# Dhaka Stock Exchange Trend Analysis Using Support Vector Regression

Phayung Meesad and Risul Islam Rasel

Faculty of Information Technology
King Mongkut's University of Technology North Bangkok, Thailand
pym@kmutnb.ac.th, rasel.kmutnb@gmail.com

**Abstract.** In this study we combine support vector machine (SVM) and windowing operator in order to predict share market trend as well as the share price. The instability of the time series data is one of the main reasons to lead to decrease of prediction accuracy in this analysis. On the other hand, some special SVM parameters such as c, ε, g should be carefully determined to gain high accuracy. In order to solve this problem mentioned above we use windowing operator as preprocess in order to feed the highly reliable input to SVM model. And train the model in iterative process such that we can find out the best combination of SVM parameters. This study is done on some listed company of Dhaka stock exchange (DSE), Bangladesh. And the training and testing data sets are real time values are collected from DSE. Four years historical data (2009-2012) are used in this analysis. And finally, we compare the output with the real time trend from DSE.

**Keywords:** SVM, Windowing operator, Stock market, Time series data.

## 1 Introduction

Stock market is the emerging sector in any country of the world now. Many people are directly related to this sector. It is important for the people who are directly related to the market to gain insight about the market trend. So, along with the development of stock market, forecasting stock has become an important topic among the people. Trend forecasting becomes an essential topic for stockholders, investors and the authority that are related to the stock market business.

Predicting stock price is regarded as a challenging task [1]. Stock market is essentially a non-linear, non-parametric, noisy and deterministically chaotic system [2][3][4]. Trend of a market depends on many things like liquid money, stocks, Human behavior, news related to stock market etc. All this together control the trend of a stock market. The goal of predicting market trend is to make assumption about the price of assets in stock market. The behavior of trend can be analyzed by using technical tools, parametric pricing methods or combination of these methods [5].

Neural Networks (NNs) and Support Vector Machines (SVMs) [6][7][8][9] are both standard, mature machine learning approaches with applications in prediction based on times series data. Many research works have been done before using these

P. Meesad et al. (Eds.): *IC²IT2013*, AISC 209, pp. 135–143.
DOI: 10.1007/978-3-642-37371-8_17 

two techniques. In some recent researches, researchers have found that SVM can produce more accurate stock prediction than the NNs can do. Support Vector Regression (SVR) is the most common application form of SVMs [10][11]. One of the main characteristics of Support Vector Regression (SVR) is that instead of minimizing the observed training error, SVR attempts to minimize the generalized error bound so as to achieve generalized performance. This generalization error bound is the combination of the training error and a regularization term that controls the complexity of the hypothesis space [12][13]. Kim [1], Lai and Liu [3] showed in their research that SVM can be more accurate in producing result if we can chose the best combination of SVM parameters. Ince & Trafalis [2] showed the way how Kernel Principal Component can be analyzed to get better result. Lai and Liu [3] compare the prediction performances of NN and SVM in predicting exact stock prices on the Hang Seng Index (HSI) over 5 days and a 22 days horizon. As preprocessing or input selection techniques for SVR and NN, they used 15 days Exponential Moving Average (EMA15) and relative difference in percentage of price (RDP) RDP-5, RDP-10, RDP-15, and RDP-20. The best MAPE result was 0.8 for the year of 2008 short term forecast (5 days). For long term prediction (22 days), the best result was 4.33 also for the 2008 dataset. The average result was 5.02.

This paper consists of five sections. Section 2 introduces the basic concept of SVM and windowing operator. Section 3 is about the research design. Section 4 is the experiment and analysis. Section 5 is the conclusions and limitations of this study.

## 2 Basic Concept of Proposed Model

### 2.1 SVM Regression

SVM regression perform linear regression in the high dimension feature space using $\varepsilon$ – insensitivity loss and, at the same time tries to reduce model complexity by minimizing $||\omega||^2$. This can be described by introducing slack variables $\xi_i$ and $\xi_i^*$ where$i = 1, \dots, n$, to measure the deviation of training sample outside $\varepsilon$- sensitive zone [3][9][13].

$$\frac{1}{2}||\omega||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{1}$$

$$\text{Min} \begin{cases} y_i - f\left(x_{i,\omega}\right) \le \varepsilon + \xi_i^* \\ f\left(x_{i,}\omega\right) - y_i \le \varepsilon + \xi_i \\ \xi_i, \xi_i^* \ge 0, i = 1, \dots, n \end{cases} \tag{2}$$

This optimization problem can transform into the dual problem and solution is given by

$$f(x) = \sum_{i=1}^{n_{sv}}(\alpha_i - \alpha_i^*)K(x_{i\,,}x) \tag{3}$$

$$\text{Subject to, } 0 \le \alpha_i^* \le C, 0 \le \alpha_i \le C,$$

Where $n_{sv}$ is the number of support vector (SVs) and the kernel function

$$K(x, x_i) = \sum_{j=1}^{m} g_j(x) g_j(x_i) \tag{4}$$

SVM generalization performance depends on a good setting of kernel parameters $C, \varepsilon$ and kernel parameters [9][13].

The following formula is the evaluation of the predicted value [3].

$$MAPE = 100 \frac{\sum_{i=1}^{n} |\frac{A-P}{A}|}{n} \tag{5}$$

Mean Absolute Percentage Error (MAPE) which is the measure of accuracy in a fitted time series value in statistics, specifically trending. $A$ and $P$ are the real close value and the predicted close value respectively and n is the time frame or number of days.

# 3 Research Design

## 3.1 Research Data

To conduct the study and verify the effectiveness of our proposed model we collect the data set from Dhaka stock exchange (DSE), Bangladesh. We separate our data set into two groups. Training data set contains data from year 2009 to 2011 (700 data) and testing data set contains data from year 2012 (124 data). There are 4 attributes of this data set. They are open price, high price, low price and close price. One special attribute is date field used as ID field for this study. Figure 1 shows the actual close

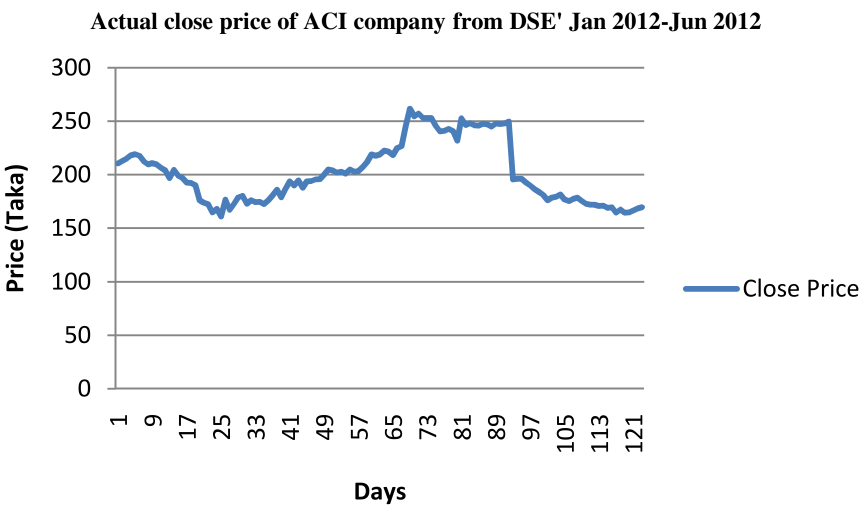


**Fig. 1.** Actual Share price from Jan 1, 2012 to Jun 28, 2012

price of ACI Company from DSE for the month of January 2012 to June 2012. Though we collect all listed company share price data from DSE but for the convenience of research analysis, we filter only one well-known company, named ACI group of company Limited's share price data in daily basis. But, the study can be apply to any kind of stock price data set which should contain 5 attributes, like date, open, high, low and close price value. Date is a special attribute used as id in this research and other 4 other attributes are used as regular attributes.

### 3.2 Analysis Steps

This study is conducted in two phases. Training phase and testing phase. The steps are given below:

**Training Phase**

Step 1: Read the training data set from local repository.
Step 2: Apply windowing operator to transform the time series data into a generic data set. This step will convert the last row of a window within the time series into a label or target variable. Last variable is treated as label.
Step 3: Accomplish a cross validation process of the produced label from windowing operator in order to feed them as input to the SVM model.
Step 4: Select kernel types and select special parameters of SVM (c, $\varepsilon$, g etc).
Step 5: Run the model and observe performance (accuracy).
Step 6: If performance accuracy is good than go to step 6, otherwise go to step 4.
Step 7: Exit from the training phase

**Testing Phase**

Step 1: Read the testing data set from local repository.
Step 2: Apply the training model to test the testing data set for price prediction.
Step 3: Produce the predicted trends and stock price.

## 4 Experiments and Analysis

### 4.1 Data Preprocess

To produce the optimized input for the SVM model, we use windowing operator to the time series data set as preprocessing step. Windowing allows us to take any time series data and transform it into a cross-sectional format. But for that, we should find out the proper windowing size and step size in order to produce the label. In our study, we use date as ID and close value as label. A Sliding Window Validation process is also applied to evaluate the output from the preprocess step. After doing the analysis, we get some proper combination of the windowing component to produce optimized input for the SVM model. Those are given in Table 1.

**Table 1.** Windowing Component

| Window size | Step size | Training window width (TWW) | Training step size (TSS) | Testing window width (tww) |
|---|---|---|---|---|
| 3 | 1 | 30 | 1 | 30 |

## 4.2 Kernel Component Analysis

In our research we use RBF kernel to predict the stock price and trend. Because the output from RBF kernel is good enough and it also takes short processing time. The most important components of RBF kernel are $C$ value, g value, and epsilon ($\varepsilon$). In this research we try to find out the best combination of these values. And finally, we got some combination that produced good prediction result.

**Table 2.** RBF kernel component

| SVM Models | Kernel | $C$ | $g$ | $\varepsilon$ | $\varepsilon+$ | $\varepsilon-$ | Avg MAPE |
|---|---|---|---|---|---|---|---|
| Model-1 | RBF | 10000 | 1 | 2 | 1 | 1 | 0.42 |
| Model-2 | RBF | 10000 | 1 | 2 | 1 | 1 | 0.27 |
| Model-3 | RBF | 10000 | 1 | 2 | 1 | 1 | 0.22 |

In Table 2, there are three models. Model-1 is for predicting 1 day horizon stock price and trend. Model-2 and Model-3 are respectively for predicting 5 days horizon and 22 days horizon stock price and trend.

Table 3 shows support vector numbers (SV), bias values (b) and weights (w) for respective regression model for deferent horizons.

**Table 3.** Proposed SVM models for 1 day, 5 days and 22 days ahead prediction

| Models | Horizon | Support Vector | Bias (offset) | Weight ($w$) | | |
|---|---|---|---|---|---|---|
| | $h$ | SV | $b$ | w1 [close-2] | w1 [close-1] | w1 [close-0] |
| Model-1 | 1 | 696 | 400.686 | 1358.881 | 627.029 | 501.037 |
| Model-2 | 5 | 692 | 381.482 | 825.014 | 734.139 | -297.092 |
| Model-3 | 22 | 675 | 421.296 | 1719.578 | 1631.468 | 805.925 |

## 4.3 Error Calculation

To evaluate the predicted stock price from our model, we apply Mean Absolute Percentage Error (MAPE). We compare our predicted stock price with real time DSE stock price for the month of January 2012 to May 2012. Table 4 shows the error calculation for 3 regression models. Model 1 is one day a-head prediction model, Model 2 is 5days a-head prediction model and Model 3 is 22 days a-head prediction model. And for evaluating prediction result and calculating MAPE we use 100 actual price data from 2$^{nd}$ January 2012 to 30$^{th}$ May 2012.

**Table 4.** MAPE calculation results for testing dataset (Jan-2012 to May-2012)

| Models | Jan-12 | Feb-12 | Mar-12 | Apr-12 | May-12 |
|---|---|---|---|---|---|
| Model-1 | 0.16 | 1.42 | 0.1 | 0.07 | 0.35 |
| Model-2 | 0.18 | 0.7 | 0.04 | 0.07 | 0.33 |
| Model-3 | 0.18 | 0.32 | 0.16 | 0.28 | 0.14 |

### 4.4 Graph Analysis

Figure 2 shows the actual and predicted share price of ACI group from DSE, for the month of January 2012 to May 2012. This is the outcome of the 1 day a-head prediction model. In this model the average error rate of predicting price value is 0.42.

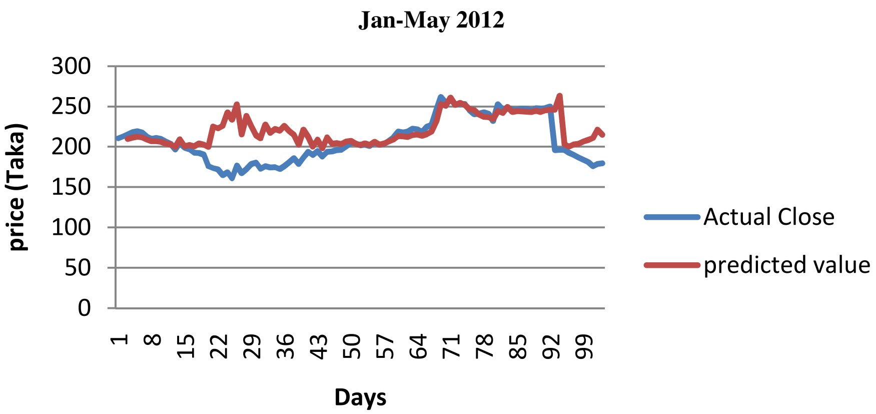


**Fig. 2.** Model-1 (1 day horizon)

Figure 3 shows the actual and predicted share price of ACI group from DSE, for the month of January 2012 to May 2012. This is the outcome of the model-2, which is based on 5 day a-head prediction. In this model the average error rate of predicting price value is 0.27.

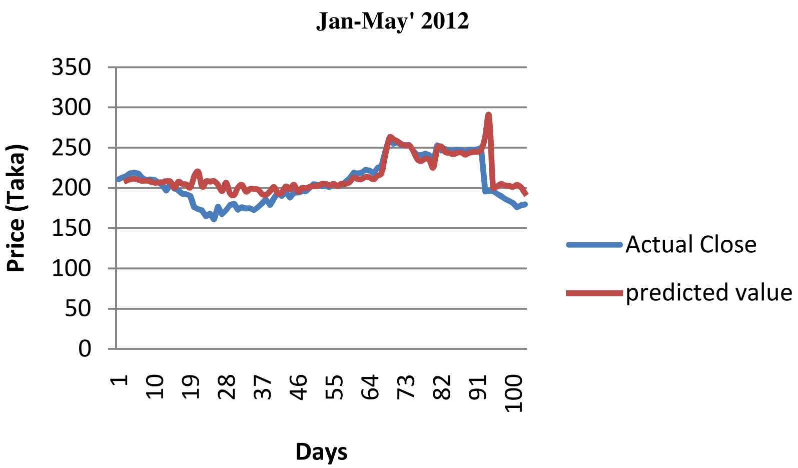


**Fig. 3.** Model-2 (5 days horizon)

Figure 4 shows the actual and predicted share price of ACI group from DSE, for the month of January 2012 to May 2012. This is the outcome of the model-3, which is based on 22 days advanced prediction. In this model the average error rate of predicting price value is 0.22.

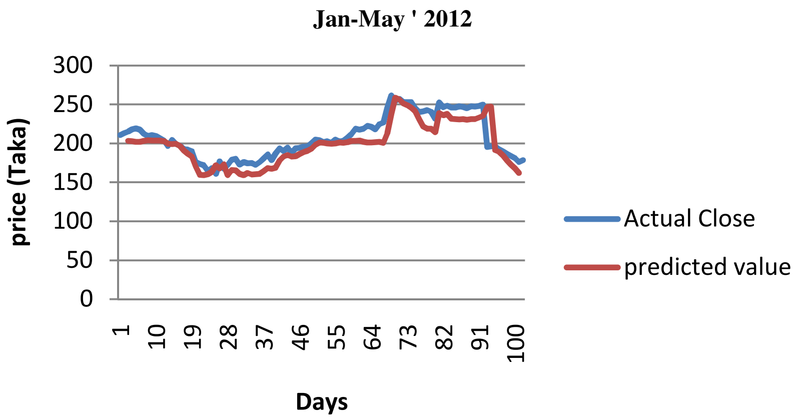


**Fig. 4.** Model-3 (22 days horizon)

### 4.5 Prediction Results

Table 5, 6, & 7 show predicted result and the error rate from 1 day, 5 days & 22 days a-head prediction model respectively, for the month of January 2012 to May 2012. In those tables, the actual price (Taka) is taken from current market trend of DSE in Bangladesh currency. And this share price is for ACI group of Company Limited, BD.

Figure 5 shows the MAPE values for the month of January 2012 to May 2012. From the figure, we see that 5days a-head and 22 days a-head regression model produce the less erroneous price value for the stock market.

**Table 5.** Prediction result from 1 day a-head model (Jan 2012 – May 2012)

| Month | Actual price (Taka) | Predicted price (Taka) | Error | Abs error value | MAPE |
|---|---|---|---|---|---|
| January | 3844.4 | 4147.85 | -303.45 | 303.45 | 0.394665 |
| February | 3341.4 | 4241.33 | -899.93 | 899.93 | 1.417512 |
| March | 4032.8 | 4110.07 | -77.27 | 77.27 | 0.095802 |
| April | 5300.1 | 5218.71 | 81.39 | 81.39 | 0.069801 |
| May | 4280.3 | 4578.76 | -298.46 | 298.46 | 0.348644 |

**Table 6.** Prediction result from 5 days a-head model (Jan 2012 – May 2012)

| Month | Actual price (Taka) | Predicted price (Taka) | Error | Abs error value | MAPE |
|---|---|---|---|---|---|
| January | 3844.4 | 4156.26 | -311.86 | 311.86 | 0.405603 |
| February | 3341.4 | 3783.34 | -441.94 | 441.94 | 0.696115 |
| March | 4032.8 | 4062.15 | -29.35 | 29.35 | 0.036389 |
| April | 5300.1 | 5220.05 | 80.05 | 80.05 | 0.068652 |
| May | 4280.3 | 4558.99 | -278.69 | 278.69 | 0.32555 |

**Table 7.** Prediction result from 22 days a-head model (Jan 2012 – May 2012)

| Month | Actual price (Taka) | Predicted price (Taka) | Error | Abs error value | MAPE |
|---|---|---|---|---|---|
| January | 3844.4 | 3876.92 | -32.52 | 32.52 | 0.042295 |
| February | 3341.4 | 3138.04 | 203.36 | 203.36 | 0.32032 |
| March | 4032.8 | 3906.77 | 126.03 | 126.03 | 0.156256 |
| April | 5300.1 | 4979.68 | 320.42 | 320.42 | 0.274798 |
| May | 3922.4 | 3825.77 | 96.63 | 96.63 | 0.136863 |

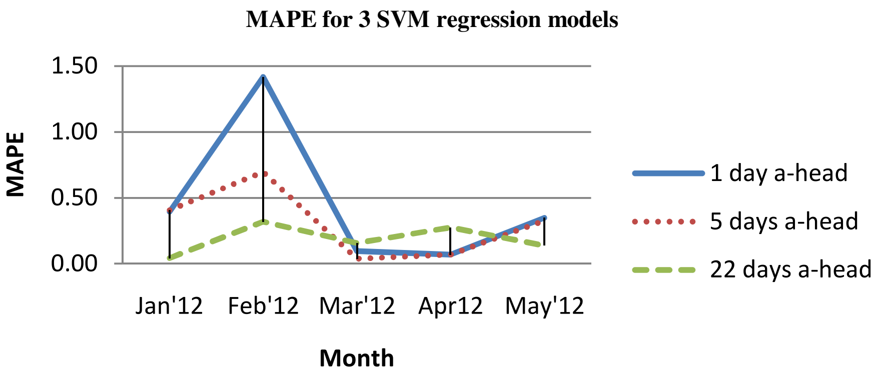


**Fig. 5.** MAPE for the month of January 2012 to May 2012

# 5 Conclusion and Limitation

## 5.1 Conclusion

Our motivation is to predict a good trend for the stock market. So, we focused on accurate stock price prediction and as well as trend prediction. After doing this study we get that, 5 days a-head and 22 days a-head prediction model produce less erroneous stock price for DSE. But, all these three models can produce a good prediction results.

### 5.2 Limitations and Future Work

In this work we only use data set from DSE and also evaluate our result comparing with DSE. In future, we will apply our model to other stock market data set and will also compare our research result with other types of data mining techniques.

**Acknowledgement.** The Authors are very grateful to Mr. Ashraf Uddin, Sr. Executive at LankaBangla Securities Ltd, for providing us the real time stock price data from Dhaka stock exchange (DSE), Bangladesh.

## References

1. Kim, K.: Financial time series forecasting using support vector machines. Neurocomputing 55, 307–319 (2003)
2. Ince, H., Trafalis, T.B.: Kernel Principal Component Analysis and Support Vector Machines for Stock Price Prediction, pp. 2053–2058 (2004) 0-7803-8359-1/04/2004 IEEE
3. Lucas, K., Lai, C., James, N., Liu, K.: Stock Forecasting Using Support Vector Machine. In: Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, pp. 1607–1614 (2010)
4. Lu, C., Chang, C., Chen, C., Chiu, C., Lee, T.: Stock Index Prediction: A Comparison of MARS, BPN and SVR in an Emerging Market. In: Proceedings of the IEEE IEEM, pp. 2343–2347 (2009)
5. Kannan, K.S., Sekar, P.S., Sathik, M.M., Arumugam, P.: Financial Stock Market Forecast using Data Mining Techniques. In: Proceedings of the International Multiconference of Engineers and Computer Scientists, pp. 555–559 (2010)
6. Hu, Y., Pang, J.: Financial crisis early warning based on support vector machine. In: International Joint Conference on Neural Networks, pp. 2435–2440 (2008)
7. Chen, K., Ho, C.: An Improved Support Vector Regression Modeling for Taiwan Stock Exchange Market Weighted Index Forecasting. In: The IEEE International Conference on Neural Networks and Brain, pp. 1633–1638 (2005)
8. Xue-Shen, S., Zhong-Ying, Q., Da-Ren, Y., Qing-Hua, H., Hui, Z.: A Novel Feature Selection Approach Using Classification Complexity for SVM of Stock Market Trend Prediction. In: 14th International Conference on Management Science & Engineering, pp. 1654–1659 (2007)
9. Debasish, B., Srimanta, P., Dipak, C.P.: Support Vector Regression. Neural Information Processing – Letters and Reviews 11(10), 203–224 (2007)
10. Hsu, C., Chang, C., Lin, C.: A Practical Guide to Support Vector Classification. Initial version (2003) Last updated version (2010)
11. Thissena, U., Brakela, R., Weijerb, A.P., Melssena, W.J., Buydensa, L.M.C.: Using support vector machines for time series prediction. Chemometrics and Intelligent Laboratory Systems 69, 35–49 (2003)
12. Cao, L.: Support vector machines experts for time series forecasting. Neurocomputing 51, 321–339 (2003)
13. Alex, J., Bernhard, S.: A tutorial on support vector regression. Statistics and Computing 14, 199–222 (2004)