ISSN:
Print - 2277 - 0593
Online - 2315 - 7461
© FUNAAB 2019

Journal of Natural Science, Engineering and Technology

STOCK PRICE TREND PREDICTION USING SUPPORT VECTOR MACHINE AND CORAL REEF OPTIMIZATION ALGORITHM

*A. DEJI-OLARERIN, O. FOLORUNSO, O.R. VINCENT AND O.M. OLAYIWOLA

Department of Computer Science, Federal University of Agriculture, Abeokuta.
***Corresponding Author:** kemmid@yahoo.com **Tel:**

ABSTRACT

Due to non-linearity and non-stationary characteristics of stock market time series data, prior approaches have not been adequate enough for predicting stock market prices. Support vector machines are classifier that have been reported in the literature as having good recognition accuracy and have been applied in the area of predicting financial stock market prices and was found efficient. It is however noted that the performance of the SVM is affected by the values of the hyper-parameters used by the SVM. There is the need to find a way for searching for the best hyper-parameters that optimizes the performance of an SVM model. Coral Reef Optimization (CRO) is one of many nature-inspired algorithms used extensively to solve optimization problems. It is very effective in solving optimization problems because it is able to achieve global optimization. This paper's contribution is the development of Coral Reef search algorithms for the improvement of the hyper-parameters of the SVM used for stock price trend prediction. The Algorithm is validated using stock data of two banks. The results obtained out-performed un-optimized SVM, and have the same performance as that of SVM optimized with the FireFly optimization algorithm.

Keywords: Stock Market, Stock Price Trend Prediction, Support Vector Machine, Coral Reef Optimization, Stock Price index

INTRODUCTION

The shares of companies that are held publicly are traded in a market known as Stock market, which could be carried out via exchanges or over the counter markets (Burton and Shan, 2017). Stock market is as well referred to as equity market. It has become one of the most vital components that make up a free-market economy, in that access to capital is provided for companies in exchange of a slice of ownership in the company (Mani *et al.*, 2017). This market creates a possibility of growing small initial sums of money into large ones, and become wealthy without taking the risk of starting a business or making the sacrifices that often accompany a high-paying career. All of these advantages have not left out global crashes which take place based on local and regional crashes in emerging economies like the stock exchange (Dutta *et al.*, 2015).

Financial market has been discovered to be complex, non-linear dynamical and evolutionary (Basher *et al.*, 2012). There are several interacting factors in finance which include political events, general economic conditions, and traders' expectation that makes stock market price index difficult. The lack of exposure to emerging stock markets crashes that take place at local level has affected investors in the developing nations. Also, it is essential that shareholders and investors make use of financial information relevant in enabling them make the right investment decision into stock market (Dutta *et al.*, 2015). The stock price trend prediction and has been regarded as one of the most challenging application of time series prediction (Kara *et al.*, 2011) and existing stock market trend prediction systems usually focus on feature selection and prediction model ignoring the evaluating features which have led to the production of optimized results which Coral Reef Optimization algorithm addresses in this research work (Lin *et al.*, 2013). Due to non-linearity and non-stationary characteristics of financial stock market price, prior approach have not been adequate enough for predicting stock market. The problem of over fitting exists in the prediction of stock market data when flexible non-linear kernels are used in high-dimensional and high-noise datasets which will be addressed using the Support Vector Machine algorithm (Kuo *et al.*, 2016).

An efficient Artificial Intelligence algorithms such as Support Vector Machine (SVM) and Coral Reef Optimization (CRO) algorithms could aid training stock data to provide effective prediction. SVM in machine learning are supervised learning models with associated learning algorithms which analyze data used for classification and regression analysis (Kazeem *et al*., 2013). SVM is aimed at minimizing an upper bound of generation error (Kuo and Li, 2016). It consists of sum of training error and a confidence interval (Guo *et al.*, 2008). CRO algorithm has an excellent property of fast convergence to optimal values as well as enhance the performance of SVM. This paper presents a framework which is an improved scheme for predicting stock market trend using Support Vector Machine (SVM) and Coral Reef Optimization (CRO) algorithms. The framework was evaluated using stock market data from two financial institutions in Nigeria with considerations of key indicators discussed in this research. This paper is organized as follows: Section 2 gives a description of the approaches and limitation of existing research works conducted in the aspect of stock market prediction. The methodology, framework, algorithms, key indicators are presented in Section 3. Section 4 describes the implementation and results which includes the procedure of implementation, evaluation metrics and results. Section 5 presents the discussion and conclusion of the entire research work.

RELATED WORKS

A number of approaches have been employed in the prediction of stock market. Each have been able to address the lapses discovered but were not left without limitations which are addressed in this research work. Some of the existing approaches will be discussed in this section.

Shaikh *et al* (2014) proposed a data mining technique to model the relationship between company stock with other companies stocks. The predictive patterns generated helped in predicting prices movements. Sliding windows process illustration was shown to generate extended dataset. The extended rules could not be applied at any given time to predict upcoming stock and price with high accuracy. Lin *et al* (2013) presented SVM based approach consisting of two parts: feature selection and prediction model. The proposed SVM-based stock market prediction system found out a good subset. The stock indicators provided useful information for investors. Also, it helped reduce dimension and noise of financial data. Yet, more influencing factors affecting stock price prediction were not considered. The setting of parameters crucially impacted the performance of resulting system. There was no means of selecting an optimal value of the parameters in the proposed prediction system.

Kavousi-Fard *et al* (2014) proposed hybrid prediction algorithm which comprises of Support Vector Regression (SVR) and a Modified Firefly Algorithm (MFA) to provide the short term electrical load forecast using datasets from the Far Electrical Power Company in Iran. The SVR models makes use of the nonlinear mapping feature to deal with nonlinear regressions while the MFA was employed to accurately and effectively obtain SVR parameters. The technique had higher performance that others (Artificial Neural Network, Auto Regression Moving Average, and Empirical Risk Minimization) in terms of relative percentage error. It was able to enhance the search ability and convergence of firefly algorithm. The single approach implemented on only one domain was not enough to draw conclusion. The result did not benchmark with the use of CRO for optimization. Kazem *et al* (2013) proposed a forecasting model that was based on chaotic mapping, firefly algorithm and support vector regression (SVR) to make stock market price prediction. The method included a delay coordinate embedded method used to reconstruct unseen phase dynamics, a chaotic firefly algorithm to optimize SVM hyper parameters. The SVR that was optimized was used to forecast stock market price. The structural risk minimization implementation gave it high prediction accuracy. The space reconstruction phase in the preprocessing procedure makes the financial time series behavior more recognized for learning machines. The approach was more suitable than other methods for the description of the data relationship. The approach was implemented on only one aspect of financial forecasting. The approach was not compared with CRO algorithm.

Salcedo-Sanz *et al* (2014) discussed the performance of a novel Coral Reefs optimization-Extreme Learning Machine (CRO-ELM) algorithm for predicting daily global solar radiation. The approach combined the prediction problem solving ability of ELM and used CRO to evolves the weights of network so as to improve the obtained solutions. The approach was able to obtain a prediction that was accurate than the classical ELM and SVR algorithm. It was very effective in solving solar radiation problem by obtaining better results than alternative methods. The CRO was not implemented on stock market prediction. The approach was not merged together with any prediction algorithm like SVM. Kara *et al* (2011) attempted developing two efficient models and compared their performances in the prediction of Stock Market movement in Istabul. The models were based on Artificial Neural Network (ANN) and Support Vector Machine (SVM). They carried out comprehensive parameter setting experiments to improve the performances of their prediction. They demonstrated and verified the predictability of stock price index direction. The adjustment of the model parameter was not done so as to know the sensitivity and comprehensiveness of the parameter setting.

Kuo and Li (2016) proposed a three-stage forecasting model with the integration of wavelet transform, firefly algorithm-based K-means algorithms and firefly algorithm-based support vector regression (SVR). Wavelet transform was employed to remove noise in data preprocessing, firefly algorithm-based K-means for cluster analysis and SVR for the development of forecasting model for individual clusters. The wavelet transform aided the decrease of the impact of noise and outliers. The clustering analysis helped improve the performance by clustering data with similar features. The searching technique was expansive which helped locate optimal parameters. The measurement of the variable weight to the forecasting model was not considered. The optimization of the model was not considered.

Kim and Han (2016) presents enhanced classification for the stock index prediction. Here the ensemble method is used to predict the stock price. The bootstrap method is used to give more weights to the instances with big price changes and less weight to the instances with little price changes. The random forest algorithm is then used as a classifier. The method is tested on KOSPI, stock index. The proposed method shows a better performance in stock price index prediction however the research uses few technical indices to test their algorithm and no feature selection were carried out.

Sezer *et al* (2017) presents the use of a deep neural-network based stock trading system based on evolutionary optimized technical indicators. Relative Stock Index (RSI) technical indicator is extracted from stock data. The features are then optimized using genetic algorithm. The optimized features are then fed into a deep neural network for buy-sell predictions. The DOW 30 stocks were used to validate the model. The stocks were trained using daily close prices between 1996 -2016 and tested between 2007-2016. The algorithm shows a better performance when compared to un-optimized neural models. The limitation of the work is that only one technical indicator was used to train the model which may not be efficient.

Some research works predicted stock market without optimizing the output with an approach with high convergence property. Also a number did not address the over fitting issue of stock market values. These lapses are looked into in this research works.

PROPOSED METHODOLOGY

The methodology proposed in this research work is based on SVM and CRO to provide stock market price trend prediction that is efficient and accurate. The proposed framework as depicted in Figure 1 consists of raw data collection, data pre-processing, feature selection, model optimization and model testing. The architectural framework shows how the process of predicting the stock market trend will be done. From the raw dataset that will be pre-processed, inputted into the MATLAB environment for training and the data through the SVM processes. The coral reef optimization process does the optimization of the output from the SVM phase.

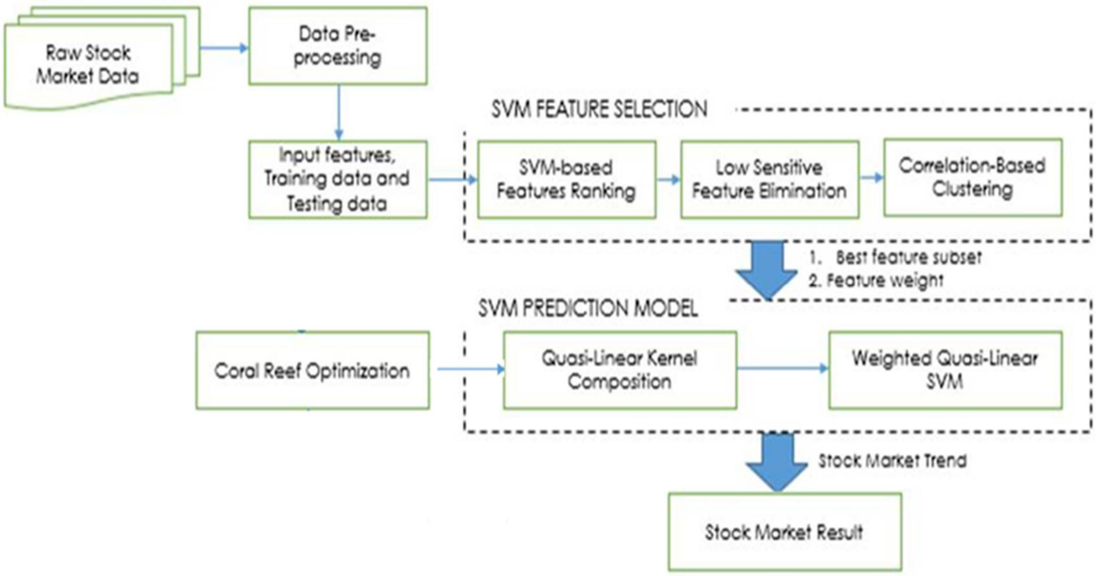

Figure 1: Architectural Framework of the SVM-Coral Reef Stock Prediction Model

Raw Stock Market Data

The research data was gotten from two financial institutions in Nigeria refers to as bank A and bank B in this paper. The information that was captured are the daily OPEN, HIGH and CLOSE prices as well as the daily traded volume. These were used to calculate the technical indicators. The Stock Index Indicators are often used as the main factor in trading. Stock traders make a decision to trade by using these indicators (Utthammajai and Leesutthipornchai, 2015).

The following 33 Technical Indicators were extracted from the dataset and used in this research. 7 days Simple Moving Average (SMA7), 7 days Exponential Moving Average (EMA7), 7 days Price Rate of Change (ROC7), 14 days Price Rate of Change (ROC14), Relative Strength Index (RSI), Money Flow Index (MFI), Stochastic Oscillator (STH), Volume Price Trend (VPT), Accumulator/Distribution (AD), Commodity Channel Index (CCI), 7 days Average True Range (ATR7), 14 days Average True Range (ATR14), ULCER index, 9 days Simple Moving Average (SMA9), 9 days Exponential Moving Average (EMA9), Mass Index Indicator (BMA), 20 days Exponential Moving Average (EMA20), 20 days Simple Moving Average (SMA20), Balance of Power (BoP), 10% WILLIAMS R indicator (10% RWil), 20% WILLIAMS R indicator (20% RWil), Plus Directional Indicator (+DI), Minus Directional Indicator (-DI), Average Directional Movement Index (ADX), Moving Average Convergence Divergence (MACDH), CLOSE, PREV, OPEN, HIGH, LOW, DEALS, VOLUME, VALUE.

Data Pre-processing

Initially, all the records with missing values were removed from the dataset in order to improve the accuracy of the prediction. Then the data was further partitioned into two parts: the training data and the testing data. The training data is the data used to train the models. The total number of the training data is 2001 for Bank A and 901 for Bank B. The testing data is the data used for the purpose of testing the models. It is used to find the accuracy rate of the model. The total number of the testing data is 228 for Bank A and 43 for Bank B. The data were first transformed into a standard form for the network to learn easily. This transforms the dataset to values between [0, 1].

Feature selection using Correlation-based SVM (C-SVM)

The SVM has a classification complexity ability which serves as a feature selection criterion that will be used in the selection of feature subset in the prediction of Stock Exchange Composite Index. The SVM approach that will be adopted will be that of Lin *et al* (2013) who made use of an hybridized correlation-based SVM (C-SVM) filter method (Li *et al.,* 2011).

This module consists of SVM-Based Feature Ranking and Affinity propagation Clustering. The use of C-SVM in this work is to select good feature sets from contain features that are highly correlated with the output, yet uncorrelated with each other. The essence of this is to remove the financial indexes that are not useful for prediction and could be irrelevant to stock market trend. C-SVM is comprised of two main modules: the SVM-based feature ranking and the affinity propagation clustering (Lin et al., 2013). The SVM-based feature ranking realizes a feature ranking using linear kernel SVM through supervised learning. The grouping of the financial indexes into clusters based on the correlation amidst the financial indexes is carried out using the affinity propagation module. In each of the clusters, a financial index could have high correlation with the other indexes but low correlation with financial indexes in the other clusters. From each clusters, a delegate financial index is selected based on its influence quantities on the output quantities on the output obtained in the SVM-based feature ranking.

Correlation-based clustering implements the affinity propagation (AP) clustering to group the financial indexes into some clusters based on the correlation among the financial indexes. In each cluster, a financial index has high correlation with the other indexes, but low correlation with the financial indexes in the other clusters. So the financial indexes from the same cluster have similar characteristics. From each cluster, one financial index is selected as the delegate based on its influence quantities on the output obtained in the SVM-based feature ranking module. In this way, the correlation-based SVM filter method are able to select a good subset of financial indexes and the selected financial indexes are ranked based on the influence on the stock market trend. In the implementation of the correlation-based SVM filter, Ratsch's methodology (Ratsch *et al*., 2001) in identifying the number of clusters in AP clustering methods. The financial indicators useful for investors are provided by the ranking of selected financial indexes. In the prediction model, the ranking will be used as weight coefficient for the model inputs.

Algorithm 1 ***Affinity Propagation***

Input x_1 through x_n a set of dataset ; S, a similarity Matrix between the data items.
$S(x_i, x_j) > S(x_i, x_k)$ iff x_j is more similar to x_j than x_k
$S(x_i, x_j) = -|| x_i - x_k ||^2$
Output: A set of exemplars i.e. the centroid of the clusters

Initialize the diagonal of S = S(i,i) Using the median of similarities of all pairs // S represents // the likelihood of an input being an exemplar
Initialize the responsibility Matrix R to Zeros NxN //N no of rows of similarity matrix
Initialize the availability Matrix R to Zeros NxN

Update Matrix R as follows:

$$r(i,k) = S(i,j) - \max_{k' \neq k}\{a(i,k') - S(i,k')$$

Update Matrix A as follows:

$$a(i,k) = min\,(0, r(i,j) + \sum max\big((i,k') - S(i',k)\big) \quad i \notin \{i,k\}\; for\; i \neq k$$

Calculate a(k,k) as

$$a(k,k) = \sum_{i \neq k} max\big((0, r(i',k)\big)$$

Loop to 1 until cluster boundaries remain unchanged or after predetermined iterations.
Finally exemplars are extracted from the matrix where diagonal

$$r(i,i) + a(i,i) > 0$$

The ranking of selected financial indexes provides the financial indicators useful for investors. And in the prediction model, the ranking will be used as weight coefficients for the model inputs. Denote S_k the sensitivities describing the ranking, the weight β_k can be calculated using (1).

$$\beta_k = \frac{S_k}{\sum_{s=1}^{d} S_s}, k = 1,2,\ldots,d \qquad (1)$$

where S_k is the sensitivity value of selected feature, feature weight vector β_k obey the following two conditions $0 \leq \beta_k \leq 1$ and $\sum_{k=1}^{d} \beta_k = 1$

Quasi-linear SVM (Q-SVM)

The purpose for applying Q-SVM in the prediction of the stock market trend is based on its multi-local linear classifier technique which could handle high-dimensional and high noise dataset which make SVM suffer over-fitting problem when there is the presence of flexible non-linear kernels which results in poor generalization performance (Chen *et al.*, 2010). Considering the uptrend and downtrend prediction problem on stock market, we will focus our attention on binary classification problem. The input space, denoted by X where $X = X.diag(\beta)$ which combines the feature weights, $\in R^n$, and the output space, denoted by $Y \in (1, -1\}$ Usually, an N-instance training dataset is expressed as.

$$D = \{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_N, y_N)\} \subseteq (X \times Y)^N$$

A non-linear separating boundary can be seen as aggregation of M local linear boundaries $\Omega_j^T x + b_j, j = 1, \dots, M$ with interpolation. The prediction model, $f_p(x)$ can be written compactly as shown in (2).

$$f_p(x) = \sum_{l=1}^{M} (\Omega_l^T x + b_l) R_l(x) + b \qquad (2)$$

Q-SVM parameters optimization using the Coral Reef Optimization (CRO)

The CRO is used in selecting an optimal value of the predicted parameters. The CRO algorithm tackles optimization problems by modeling and simulating all the distinct processes. Let Λ be a model of reef, consisting of a N × M square grid. We assume that each square (i, j) of Λ is able to allocate a coral (or colony of corals) Ξi, representing different solutions to our problem, encoded as strings of numbers in a given alphabet I. The CRO algorithm is first initialized at random by assigning some squares in Λ to be occupied by corals (i.e., solutions to the problem) and some other squares in the grid to be empty; that is, holes in the reef where new corals can freely settle and grow. The rate between free/ occupied squares in Λ at the beginning of the algorithm is an important parameter of the CRO algorithm, which will be denoted in what follows as 0<p0<1. Each coral is labeled with an associated health function (Ξij): I → R that represents the problem's objective function. Note that the reef will progress as long as healthier (stronger) corals (which represent better solutions to the problem at hand) survive, while less healthy corals perish.

After the reef initialization described above, a second phase of reef formation is carried out by the CRO algorithm. To this end, a simulation of the corals' reproduction in the reef is done by sequentially applying different operators. This sequential set of operators is then applied iteratively until a given stop criteria is met. Thus, we define different operators for modeling sexual reproduction (broadcast spawning and brooding), asexual reproduction (budding), and polyps depredation. In both sexual and asexual reproduction we give the conditions under which new corals effectively get attached to the reef or are depredated while at the larvae phase, it is as follows:

Broadcast Spawning (External Sexual Reproduction)

The modeling of coral reproduction by broadcast spawning consists of the following steps.

(1a) In a given step k of the reef formation phase, select uniformly at random a fraction of the existing corals pk in the reef to be broadcast spawners. The fraction of broadcast spawners with respect to the overall amount of existing corals in the reef will be denoted as Fb. Corals that are not selected to be broadcast spawners (i.e., 1−Fb) will reproduce by brooding later on, in the algorithm.

(1b) Select couples out of the pool of broadcast spawner corals in step k. Each of such couples will form a coral larva by sexual crossover, which is then released out to the water. Note that, once two corals have been selected to be the parents of a larva, they are not chosen anymore in step (i.e., two corals are parents only once in a given step). These couple selection can be done uniformly at random or by resorting to any fitness proportionate selection approach (e.g., roulette wheel).

Brooding (Internal Sexual Reproduction)

As previously mentioned, at each step k of the reef formation phase in the CRO algorithm, the fraction of corals that will reproduce by brooding is 1−Fb. The brooding modeling consists of the formation of a coral larva by means of a random mutation of the brooding-reproductive coral (self-fertilization considering hermaphrodite corals). The produced larva is then released out to the water in a similar fashion than that of the larva generated in step (1b).

Larvae Setting

Once all the larvae are formed at step k either through broadcast spawning (1) or by brooding (2), they will try to set and grow in the reef. First, the health function of each coral larvae is computed. Second, each larva will randomly try to set in a square (i, j) of the reef. If the square is empty (free space in the reef), the coral grows therein no matter the value of its health function. By contrast, if a coral is already occupying the square at hand, the new larva will set only if its health function is better than that of the existing coral. We define a number κ of attempts for a larva to set in the reef: after κ unsuccessful tries, it will be depredated by animals in the reef.

Asexual Reproduction

In the modeling of asexual reproduction (budding or fragmentation), the overall set of existing corals in the reef are sorted as a function of their level of healthiness (given by $f(\Xi ij)$), from which a fraction F a duplicates itself and tries to settle in a different part of the reef by following the setting process described in Step(3).

Depredation in Polyp Phase

Corals may die during the reef formation phase of the CRO algorithm. At the end of each reproduction step k, a small number of corals in the reef can be depredated, thus liberating space in the reef for next coral generation. The depredation operator is applied with a very small probability Pd at each step k, and exclusively to a fraction Fd of the worse health corals in Λ. For the sake of simplicity in the parameter setting of the CRO algorithm, the value of this fraction may be set to Fd = Fa. Any other assignment may also apply provided that Fd+Fa≤1 (i.e., no overlap between the asexually reproduced and the depredated coral sets).

The SVM-CRO algorithm is presented in algorithm 2.

Algorithm 2 **CRO-SVM algorithm**
Input:. Training dataset, Test dataset, initial SVM model
Output:. Optimized SVM model

Initialize CORAL REEF with corals σ_{1i}, σ_{2i} , σ_{3i} C_i // the hyper-parameters of SVM
Obtain health function of the corals using $f(x_i)$ =SVMpredict(SVMTrain(σ_{1i}, σ_{2i} , σ_{3i} C_i))
  While iteration < Maximum Iteration
    Perform external sexual reproduction by some of the CORALS
    Perform internal sexual reproduction by the remaining CORALS
    Obtain health function of each coral using $f(x_i)$ =SVMpredict(SVMTrain(σ_{1i}, σ_{2i} , σ_{3i} C_i))
    Perform settlement of new CORALS

By placing the new CORALS in any available empty space in the Reef or
By displacing the CORALS with low fitness function by the CORALS of higher fitness function in the REEF
Perform asexual reproduction of some of the CORALS
Perform predation process
loop
End while
OBTAIN BEST SVM parameters (σ_{1i}, σ_{2i} , σ_{3i} C_i)
SVMPREDICT() // using test dataset

Table 1 : The SVM parameters

Parameters	Description
C	Penalty parameters
α	The Lagrange parameters
σ	Kernel parameter

Table 2 : The Quasi-linear SVM parameters

Parameters	Description
C	Penalty parameters
α	The Lagrange parameters
σ_1, σ_2, σ_3	Kernel parameter

Table 3 : The optimization algorithms

Optimization Algorithms	Description
CORAL REEF	Uses the concept of CORAL settlement in optimization
FIRE-FLY	Uses the concept of firefly light shining for optimization

Table 4 The CRO optimization parameters

Parameters	Description
Ngen	Number of generations
N1	Reef Width
N2	Reef Height
Fb	Broadcast probability
Fa	Asexual Reproduction probability

Fd	Fraction of Corals to be eliminated
r0	Free/Total initial settlement
k	No of trials for coral's settlement
Pd	Depredation probability

Table 5 The Fire-Fly parameters

Parameters	Description
MaxIt	Maximum Number of Iterations
nPop	Number of Fireflies
gamma	Light absorption coefficient
Beta	Attraction coefficient
alpha	Mutation Coefficient
Alpha_damp	Mutation Coefficient damping ratio

IMPLEMENTATION AND RESULTS

The system that employed the hybridized model was implemented using Java and MATLAB programming languages. The two optimization algorithms are depicted in Table 3. The parameters of CRO are shown in Table 4. The classification system was implemented using Support Vector Machine. The two support vector models adopted are shown in Tables 1-2. Coral Reef algorithm is used to optimize the parameters of the SVM model. The parameters of the CRO are shown in Table 4. The SVM model makes use of LiBSVM Java version (Cheng and Lin 2011). The benchmark optimization algorithm is Fire-Fly algorithm shown Table 5. The system is tested using Bank A stock dataset and Bank B stock dataset. The implementation of the proposed framework was carried out as follows: data collection and preparation using Microsoft excel, data acquisition, data cleaning, feature extraction, SVM training and evaluation, SVM training using CRO, SVM training using Firefly which will be described in this section.

Data collection

The first dataset used to validate this model is the stock of Bank A. Gotten from the Nigerian Stock Exchange (NSE). The dataset is for almost 11 years and contains data from the period of 12th February 2005 to 4th May 2016. The dataset include the attributes: Date, Opening price, Closing price, previous close price, High, Low, Average, Deals, Volume, and Value. Table 6 shows the attributes of the dataset with their descriptions. The dataset is made up of two classes viz: UP which denotes that the price of next day will go up and DOWN which denotes that the price of next day will come down. The total number of data items selected and used is 2229. There are 1148 data that constitutes upward price movement while 1081 constitutes downward trend movement. Table 7 presents this distribution. The second dataset used for the research work is the stock of Bank B. The dataset is for almost 4 years and contains data from the period of 30th September, 2013 to 21st September, 2017. The dataset include same the attributes: Date, Opening price, Closing price, and previous close price, High, Low, Average, Deals, Volume, and Value as shown in Table 6. The dataset is made up of two classes viz: UP which denotes that the price of next day will go up and DOWN which denotes that the price of next day will come down. The total number of data items selected is 944. There are 478 data that constitutes upward price movement while 466 constitutes downward trend movement. Table 8 presents this distribution.

Table 6: Attributes with their descriptions

Attribute	Description
Close	The last price traded during the day
Open	The first price traded during the day or in the morning
Previous	The Previous day close price of the stock
High	The highest traded price during the day
Low	The lowest price traded during the day
Average	The average price of the stock during the day
Deals	Deals made
Volume	No of volume of stock
Value	Value of stock
Date	The date for which the above attributes are collected

Table 7: Dataset Distribution for Bank A

Dataset	Total samples	Number of Attributes	Class	Class-wise samples
Bank A	2229	33	UP	1148
			DOWN	1081

Table 8: Dataset Distribution for Bank B

Dataset	Total samples	Number of Attributes	Class	Class-wise samples
Bank B	944	33	UP	478
			DOWN	466

Load Data and extract Features

Data loading

The data loading prepares and load data into memory in a form that can be processed by Support Vector Machine. The user loads the training data as well as the testing data. Each data item is labeled as +1 if today's price is greater than that of yesterday price and -1 otherwise. The data when loaded is then subjected to a linear SVM for feature reduction and then passed to affinity propagation for further feature reduction.

Price Trend Prediction using SVM

For the Bank A, a support vector machine is trained using 2001 data items consisting of 4 attributes obtained during feature selection as shown in Table 9. The model is then trained using Libsvm package (Chang and Lin, 2001) developed in Java and called in Matlab. The model's parameter obtained from the training are then saved. For testing, 228 data items are used to test the system. Out of this, 172 are classified correctly while 56 were misclassified. Error in forecast is therefore 24.5614% and the prediction accuracy 75.4386%.

Table 9: Features Selected for Bank A

Indicators	Money Flow Index (MFI)	10% WILLIAMS R indicator (10%RWil)	LOW	VOLUME
Weights	0.049	0.214	0.672	0.0586

For the Bank B, SVM is trained using 880 data items consisting of 3 attributes obtained during feature selection as indicated in Table 10. The model's parameter obtained from the training are then saved. For testing, 64 data items are used to test the system. Out of this 27 are classified correctly while 37 were misclassified. Error in forecast is therefore 57.8125% and the prediction accuracy 42.1875%.

Table 10: Features Selected for Bank B

Indicators	ULCER index	10% WILLIAMS R indicator (10%RWil)	Plus Directional Indicator (+DI)
Weights	0.15	0.8188	0.0283

Price Trend prediction CRO-SVM

On clicking Train and evaluate with CRO-SVM, a user interface is presented. The parameters of the training is as shown in Figure 2.

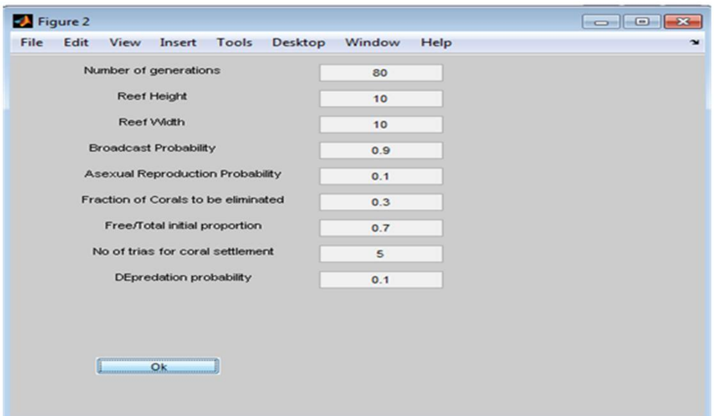

Figure 2: CRO User Interface

For the Bank A, a CRO-SVM is trained using 2001 data items. The model is first trained using CRO algorithm and the optimal parameters are then saved. Figure 3 shows the fitness function against iteration. These optimal parameters from CRO are fed to a quasi-linear SVM model to evaluate the test data. For testing, 228 data items are used to test the system. Out of this, 176 were classified correctly while 52 data items were misclassified. Error in forecast is 22.807% and the accuracy is 77.193%

For the Bank B, a CRO-SVM is also trained using 880 data items. The model is first trained using CRO algorithm and the optimal parameters are then saved. These optimal parameters from CRO are fed to a quasi -linear SVM model to evaluate the test data. For testing, 64 data items are also used to test the system. Out of this, 37 were classified correctly while 27 data items were misclassified. Error in forecast is 42.1875% and the accuracy is 57.8125%

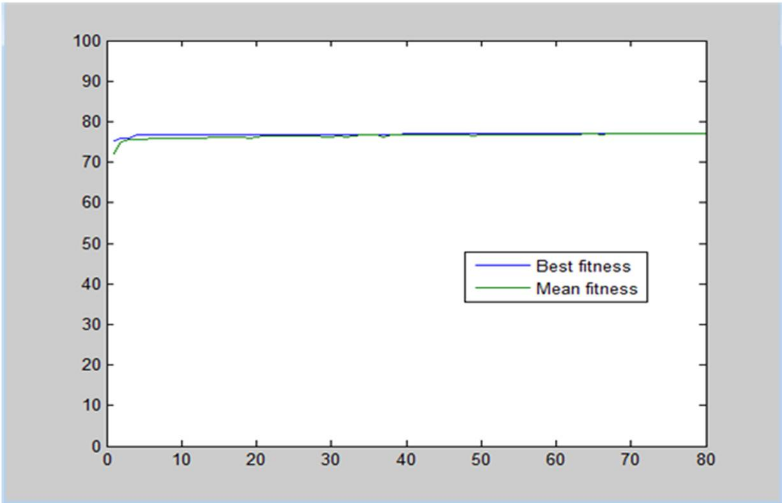

Figure 3: Coral Reef fitness function

Price Trend prediction FireFly-SVM

On clicking Train and evaluate with FirerFly, a user interface is presented. The parameters of the training is as shown in Figure 4.

Using Bank A dataset, a FireFly-SVM is trained using 2001 data items. The model is first trained using FireFly algorithm and the optimal parameters are then saved. Figure 5 shows the fitness function against iteration. These optimal parameters from FireFly are fed to a quasi-linear SVM model to evaluate the test data. For testing, 228 data items are used to test the system. Out of this, 176 were classified correctly while 52 data items were misclassified. Error in prediction is 22.807% and the accuracy is 77.193%.

For the Bank B, a FireFly-SVM is also trained using 880 data items. The model is first trained using FireFly algorithm and the optimal parameters are then saved. These optimal parameters from FireFly are fed to a quasi-linear SVM model to evaluate the test data. For testing, 64 data items are also used to test the system. Out of this, 37 were classified correctly while 27 data items were misclassified. Error in forecast is 42.1875% and the accuracy is 57.8125%.

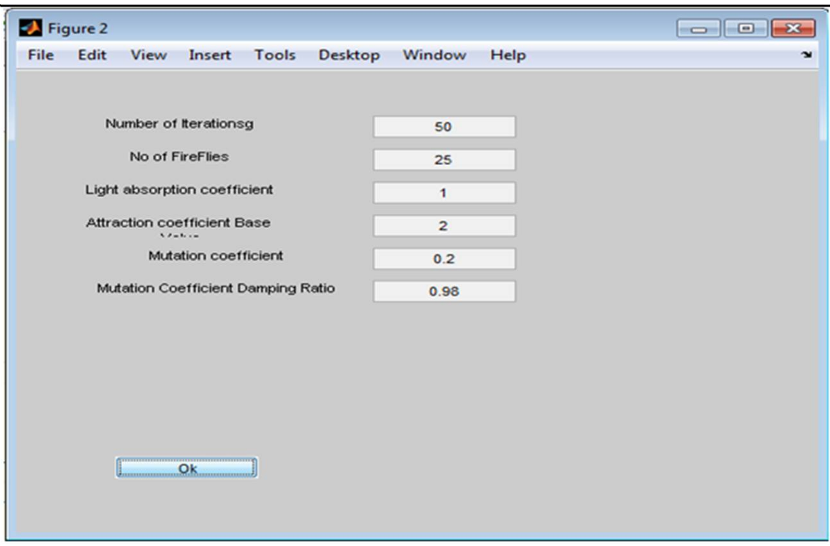

Figure 4: FireFly user interface

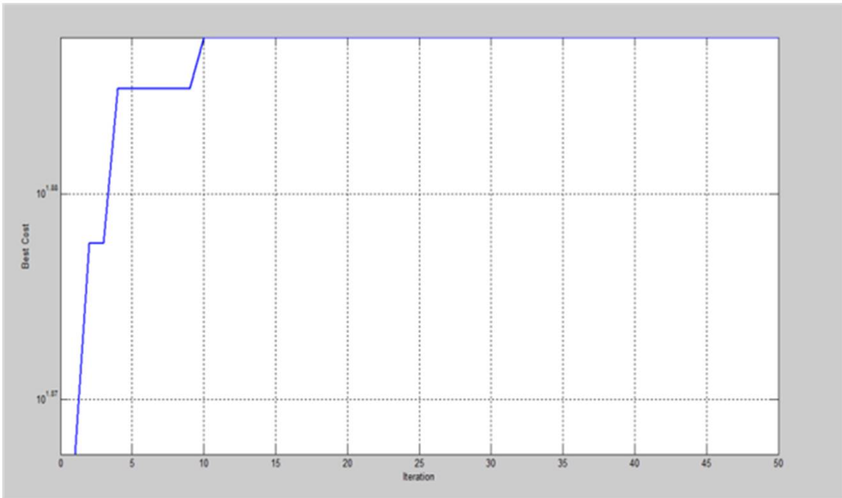

Figure 5: FireFly fitness function

Performance Evaluation

As a means of ascertaining the performance of the proposed approach, a benchmark was done by evaluating SVM+CRO against SVM only and SVM+Firefly techniques. The evaluation metric used in evaluating the technique is the hit ratio.

Hit Ratio (Lin *et al.*, 2013): This the ratio of the outputs produced by the model and the actual outputs that is expected.

$$Hit\ ratio = \frac{1}{N}\sum_{i=1}^{N} H_i \qquad (3)$$

Where $H_i = 1$ if $MO_i = AO_i$; $H_i = 0$, otherwise MO_i is the model output AO_i is the actual output, and N is the number of testing samples.

Recognition Accuracy of the Classifiers

Tables 11-13 shows the summary of the performance of the three classifiers used for stock price trend prediction while Figure 6 is the chart of the classification accuracy of the classifiers on Bank A. It is known from the Tables 11-13 that CRO-SVM and Fire-Fly SVM have the same classification accuracy (77.193%) while SVM has the lowest performance of (75.4386%).

Table 11 : SVM Result

Bank Dataset	Hit-Ratio
Bank A Data	0.754386
Bank B Data	0.421875

Table 12 : CRO-SVM Result

Bank Dataset	Hit-Ratio
Bank A Data	0.77193
Bank B Data	0. 578125

Table 13 : FireFly-SVM Result

Bank Dataset	Hit-Ratio
Bank A Data	0.77193
Bank B Data	0. 578125

Tables 11-13 also show the classification accuracy of the classifiers on Bank B. It is known from the Tables 11-13 that CRO-SVM and Firefly SVM have the same classification accuracy (57.8125%) while SVM has the lowest performance of (42.1875%).

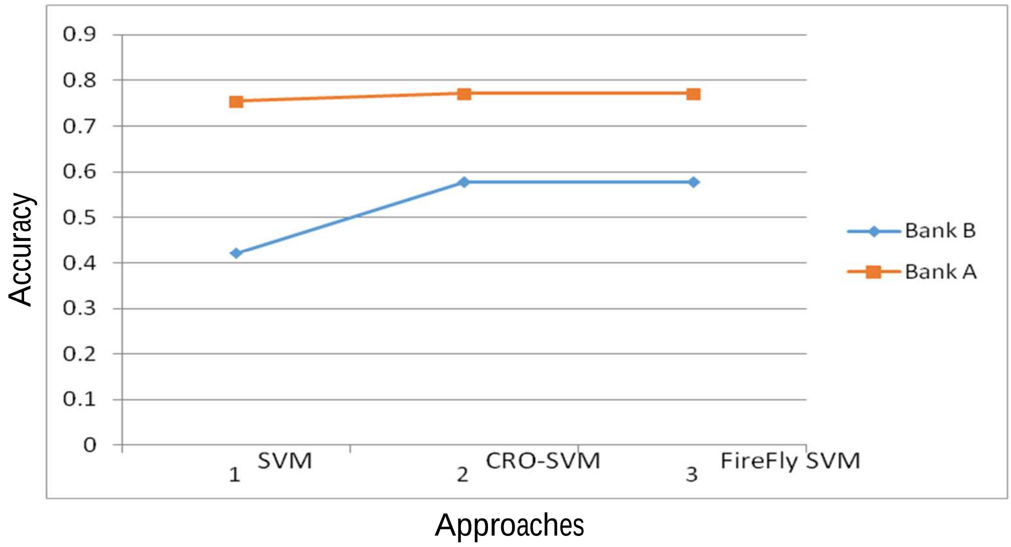

Figure 6: Recognition Accuracy of Classifiers

Table 14 shows the accuracies of each of the approaches based on the two banks data used for testing and training. The chart in Figure 7 is a diagrammatic representation of this result.

Table 14: Table showing the accuracy output of each approach

	SVM	SVM + Coral Reef	SVM + Firefly
Bank A	75.44%	77.19%	77.19%
Bank B	42.19%	57.81%	57.81%

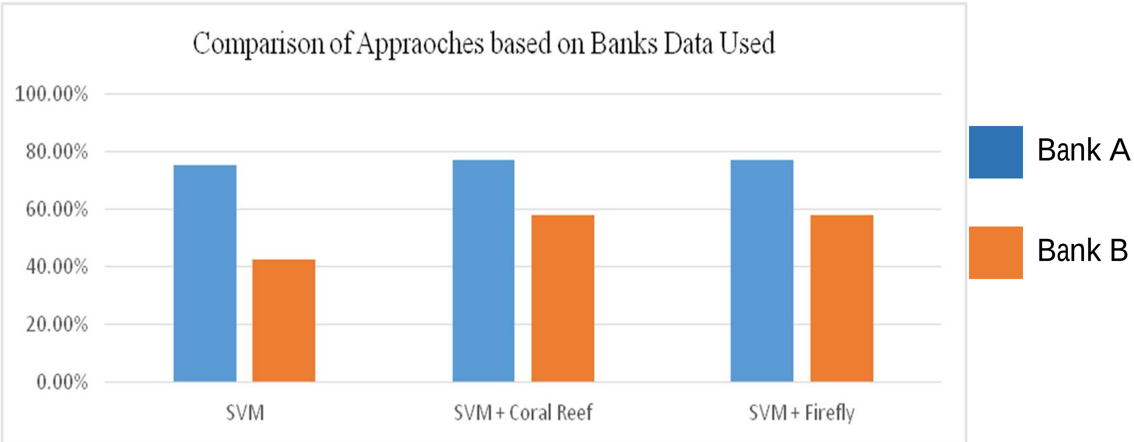

Figure 7: Chart of the outcome of approaches based on the banks data

DISCUSSION

Various experiments carried out with the models developed in this work are reported to know the efficiency of the models as regards to the data collected from two banks (Bank A and Bank B). The experiment investigates various two optimization algorithms on SVM model and shows the best in terms of recognition accuracy. The test dataset were used to investigate the performance of Support Vector Machine, Support Vector Machine optimized by Coral Reef and Support Vector Machine optimized by FireFly. It is observed that the proposed CRO-SVM performs more than un-optimized Support Vector Machine but has the same performance with FireFly-SVM in terms of recognition accuracy. The SVM returned accuracy for 75.44% for Bank A and 42.19% for Bank B. The SVM+Coral Reef approach returned accuracies of 77.19% and 57.81% for Bank A and Bank B respectively. The SVM+Firefly approach returned accuracies of 77.19% and 57.81% for Bank A and Bank B respectively. The Hit ration of SVM+Coral Reef optimization and SVM+firefly outperformed that of SVM alone. It is observed that the performance of the models depends on the selection of suitable features. The sensitivity of the variables and feature selected varies with the stock.

CONCLUSION

The problem of stock price trend prediction has been the research focus in the past few decades. The outcome of the research can assist humans in making decision on acquiring and selling stocks. Stock market has several challenges such as the erratic behavior and volatility of stocks and natural disasters. Many systems have been developed to tackle this problem. Many of these systems do not looked into efficient feature selection and optimization of the models hence their performance is inadequate. This research work therefore provides an efficient framework for stock price trend prediction. Two banks dataset from Nigeria Stock Exchange are used to validate the various models developed. The system reported in this research work is able to perform the technical indicator extraction, feature selection, prediction of stock price trend and analysis of the performance of the models. An important part of prediction systems is the efficient selection of important features. A lot of work has been done in the area of feature selection. We have the wrapper methods and filter methods. In this work a filter approach to feature selection is adopted. The framework adopted a model of correlation-SVM for efficient stock feature selection. The proposed method has been tested and found to be efficient in terms of selecting quality features. Accurate prediction of stock price trends has always been a problem. It has been reported in the literature that SVM performs efficiently at prediction. Soft-computing have been able to adapt to imprecise data. This research work has adopted a model of a quasi-linear SVM for stock price trend prediction. The Q-SVM model was optimized with CRO and then Firefly algorithms with the goal of determining the most effective one in terms of recognition accuracy. It is observed that CRO-SVM outperforms the un-optimized Q-SVM while it has comparative recognition accuracy with the FireFly+SVM. Three fundamental contributions to knowledge made in this research work are as follows: hybridized models to enhance stock feature selection using feature sensitivity and correlation information, a framework for stock price trend prediction using quasi-linear SVM and a Coral –Reef algorithm for the optimization of the parameters of the quasi-linear SVM.

This research work can be investigated further by investigating other feature Selection algorithms because the performance of this research depends on the quality of features obtained from the preprocessing stage. In future, there is the need to investigate the use of other hybrid models such as Neuro-Fuzzy systems or Neuro-SVM models in order to improve the prediction accuracy.

REFERENCES

AraúJo, R.D.A. and Ferreira, T.A. 2013. A morphological-rank-linear evolutionary method for stock market prediction. *Information Sciences*, 237: 3-17.

Basher, S.A., Haug, A. A. and Sadorsky, P. 2012. Oil Prices, Exchange Rates nd Emerging Stock Markets. *Energy Economics*, 34(1): 227-240.

Boyacioglu, M.A. and Avci, D. 2010. An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: the case of the Istanbul stock exchange. *Expert Systems with Applications*, *37* (12): 7908-7912.

Burton, F.E.T. and Shah, S.N. 2017. Efficient Market Hypothesis. *CMT Level I 2017: An Introduction to Technical Analysis*.

Chakravarty, S. and Dash, P.K. 2012. A PSO based integrated functional link net and interval type-2 fuzzy logic system for predicting stock market indices. *Applied Soft Computing*, 12(2): 931-941.

Chang C-C and Lin C-J LIBSVM 2011. A library for Support Vector Machines. ACM Transactions Intelligent Systems and Technology, 2(27): 1-27.

Chen, W. and Shih, J. 2006. Comparison of support vector machines and back propagation neural networks in forecasting the six major Asian stock markets. *Journals Electronics Finance*, (1): 49-67.

Chauhan, B., Bidave, U., Gangathade, A. and Kale, S. 2014. Stock Market Prediction Using Artificial Neural Networks.

Chen, B., Sun, F. and Hu, J. 2010. Local linear multi-SVM method for gene function classification. *Proc of the World Congress on Nature and Biologically Inspired Computing* (NaBIC'10) (Kitakyushu), 183–188.

de Oliveira, F.A., Nobre, C.N. and Zarate, L.E. 2013. Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index–Case study of PETR4, Petrobras, Brazil. *Expert Systems with Applications*, 40 (18): 7596-7606.

Dutta, A., Bandopadhyay, G. and Sengupta, S. 2015. Prediction of Stock Performance In Indian Stock Market Using Logistic Regression. *International Journal of Business and Information*, 7(1): 105-136.

Gharehchopogh, F.S., Bonab, T.H. and Khaze, S.R. 2013. A linear regression approach to prediction of stock market trading volume: a case study. *International Journal of Managing Value and Supply Chains*, 4(3): 25.

Grigoryan H. 2017. Stock market prediction using support vector machine and variable selection method. Advances in intelligent system research (AISR). *International conference on Applied Mathematics Modelling and Statistics Application (AMMSA 2017)*, 141: 141-213.

Guo, X., Sun, L., Li, G. and Wang, S. 2008. A hybrid wavelet analysis and support vector machines in forecasting development of manufacturing. *Expert Systems with Applications*, 35(1–2): 415–422.

Herz, F.S., Ungar, L.H., Eisner, J.M., Labys, W.P. and Fred Herz Patents, LLC, 2012. Stock market prediction using natural language processing. *U.S. Patent* 8: 285- 619.

Kara, Y., Boyacioglu, M. and Baykan, O. 2011. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange, *Expert Systems with Applications,* 38(11): 5311–5319.

Kavousi-Fard, A., Samet, H. and Marzbani, F. 2014. A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. *Expert systems with applications*, 41(13): 6047-6056.

Kazeem, A., Sharifi, E., Hussain, F.K., Saberi, M. and Hussain, O.K. 2013. Support vector regression with chaos-based firefly algorithm for stock market price forecasting. *Applied soft computing*, 13(2): 947-958.

Kim, H., and Han, S. T. 2016. The Enhanced Classification for the Stock Index Prediction. *Procedia Computer Science*, 91: 284-286.

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

Li, B., Wang Q. and Hu, J. 2011. Feature subset selection: A correlation based SVM filter approach. *IEEJ Trans. on Electrical and Electronic Engineering*, 6(2): 173–179.

Li, F. and Liu, C. 2009. Application Study of BP Neural Network on Stock Market Prediction. *Ninth International Conference on Hybrid Intelligent Systems*, 174-178.

Lin, Y., Guo, H. and Hu, J. 2013. August. An SVM-based approach for stock market trend prediction. *Neural Networks (IJCNN), The 2013 International Joint Conference*, 1-7. IEEE.

Liu, F. and Wang, J. 2012. Fluctuation prediction of stock market index by Legendre neural network with random time strength function. *Neurocomputing*, 83: 12-21.

Mani, S., Kothandaraman, P., Kashyap, R. and Ashnai, B. 2017. Stock Market and Network Influence on Alliance Formation: Evidence from the Biopharmaceutical Industry. *Journal of Marketing Theory and Practice*, 25(1): 69-85.

Patil. S.S., Patidar K. and Jain, M. 2016. Stock market prediction using support vector machine. *International journal of current trends in Engineering and Technology* 2: 18-25.

Ratsch, G., Onoda, T. and Muller, K. 2001. Soft margins for AdaBoost. *Machine Learning*, 42(3): 287–320.

Salcedo-Sanz, S., Del Ser, J., Landa-Torres, I., Gil-López, S. and Portilla-Figueras, J.A. 2014. The coral reefs optimization algorithm: a novel metaheuristic for efficiently solving optimization problems. *The Scientific World Journal*, *2014.*

Shaikh, A., Pritam, T., Ankita, P., Shital, W. and Pooja, T. 2014. Stock Exchange Market Prediction. *International Journal of Advance in Computer Science and Technology, India*, 3 (5): 349-351.

Utthammajai K. and Leesutthipornchai P. 2015. Association Mining on Stock Index Indicators. *International Journal of Computer and Communication Engineering,* 4(1): 46-49.

Vohandu L. and Krusberg H. 1977. A direct factor analysis method. The proceedings of TTU, 426: 11-21.

Xu, X., Zhou, C. and Wang, Z. 2009. Credit scoring algorithm based on link analysis ranking with support vector machine. *Expert Systems with Applications,* 36: 2625-2632.

(Manuscript received: 14th June, 2018; accepted: 29th May, 2020).