accepted machine learning technique for forecasting stock prices. However, a crucial drawback of Neural Networks which implement the empirical risk minimization principle is the over – fitting problem leading to a poor level of generalization. A recently introduced machine learning method called Support Vector Machine (SVM) has been gaining increasing popularity in this area as it is said to improve the generalization property of neural networks; many papers affirm that SVM is a superior technique as SVM decreases the level of risk in information data and leads to the more degree of accuracy by using a structural method (Sap and Awan, 2005).

Literature survey witnesses that most of the previous research efforts on stock market prediction were limited to either using technical indicators or fundamental ratios as input features to the machine learning techniques, which arguably fail to provide a complete picture. Therefore, this paper explores the hybrid model i.e. the combination of both technical and fundamental parameters and compares its performance with that of the standalone technical analysis model and fundamental analysis model with the objective of testing which practice attains improved stock movement prediction in the Indian context. Integrated method of Genetic algorithm with the two classification techniques i.e. ANN and SVM is also investigated, wherein GA is used to select the set of most informative input features from among all the technical and financial indicators. Many empirical researches suggest that examining GA-SVM and GA-ANN has the potential of offering more accurate results over SVM (Choudhry and Garg, 2008) and ANN (Whitley, 1995) respectively.

In the next section the paper reviews the existing literature concerning market indicators and machine learning techniques. Section 3 describes the theoretical background while section 4 lays down the data and methodology. Section 5 discusses the experimental results and section 6 concludes the paper.

2. LITERATURE REVIEW

A host of researches have been conducted in past to address the prediction of future stock movements using statistics, technical analysis, fundamental analysis, time series analysis, chaos theory, and linear regression, but owing to the enormous noise and complex dimensionality in the data, there has been a shift in the focus of the researchers towards artificial intelligence (AI) techniques. However, AI techniques alone have not been able to consistently produce correct prediction of the stock market, which is why recent researches have hybridized several artificial intelligence (AI) techniques with technical indicators and/or fundamental

analysis with the intention of improving the forecasting accuracy.

Lee and Chen (2007) put in technical indicators as inputs to a feedforward neural network for prediction of NASDAQ and Taiwan Stock Exchange. Nagarajan et al. (2005) also used technical indicators and neural networks to predict the US Dollar Vs British Pound exchange rates. Mitra (2009) carried out an experiment of combining signals generated from moving averages of different window size using ANN. Kim and Han (2000) used neural network modified by Genetic Algorithm. Zhang and Bai (2012) utilized ϵ -SVM to build a stock price prediction model and the empirical analysis proved that ϵ -SVM model has a high prediction accuracy to predict the stock price. Sap and Awan (2005) studied the use of SVM in financial forecasting and concluded that SVM provides a promising alternative to time series forecasting. Pantazopoulos et al. (1998) employed a neurofuzzy approach for predicting the prices of IBM stock. Kim and Chun (1998) exhibited refined Probabilistic Neural Network (PNN) to predict the stock market index. Zahid Iqbal et al. (2013) described different state of the art techniques used for stock forecasting, their extensive studies show that Neural Network is the best technique to predict stock prices especially when some de-noising schemes are applied with neural network and their study also concluded that Layered Recurrent Neural Network performs better than feed forward NN and WSMPCA-NN performs better than LRNN and NN on Pakistan State Oil and S&P500 data sets. Zhu and Chung (2003), Zhu et al. (2008), Tsanga et al. (2007), Avci (2007), and Stansel and Eakins (2004) also applied ANN models with technical indexes for forecasting financial time series.

From the above literature review it is evident that technical indicators with AI models have been extensively used, while there are only a few cases of the use of fundamental indicators. O'Connor and Madden (2006) evaluated the effectiveness of using external indicators in predicting movements in the Dow Jones Industrial Average index using Neural Network Approach. Wu and Xu (2006) found that fundamental analysis can be powerful to predict stock price, especially with the aid of neural networks approach and rough set theory. Atiya et al. (1997), Quah and Srinivasan (1999), Raposo and Cruz (2002) also roped in fundamental indicators to forecast stock prices using Neural Networks. Schumaker and Chen (2006) tried to find out influence of news article on stock price using Support Vector Machine (SVM). Han and Chen (2007) proposed using SVM with financial statement analysis for prediction of stocks and the experimental results show that the method not only improved the accuracy rate, but also met different stockholders' expectations.

Chavan and Patil (2013) surveyed different input parameters that can be used for stock market prediction with ANN and concluded that hybridized parameters like combination of technical and fundamental variables gave better prediction accuracy over application of standalone parameters after surveying. The combination of parameters or AI techniques has been proposed by several researchers and their studies indicated that the integrated forecasting techniques outperformed the individual parameters or techniques. Adebisi et al. (2012a) and Adebisi et al. (2012b) presented a hybridized and an improved predictive model, wherein, the former combined the technical and fundamental variables; and the latter combined experts' opinion, technical and fundamental indices using ANN. The empirical results of both the studies confirmed superior performance of their proposed

models over the conventional approach of using ANN with technical indicators.

Choudhry and Garg (2008), and Khatibi and Rasouli (2011) proposed a hybrid GA-SVM system for predicting the future direction of stock prices using technical indicators. Yanshan Wang and In-Chan Choi (2013) proposed a PCA-SVM integrated model to forecast the directions of the stock market indices and the individual stock prices and the results provided markedly high hit ratios for forecasting movement directions of the constituents in the KOSPI and HSI. The work of Nair, Mohandas, and Sakthivel (2010) presented the design and performance evaluation of a hybrid decision tree-rough set based system for predicting the next day's "trend in the Bombay Stock Exchange (BSESENSEX)". The research conducted by Tsai and Wang (2009) combined ANN techniques with the great explanation ability of decision tree to help investors with better decision support system.

Kim and Shin (2007) investigated the effectiveness of a hybrid approach based on the artificial neural networks (ANNs) for time series properties, such as the adaptive time delay neural networks (ATNNs) and the time delay neural networks (TDNNs), with the genetic algorithms (GAs) in detecting temporal patterns for stock market prediction tasks. The results indicated that the accuracy of the integrated approach proposed for the study was higher than that of the standard ATNN, TDNN and the recurrent neural network (RNN). Chang, Wang, and Yang (2004) attempted to develop a hybrid model that integrated the technical and basic analyses to forecast the stock market index more accurately. Khoa, Sakakibara, and Nishikawa (2006) proposed an algorithm to forecast the stock price by using indicators such as inflation rate, GDP, relative strength index, directional index, daily high, daily low, closing price and moving averages as inputs to feed-forward neural network and simple recurrent neural network.

Diğer, Timor & Emir (2012) developed a novel hybrid model by employing data set of technical and fundamental analysis variables obtained after dimension reduction process to SVM & ANN and their study reported that the models created by means of the SVM method presented more successful results over ANN. On the contrary, the findings of Kara, Boyacıoğlu, and Baykan (2011) suggested that the average prediction performance of the ANN model was significantly better than that of the SVM model.

Artificial Intelligence for stock market prediction has grabbed fewer eyeballs in the Indian financial market. Therefore, this paper encapsulates market indicators with AI techniques to generate useful extracts to improve decisions under conditions of uncertainty. Three approaches (fundamental model, technical indicators model and hybrid model) have been tested using the standalone and integrated machine learning algorithms viz. SVM, ANN, GA-SVM, and GA-ANN and the results of all the three approaches have been compared in the four abovementioned methods.

Thus, the core objective of this paper is to identify a system from the four most popular methods so far, that best predict the Indian stocks price movement using quarterly data.

3. THEORETICAL BACKGROUND

3.1 Genetic Algorithm

Sometimes the data variables may be quite a lot in number for many pragmatic problems and with a large data set there may be presence of some redundancy among the variables. Furthermore, too many input features may lead to "curse of

dimension” (Yang and Hanover, 1998). This is where feature selection techniques come to rescue. Feature selection is the process of selecting a subset of relevant features for use in model construction. The main idea is to remove redundant or irrelevant features from the data set as they lead to a reduction of the classification accuracy or clustering quality and to an unnecessary increase of computational cost (Blum and Langley, 1997). The advantage of feature selection is that no important information of any feature is lost.

Genetic Algorithms are by far the most popular stochastic optimization method as they often succeed in finding the best optimum in contrast to most common optimization algorithms. Genetic algorithms are complex problem solving techniques that utilize the concepts of natural selection to obtain an optimal subset of features which produce the best possible results. In a genetic algorithm, solutions with best fitness are taken from a population to form new solutions. Each solution has a set of properties called chromosomes and solutions are represented in binary as strings of 0s and 1s called genes which, in turn, are mutated and altered to ensure that the new population is better than the old one. A chromosome consists of as many genes as parameters to be optimized (Lean Yu et al, 2006).

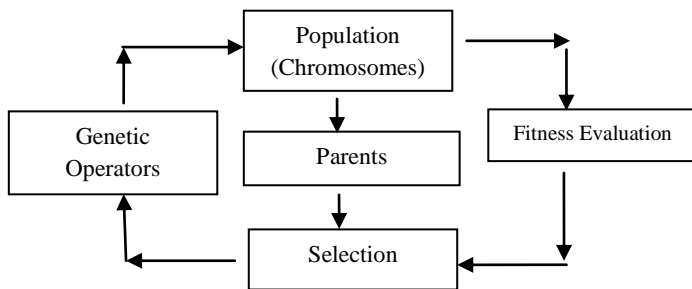


Figure 1: Cycle of a Genetic Algorithm’s operation

Genetic Algorithm Cycle - a) Creation of a random population of chromosomes, b) Evaluation of fitness of each chromosome, c) Selection of best chromosomes and d) Genetic manipulation to create new population of chromosomes.

Each cycle (Fig.1) in Genetic Algorithm produces a new generation of possible solutions for a given problem. Genetic algorithm begins with randomly generated set of individual solutions represented by chromosomes called population. The population size contains hundreds or thousands of chromosomes depending on the nature of the problem. The fitness or the performance of the chromosome is then evaluated with the help of fitness based process which ensures that only the candidates with higher fitness value are selected for further genetic manipulation process. After selection of fitter solution is over, the genetic manipulation process consisting of two steps is carried out. In the first step, the crossover operation recombines genes of two strings (chromosomes) to get better and different individuals in the successive generations which possess characteristics similar to both the parents. The second step in the process is mutation, where the bits at one or more randomly selected positions of the chromosomes are altered to avoid getting trapped local optima. Mutation also introduces diversity in the population whenever the population tends to become homogeneous. The offsprings produced by the genetic manipulation process are the next population to be evaluated. The whole process is repeated until specific termination conditions are satisfied.

Genetic Algorithm is selected over traditional optimization techniques because (Sahil Kumar, 2012) -

- i. GA works with a string coding of variables instead of variables
- ii. GA processes more than one string simultaneously and thus works with a population of solutions instead of one solution at a time
- iii. GA operators find the similarities in string structures and hence may find the global best solution

So for the purpose of this study, dimension reduction using genetic algorithm is implemented prior to the classification procedure to determine the relevance of attributes with the intention of improving the accuracy of prediction.

3.2 Support Vector Machine

Support vector machine is a useful machine learning technique for data classification based on statistical theory. It is a supervised learning method used for binary classification that constructs a hyper plane to separate the N-dimensional data into two categories. Initially, SVMs were developed for pattern recognition problems but recently with Vapnik’s proposed ϵ -insensitive loss function, SVMs have also been rendered to solve non-linear regression problems. Built on the theory of structural risk minimization principle, it estimates a function by minimizing an upper bound of the generalization error rather than minimizing the mean square error. SVM is shown to be very resistant to the over-fitting problem, eventually achieving a high generalization performance (Huang et al., 2004).

SVM algorithm functions by identifying the hyperplane that gives the largest minimum distance to the training examples. The dashed lines drawn parallel to the hyperplane indicate the difference between the separating line and the closest feature to the line, and the difference between the dashed lines is termed as “margin”. The optimal separating hyperplane maximizes the margin of the training data.

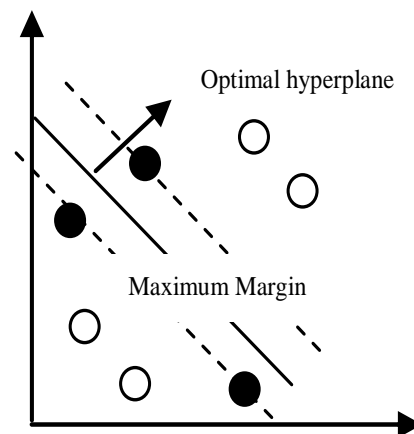


Figure 2: Support vector Machine

- **Support Vectors**

The data points that lie closest to the decision surface are called support vectors. They have a direct bearing on the optimum location of the decision surface as they are the most difficult ones to classify (Yang, 2003). In fig. 2, support vectors are shown as the black circles.

• **Kernel Function**

Though the basic idea of SVM is to find an optimal hyperplane for linearly separable patterns, its use can also be extended to non-linearly separable patterns by using kernel functions which transform the original data to map into new space.

The kernel functions return the inner product between two points in a suitable feature space, thus defining a notion of similarity with little computational cost even in very high-dimensional spaces (Karatzoglou and Meyer, 2006). SVM transforms the inputs into the high dimension indicators range using a kernel function. Various kernels are used to generate the internal products to build SVM with various models of nonlinear equations in the input range (Khatibi et al., 2011) but the most commonly used ones are-

Gaussian RBF:

$$K(a; b) = \exp\left(-\frac{1}{\delta^2} (a - b)^2\right) \quad (1)$$

Polynomial function: $K(a; b) = (ab+1)^n$ (2)

Where n is the degree of the polynomial kernel and δ^2 is the bandwidth of the Gaussian RBF kernel.

• **C and Epsilon value**

Other SVM parameter settings include C value and epsilon value, the former tells the SVM optimization the extent to which misclassifying of each training example is to be avoided and the latter determines the level of accuracy of the approximated function.

Determining the value of C and epsilon is an important part of the process because a large value of C chooses a smaller-margin hyperplane which enables correct classification of the training points but an infinitely large value of C would not allow the occurrence of any error and result in a complex model. Conversely, a very small value of C causes the optimizer to look for a larger-margin separating hyperplane, even if that hyperplane misclassifies more points.

In case of epsilon, if the value is larger than the range of the target values, a good result cannot be expected and if epsilon is zero, overfitting problems can be experienced.

The methodology of SVM for a two class classification problem is stated briefly as follows: Assuming that there are a set of input vectors $x_i \in R^d (i = 1, 2, \dots, N)$ with corresponding labels $y_i \in \{+1, -1\} (i = 1, 2, \dots, N)$. Here, +1 and -1 indicate the two classes. In a learning problem, the aim is to provide generalization to unseen data points. This means that given some new $x_i \in X$, corresponding label $y_i \in \{+1, -1\}$ is to be predicted.

Decision function of the SVM classifier is given as:

$$f(x) = \text{sgn}(\sum_{i=1}^N y_i \alpha_i \cdot K(x, x_i) + b) \quad (3)$$

The quadratic programming problem to determine the coefficients α_i (Hua & Sun, 2001) is –

$$\text{Max} \sum_{i=1}^N \alpha_i - \frac{1}{2 \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \cdot y_i y_j \cdot K(x_i, x_j)} \quad (4)$$

Subject to $0 \leq \alpha_i \leq c$

$$\sum_{i=1}^N \alpha_i y_i = 0, \quad \text{where, } i = 1, 2, \dots, N$$

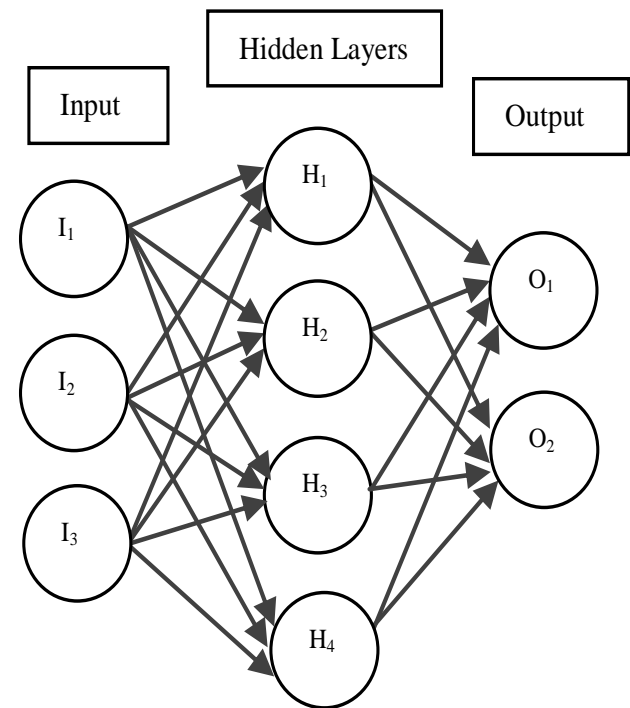
Here C is called regularization parameter which controls the trade-off between classification error minimization and margin maximization. The α_i coefficients are called Lagrange

multipliers and represent support vectors which define optimal separating hyper plane (Timor et al., 2012).

3.3 Artificial Neural Network

Artificial Neural Network is a computational model that simulates a biological brain. ANNs are regarded as nonlinear statistical data modeling tools in which the patterns between the inputs and the outputs are found. A large number of input patterns are fed to NN and the resultant output of each input pattern is showed in order to train the neural network. Training helps the NN in recognizing similarities when presented with a new input pattern, resulting in a predicted output pattern.

This study utilizes three-layer multilayer perceptron model (one hidden layer), also called as feedforward neural network model. This type of neural network has the perceptrons arranged in layers, with the first layer taking in the inputs which sends data on to the second layer, which inturn sends the output neurons to the last layer producing outputs. Generally no computation is done at the input layer; rather all the data processing activities take place at the middle (hidden) layer. Each perceptron in one layer is connected to every perceptron on the next layer. Hence the information is constantly “fed forward” from one layer to the next.



forward neural network

Similar to econometric models, input layer includes independent variable(s) and output layer includes dependent variable and bias term. Each connection from the input layer to the hidden layer or from the hidden layer to the output layer has a weight. These weights represent the coefficients or the parameters of the model and the size of each weight represents the relative strength of the connection (Mazibaş and Küçüközmen, 2007).

The input to each neuron in the hidden layer is calculated by

$$a = \sum_{i=0}^n w_i x_i \quad (5)$$

Where x_i represents the inputs to the neuron and w_i represents the weights of the neuron.

Now in order to map the input variables to the output variables i.e. to normalize the sum into a standard range, a special function known as sigmoid threshold function is used as it performs better than other activation functions like the Unit Step function, Piecewise linear function, Binary Transfer function, and the Gaussian function.

A sigmoid function is defined as $f(a)=1/1+e^{-a}$

The output layer neurons then receive a weighted sum of the hidden unit's output and apply an activation function to this sum. The architecture of the three-layered feedforward ANN is illustrated in Fig. 3.

Important parameters of NN are –

- **Learning rate** - determines how fast the neural network learns the pattern in the training data. Too small a value makes the learning process slow, but a large one might lead to divergence.
- **Momentum** - influences the way past weight changes affect the current ones. It helps in preventing the system from converging to a local minimum.
- **Epoch** – It is a measure of the number of times all of the training vectors are used to update the weights.

3.4 Fundamental Analysis

The Fundamental factors sourcing from the micro-environment of companies i.e. financial ratios have been used to model the stock price movement. Financial Ratios are useful indicators of a firm's financial situation as they evaluate the performance of the business and identify potential problems. Over the years, ratios have successfully been employed as a management tool to gauge the effectiveness of firms operations by illustrating strengths and weaknesses of the business during the performance management process. They have proven to be reliable and useful in the financial analysis and forecasting process, thereby, helping an organization in forming strategies and investors in making profitable portfolios of securities.

3.5 Technical Analysis

Technical analysis is a security analysis methodology for forecasting the direction of prices by analyzing statistics generated by market activity viz. past prices and volume. The basic premise is to look at past price behavior in an attempt to determine where prices are headed to in the future with a fair amount of classification accuracy. Technical analysis assumes that stock prices move in trends, and that the information which affects prices enters the market over a finite period of time, not instantaneously (Choudhry and Garg, 2008).

Technical indicators are very instrumental in maximizing trading profits and minimizing losses and are commonly used to enter and exit trades.

4. DATA AND METHODOLOGY

The study utilizes quarterly financial ratios and technical indicators data from April 1, 2001 to December 31, 2013 for the purpose of predicting the movement of 25 stocks of S&P CNX Nifty. The selection criterion of the companies is based on timely availability of data on Bombay Stock Exchange Website. Indian Companies have been chosen because most of the work in this area in past has predominantly engaged developed stock markets for the purpose of forecasting and there exists little published work on Indian stocks movement prediction using AI techniques, thus, making it all the more indispensable as previous studies have witnessed that different

stock markets have different characteristics and results obtained for one may not necessarily hold true for another (Chen and Shih, 2006). The data is not adjusted for splits and bonuses.

The 25 companies that have been included for the purpose of carrying out the research are – ONGC Ltd., Reliance Industries Ltd., ITC Ltd., Infosys Ltd., HUL Ltd., Sun Pharmaceutical Industries Ltd., Larsen Toubro Ltd., Mahindra Mahindra Ltd., Bharat Heavy Electricals Ltd., Asian Paints Ltd., NMDC Ltd., Gail (India) Ltd., Dr Reddy's Laboratories Ltd., Grasim Industries Ltd., Bharat Petroleum Corporation Ltd., Hero Motocorp Ltd., Hindalco Industries Ltd., Lupin Ltd., Cipla Ltd., Ranbaxy Laboratories Ltd., Reliance Infrastructure Ltd., Sesa Goa Ltd., HDFC Ltd., HCL Technologies Ltd., Axis Bank Ltd.

Technical parameters (n=4 in all the indicators) - Mass Index (MI), Average True Range (ATR), Momentum, Chaikin Money Flow (CMF), Commodity Channel Index (CCI), Exponential Moving Average (EMA), Relative Strength Index (RSI), Stochastic Oscillator, William %R, Rate of Change (ROC), Price and Volume Trend (PVT), On Balance Volume (OBV), MACD(6,13).

Fundamental Variables – Growth in net sales, Growth in net profit, Return on Equity, P/E ratio, Net Profit Margin (NPM), Price/Sales ratio.

The stocks direction problem is modeled as a two class classification problem. The directions are classified as positive and negative in the dataset which indicate a stock's gaining and losing status respectively in a particular quarter. The performance of the stock movement prediction system is evaluated using the 10- folds cross validation method in all the aforementioned AI techniques.

4.1 Proposed System

Three different models are investigated with GA, SVM and ANN for the predictability of financial movement direction. The inputs to the first model purely consist of financial ratios, the inputs to the second model comprise of technical indicators, while, the inputs to the third model combine both fundamental and technical variables. All the models are tested in two sets i.e. the standalone approach (SVM, ANN) and the integrated approach (GA-SVM, GA-ANN). Comparison of both the approaches is carried out with the purpose of examining whether the application of data dimensionality reduction pre-processing step prior to the classification procedure improves the overall classification performance or not. The AI techniques are implemented using WEKA. The procedure is as follows –

- Model 1 – Fundamental data
 - ✓ Model 1.1 – Runs SVM and ANN with all their parameter configurations on the entire fundamental data set to obtain the accuracy % of stock movement prediction
 - ✓ Model 1.2 – only the optimal fundamental features selected by GA are fed to SVM and ANN
- Model 2 – Technical data
 - ✓ Model 2.1 – Applies SVM and ANN to all the technical variables to evaluate their prediction ability

- ✓ Model 2.2 – Employs SVM and ANN only to GA optimized technical variables
- Model 3 – Explores the combination of technical and fundamental indicators to check if improved stock movement prediction is obtained.
 - ✓ Model 3.1 – inputs ‘all’ the technical and fundamental indicators to standalone SVM and ANN
 - ✓ Model 3.2 – inputs ‘optimal set’ of fundamental and technical features selected by GA to SVM and Neural Networks.

The parameter combinations that result in the best classification accuracy in all the aforementioned models are noted, recorded and compared.

4.2 AI Parameter Settings

The optimum combinations of parameters of the algorithms need to be determined for best accuracy. An exhaustive search is carried out with different ANN and SVM configurations to determine the best performance in each model. Selection of Genetic Algorithm parameters needs to be paid heed to for finding good solutions because choosing unsuitable parameters might result in bad optimization results. Traditionally, Trial and error method has been used to tailor the search for a specific problem space but there are some popular standard settings like DeJong settings (DeJong and Spears, 1990), Grefenstette settings (Grefenstette, 1986) and MicroGA settings (Carroll, 1996) which have been proposed to give fairly good results. The study utilizes all the 3 standard parameters theories and compares the results to be able to reach to that one theory which gives best results. For Neural Networks, Ep = 500, 1000, 2000, 5000 and momentum = 0.1, 0.2,0.9 and lr = 0.1 are used since they are tested in previous studies conducted by Adebisi et al. (2012) and Kara et al. (2011) respectively. Different parameter settings of Genetic Algorithm, SVM and ANN are summarized in the table 1, table 2 and table 3 respectively.

Table 1. Genetic Algorithm

Parameters	DeJong settings	Grefenstette settings	MicroGA settings
Population size	50	30	5
Number of generations	1000	1000	100
Crossover rate	0.6	0.9	0.5
Mutation rate	0.001	0.01	0.2

Table 2. Support Vector Machine (SVM)

Parameters	Configurations
C	1,10,100
Epsilon (ϵ)	1.0E-12, 1.0E-6, 1.0E-3
Kernel Functions	Polynomial, Radial basis

Table 3. Artificial Neural Network (ANN)

Parameters	Configurations
Epochs (Ep)	500, 1000, 2000, 5000
Momentum (m)	0.1, 0.2, , 0.9
Learning rate (lr)	0.1

A total of 441 combinations are explored to gauge how various architecture types of GA, SVM and ANN along with different data inputs affect quarterly stock movement predictions. Population size, Number of generations, Crossover rates and Mutation rates have been adjusted according to the standard parameters proposed in GA for fundamental, technical and hybrid model, yielding a total of 9 treatments for the same. The choice of 3 levels of Cost of error (C), 3 levels of epsilon (ϵ) and 2 kernel functions for all the three aforementioned models in SVM have resulted in a total of 144 combinations, while a figure of 288 treatments is obtained in ANN by employing 9 levels of momentum constant (mc) and 4 levels of epochs in the parameter setting experiments. A small value of 0.1 for lr has been fixed because the prediction performance of ANN model decreased when higher values of lr (Fig. 4) were experimented with.

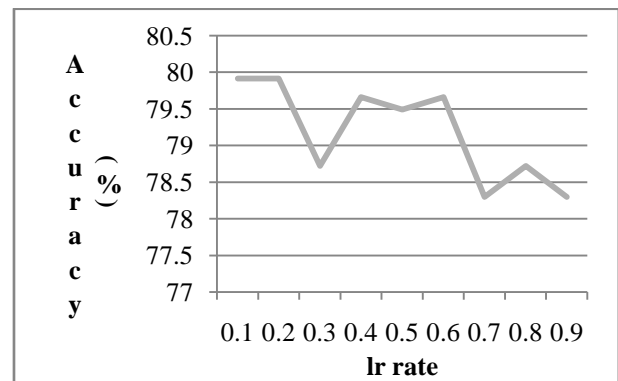


Figure 4: Decreasing accuracy with increasing lr rate

5. EXPERIMENTAL RESULTS

• Genetic algorithm

GA is based on the survival of the fittest tenet of Darwinian evolution which is why the study applies GA to identify the fittest variables in all the models using standard parameters settings. The variables suggested by different standard settings of GA to improve the classification performance are as follows –

Table 4. Model 1, Model 2 and Model 3 variables

Standard Settings	Model 1	Model 2	Model 3
Dejong settings	EPS	CCI, Stochastic Oscillator and PVT	CCI, Stochastic Oscillator, PVT, RSI, Williams %R and MI
Grefenstette settings	NPM	CCI, Stochastic Oscillator, PVT, RSI, Williams %R	CCI, Stochastic Oscillator, PVT, RSI, Williams %R and MI
MicroGA settings	NPM	CCI, Stochastic Oscillator, PVT, RSI, Williams %R	CCI, Stochastic Oscillator, PVT, RSI, Williams %R and MI

It is evident from table 4 that Grefenstette settings and Micro GA settings gave the same set of output for Model 1 and Model 2 while DeJong chose a different set of variables, but all the three settings chose exactly same variables for Model 3. Fig. 5 shows that the variables chosen by Grefenstette and MicroGA give better accuracy results than DeJong, the reason can be attributed to the fact that small population size and relatively large mutation rate is far superior to the large population sizes and low mutation rates (Haupt, 2000). Therefore, for the purpose of this study, the variables suggested by Grefenstette and MicroGA are taken forward for use in SVM and ANN.

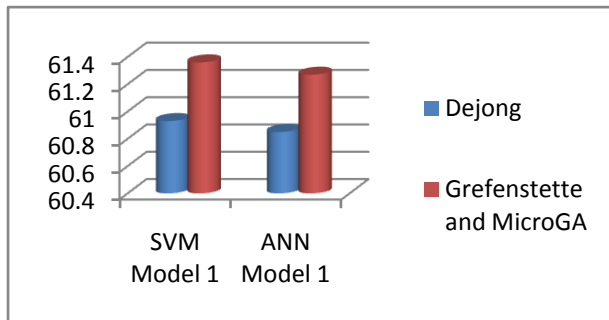


Figure 5: Comparison of DeJong settings with Grefenstette and MicroGA Settings

It is interesting to note here that in the hybrid model i.e. Model 3; Genetic Algorithm has picked only the technical indicators as the fittest variables to predict the stocks movements out of the input set of technical + fundamental indicators. The possible reason can be the time interval for which the data has been extracted i.e. quarterly data, technical data is by far considered to be the best indicator of where the stocks are headed to in short term.

Another fascinating thing to be observed here is that the GA has selected only the momentum indicators in Model 2 and Model 3 because momentum indicators are the ones that come closest to predicting the changes in prices of stocks by using historical price and volume data. The GA elected indicators have OHLCV data covered that identify new trends, measure speed and change of price movements, identify trend reversals and provide excellent entry and exit signals which basically form the crux of technical analysis.

The best prediction performances as suggested by the empirical experimentation are obtained at the parameter combinations listed in table 5 and 6 for each model.

Table 5. Performance of all the models in SVM

Models	C	Kernel	Epsilon	Highest Accuracy
1.1	10	Poly	0.001	61.02%
1.2	100	RBF	1.0E-6	61.36%
2.1	10	Poly	1.0E-12	79.40%
2.2	100	Poly	1.0E-12	78.21%
3.1	10	Poly	1.0E-12	79.23%
3.2	100	Poly	1.0E-12	78.97%

Table 6. Performance of all the models in ANN

Models	Ep	M	Lr	Highest Accuracy
1.1	500	0.1	0.1	61.27%
1.2	2000	0.1	0.1	61.27%
2.1	500	0.2	0.1	78.46%
2.2	1000	0.1	0.1	80.25%
3.1	500	0.1	0.1	77.61%
3.2	5000	0.1	0.1	80.51%

Table 5 and 6 illustrate the level of best accuracy achieved by each model in SVM and ANN respectively. The prediction performance of the ANN configurations in model 1.1 and model 1.2 varied from 60.4255% to 61.2766% and 60.766% to 61.27% respectively while the variation ranged from 76.9362% to 78.4681% and 77.7% to 80.25% in model 2.1 and model 2.2 respectively. The variation for Model 3.1 stood between 73.7021% and 77.617% and for Model 3.2 from 79.06% to 80.51%.

The accuracy rates of SVM parameter settings in model 1.1 and model 1.2 fluctuated from 60.5106% to 61.0213% and 55.74% to 61.36% respectively, while the fluctuation figure ranged from 72.766% to 79.404% and 73.1% to 78.21% in model 2.1 and model 2.2 respectively. Model 3.1 and Model 3.2 produced a variation from 72.766% to 79.234% and 77.78% to 78.97% respectively.

As is evident above, numerous combinations were employed, only the best results of each model are listed in table 5 and 6. The parameter structure that returns the highest accuracy % is noted to give the best forecast. The larger the accuracy rates, the closer are the predicted values to that of the actual values. Though, it is important to note here that the best performances were not obtained at the same parameter combinations for the models. After performing several experiments with different configurations of SVM and ANN, the most accurate results are obtained at -

• **SVM**

The investigation shows that C and ϵ value of 10 and 1.0E-12 are found to produce the best possible results. However, C=100 gives best results when GA features are fed to SVM. Empirical results also witness that polynomial SVM has better forecasting ability than radial basis SVM.

• **ANN**

After testing four levels of epochs i.e. 500, 1000, 2000, 5000 and nine levels of momentum rates, table 6 makes it apparent that momentum rates work best on the lower end and epoch =5000 results in higher classification accuracy.

The empirical results are analyzed in 3 instances –

5.1 Comparison between model 1.1, model 1.2, model 2.1, model 2.2, model 3.1 and model 3.2

When the study is considered according to methods, it can be seen that the empirical outputs are in the favor of hybrid model for which GA has chosen only the technical variables owing to high volatility in Indian stock market. Financial market in India has witnessed spectacular growth in recent years with the rising interest from foreign as well as domestic investors; it has presented opportunities to make money with every passing second which have been availed using technical analysis techniques. It is clear from table 5 and 6 that ANN-model 3.2 is best of all the twelve predictive models; it shows a remarkable improvement over the use of model 1, though the results are very close to that of ANN - Model 2.2. It is important to note here that though GA has narrowed down only the technical indicators for the hybrid model but the selected set is not the same as the one in technical model, the GA proposed variables for hybrid model have one extra variable along with the ones chosen in technical model. Hybrid model has Mass Index as an addition in its basket which is probably responsible for the increased accuracy over technical model. Thus, the classification accuracy of technical indicators/integrated indicators beats that of the fundamental model and hence can be successfully used to enhance the profits of investors or traders who like to hold their shares for a period of 3 months.

Chavan and Patil (2013), and Adebisi et al. (2012b) also show that the hybridized approach has the potential of enhancing the quality of decision making of investors in the stock market by offering more accurate stock prediction compared to existing technical analysis based approach.

5.2 Comparison of standalone SVM and ANN with GA-SVM and GA-ANN respectively

When the results of standalone SVM and ANN are compared with GA-SVM and GA-ANN, the results are found to match very closely, but, GA-ANN model has outperformed standalone ANN model as against SVM wherein standalone SVM has a better accuracy rate over GA-SVM. GA didn't prove to be useful with SVM because SVM treats higher dimensional data with a relatively low amount of training set (Thissen, 2003). However, GA-ANN transcends standalone SVM also. Thus, in ANN, GA proves to be a critical step in the design of pattern classifier for it assists in selecting a pre specified number of features leading to the best possible performance of the classifier. Thus, the use of hybridized indicators along with hybridized AI techniques very much holds in the Indian context too.

The results are in sync with the results of existing studies in the literature (Khatibi and Rasouli, 2011; Wei et al., 2010; Kim and Han, 2000; Choudhry and Garg, 2008), where it is argued that integrated machine learning forecasting techniques outperform stand alone machine learning forecasting techniques.

The accuracy rates of the model 1.1, 1.2, 2.1, 2.2, 3.1 and 3.2 for both the algorithms are presented in table 5 and table 6 respectively.

5.3 Comparison of the best accuracy model created by means of SVM with the one created by means of ANN

Both ANN and SVM models have shown significant performance in predicting the direction of stock price movement. The findings show that ANN – model 3.2 returns the smallest RMSE (root mean squared error) along with the highest classification accuracy (table 7) on account of “deep architectures” which can represent intelligent behavior or functions more efficiently than “shallow architectures” like that of SVMs, thereby, confirming performance of ANN over SVM model. Many advantages of SVM over ANN from theoretical point of view have been identified; however, such drawbacks of ANN are not always identified as practical issues. Thus, ANN provides a promising alternative for stock movement prediction over SVM for ANNs are universal approximators.

A study conducted by Kara et al. (2011) also affirmed the average prediction performance of the ANN model (75.74%) was found significantly better than that of the SVM model (71.52%) as opposed to the study of Dinçer, Timor & Emir (2012) which reported that the models created by means of the SVM method presented more successful results over ANN.

Table 7. Best Model performers in both the algorithms

Models	Approach	Highest Accuracy	RMSE
Model 2.1	SVM	79.40 %	0.3705
Model 3.2	ANN	80.51%	0.3694

6. CONCLUSION

The study investigated the use of 3 methods viz. financial ratios, technical indicators and hybridized approach along with 4 forecasting tools i.e. SVM, ANN, GA-SVM, GA-ANN to predict stock movement since traditional regression models have proven to be ineffective in out – of – sample prediction tests. The paper utilizes Support vector machine and artificial neural network for classification, and genetic algorithm for feature selection. All the three methods have been tested in the aforementioned four tools using the data of 25 companies of S&P CNX Nifty.

The study is significant in several ways. First, most of the studies in this area have been focused on predicting American and Korean stock prices; this paper sheds light on the method - tool combination that works best for Indian stocks. Second, the results obtained in this study favor hybrid model, where in, GA has included only the technical indicators, thereby, reporting that the performance levels of the technical model are very useful for the investors and the regulators in a highly volatile market like Indian stock market. Finally, not much of the research efforts have been put in examining combined fundamental and technical analysis parameters in the Indian context.

The findings can be summarized as –

- It is observed that the performance of ANN method created by means of genetic algorithm improved significantly. Therefore, it can be said that GA selects the informative features and discards the redundant ones
- Hybrid indicators with hybrid AI techniques offer superior performance over other methods.

- Empirical evidence shows that both SVM and ANN have performed significantly in predicting the direction of stock price movement, but ANN is found to offer better performance when compared with SVM.

The findings can aid financial practitioners in enhancing their investment performance by suggesting proper asset allocation, and academicians in understanding the nature of stock return predictability in order to produce more realistic asset pricing models that better explain the data.

The scope in this area can be further extended to including fundamental factors sourcing from macro – economic environment that affect stock movements and have the potential of enhancing the quality of prediction. Furthermore, considering the use of other AI techniques might yield better results in a model of stock selection.

It can be concluded from the present study that using hybrid indicators/technical analysis with GA-ANN is well suited for Indian stocks and helps the investors and traders in maximizing their profits by predicting accurately in advance the stocks making their way in the green basket on quarterly basis.

7. REFERENCES

- [1] Adebisi, A.A., Ayo, C.K., Adebisi, M.O., and Otokiti, S.O. 2012a. Stock Price Prediction using Neural Network with Hybridized Market Indicators. *Journal of Emerging Trends in Computing and Information Science*. 3(1), 1-9.
- [2] Adebisi, A.A., Ayo, C.K., Adebisi, M.O., and Otokiti, S.O. 2012b. An Improved Stock Price Prediction using Hybrid Market Indicators. *African Journal of Computing & ICT*. 5(5), 124–135.
- [3] Atiya, A., Noha, T., and Samir, S. 1997. An efficient stock market forecasting model using neural networks. *Proceedings of the IEEE International Conference on Neural Networks*. 4, 2112-2115.
- [4] Avci, E. 2007. Forecasting Daily and Sessional Returns of the Ise-100 Index with Neural Network Models. *Dogus Universitesi Dergisi*. 2(8), 128-142.
- [5] Blum, A.L. and Langley, P. 1997. Selection of relevant features and examples in machine learning. *Artificial Intelligence*. 97, 245-27.
- [6] Chang, P.C., Wang, T.W., and Yang, W.N. 2004. An investigation of the hybrid forecasting models for stock price variation in Taiwan. *Journal of the Chinese Institute of Industrial Engineering*. 21(4), 358–368.
- [7] Chavan, P.S. and Patil, S.T. 2013. Parameters for Stock Market Prediction. *International Journal of Computer Technology and Applications*. 4(2), 337-340.
- [8] Chen, W.H., Shih, J.Y. 2006. Comparison of support-vector machines and back propagation neural networks in forecasting the six major Asian stock markets. *International Journal of Electronic Finance*. 1, 49-67.
- [9] Choudhry, R. and Garg, K. 2008. A Hybrid Machine Learning System for Stock Market Forecasting. *World Academy of Science, Engineering and Technology*. 4, 315 – 318.
- [10] Chung, H.K., Phua, P.K.H., and Zhu, X. 2003. Forecasting Stock Index Increments using Neural Networks with Trust Region Methods. *International Symposium of Neural Networks*. 1, 260-265.
- [11] Dinçer, H., Timor, M., and Emir, Ş. 2012. A Stock Selection Model Based on Fundamental and Technical Analysis Variables by Using Artificial Neural Networks and Support Vector Machines. *Review of Economics & Finance*. 2, 106-122.
- [12] Han, S. and Chen, R-C. 2007. Using SVM with Financial Statement Analysis for Prediction of Stocks. *Communications of the IIMA*. 7(4), 63-72.
- [13] Haupt, R.L. 2000. Optimum population size and mutation rate for a simple real genetic algorithm that optimizes array factors. *Antennas and Propagation Society International Symposium. IEEE*. 2, 1034-1037.
- [14] Iqbal, Z., Ilyas, R., Shahzad, W., Mahmood, Z., and Anjum, J. 2013. Efficient Machine Learning Techniques for Stock Market Prediction. *Journal of Engineering Research and Applications*. 3(6), 855-867.
- [15] Kara, Y., Boyacioglu, M.C., Baykan, O.K. 2011. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Journal of Expert Systems with Applications*. 38(5), 5311-5319.
- [16] Karatzoglou, A. and Meyer, D. 2006. Support Vector Machines in R. *Journal of Statistical Software*. 15(9), 1-28.
- [17] Khatibi, E., Rasouli, A., and Khatibi, V. 2011. A New Support Vector Machine – Genetic Algorithm (SVM-GA) based method for Stock Market Forecasting. *International Journal of the Physical Sciences*. 6(25), 6091-6097.
- [18] Khoa, N.L.D., Sakakibara, K., Nishikawa, I. 2006. Stock Price Forecasting using Back Propagation Neural Networks with Time and Profit Based Adjusted Weight Factors. *SICE-ICASE International Joint Conference*, 5484 – 5488.
- [19] Kim, H-J. and Shin, K-S. 2007. A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. *Applied Soft Computing*. 7(2), 569- 576.
- [20] Kim, K.J. and Han, I. 2000. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Journal of Expert Systems with Applications*. 19(2), 125-132.
- [21] Kim, S.K. and Chun, S.H. 1998. Graded forecasting using an array of bipolar predictions: Application of probabilistic neural networks to a stock market index. *International Journal of Forecasting*. 14(3), 323-337.
- [22] Lee, C-T. and Chen, Y-P. 2007. The efficacy of neural networks and simple technical indicators in predicting stock markets. *International Conference on Convergence Information Technology*, 2292-2297.
- [23] Mazibaş, M. and Küçüközmen, C.C. 2007. Forecasting the change and direction of change in ISE Stock Indices: A Neural Network Application. *Proceedings Book of the Fourth International Conference on Business, Management and Economics*.

- [24] Mitra, S.K. 2009. Optimal Combination of Trading Rules Using Neural Networks. *International Journal of Business Research*. 2(1), 86-99.
- [25] Nair, B.B., Mohandas, V.P., and Sakthivel, N.R. 2010. A Decision Tree – Rough Set Hybrid System for Stock Market Trend Prediction. *International Journal of Computer Applications*. 6(9), 1-6.
- [26] Nagarajan, V., Wu, Y., Liu, M., and Wang, Q-G. 2005. Forecast Studies for Financial Markets using Technical Analysis. *International Conference on Control and Automation (ICCA2005)*. 1, 259-264.
- [27] O'Connor, N. and Madden, M.G. 2006. A Neural Network Approach to Predicting Stock Exchange Movements using External Factors. *International Conference on Innovative Techniques and Applications of Artificial Intelligence*. *Journal of Knowledge-based systems*. 19(5), 371-378.
- [28] Pantazopoulos, K.N., Tsoukalas, L.H., Bourbakis, N.G., Brün, M.J., and Houstis, E.N. 1998. Financial Prediction and Trading Strategies using Neurofuzzy Approaches. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics* 28(4): 520-531, August.
- [29] Quah, T.S. and Srinivasan, B. 1999. Improving returns on stock investment through neural network selection. *Journal of Expert Systems with Applications*. 17(4), 295-301.
- [30] Raposo, R.D.C.T. and Cruz, A.J.D.O. 2002. Stock Market Prediction based on Fundamental Analysis with Fuzzy Neural Networks. *Proceedings of 3rd WSES International Conference on Fuzzy Sets & Fuzzy Systems (FSFS'02)*. *Neural Networks and Applications (NNA'02)*. *Evolutionary Computation (EC'02)*.
- [31] Sap, M. and Awan, A.M. 2005. Stock Market Prediction using Support Vector Machines. *Jurnal Teknologi Maklumat*. 17(2), 27-35.
- [32] Schumaker, R.P. and Chen, H. 2008. Evaluating a News-Aware Quantitative Trader: The Effect of Momentum and Contrarian Stock Selection Strategies. *Journal of the American Society for Information Science and Technology*. 59(2), 247-255.
- [33] Stansel, S. and Eakins, S. 2004. Forecasting the direction of change in sector stock indexes: An application of neural networks. *Journal of Asset Management*. 5(1), 37-48.
- [34] Tsai, C.-F. and Wang, S.-P. 2009. Stock Price Forecasting by Hybrid Machine Learning Techniques. *Proceedings of the International MultiConference of Engineers and Computer Scientists*. 1, 755-760.
- [35] Tsanga, P.M., Kwoka, P., Choya, S.O., Kwanb, R., Nga, S.C., Maka, J., Tsangc, J., Koongd, K., and Wong, T.L. 2007. Design and Implementation of NN5 for Hong Stock Price Forecasting. *Journal of Engineering Applications of Artificial Intelligence*. 20, 453-461.
- [36] Wang, Y. and Choi, I. 2013. Market Index and Stock Price Direction Prediction using Machine Learning Techniques: An empirical study on the KOSPI and HIS. *Technical Report*.
- [37] Whitley, D. 1995. Genetic Algorithms and Neural Networks. *Genetic Algorithms in Engineering and Computer Science*. 3, 191-201.
- [38] www.slideshare.net/SahilKumar2/genetic-algorithms-artificial-intellig
- [39] Wu, W. and Xu, J. 2006. Fundamental Analysis of Stock Price by Artificial Neural Networks Model based on Rough Set Theory. *World Journal of Modelling and Simulation*. 2(1), 36-44.
- [40] Yang, J. and Hanovar, V. 1998. Feature subset selection using genetic algorithm. *Journal of IEEE Intelligent Systems*. 13, 44-49.
- [41] Yu, L., Wang, S., Lai, K.K., Huang, T., Zhou, C. 2006. Selecting Valuable Stock Using Genetic algorithm. *Simulated Evolution and Learning*. 688-694.
- [42] Zhang, Z., Li, M., Bai, R. 2012. An Integrated Model for Stock Price Prediction based on SVM and ARMA. *Journal of Theoretical and Applied Information Technology*. 44(1), 51-54.
- [43] Zhu, X., Wang, H., Xu, L., and Li, H. 2008. Predicting stock index increments by neural networks: The role of trading volume under different horizons. *Journal of Expert Systems with Applications*. 34(4), 3043-3054.