Internet of Things and Cloud Computing
2020; 8(4): 46-51
http://www.sciencepublishinggroup.com/j/iotcc
doi: 10.11648/j.iotcc.20200804.12
ISSN: 2376-7715 (Print); ISSN: 2376-7731 (Online)

Predicting Daily Closing Prices of Selected Shares of Dhaka Stock Exchange (DSE) Using Support Vector Machines

Md. Farhad Hossain[1, *], Sharmin Islam[2], Partha Chakraborty[3], Ajit Kumar Majumder[4]

[1]Department of Statistics, Comilla University, Cumilla, Chattogram, Bangladesh

[2]Department of Statistics, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Dhaka, Bangladesh

[3]Department of Computer Science and Engineering, Comilla University, Cumilla, Chattogram, Bangladesh

[4]Department of Statistics, Jahangirnagar University, Savar, Dhaka, Bangladesh

Email address:
farhad390ju@gmail.com (Md. F. Hossain), sharminislameity@gmail.com (S. Islam), partha.chak@cou.ac.bd (P. Chakraborty), ajit@juniv.edu (A. K. Majumder)

[*]Corresponding author

To cite this article:
Md. Farhad Hossain, Sharmin Islam, Partha Chakraborty, Ajit Kumar Majumder. Predicting Daily Closing Prices of Selected Shares of Dhaka Stock Exchange (DSE) Using Support Vector Machines. *Internet of Things and Cloud Computing*. Vol. 8, No. 4, 2020, pp. 46-51. doi: 10.11648/j.iotcc.20200804.12

Received: August 26, 2020; **Accepted**: December 23, 2020; **Published**: December 31, 2020

Abstract: Support Vector Machines (SVM) has been a naval research field in scientific research for forecasting. This study deals with the application of SVM in financial time series predicting. This paper suggests a model of stock market prediction based on SVMs with appropriate parameter values. A data set of daily closing prices of five selected companies such as Alhaj Textiles Limited, Apex Tannery Limited, Jamuna Bank Limited, Padma Oil Company, and Square Pharmaceuticals Limited of the Dhaka Stock Exchange (DSE) from 01 January 2017 to 13 August 2019 was selected and uses these data to train the model and checks the predictive power of the model. The obtained results show that all the companies closing stock prices are non-stationary. Also the number of support vectors and mean square error is decreasing pattern with the increase of kernel parameter. It is also found that original data and predicted data are very much identical. The result shows that in all the cases SVM model has some predictive power it can be used to forecast financial time series. Several methods, such as SVM, ARIMA, single exponential smoothing, and double exponential smoothing, were performed to predict Bangladesh's stock market. Amazingly, the outcome shows the most efficient method to be Support Vector Machine because of its lowest forecasting errors.

Keywords: Time series Forecasting, Financial Market, Support Vector Machines, Dhaka Stock Exchange, Machine Learning

1. Introduction

Financial time series forecasting is one of the most challenging applications of modern time series forecasting [1, 4]. Stock price time series are data-intensive, noisy, dynamic, unstructured, and highly uncertain [20]. There have been many studies on forecasting time-series data. In recent years, the neural network has been successfully applied to financial time series modeling from Stock Price Index [9, 12] to the option price [11]. Over the past decade, neural networks have been successfully used for modeling financial time series [1, 19]. Recently, Support Vector Machines (SVM), a novel neural network algorithm developed by Vapnik and his colleagues is a focus research field in the world [14, 16]. SVM method, which was first suggested by Vapnik has recently been used in a range of applications such as in data mining, classification, regression, and time series forecasting [10, 7, 13]. The SVM has become a hot topic of intensive study due to its successful application in classification tasks [17, 3] and regression tasks [8, 18], especially on time series prediction [2]. SVM is a training algorithm for learning classification and regression rules from data [3]. This research is to explore the use of SVM for stock price prediction by performing a comprehensive experimental study for predicting stock prices. This paper navels on the application of SVMs in regression tasks to predict the financial time series in the D.S.E. This study develops the SVM model for the different regularization constant (C) and different kernel parameters of radial basis function. Determine the validity of the model, different performance metrics were calculated.

2. Material and Methodology

2.1. Research Data

This study considered the daily closing price index of the Dhaka Stock Exchange conveniently from five different sectors for the selected companies such as Alhaj Textiles Limited, Apex Tannery Limited, Jamuna Bank Limited, Padma Oil Company, and Square Pharmaceuticals Limited within 01-01-2017 to 13-08-2019 from the DSE website based on their historical prices. Among these data, the training data cover the period from 01 January 2017 up to the end of 30 December 2018, while the data starting from 01 January 2019 up to 13 August 2019 has used as the test data. All the programs used to generate the results in this study have been written in R- package version 3.0.2. and the figures are produced in M.S. Excel.

2.2. Prediction Theory of SVM

Suppose the training nonlinear time series sample is $\{(x_1, y_1), (x_2 y_2),\ \Lambda, (x_k, y_k)\}$ $(x_i \in R^n,\ y_i \in R,\ i = 1, 2,\ \Lambda\, k,$ is the number of training sample data). The basic idea of SVM for time series prediction is a nonlinear mapping φ that transfers time series to the high dimensional feature space F. It then constructs the optimized linear regression function in the ample feature space, and the expression of the linear regression function as follows:

$$f(x) = w\varphi(x) + b. \tag{1}$$

In the expression above, w and $\varphi(x)$ are both m-dimension vector, and b is the offset value. SVM adopts the structural risk minimization principle to determine the importance of w and b. Namely

$$\min R_{str} = \frac{1}{2}\|w\|^2 + CR_{emp}. \tag{2}$$

In the expression (2) $\|w\|^2$ is the complexity of control and c is the weight thathas used to control the punishment degree that exceeds the error sample.

$R_{emp} = \dfrac{1}{k}\sum_{i=1}^{k} L_t[x_i, y_i - f(x_i)\ \ ]$ is the error control function, which is usually measured by the ε insensitive loss function, and the insensitive loss function is defined as follows:

$$L_e = \begin{cases} \left|y - f(x)\right| - \varepsilon & \left|y - f(x)\right| \ge \varepsilon \\ 0 & \left|y - f(x)\right| < \varepsilon. \end{cases}$$

According to the structural risk minimization principle considering the complexity of the regression model obtained from the training set, regression based on the SVM essentially is a solution of an optimized question, and the optimized question is in the following.

$$\min\ \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{k}\left(\xi_i + \zeta_i^*\right).$$

In the question ζ_i and ζ_i^* are slack variables and the problem is described as the original question of the SVM. For the value of a dimension, w is vast, to conveniently solve the issue, introduces the Lagrange multiplier α_i and α_i^* according to the duality theorem, and establishes a Lagrange function. The optimized question is converted to the dual space and acquires the twofoldissue of the original question; the formula has shown as the (2) expression.

$$\min\ \frac{1}{2}\sum_{i,j=1}^{k}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)k(x_i, x_j) + \varepsilon\sum_{i=1}^{k}(\alpha_i^* + \alpha_i) - \sum_{i=1}^{k} y_i(\alpha_i^* - \alpha_i)$$

$$s.t. \begin{cases} \sum(\alpha_i^* - \alpha_i) = 0 \\ 0 \le \alpha_i^*, \alpha_i \le c(I = 1, 2 \wedge k). \end{cases}$$

In the expression $k(x_i, x_j) = [\varphi(x_i) \bullet \varphi(x_j)]$ are the kernel function, and the most commonly used optimized kernel function is the Gauss function and the concrete formula

$$k(x_i, x_j) = \exp(-\frac{\left\|x_i - x_j\right\|^2}{\sigma^2}).$$

Based on the nature of quadratic programming, only several coefficients $a_i, a_i^{'}$ assume as nonzero, and the data points associated with them could be referred to as support vectors.

Kernel Function

For SVM, the following four underlying kernels are used

1) Linear: $K\left(x_i, x_j\right) = x_i^T x_j.$,

2) Polynomial: $\left(\gamma x_i^T x_j + r\right)^d, \gamma > 0$

3) Radial basis function (RBF): $\mathrm{Exp}\left(-\gamma\left\|x_i - x_i\right\|^2\right),\ \gamma > 0$

and

4) Sigmoid: $K\left(x_i, x_j\right) =$ tanh $\left(\gamma x_i^T x_j + \gamma\right)$. Where, $\gamma,\ r$, and d are kernel parameters.

3. Results and Discussions

The use of SVMs of the selected companies' stock price prediction has studied in this paper. This study considers the daily closing price index of the Dhaka stock index from five different companies, such as Alhaj Textiles Limited, Apex Tannery Limited, Jamuna Bank Limited, Padma Oil Company, and Square Pharmaceuticals Limited between 01-01-2017 to 13-08-2019. Among the total of 629 data, 481 data are used for training data, and the rest 148 data are used for test data. The experiments will emphasize on verifying the performance and effectiveness of the SVMs. In our operations, the kernel parameter, and are selected based on the validation set. In the next paragraph, M.S.E., RMSE, NRMSE, and the number of support vectors concerning the three free parameters have investigated. Only the results of the kernel parameter have illustrated; the same has applied to the other two settings. In this investigation, the Radial basis function is used as the kernel function of the SVMs because Radial basis kernels tend to giveexcellent performance under general smoothness assumptions [6].

Consequently, they are especially useful if no additional knowledge of the data is available. This study compares the prediction performance concerning various kernel parameters and constants. According to Tay and Cao [15], an appropriate range was between 1 and 100. Besides, they proposed that a properscale for C was between 10 and 100. The LibSVM algorithm hasused in this study. This algorithm needs to determine some parameters they are regularization constant, setting of the kernel functions, and the margins of error. After performing cross-validation in the first training data, themetrics and are chosen to be and respectively, because these values produced the best possible results according to the validation set. An appropriate amount of the kernel parameter ( ) would be between 10 and 100 [15]. The prediction performance is evaluated using the statistical matrices, Mean Squared Error (M.S.E.), a measure of the deviation between actual and predicted values. The smaller amounts of M.S.E., RMSE, NRMSE, the closer is the expected time series values to that of the actual costs [5].

The study has been calculated different performance metrics for varying parameters of the kernel ( ) for the selected companies that illustrate in the below tables.

Table 1. *Performance metrics of Alhaj Textiles Company Limited.*

γ	Number of support vector (N.S.V.)	MSE	RMSE	NRMSE
.1	81	230.29	15.175	.02741
1	77	227.31	15.076	.02723
10	54	241.32	15.534	.02806
100	30	221.28	14.875	.02687
1000	30	221.21	14.873	.02686
10000	28	209.28	14.466	.02613

Table 2. *Performance metrics of Apex Tannery Limited.*

γ	Number of support vector	MSE	RMSE	NRMSE
.1	348	180.32	13.42	0.2968
1	331	175.86	13.26	0.315
10	269	190.28	13.79	0.374
100	176	150.85	12.28	0.349
1000	118	148.62	12.19	0.345
10000	118	141.86	11.91	0.328

Table 3. *Performance metrics Jamuna Bank Limited.*

γ	Number of support vector	MSE	RMSE	NRMSE
0.1	194	228.82	15.126	0.358
1	186	225.32	15.010	0.364
10	184	238.82	15.45	0.374
100	181	219.66	14.820	0.381
1000	180	218.28	14.774	0.38
10000	180	208.75	14.448	0.363

Table 4. *Performance metrics ofPadma Oil Company Limited.*

γ	Number of support vector	MSE	RMSE	NRMSE
0.1	212	825.68	28.73	0.084
1	198	840.29	28.98	0.082
10	110	827.62	28.76	0.08
100	95	820.64	28.64	0.078
1000	90	818.18	28.60	0.075
10000	81	815.13	28.55	0.071

Table 5. *Performance metrics ofSquare Pharmaceuticals Company limited.*

γ	Number of support vector	MSE	RMSE	NRMSE
0.1	276	195.65	13.98	0.0198
1	210	198.98	14.10	0.0197
10	190	205.63	14.33	0.0194
100	185	211.82	14.55	0.0182
1000	171	210.12	14.49	0.0174
10000	160	208.68	14.44	0.0161

The above tables give the Mean Square Error, Number of Support Vector, Root Mean Square Error, and Normalized Root Mean Square Error with various kernel parameters $\gamma((0.1,10000))$, respectively, in which c and ε has fixed at 10 and .001. The result of these tables shows that in all the cases, the Mean Square Error, Number of Support Vector, Normalized Root Mean Square Error is decreasing with the increasing of kernel parameter γ, So most of the data points are converging to the support vectors. Therefore, the support vector model fits the actual data well.

After the training with SVM, the predicted price and the actual price for the test data has exhibited in the following figures.

Table 6. *Comparison of SVM with other time series method in terms of forecasting errors by four model for five companies.*

Model	Alhaj Textiles			Apex tannery			Jamuna bank		
	MAPE	MAD	RMSE	MAPE	MAD	RMSE	MAPE	MAD	RMSE
ARIMA	4.32	1.97	19.06	3.45	1.79	12.95	3.59	1.17	15.11
SES	3.27	1.95	15.87	3.47	1.62	10.77	3.46	1.68	12.92
DES	3.87	0.987	15.27	3.25	1.50	10.65	2.62	1.60	12.81
SVM	2.29	0.613	10.34	2.85	1.39	7.76	2.02	1.51	10.43

Model	Padma oil company			Square pharmaceuticals limited		
	MAPE	MAD	RMSE	MAPE	MAD	RMSE
ARIMA	2.28	4.07	18.61	2.14	15.65	16.71
SES	2.89	4.35	18.23	2.21	14.59	15.11
DES	2.97	4.53	17.49	2.34	14.72	15.43
SVM	1.28	2.13	14.23	1.38	10.23	11.87

*Mean absolute percentile error (MAPE), Mean absolute deviation (MAD), and Root mean square error (RMSE).
**Auto regressive integrated moving average (ARIMA), Single exponential smoothing (SES), Double exponential smoothing (DES) and Support Vector Machines (SVM).

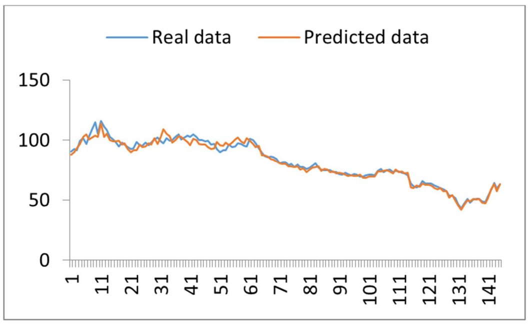

Figure 1. *Real data and predicted data of Alhaj Textiles Limited.*

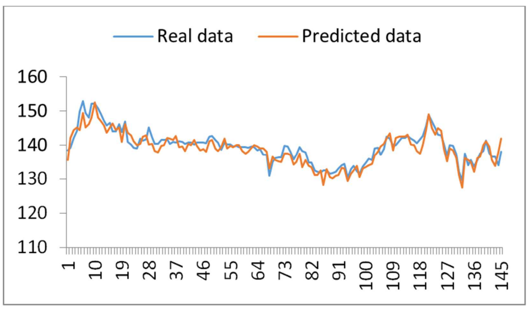

Figure 2. *Real data and predicted data of Apex Tannery Limited.*

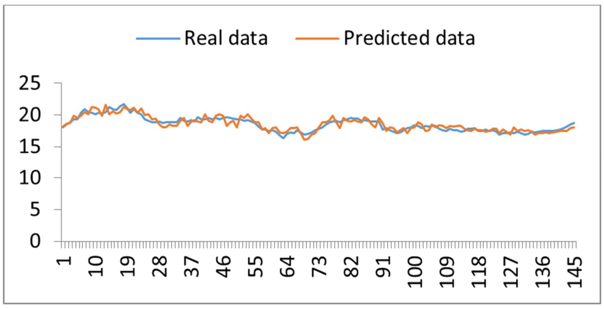

Figure 3. *Real data and predicted data of Jamuna Bank Limited.*

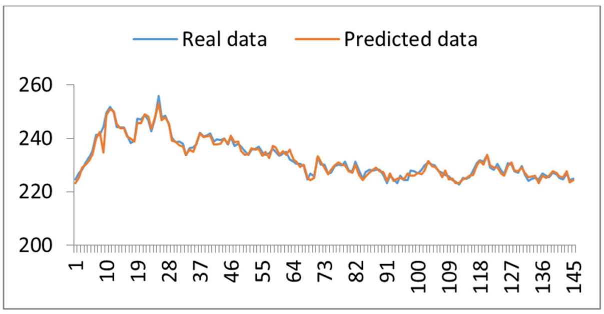

Figure 4. *Real data and predicted data of Padma Oil Company Limited.*

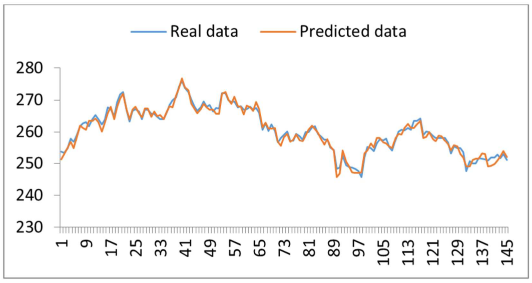

Figure 5. *Real data and predicted data of Square Pharmaceutical Company Limited.*

The above five figures illustrate the Real data and predicted data. The blue line is the actual value, and the red line is the predicted value of our method. From the figures, it can be observed that the predicted value curve's tendencies are identical to that of the actual value curve in all the cases. The predicted curve fits the real curve in most of the period. Therefore, SVMs predict the actual data very well. The experimental results show that this method is more effective and efficient in predicting closing stock prices. The study has also calculated different forecasting errors of Support Vector Machines with Classical time series methods.

Table 6 suggests that, Support Vector Machines returns lowest forecasting errors such as MAPE, MAD, RMSE, NRMSE in all the selected companies, with compare to the forecasting errors obtains from ARIMA, Single Exponential Smoothing, and Double Exponential Smoothing method. So, in terms of minimum forecasting error Support Vector Machines has performed well.

4. Conclusions

The use of Support Vector Machines (SVM) in financial forecasting has been studied in this paper. This work establishes a model of stock market prediction based on SVM. The experimental result showed that the SVM model has some predictive power; it has been used to predict stock prices. The study also reveals that the predicted value curve is identical to the actual value curve. So SVM predicts the actual value very well. It has also been found that in all cases, SVMs are more appropriate to predict stock prices. SVMs predict better, as SVMs provide a smaller M.S.E., NRMSE, and RMSE compares to other classical time series methods. Therefore SVMs shift a hopeful alternative to time series prediction in Bangladesh.

To sum up, as the stock market is an important sector, this study can help determine whether to buy a stock or sell it, and this crucial purpose can be served with the help of stock market predictions. This study has been done for five selected companies, and it has extended with more stock companies on DSE. The other prospect of taking this research further is by comparing the results with other machine learning techniques.

Data Availability Statement

The data that support the findings of this study are openly available on the Dhaka Stock Exchange (D.S.E.) website www.dsebd.org.

References

[1] Abecasis, S. M., Lapenta, E. S., and Pedreira, C. E. (1999). Performance metrics for financial time series forecasting, Journal of computational intelligence in finance, Vol. 7, No. 4: 5-23.

[2] A. Smola and B. Scholkopf. A Tutorial on Support Vector Regression. 1998, Technical Report NeuroCOLT NC-TR-98-030.

[3] Burges, C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discovery 2*, 121-167.

[4] Cao, L. J. & E. H., Tay (2001). Support vector with adaptive parameters in financial time series forecasting, *IEEE Trans. Neural Network*, Volume 14, pp. 1506-1518.

[5] Das, S. P., and Padhy, S., Support Vector Machines for Prediction of Futures Prices in Indian Stock Market, *International Journal of Computer Applications (0975 – 8887), Volume 41 – No. 3, March 2012.*

[6] David J. and Sally L. (Eds), (2012); Advances in Computer Science and Information Engineering. New York, Springer.

[7] Dunham, M. H. (2006). *Data Mining: Introductory and Advanced Topics*. India: Pearson Education.

[8] Edgar Osuma and Robert Freund and Federico Girosi. Support Vector Machines: Training and Applications. AIM-1602, M. I. T., 38, 1997.

[9] Kim, K. J., Han, I.: Genetic Algorithms Approach Feature Discretization in Artificial Neural Networks for the Prediction of Stock Price Index. J. Expert, Syst. Appl. 19, 125–132 (2000).

[10] K. R. Muller, A. Smola, G. Ratsch, B. Scholkopf, J. Kohlmorgen and V. N. Vapnik. Predicting time series with support vector machines. ICANN, 999-1004, 1997.

[11] Liang, X., Zhang, H. S., Xiao, J. G., Chen, Y.: Improving Option Price Forecasts with Neural Networks and Support Vector Regressions. J. Neural Comput. Appl. 72, 3055–3065 (2009).

[12] Liao, Z., Wang, J.: Forecasting Model of Global Stock Index by Stochastic Time Effective Neural Network. J. Expert, Syst. Appl. 37, 834–841 (2010).

[13] Olson, David L., Delen**,** Dursun. (2008). *Advanced-Data Mining Techniques*. New York: Springer.

[14] R. Sharda, R. Pati, A connectionist approach to time series prediction: an empirical test. Neural Networks in Finance and Investing, Chicago: Probus Publishing, 1994, pp. 451-464.

[15] Tay, F. E. H. and Cao, L., "Application of support vector machines in financial time-series forecasting," Omega, Vol. 29, 2001, pp. 309–317.

[16] V. N. Vapnik. The Nature of Statistical Learning Theory. Springer, New York, pp. 1995.

[17] V. N. Vapnik. Statistical Learning Theory. Wiley, New York, 1998.

[18] V. N. Vapink, S. Golowich, and A. Smola. Support vector method for function approximation, regression estimation, and signal processing.

[19] W. Cheng, L. Wanger, Forecasting the 30-year U.S. Treasury bond with a system of neural networks. Journal of Computational Intelligence in Finance, 4 (1996), pp. 10-16.

[20] Yaser, S. A. M., Atiya, A. F.: Introduction to Financial Forecasting. J. Appl. Intel. 6, 205–213 (1996).