Enhancing Support Vector Machine Model for Stock Trading using Optimization Techniques

Gangamol Jaiwang
School of Information Technology
Mae Fah Luang University
Muang, Chiang Rai
gangamol.jai@lamduan.mfu.ac.th

Piyasak Jeatrakul
School of Information Technology
Mae Fah Luang University
Muang, Chiang Rai
piyasak.jea@mfu.ac.th

***Abstract*—This study aims to optimize the stock market forecast model using the Support Vector Machine. Two types of SVM optimization techniques were tested, Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), in order to find which optimization technique provides the most appropriate parameters for the SVM Dot Kernel and results in improved stock market forecast efficiency. The experiment uses data from five stocks from the Stock Exchange of Thailand, which are tested and compared with results from before and after application of the optimization techniques. In addition, the SVM model, which was enhanced by the optimization technique, was backtested for trading in order to analyze the return on investment by comparing trading results with technical analysis techniques that are popular with general investors, such as Relative Strength Index (RSI), Stochastic Oscillator (STO), and Moving Average Convergence/Divergence (MACD). The results showed that genetic algorithm can improve the SVM Model for increased forecasting accuracy, and the average return results were higher than the compared results from the technical analysis indicators. In addition, this study also analyzes the SVM model trading performance based on price trends, which include uptrend, downtrend and sideways trend. This concludes that SVM with Dot Kernel optimized by GA is able to provide effective forecasting for all trends, with the best performance during the uptrend.**

Keywords—Stock market, prediction model, support vector machine, optimization technique, genetic algorithm, particle swarm algorithm

I. INTRODUCTION

Stock market investments are risky as stock prices are constantly changing depending on many factors, including exchange rates, economic situations and the flow of capital. However, investors may have a higher return if they invest at the right time. In general, there are two stock analysis techniques, analysis of fundamental data and analysis of technical data [1]. The fundamental analysis uses fundamental data from financial budgets before making a decision about trading. Fundamental data examples are income data, profit and loss data and dividend data. However, analysis of fundamental factors does not provide direct indicators for trading but provides only indicators of the company's ability and efficiency, so fundamental analytical techniques are often used for long-term investments [2]. The second method is a technical analysis using statistical data from opening prices, closing prices and daily trading volumes for assessment. It can provide investors with trading indicators. In addition, technical analysis has developed technical indicators with the ability to understand stock price trends and momentum in trading shares, which is illustrated in stock charts. There are several technical indicators used in forecasting, such as Relative Strength Index (RSI), Stochastic Oscillator and Moving Average Convergence/Divergence (MACD) [2]. In general, technical analysis methods using technical indicators are used by short-term investors wishing to speculate the profit [2].

At present, machine learning is being introduced to assist data analysis and result forecasting. Support vector machines (SVM) are one of the most widely-used machine learning models due to high performance [3][4][5]. Researchers have tried to use SVM to forecast shares in the past [1][6][7][8][9]. In the research experiment [1], SVM was proven successful in forecasting buying and selling points used in previous studies [1], which demonstrates the effectiveness of SVM with Dot Kernel combined with ready-to-use technically prepared data. Feature selection helps to reduce attributes and can improve classification by majority voting techniques. In past projects, support vector machines have provided better results when appropriate parameters are used for forecasting [1][6][7][8][9]. SVM has several parameters that can be used in combinations, and past studies have yet to provide clear conclusions as to which SVM parameter value sets are optimal for stock market forecasting.

In this study, we have attempted to optimize buying and selling point forecasting by using the SVM Model with Dot Kernel [1]. As to this, parameter optimization techniques can be applied to significant parameter tuning values of the SVM with dot Kernel to optimize forecasting [10]. In this study, previously proposed optimization techniques were used [1][10][11][12][13], which were selected from optimization techniques popularly used in literature, such as Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). The researcher aims to compare and find optimization techniques that are able to optimize stock trading point forecasting.

II. Optimization Techniques

In order to design an effective SVM model, the SVM parameters must be carefully selected. These parameters are comprised of the regularization parameter *C* which determines the trade-off cost between minimizing the training error and minimizing the complexity of the model, the sigma parameter (σ or d) of the kernel function which defines the non-linear mapping from the input space to some high-dimensional feature space, and the kernel used in SVM which constructs a non-linear decision hypersurface in an input space [10]. This section will present the popular techniques used for parameter optimization that will be used in this study.

A. The Genetic Algorithm (GA)

GA is an evolutionary algorithm invented by John Holland [1] in the 1960s and was based on the genetic evolution of living beings, also known as “the Theory of Evolution” invented by Charles Darwin [1]. GA can be used to generate solutions to many problems, such as in finding the lowest possible answer for specific mathematical equations or in problems similar to the traveling salesman question. GA is suitable for solving difficult problems or for problems that do not have a fixed method for finding answers [12]. The exploration predictability and efficiency of the SVM algorithm are highly impacted by the selection of *C* and the parameter *g* of the kernel function. These parameters are typically selected under professional supervision. This results in the parameters being selected by individual preference, which is clearly subjective and inefficient. This method makes it difficult to achieve the correct selection, and the selected parameters may not be optimally matched for the various degrees, factors and complex problems of each instance. In the end, the development of the SVM parameter is greatly hindered by this limitation. However, among its many optimization advantages, the GA is able to decipher complicated problems and achieve optimization goals based on its ability assess problems by applying methods from biological evolution-based roots. As in “the Survival of the Fittest”, the GA is able to use the rules as pertaining to natural selection to assess the chromosome information exchange mechanism and find the optimal match. Therefore, this paper proposes a GA-based SVM method to find optimal SVM parameters, using the procedure as following [5]:

Step 1. Input and Initialization: Specify the data error penalty coefficient *C* for the solution space. Specify the kernel function parameter *g* as genotype data for the genetic space. Genotype data arranged in different combinations make up different points. N initial genotype data is then generated, with each genotype indicated to as an individual. *N* individuals form a group or string, which are then used by the GA as the starting point to initiate iteration.

Step 2. Individual fitness calculation: Calculate individual fitness and assess whether or not each optimization criteria are fulfilled. Fulfilled optimal individuals and their represented solutions will then comprise the output, completing the calculation. If not fulfilled, proceed to step 3.

Step 3. Selection: Analyze each individual’s fitness value, then move the individuals with higher fitness values to the next group. The individuals with a higher fitness level have higher adaptability and are therefore more likely to be selected. Individuals with lower fitness values are more likely to be removed.

Step 4. Crossover: Individuals in the group are matched and a part of an individual’s chromosome is exchanged with its pair based on crossover probability conditions.

Step 5. Variation: Variation probability is applied to change some or specific genes to allelic genes. This is performed on all individuals in the group.

Step 6. Calculation judgments: Determine whether or not the individuals of the newly generated group fulfill the end conditions. If the end conditions are not met, return to step 3 and repeat the calculations. Stop once the end conditions are met.

Step 7. Application: The GA-SVM model is well trained and can be applied to find solutions to real-world problems, as shown by the calculation flow in Figure. 1

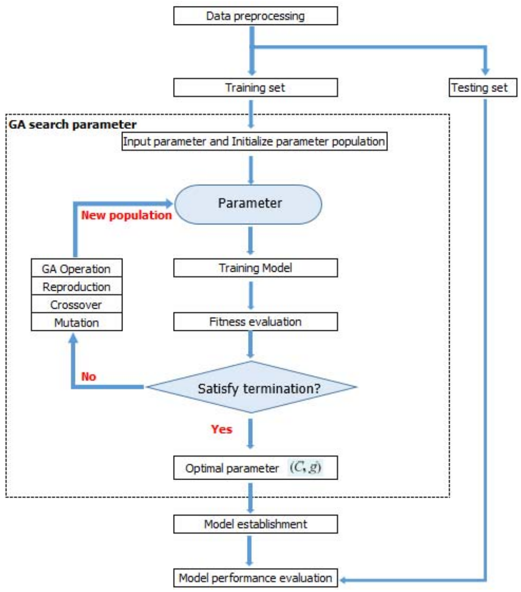

Figure 1. Flow diagram of Genetic Algorithm Optimization [5]

B. Particle Swarm Optimization (PSO)

PSO is the best method for finding solutions and was invented by Eberhart and Kennedy in 1995 [11][12]. This method was inspired by observing the movements of birds in flocks when searching for food [12]. PSO finds the answer by using a large number of particles to move into the search space to find the best solution. The PSO + SVM approach is applied to eliminate unnecessary or insignificant features and effectively determines the parameter values, in turn improving the overall classification results [8].

Let $P_i^t = \{P_{i1}^t, P_{i2}^t, \dots, P_{iD}^t\}$ represent the best solution that particle i has obtained until iteration t, and $P_g^t = \{P_{g1}^t, P_{g2}^t, \dots, P_{gn}^t\}$ denote the best solution obtained from P_i^t in the population at iteration t. To search for the optimal solution, each particle changes its velocity according to the cognition and social parts as follows:

$$V_{id}^t = V_{id}^{t-1} + c_1 r_1 (P_{id}^t - x_{id}^t) + c_2 r_2 (P_{gd}^t - x_{gd}^t), \quad d = 1,2,\dots,D \tag{1}$$

where c_1 indicates the cognition learning factor, c_2 indicates the social learning factor, r_1 and r_2 are random numbers uniformly distributed in $U(0,1)$.

Each particle then moves to a new potential solution based on the following equation:

$$X_{id}^{t+1} = X_{id}^t + V_{id}^t, \quad d = 1,2,\dots,D \tag{2}$$

Figure 2 shows the flowchart for Particle Swarm Optimization (PSO). The particle population is first initialized at a random position in the D-dimensional space and a random velocity is assigned for each dimension. Next, fitness evaluations are performed for each particle. In this study, the fitness value of each particle determines the classification accuracy. Particles with higher fitness values than the best fitness are noted and the position vector is saved for that particle. The position vector is saved for global best if the fitness value of that particle is higher than the global best fitness. Lastly, the particle positions and velocities are updated and adapted until finally fulfilling the termination condition.

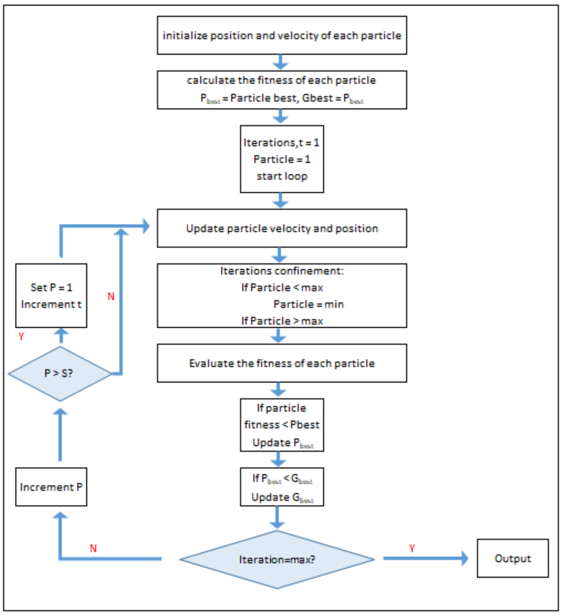

Figure 2. Flow diagram of Particle Swarm Optimization [14]

The concept in which modulation applies the Particle Swarm Optimization technique to find good fit points is shown in Figure 3.

III. DATA PREPARATION

Data selection is performed from individual stocks from the past ten years (from 04/01/2005 to 30/12/2015). The stocks selected for testing and data preparation consist of CPALL, ICC, MINT, SC, and SCC, which were chosen from different industries within the Thai Stock Market. A total of 60 technical analysis indicators, such as RSI, MACD, STO, DI+, DI-, Volume, and MFI, were used as input attributes. Then, all input attributes were selected by feature selection using the majority voting technique [1] to obtain input attributes that were related to the study results. There were three output classes, consisting of buying, holding and selling class [2] as shown in Figure 4. The reversal point can be found by applying the pivot technique, which calculates the average stock price of the previous day (called the pivot value) and it can be used for calculating the market pivot point (P). Most commonly, it is the arithmetic average of the high (H), low (L), and closing (C) prices of the market in the prior trading period [2]. If the pivot value is lower than the current price, this means that it is a buying point. If the pivot value is less than the current price, then this means it is a selling point. Other points that are not listed here are holding points [2].

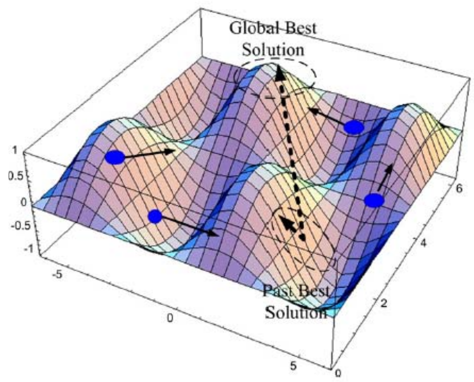

Figure 3. Assigned each instance based on a turning point [15]

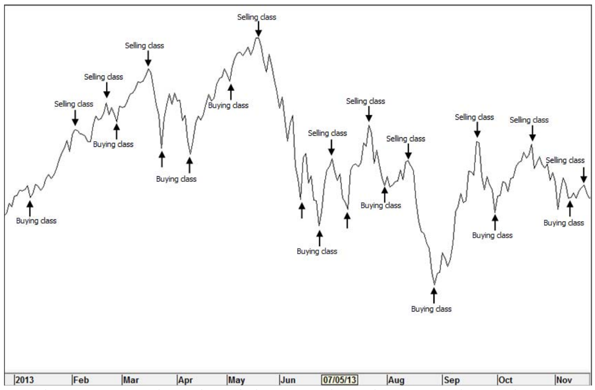

Figure 4. Assigned each instance based on a turning point [1]

The data from this study will be separated into two groups: Data from the stock index from the past ten years (from 04/01/2005 to 29/08/2014) and from stock data from the past five years (from 04/01/2012 to 15/03/2017). Then, after collecting previous data, the results will be assessed to determine the turning point. Group one data will collect data from 300 records, separated into buying data (100 records), holding data (100 records) and selling data (100 records) for application as a training SVM Model. Group two data will use all data in backtesting in order to find the return from the true trading point at that time.

IV. Experimental And Results

In Figure 5, the SVM model with Dot Kernel, which is a prediction model for stock market, is trained by training data. Two optimization techniques, Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), are applied to the model before the prediction results on testing data are compared.

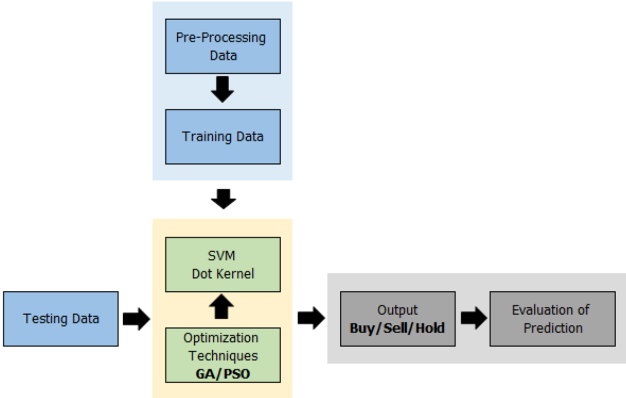

Figure 5. The SVM model enhanced by optimization techniques

In this paper, two experiments are conducted in order to investigate optimization techniques. Firstly, Experiment I compared two SVM parameter optimization techniques, Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), in order to find the most suitable C for stock market forecasting. Normally, C controls the trade-off between error on the training set data and the margin, which essentially affects forecasting accuracy (the accuracy of the testing set). This experiment focuses only optimizing C value. For other parameters, for instance, gamma is set at 0.11, which has been determined to be the best parameter by past studies [1]. Secondly, Experiment II compares return rate forecasting ability between the SVM dot kernel with optimization technique and commonly used technical analysis indicators, consisting of RSI, MACD, and STO, to determine whether the SVM dot kernel with optimization technique outputs more accurate results (in terms of return) than when using these indicators.

A. Experiment I: Comparison of Optimization Techniques on SVM

The objective of Experiment aims to compare the Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) optimization techniques in order to optimize and find which technique yields the most suitable *C* for SVM Dot Kernel stock forecasting and outputs the most accurate forecasting results. The range of *C* chosen for optimization was 1-20 and other initial data was selected as shown in Table I.

TABLE I. Initial Data Values for Each Optimization Technique

Initial Parameters	**Initial data values**	
	GA	***PSO***
Inertial weight		0 - 1
Acceleration factors (C1 and C2)	1 - 20	1 - 20
Population size	10 – 100	10 – 100
Maximum iteration (Maxite)		1 -10
Number of validation	2 - 5	2 - 5

Table II shows the performance results on the testing set for five stocks in comparison with the results from the SVM dot kernel using different *C* parameters. These results show that both the GA and PSO optimization techniques can find suitable values for C and effectively optimize the SVM Model, and yields more accurate average forecasting results than SVM with default parameters. The SVM-GA average accuracy was 78.22%, while SVM-PSO average accuracy was 73.11%. The average accuracy of SVM with default parameter values were 72.45%. From this data, we can conclude that SVM-GA is more accurate than SVM-PSO by 5.11% and more accurate than SVM with default parameter values by 5.77%. For the forecasting results of all five stocks, SVM-GA was more accurate than SVM with PSO and SVM with default parameters for almost all stocks. SCC stocks obtained the most accurate results (83.33% accuracy) with a C value of 20. SVM-PSO was most accurate (81.11% accuracy) with C set at 18.42, while SVM-with default parameters yielded 74.44% accuracy for SCC stocks with C set at 14.01.

B. Experiment II: The comparison between SVM-GA technique and indicator analysis techniques

SVM Dot kernel with Genetic Algorithm (GA) obtained the best results from the first experiment was used for further experimenting in Experiment II, which uses the backtesting method to evaluate forecasting ability in terms of return on investment. In addition, the return rate results from SVM Dot kernel with GA will then be compared with other buying and selling techniques that are popular among investors by using indicator analysis techniques such as RSI, Stochastic and MACD [2]. Backtesting will use stock data for the past five years (04/01/2012 - 30/12/2017) from the Stock Exchange of Thailand (SET). This time period covers all buying and selling trends, including uptrend, downtrend and sideways trend. This experiment simulates true investment, using one million Baht to purchase stocks and to compare the resulting data from the experiment with the true values of that time.

TABLE II. The Performance Results (%Accuracy) of SVM Dot Kernel Models With Optimized C Values

Stock	**SVM Dot Kernel**					
	Optimization parameters with GA		***Optimization parameters with PSO***		***Default parameter***	
	C Value	***% Accuracy***	***C Value***	***% Accuracy***	***C Value***	***% Accuracy***
CPALL	20	77.78	1	64.40	14.01	67.78
ICC	20	76.67	12.54	66.67	14.01	67.78
MINT	18	71.11	13.68	67.78	14.01	75.56
SCC	20	83.33	18.42	81.11	14.01	74.44
SC	10	82.22	4.45	85.56	14.01	76.67
Avg. Accuracy (%)		78.22		73.10		72.45

TABLE III. THE COMPARISON OF BACK TESTING: RETURN ON INVESTMENT ON FIVE STOCKS

Stock	SVM - GA		RSI		MACD		STO	
	Profit (Baht)	*% Return*	*Profit (Baht)*	*% Return*	*Profit (Baht)*	*% Return*	*Profit (Baht)*	*% Return*
CPALL	454,080	45.41	128,828	12.88	-69,722	-6.97	399,319	39.93
ICC	14,345	1.43	-35,579	-3.56	19,194	1.92	136,407	13.64
MINT	241,573	24.16	161,434	16.14	207,123	20.71	212,231	21.22
SCC	266,634	26.66	37,299	3.73	192,680	19.27	78,686	7.78
SC	-157,954	-15.8	-224,044	-22.4	-191,054	-19.11	-468,434	-46.84
Avg. Return	818,678	16.37	67,938	1.36	158,222	3.16	358,209	7.16

Table III shows the comparison of return rate results from backtesting for the SVM model with GA optimization technique with the results of technical analysis indicators. The results show that SVM-GA provides the highest return rate, with an average return of 16.37%. In comparison, the technical analysis indicators RSI, MACD and STO return is at 1.36%, 3.16%, and 7.16% respectively. Return rates for SVM-GA are higher than STO by 9.21%, higher than MACD by 13.21%, and higher than RSI by 15.01%. Analysis of each stock return rate shows that SVM-GA yields better return rates than technical analysis indicators for 4 of 5 stocks, comprising CPALL, MINT, SCC, and SC, in which CPALL has the highest profit at 454,080 baht or 45.41%. SCC is second, with a profit of 266,634 baht or 26.66%, and MINT is third with a profit of 241,572 baht or 24.16%. This shows that SVM-GA provides the best results for return from buying and selling stocks in comparison with RSI, MACD, and STO.

To study the correlation between return and price trends, the return was separated into three classes, consisting of the uptrend, downtrend, and sideways trend. The results in Figure 6 show that the SVM-GA Model yielded the highest return at 9.73% during the uptrend, followed by STO with the return at 7.25%. During downtrend and sideways trend, SVM-GA results were similar to the other techniques. During the downtrend, SVM-GA yielded 1.53% return while RSI had the best return at 1.56%. During sideways trend, the SVM-GA Model yielded an average 4.31% return while MACD showed the best results at 4.55% return.

Analysis of the results of return during each trend shows that SVM-GA is the best technique for stock forecasting, as SVM-GA is most effective during the uptrend and provides results that are similar to the other techniques during the other trends. Therefore, SVM-GA provides satisfactory results for all trends.

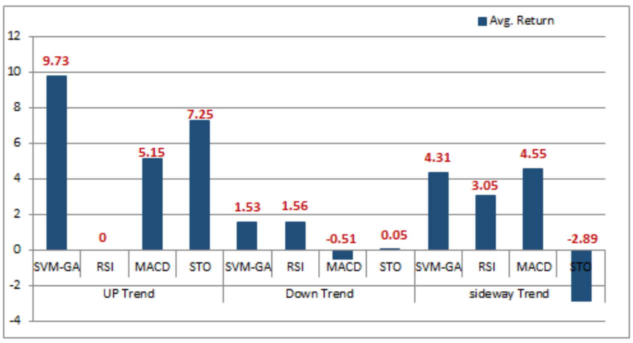

Figure 6. The comparison of investment return in each trend

V. CONCLUSION

This study assessed the stock market forecasting ability of the SVM Dot Kernel by comparing between optimization techniques to find which technique can optimize C parameter value to be most suitable for effective forecasting, between Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), and comparing accuracy and precision for highest return between SVM Dot kernel, RSI, MACD, and STO. Results show that Genetic Algorithm (GA) is more suitable for optimizing C and can support the SVM Model to yield more accurate results than Particle Swarm Optimization (PSO). In the comparison by backtesting, SVM-GA was shown to obtain better results for profit than other three popularly technical analysis indicators, consisting of RSI, STO, and MACD. In addition, SVM-GA results are satisfactory for all trends, with the highest return in the uptrend. For the future work, in order to generalize the current study, a number of stocks are needed to perform backtesting. Other alternative optimization techniques which are related to data modeling can be tested and compared the results with this study.

REFERENCES

[1] G. Jaiwang and P. Jeatrakul, “A Forecast Model for Stock Trading using Support Vector Machine,” 20th International Computer Science and Engineering Conference Smart Ubiquitous Computing And Knowledge(ICSEC), Chiang Mai, Thailand, 2016, pp. 1-6.

[2] S. Boonpeng and P. Jeatrakul, “Decision Support System for Investing in Stock Market by using OAA-Neural Network,” 8th International Conference on Advanced Computational Intelligence, Chiang Mai, Thailand, 2016, pp. 1-7.

[3] S.-W. Lin, Z.-J. Lee, S-C Chen, and T.-Y. Tseng, "Parameter detrermination of support vector machine and feature selection using simulated annealing approach," Journal of Soft Computing for Dynamic Data Mining, vol. 8, pp. 1505-1512, 2008.

[4] Y. Y, Meng, "Parameters Selection of Hybrid Kernel Based on GA," International Workshop on Intelligent Systems and Applications (ISA), TBD Wuhan, China, 2009 , pp. 1-4.

[5] L. Liu, X. Lai, K.-I. Song, and D. Lao, “Intelligent Prediction Model base on Genetic Algorithm and Support Vector Machine for Evaluation of Mining-Induced Building damage,” Journals in Technology and Engineering of Technical Gazette, Vol. 22, No. 3, pp. 743-753, 2015.

[6] R. Rajabioun and A. R.-Kian., “A Genetic Programming Based Stock Price Predictor together with Mean-Variance Based Sell/Buy Actions,” in Proceedings of the World Congress on Engineering, WCE 2008, U.K., 2008, pp. 1136-1141.

[7] G. Bonde and R. Khaled, “Stock price prediction using genetic algorithms and evolution strategies,” in Proceedings of the 2012

International Conference on Genetic and Evolutionary Methods(GEM 2012), Las Vegas, United States, 2012, pp. 1-6.

[8] S.-W. Lin, K.-C. Ying, S.-C. Chen, and Z.-J. Lee, "Particle swarm optimization for parameter determination and feature selection of support vector macvhines," International Journal of Expert Systems with Applications, Vol. 35, Issue 4, pp. 1817-1824, 2008.

[9] S. Lessmann, R. Stahlbock, and S. F. Crone, "Genetic Algorithms for Support Vector Machine Model Selection," International Joint Conference on Neural Networks, Vancouver, BC, Canada, 2006, pp. 3063-3069.

[10] C.-H. Wu, G.-H. Tzeng, Y.-J. Goo, and W.-C. Fang, "A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy," International Journal of Expert Systems with Applications, Vol. 32, pp. 397-408, 2007.

[11] C.-L.Huang and J.-F. Dun, "A distributed PSO-SVM hybrid system with feature selection and parameter optimization," Journal of Soft Computing for Dynamic Data Mining, vol. 8, pp. 1389-1391, 2008.

[12] A. Wongjak, K. Siriboon, and B. Kruatrachue, "Improve Neighbor Search Ability of Genetic Algorithms Using Particle Swarm Optimization Technique," Journal of Information Technology, vol. 7, No.14, pp.1-6, 2011.

[13] D. S. Kim, H.-N. Nguyen, and J. S. Park, “Genetic Algorithm to Improve SVM Based Network Intrusion Detection System,” in Proceedings of the 19th International Conference on Advanced Information Networking and Applications (AINA’05), Tamkang University, Taiwan, 2005, pp. 1-4.

[14] M. N. Alam, Particle Swarm Optimization: Algorithm and its Codes in MATLAB. ResearchGate (2016). Available: https://www.researchgate.net/publication/297245624. [Accessed: Mar-2016].

[15] P. Biswas, Particle Swarm Optimization (PSO). 2016. Available: https://www.mathworks.com/matlabcentral/fileexchange/43541-particle-swarm-optimization--pso-. [Accessed: 15-Jun- 2016].