

# Stock Price Prediction Using Support Vector Machine Approach

**Naliniprava Tripathy**

Indian Institute of Management

## **Abstract**

The present study predicts the direction of the movement of the closing price of S&P BSE TECK index from January 2008 to January 2018 by using Support Vector Machines model. Further, the study uses performance measurement 'Hit ratio' to determine the accuracy of the SVM model. The outcomes of the study indicate that the average prediction accuracy is 60.2% after financial crises period 2008. The study also finds that the direction of the market movement is positive when closing price is higher than previous day's closing price. The study suggests that SVM model has better prediction performance in short and medium term compared to long term. The study indicates that an investor can make profit by investing in the Indian stock market.

**Keywords:** Support Vector Machine, Stock Market Prediction, BSE TECK index

**JEL classifications: G1, C32, C53**

### **Introduction**

Prediction of the stock market is a challenging task in financial time series today. Efficient Market Hypothesis (EMH) advocates that stock price cannot be predicted by taking the past trading information of stock price. However, researchers have proposed models to capture the nonlinear behavior of the stock market and debated that stock price can be predicted. Recently, a new technique called Support Vector Machines (SVM) learning approach is gaining prominence for stock price forecasting. SVM is a mathematical model with solid theoretical foundation and substantially developed in pattern recognition, function estimation, and time series prediction. Accurate prediction of stock prices helps to take better investment decisions with minimum risk. However, no such prominent research work has been undertaken especially after financial crises 2008 to predict the movement of stock market price in India. The present research study seeks to address this gap. The present study uses SVM model to predict the direction of S&P BSE TECK. Secondly, the study also determines the accuracy of the model using Hit Ratio. The rest of the paper is planned as follows: Section two describes the literature review, section three briefly explain the Machine Learning approach. Section four presents the data and methodology, the experimental results are presented in section five. The concluding observations is provided in section six.

### **Literature Review**

Support Vector Machines (SVMs) have been extensively researched in machine learning community for the last decade. Yang et al. (2002) used SVM to find the varying in the margins of SVM regression and change in volatility of financial data. The study also analyzed the effect of asymmetrical margins to allow for the reduction of the downside risk. The study found that the former approach produced the lowest error when predicting the daily closing price of Hong Kong's Hang Seng Index. Kim (2003) predicted the direction of change in the daily Korean composite Stock Price Index 200 (KOSPI 200) by using SVM and ANN model. The experiment showed that the SVM outperformed the ANNs model in predicting the future direction of a stock market. The study reported that the best prediction performance obtained with SVM is 57.8 % significantly above the 50% threshold. Huang et al. (2004) used SVM for credit rating analysis and Back Propagation Neural Network (BNN) as a benchmark. The study obtained prediction accuracy around 80 % for the United States and Taiwan markets by using both BNN and SVM models. However, SVM model shows only a slight improvement than BNN model. Shin et al. (2005) demonstrated that SVM performs better than Back-Propagation Neural Networks when applied to corporate bankruptcy prediction. The accuracy and performance of SVM model are better than BPN as the

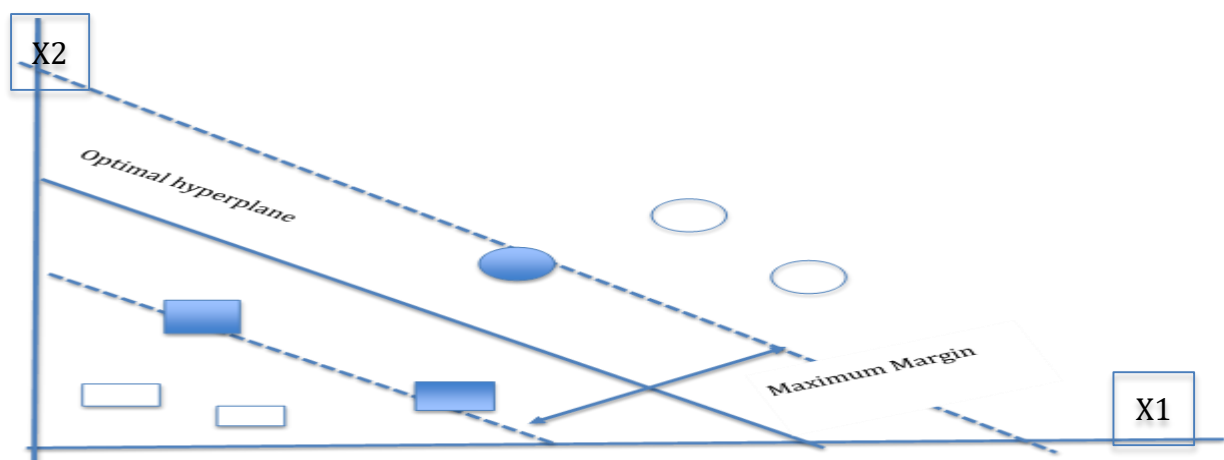
training set size is getting smaller. Shah (2007) conducted a study on stock prediction using various machine learning models and found that the best results achieved with SVM are 60%. The study confirmed with Kim's conclusion. Soni and Srivastava (2010) used Machine-learning algorithms like Classification and Regression Tree (CART), Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) for taking investment decisions in the stock market. The results and comparison of all the models reveal that classification and regression misclassification rate, is only 56.11% whereas LDA and QDA show 74.26% and 76.57% respectively. The study concludes that CART algorithm performs better in comparison to LDA and QDA algorithms in Indian stock market. Yakup et al. (2011) predicted the direction of daily Istanbul Stock Exchange (ISE) National 100 Index using Artificial Neural Networks (ANN) and Support Vector Machine (SVM), model. The study indicated that the performance of ANN model is significantly better than SVM model. Yanshan Wang (2014) predicted the direction of the Korean stock exchange and Hangseng index by using an integrated model of Principal component analysis and Support Vector Machine (PCA-SVM). The results of the study indicated that the proposed model provides distinctly high hit ratios for predicting the movement of the directions of Korean and Hong Kong stock market.

#### **The Machine Learning Approach and their Applications in finance:**

Machine learning methods allow a machine to decide with explicit programming (Avrim and Langley, 1997). Machine learning techniques broadly divided into two categories – supervised learning and unsupervised learning. A set of training data supplied to the machine in supervised learning to learn them whereas, in unsupervised learning, no training data provided. The unsupervised learning algorithm tries to find a similarity between the data without any given labels (Avrim and Langley, 1997) Vapnik and co-workers developed support Vector Machine in 1992. It is a supervised machine-learning algorithm based on statistical learning theory. SVM is a useful method for data classification and regression analysis. It is also an effective method for pattern recognition and regression. SVM mostly used in classification problems such as classification of linear and nonlinear data. SVM creates a boundary and the data points on either side of the boundary labeled differently (Keerthi et al., 2005). The boundary in the multidimensional case called a hyperplane. The most critical elements in this technique are the data points closest to the hyperplane. The optimal hyperplane maximizes the distance of the hyperplane from the extreme points on either side of the hyperplane. The optimal hyperplane referred to as the maximum margin hyperplane (MMH). The hyperplane selection only based on these extreme points. These extreme data points called the support vectors, and the maximum margin hyperplane known as the Support Vector Classifier (SVC). Figure-1 depicts the idea of the optimal

hyperplane in SVM that separate two classes. In figure-1, support vectors shown as a rectangle. Support vector regression is different from conventional regression techniques. In ordinary least squares regression, the optimal line is the one for which the sum of error is minimized. In SVM, the SVC is the hyperplane for which the sum of the distance between the hyperplane and the support vector is the maximum. SVM provides not only linear boundaries but also models nonlinear hyperplane. SVM does not reduce the empirical risk of making a few mistakes but pretends to build reliable models with future data. This principle called Structural Risk Minimization (SRM) in statistical learning theory. SVM uses the Structural Risk Minimization (SRM) but not the Empirical Risk Minimization (ERM) orientation principle. SVM seeks to minimize an upper bound of generalization error and not the training error. Hence, it expected to perform better than conventional techniques. SVM is usually resistant to the over-fitting problem.

**Figure-1 Optimal hyper plane in support vector machine**



The problem found in regression is to estimate a function based on a given data set.

Let us take a data set  $S = \{(x_i, d_i)\}_{i=1}^N \dots \dots \dots (1)$

Where  $x_i$  is the input vector,  $d_i$  is the desired result, and  $N$  corresponds to the size of the data set. The general form of Support Vector Regression estimating function is

$$f(x) = (w \cdot \phi(x)) + b \dots \dots \dots (2)$$

Where  $w$  and  $b$  are the coefficients that estimated from data.  $\phi(x)$  is the nonlinear function in feature space.

$$R(C) = \frac{1}{2} \|w\|^2 + C \frac{1}{n} \sum_{i=1}^N L_{\epsilon}(d_i y_i) \dots \dots \dots (3)$$

$$\text{Where } L_{\epsilon}(d, y) = \begin{cases} |d - y| - \epsilon, & |d - y| \geq \epsilon \\ 0, & \text{others} \end{cases} \quad \left\{ L_{\epsilon}(d, y) = \begin{cases} |d - y| - \epsilon, & |d - y| \geq \epsilon \\ 0, & \text{others} \end{cases} \right.$$

After slack variables introduced, the risk function articulated in the following constrained form:

$$\text{minimize } R(w, \vartheta) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\vartheta_i + \vartheta_i^*) \dots \dots \dots (4)$$

$$\begin{aligned} \text{Subject to } & d_i - w\phi(x_i) - b_i \leq \epsilon + \vartheta_i, \\ & (w \cdot \phi(x)) + b - d_i \leq \epsilon + \vartheta_i^* \end{aligned}$$

Where  $\vartheta_i, \vartheta_i^* \geq 0$

The dual form of nonlinear SVM stated as

$$Q(\alpha_i, \alpha_i^*) = \sum_{i=1}^N d_i (\alpha_i - \alpha_i^*) - \epsilon \sum_{i=1}^N (\alpha_i - \alpha_i^*) - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i \alpha_j^*) (K(X_i X_j) \dots \dots \dots (5)$$

Subject to

$$\begin{aligned} & \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\ & 0 \leq \alpha_i, \alpha_j^* \leq C, i = 1, 2, \dots, n \end{aligned}$$

$\alpha_i$  and  $\alpha_i^*$  are the Lagrange multipliers which act as forces to predict towards target value  $d$ .

The input space achieved using a kernel function  $K(X_i X_j) = \phi(X_i) \cdot \phi(X_j)$  in feature space. Any function that contents Mercer's theorem used as a kernel function. Some of the kernel function is as follows:

**Table-1 Kernel functions that Satisfy Mercer's Theorem**

kernel	Function	Comment
Polynomial	$[1 + (X \cdot X_i)]^p$	$P$ is power
Radial Basis function	$\exp(-\frac{1}{2\sigma^2} \ X - X_i\ ^2)$	$\sigma^2$ is the measurement

C controls the flatness of the approximating function and regulates the margin within which the error tolerated. The SVR function stated as follows:

$$f(x) = \sum_{i=1}^{nsv} (\alpha_i - \alpha_i^*) K(X, X_i) + b \dots \dots \dots (6)$$

Where *nsv* is the number of support vectors

In this paper, the SVM model with RBF Kernel function used for price forecasting. The function represented as follows:

$$K(x_i, x_j) = \exp\left(-\frac{1}{\sigma^2} \sum_{j=1}^N (x_{ij} - x_k)^2\right) \dots \dots \dots (7)$$

Where  $\sigma$  is the bandwidth of the kernel function, A bandwidth of radial basis function is represented by  $\sigma$ . If the value of  $\sigma$  fall in-between 0.1 and 0.5, the SVR model achieve the best performance. In this paper,  $\sigma$  value is determined as 0.5. The R software used to accomplish the experiment.

In this paper, the prediction of stock price problem treated as a classification problem. The dependent variable captures the direction of the market movement. The direction labeled positive when the return on a day is positive, and the path is marked negative on days with negative returns. The problem modeled in this paper is a binary classification problem with the objective of predicting the direction of market movement. The feature set for the classification model is derived using historical price data on the stock. The input parameters such as stock price volatility, stock momentum, index volatility, and index momentum are used for prediction to know the stock's price 'm' days in the future will be higher or lower than the current day's price.

The study predicts the direction of daily change of the S&P BSE Teck index. This trend prediction modeled as a two-class classification problem. Hence, the classes are leveled with 0 and 1. For each day, an upward and downward change is calculated. It means "0" indicates the daily closing price is lower than the closing of the previous day ( $U_t = 0, D_t = S_{t-1} - S_t$ ) i.e. fall in the stock price and "1" means "Up days" are characterized by the daily closing price  $S_t$  being higher than the closing of previous day  $S_{t-1}$  ( $U_t = S_t - S_{t-1}, D_t = 0$ ) i.e. a rise in the stock price. It confirms that the more significant value input attributes do not beat smaller value input then helping to reduce the prediction error.



SVM provides not only linear boundaries but also models nonlinear hyperplane. The non-kernel functions used to model nonlinear boundaries. The volatility of the stock returns, the momentum of stock returns and index returns used as features in the study. The study is conducted by changing the parameter “N” (past N days) to know how the trends of volatility and momentum can be used to predict future changes in the price. For transforming the input space to the higher dimension space, the radial basis kernel function used. The advantage of the radial kernel function is that it can output a nonlinear boundary function. It also matches the test data points to the training data points using a minimum Euclidean distance. It weights closer training data points heavily and then outputs the predicted label.

### **Input Features used in SVM:**

The study has used four features to predict the direction of the stock price movement. The four elements are index volatility, index momentum, stock price volatility, and stock market momentum. Volatility generates a market return experience of investors. The higher the probability of a rising /declining market leads to lower/higher volatility. If volatility index high/low, market will continue to go up/down, and the market's performance will tend to increase/decrease — accordingly, risk and returns increase /decreases. If the volatility predicted, investors can understand the volatility of the market to align their portfolios with the expected returns. Therefore, we have taken stock price volatility and index volatility as an input feature in our study. Stock price volatility shows the average over the past N days of the percentage change in the stock price per day. Index volatility displays the average over the past N days of the percentage change in the index price per day.

Momentum is the average rate of change of price movements of a stock. It is the measurement of the velocity of price changes. Momentum used to identify trend lines. The investor/trader often take a long or short position in the stock with the expectation that its momentum will continue in either an upward or a downward direction. In this way, momentum investing is purely a technical indicator. Therefore, we rely on the technical signs of momentum in our study that measures the speed and change of price movements. The stock momentum indicator is the average of the stock's momentum over the past N days, and index momentum is an average of the index's momentum over the past N days. The features computed using the formula displayed in Table 2.

**Table-2 Input Features used in SVM**

Technical indicator	Description	Formula
Stock price volatility	Shows the Percentage change between the most recent price and the price "n" periods in the past	$\frac{\sum_{i=t-n+1}^t \frac{P_i - P_{i-1}}{P_{i-1}}}{n}$
Stock Momentum	Measures the change in price. Each day labeled 1 if closing price recent day is higher than the day before, and -1 if the price is lower than the day before.	$\frac{\sum_{i=t-n+1}^t y}{n}$
Index volatility	Shows the Percentage change between the most recent index's price and the price "n" periods in the past	$\frac{\sum_{i=t-n+1}^t \frac{I_i - I_{i-1}}{I_{i-1}}}{n}$
Index Momentum	Measures the security's closing price to its price range over a given period. Each day is labeled 1 if closing price recent day higher than the day before, and -1 if the price is lower than the day before.	$\frac{\sum_{i=t-n+1}^t d}{n}$

where  $P_i$  is the closing stock price at time  $t$ .  $P_{i-1}$  is the previous day price.  $I_i$  is the index's closing price. The directional change of stock price is considered as  $y \in \{-1, 1\}$  and directional change of Index is  $d \in \{-1, 1\}$ . These feature is used to predict the direction of stock price change between  $t$  and  $t + M$  where  $M \in \{1; 5; 10; 20; 90; 270\}$

The four features are calculated using different periods over the past N days. The study has conducted to know how the trends of volatility and momentum of stock and index be used to forecast the future changes of that stock and index.  $N_1$  is the period for the stock price features, and  $N_2$  is the period for index-based features, where  $N_1, N_2 \in$



{1; 5; 10; 20; 90; 270}. it signifies one week, two-week, one month, four-month, and one year. In each iteration, combination N1 and N2 parameters calculated the feature set, train on the training data and predict the testing set. The direction of the index movement predicted using all the possible combination of N1 and N2 and for every 25 combinations. The four features calculated at every trading date from 2008 to 2018, and all these features on that date collected as one vector. The study run iterations of each combination of N1 and N2. Since volatility and momentum are estimated on the previous day data, the feature vectors on the d are calculated as follows:  $[Max(N_1 + N_2) + 1] - th \text{ date}$ . The price direction 'M' day are predicted in the future based on  $M \in \{1; 5; 10; 20; 90; 270\}$ . There are 1820 trading days. Therefore, the total number of days is 1593-d-M. The complete set of the feature vector is X, and the output vector is Y. The output vector also calculated on each of 1820-d-M days. The feature vector X train, Y test and X test, Y train supply to SVM model for training time. The testing feature vector X test supply to SVM for predicting the corresponding output vector for comparing the said output with Y train.

The training set consists of data from January 2008 to January 2018, and the SVM model developed based on training data. The model tested on the test data. The accuracy of the model computed based on the prediction of the test data. After training the SVM, the forecasted price and the actual price used for finding the accuracy ratio. The degree of accuracy calculated by estimating the deviations from observed values. At present, more emphasis given to turning point prediction capability using sign and direction test. The most popular method used as a performance metric is the Hit ratio. Hit ratio measures the directional accuracy and the trading return. Hit ratio is one of the metric performances measures the directional efficiency of the stock price. Directional accuracy indicates the degree to which prediction correctly forecasts the direction of change in the actual stock market prices. Hit ratio measures the number of correct predictions of the stock movement direction to the total number of predictions. The predicted performance Hit ratio estimated using the following equations:

$$Hit \text{ Ratio} = \frac{1}{n} \sum_{i=1}^n P_i \quad (i = 1, 2, \dots, n) \text{-----} (8)$$

Where  $P_i$  is the prediction result for the  $i$ th trading day defined by equation 8

$$P_i \begin{cases} 1 & \text{if } (y_{t+1} - y_t)(\hat{y}_{t+1} - \hat{y}_t) > 0, \\ 0 & \text{otherwise} \end{cases}$$

Where  $y_t$  is the actual value, and  $\hat{y}_t$  is predicted the value for the  $i^{\text{th}}$  trading day, the variable  $n$  is the number of test examples. The variable  $y_t$  denotes the actual value of

the closing stock index for the  $i^{\text{th}}$  trading day, and  $\hat{y}_i$  is the predicted value for the  $i^{\text{th}}$  trading day. The variable  $n$  denotes the number of test samples.

### **Results and Discussion:**

The data alienated into training, validation and test sets. The training set used to build the model, the validation set used for parameter optimization, and the test set used to evaluate the model. The study adopted a Radial Basis Function (RBF) on sample data. The RBF kernel function requires the setting of the parameter  $\sigma$  in addition to the regular parameters  $C$  and  $\epsilon$ . The performance of the model evaluated after renormalizing the output generated by the models. Hit ratio taken as a performance measure in this work.

The features are calculated using different periods. N1 is the period for the stock market-based features, and N2 is the period for index-based features. N1, N2 can take the following values such as  $N1, N2 \in (5, 10, 20, 90 \text{ and } 270)$ . It signifies one week, two weeks, one month, and one-quarter and one year. The direction of the index movement predicted using all the possible combination of N1 and N2. These combinations of N1, N2 parameters used to estimate the features set, trained them, then predicted the testing set and check the accuracy of results. There are 25 iterations run for each of these combinations.

The features are used to calculate the direction of movement for one day -ahead of the market price (M=prediction accuracy). M takes the following values such as  $\in (1, 5, 10, 20, 90 \text{ and } 270)$ . The study has developed several combinations of M. Table-3 reveals that the accuracy ratio increased when  $M=1, 5, 10$  and  $20$  and the accuracy ratio decreases when  $M=90$  and  $270$ . The table-2 exhibits that When  $M=1, N1=5,$  and  $N2=20$ , the mean accuracy is  $64.71\%$ . It indicates that the stock's recent momentum is better than other indicators since it replicates the latest information. When the combination of  $N1=270, N2=270$  and  $M=270$ , the prediction accuracy is  $60.43\%$ . The results suggest that the model can predict price direction with  $60.43\%$  accuracy. The accuracy decreases for forecast far in the future. The accuracy is high for lower values of M while it decreases as M increases. The study specifies that long run stock direction predicted, and the price reacts not only to new information but also to existing news. This study violating the efficient market hypothesis since the research indicates that the mean price not only responded to new information transmitted to the market instantaneously but also to the past data. However, it is evident from the experimental results that the prediction performance accuracy is  $60.19\%$  after the financial crises period 2008. Under such circumstances, a  $60.19\%$  prediction performance of technical indicators can be acceptable.

**Table 3: HIT Ratios of each combination of M, N1 and N2**

M	N1	N2	Accuracy
1	5	5	61.31%
	5	10	60.11%
	5	20	64.71%
	5	90	60.02%
	5	270	61.03%
	10	5	62.22%
	10	10	60.01%
	10	20	64.44%
	10	90	62.15%
	10	270	59.00%
	20	5	59.01%
	20	10	61.32%
	20	20	58.61%
	20	90	60.43%
	20	270	57.31%
	90	5	61.31%
	90	10	59.00%
	90	20	59.31%
	90	90	62.04%
	90	270	61.02%
5	270	5	62.15%
	270	10	60.14%
	270	20	60.43%
	270	90	61.00%
	270	270	70.02%
	5	5	62.44%
	5	10	58.04%
	5	20	60.14%
	5	90	62.04%
	5	270	60.01%
	10	5	59.00%
	10	10	62.03%
10	20	62.01%	
10	90	65.03%	
10	270	62.44%	

**International Academic Conference on Management & Economics**
**8 - 10 November, 2019**
**Oxford, United Kingdom**

	20	5	60.43%
	20	10	60.14%
	20	20	63.02%
	20	90	60.43%
	20	270	60.04%
	90	5	59.00%
	90	10	58.42%
	90	20	57.28%
	90	90	60.14%
	90	270	61.02%
	270	5	63.01%
	270	10	62.10%
	270	20	59.28%
	270	90	59.01%
	270	270	60.43%
	5	5	61.00%
	5	10	59.00%
	5	20	62.15%
	5	90	60.14%
	5	270	60.03%
	10	5	62.04%
	10	10	61.02%
	10	20	60.14%
	10	90	57.00%
	10	270	60.43%
	20	5	65.30%
	20	10	62.15%
	20	20	60.72%
	20	90	60.72%
	20	270	59.11%
	90	5	58.04%
	90	10	61.00%
	90	20	60.02%
	90	90	61.00%
	90	270	60.14%
	270	5	57.00%
	270	10	61.00%
	270	20	62.03%
	270	90	60.43%
	270	270	61.43%
20	5	5	61.29%

[info@conferenceme.org](mailto:info@conferenceme.org)

**International Academic Conference on Management & Economics**
**8 - 10 November, 2019**
**Oxford, United Kingdom**

	5	10	60.40%
	5	20	58.14%
	5	90	58.42%
	5	270	58.14%
	10	5	51.22%
	10	10	60.04%
	10	20	58.42%
	10	90	59.28%
	10	270	61.00%
	20	5	55.00%
	20	10	60.14%
	20	20	60.43%
	20	90	60.14%
	20	270	60.14%
	90	5	58.14%
	90	10	58.13%
	90	20	60.01%
	90	90	59.00%
	90	270	58.42%
	270	5	60.14%
	270	10	60.06%
	270	20	60.14%
	270	90	59.00%
	270	270	60.14%
90	5	5	65.01%
	5	10	63.01%
	5	20	63.01%
	5	90	59.28%
	5	270	58.06%
	10	5	64.40%
	10	10	60.43%
	10	20	58.14%
	10	90	58.42%
	10	270	60.14%
	20	5	60.43%
	20	10	58.05%
	20	20	56.40%
	20	90	55.00%
	20	270	59.11%
	90	5	58.42%
	90	10	61.00%

[info@conferenceme.org](mailto:info@conferenceme.org)

**International Academic Conference on Management & Economics**

8 - 10 November, 2019

Oxford, United Kingdom

	90	20	60.40%
	90	90	58.40%
	90	270	60.06%
	270	5	60.34%
	270	10	57.00%
	270	20	60.06%
	270	90	60.02%
	270	270	59.00%
270	5	5	60.43%
	5	10	63.01%
	5	20	59.00%
	5	90	64.07%
	5	270	62.44%
	10	5	62.15%
	10	10	62.06%
	10	20	53.01%
	10	90	59.28%
	10	270	59.00%
	20	5	60.14%
	20	10	60.43%
	20	20	60.43%
	20	90	56.42%
	20	270	60.07%
	90	5	61.00%
	90	10	58.14%
	90	20	60.11%
	90	90	58.42%
	90	270	60.17%
	270	5	60.43%
	270	10	60.14%
270	20	60.01%	
270	90	60.14%	
270	270	60.43%	

The results of the study suggest that forecasting long-term change of price direction may depend more on the overall market trends and macro conditions of the economy.

**Concluding observations:**

The present study predicts the direction of S&P BSE TECK price movement in the Indian stock market. The study finds that the average prediction performance of SVM

[info@conferenceme.org](mailto:info@conferenceme.org)

[www.conferenceme.org](http://www.conferenceme.org)

models shows 60.2%% after financial crises period 2008. Different combinations of the feature set are used to find the efficient combination. The prediction performance of the proposed model outperforms more or less similar results in the studies of literature also. This study will help to develop an efficient market trading strategy and enable to take buy, hold, and sell decisions before making investment decisions. The SVM model is beneficial for the investors and the regulators in a highly volatile market like Indian stock market. However, future research can be explored by including other macroeconomic variables such as foreign exchange rates, interest rates, and consumer price index that are also highly influencing the stock market.

**References:**

Avrim L, and Pat Langley, 1997. Selection of relevant features and examples in machine learning, *Artificial Intelligence*, 97(1–2):.245-271

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

Huang, W., Nakamori, Y. and Wang, S.Y. 2005. Forecasting Stock Market Movement Direction with Support Vector Machine, *Computers and Operations Research*, 32(10):2513–2522

Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., and Wu, S. 2004. Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision support systems*, 37(4):543-558

Hsu S. H., Hsieh J. J. P. A., Chih, T. C., and Hsu, K. C. 2009. A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression. *Expert Systems with Applications*, 36(4):7947–7951

Gavrishchaka Valeriy V. and Supriya B. Ganguli, 2003. Volatility forecasting from multiscale and high-dimensional market data, *Neurocomputing*, 55(1-2):285-305

Fan, A. and Palaniswami, M., 2001. Stock selection using support vector machines, *IJCNN'01: International Joint Conference on Neural Networks*, 3:1793-1798

Keerthi S. S. Duan, K.B., Shevade, S.K., Poo, A.N., 2005. A Fast Dual Algorithm for Kernel Logistic Regression, *Machine Learning*, 61(1–3):151–165.

Kim, K.2003. Financial time series forecasting using support vector machines, *Neurocomputing*, 55(1):307-319

Khalid Alkhatib, Hassan Najadat, Ismail Hmeidi., Mohammed K. and Ali Shatnawi , 2013. Stock Price Prediction Using K-Nearest Neighbor (kNN) Algorithm, *International Journal of Business, Humanities and Technology*,3(3):34-44

Ongsritrakul, P. and Soonthornphisaj,N., 2003. Apply decision tree and support vector regression to predict the gold price, *Proceedings of the International Joint Conference on Neural Networks*, 4:2488-2492

Sansom, D.C., Downs, T., Saha, T.K., 2002. Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants. *Journal of Electrical and Electronics Engineering*, Australia, 22(3):227-233.

Shin Kyung-Shik, Taik Soo Lee and Hyun-jung Kim, 2005. An application of support vector machines in bankruptcy prediction model, *Expert Systems with Applications*, 28(1):127-135

Shah H.V., 2007. Machine Learning Techniques for Stock Prediction, *Foundations of Machine Learning*, Spring

Soni S. and Shrivastava, S., 2010. Classification of Indian Stock Market Data Using Machine Learning Algorithms, *International Journal on Computer Science and Engineering*, 02(09):2942-2946.

Tay, Francis E. H., and Cao, Lijuan 2001. Application of support vector machines in financial time series forecasting” *Omega*, 29(4):309-317

Yakup Kara, Melek Acar Boyacioglu , Ömer Kaan Baykan, 2011. Predicting direction of stock price index movement using artificial neural networks and support vector machines, *Expert Systems with Applications: An International Journal*, 38(5):5311-5319

Yanshan Wang.2014. Stock price direction prediction by directly using prices data: an empirical study on the KOSPI and HIS, *International Journal of Business Intelligence and Data Mining*, 9(2):145-160.