Stock Price Prediction Using Support Vector Regression

Abirami R. and Vijaya M.S.

PSGR Krishnammal College for Women, Coimbatore - 4
rr.abirami@gmail.com, msvijaya@grgsact.com

Abstract. Forecasting stock price is an important task as well as difficult problem. Stock price prediction depends on various factors and their complex relationships. Prediction of stock price is an important issue in finance. Stock price prediction is the act of trying to determine the future value of a company stock. The successful prediction of a stock future price could yield significant profit. Hence an efficient automated prediction system is highly essential for stock forecasting. This paper demonstrates the applicability of support vector regression, a machine learning technique, for predicting the stock price by learning the historic data. The stock data for the period of four years is collected and trained with various parameter settings. The performance of the trained model is evaluated by 10-fold cross validation for its predictive accuracy. It has been observed that the support vector regression model with RBF kernel shows better performance when compared with other models.

Keywords: Linear regression, Machine learning, Stock price prediction, Support vector regression.

1 Introduction

A stock market is a public entity for the trading of company stock or shares and derivatives at an agreed price. These are securities listed on a stock exchange as well as those only traded privately. A stock market is a platform to buy and sell stocks. The stock market is driven by supply and demand. The number of shares or stock dictates the supply and the number of shares that investors want to buy dictates the demand. The stock market will climb and climb faster than almost any other traditional investment. The stock market is really just a big, automated superstore where everyone goes to buy and sell the stock. The main players in the stock market are the exchanges. Exchanges are where the sellers are matched with buyers to both facilitate trading and to help set the price of the shares. Among the stock market basics, the share is the smallest unit of ownership in a company.

The size of a share varies from company to company. It can even vary within one company. If the person owns a share in a company, the shareholders are a part of the owners in that company. The total sum of an owner's stock is called stock portfolio. If a company issues dividends, or profits, to the shareholders, owners will receive

P.V. Krishna, M.R. Babu, and E. Ariwa (Eds.): ObCom 2011, Part I, CCIS 269, pp. 588–597, 2012.

money based on the number of shares they owned. Every year, a company issues corporate results and basic stock information for their stockholders to review. Stock market allows corporations looking to expand to raise capital from investors in the primary market and facilitate trade between buyers and sellers of stock in the secondary market. Stock exchange is the efficient way to create a favorable climate for an active and growing for new issues. Hence it is extremely significant, to predict the stock value more accurately in order to facilitate the financial growth of a company [1].

Traditionally, statistical algorithms and time series analysis have been used to predict the stock market. Currently machine learning technology is being employed in computational finance since machine learning deals with techniques that allow computers automatically learn to make accurate predictions based on past observations. In [2] time series analysis system has been used for learning the model and to predict the stock market movement. H.watanabe, B. Chakra borty and G. Chakra borty, developed a stock price trend prediction model based on neural networks and time series analysis for forecasting the stock trends [3]. Vatsal H Shan investigated the application of yale and weka system for stock price prediction. In [5] genetic algorithm has been employed to select the set of most informative input features from all the technical indicators for generating the model.

In this paper, machine learning algorithm namely support vector regression is used for modeling the prediction task. In machine learning, training set is used to fit a model that can be used to predict a 'response' value from one or more 'predictors'. The model is used to forecast the stock value. Training the features of historical data creates the learned model. The process of closing stock price prediction is described in the following section and the various experiments carried out to discover the performance of the models are demonstrated in the rest of this paper. The linear regression and support vector regression algorithms are applied for training the dataset and predict the future stock price.

2 Proposed Stock Price Prediction System

The proposed stock prediction system is based on support vector regression, a supervised learning in order to predict the closing price based on the historical values. The historical data are collected from Aditya Birla money limited, Erode. The historical stock data include opening price of the stock, previous closing price, highest price of stock, lowest price of stock, last price of the stock in a day and average price; these are all the input parameters or predictors. The input parameters are applied to learn the model, which is then used to predict the closing price of the stock.

Opening price

The price at which, a stock first trades upon the opening of an exchange on a given trading day. This is normally determined by the price at which a stock was sold on the previous day. The opening price of shares serves as a benchmark to the company for the day. The calculation of opening price is differs depending on the stock exchange. Opening price is calculated by a special list that looks the book of orders or calculated

by supply and demand method. The opening price is one of the attributes for forecasting the future stock price.

Previous closing price

Previous closing price is the stock price of last transaction in the previous day. It is used for the purpose of market verification. The previous closing price is used to calculate the future stock price. The future price prediction is based on technical indicators and sometimes the previous closing price of a stock becomes the current day opening price.

Highest price

The highest price of stocks can make easy to trade profitable. There will be higher demand for the stock with the highest price. The highest historical price level reached by shares, commodity or index during the trading time is called as the highest price of the company. Highest price is also one of important attribute to predict the future stock price.

Lowest price

The lowest price of stocks can make it much more difficult to trade profitable. The lowest price shares are avoided when acquiring a share as it can result in liquidity problems because the share price can easily drop. The stock exchange has limits to open the market because; the increase or decrease in the stock trade depends on the current trading day. If the stock gaps, up or down more than $5/8^{th}$, the stock cannot enter for the trade.

Last Price

The last price is different from the closing price because the last price in this instance represents the last transaction that occurred in the trading day. Last price is a stock price at which the stock was sold from seller to buyer.

Average price

The average price per share is determined by dividing the total cost of the shares by the number of shares purchased. The average price per share can seem complicated to determine if an investor has purchased different quantities of a stock at different prices**.** The average price reduces the range into a single value, which can be compared to any point, to determine if the actual value is higher or lower than the expected value.

Closing price

The closing price generally refers to the last price at which a stock trades during a regular trading session. The closing price represents the most up to date valuation of a stock until trading on the next day. It is the end of trading session and final procedure in a sale in which documents is signed and recorded. In that time the ownership of the property is transferred. Some time market centers are offered after-hours trading.

Financial publications and market data vendors use the last trade in after-hours markets as the closing price for the day. This parameter is called 'response value' which is to be predicted once the model is learnt.

Price Earning (P/E Ratio)

Price earnings ratio, the most widely used investment measure, is calculated on the previous day's closing share price divided by the earning per share basis. PE ratio is used to appraise a company's profit performance. Where a company's prospects are considered by the stock market is to be good, and then the company's share price will be rise.

The data set with 970 stock records is prepared for training. The learned model is generated through implementing support vector regression. The learned SVR model is used for predicting the future stock price of the company. The architecture of the automated stock prediction system is shown in Figure 1.

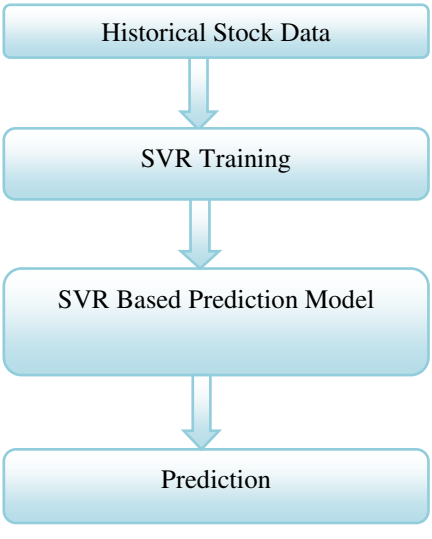

Fig. 1. Proposed SVR based stock price prediction model

3 Methodology

Modeling of stock properties by deciphering the functional relationship between the demand and supply properties is one of the most fascinating topics in forecasting stock market research. By and large, there are essentially three distinguished modeling tools for predicting the stock price namely mathematical models, statistical regression models and machine learning. Mathematical models are very appealing as they are based on the theories of basic sciences and give good understanding about the mechanics of the process. However, the prediction accuracy of mathematical models is not very encouraging due to the assumptions used while building the model.

Statistical regression models are very easy to develop and data coefficient analysis give an indication about the relative importance of various inputs. To counteract this problem, stock researchers have proposed various models relating to price and movement direction. This study employs data driven methodology to generate the predictive model for determining the stock price. Support vector machines (SVMs), based on supervised learning theory, and are gaining applications in the areas of machine learning and pattern recognition because of the high accuracy and good generalization capability. This study briefly introduces the SVM regression algorithms, and presents the SVR model for forecasting stock price.

3.1 Support Vector Regression

Support vector regression is the natural extension of large margin kernel methods used for classification to regression analysis. It retains all the properties that characterize maximal margin algorithms of support vector machines such as duality, sparseness, kernel and convexity. It has become a powerful technique for predicting data analysis with many applications in varied areas of study like biological contexts, drug discovery, civil engineering, stock market prediction, image compression etc., [7] [8]. The problem of regression is that of finding a function, which approximates mapping from an input domain to the real numbers on the basis of a training sample. This refers to the difference between the hypothesis output and its training value as the residual of the output, an indication of the accuracy to fit at the prediction point. To decide how to measure the importance of this accuracy, as small residuals may be inevitable even need to avoid large ones. The loss function determines this prediction measure.

Support vector regression performs linear regression in the feature space using ε insensitive loss function.

$$Ls(y, f(x,w)) = \begin{cases} 0 & if|y - f(x,w)| \leq \varepsilon \\ |y - (x,w)| - \varepsilon & otherwise \end{cases} \quad (1)$$

Empirical risk is:

$$R_{emp}(w) = \frac{1}{n}\sum_{i=1}^{n} Ls(Yi, f(Xi, W)) \quad (2)$$

SVM regression performs linear regression in the high dimension feature space using ε insensitive loss and, at the same time, tries to reduce model complexity by minimizing $\|w\|^2$. This can be done by using slack variables.

$$\xi_i, \xi_i^*, i = 1, \ldots, n. \quad (3)$$

To measure the deviation of training samples outside Ɛ- insensitive zone, thus SVM regression is formulated as minimization of the following functional:

Min

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{4}$$

Such that

$$\begin{cases} y_i - f(x_i w) \le \xi + \xi_i \\ f(x_i, w) - y_i \le \xi + \xi_i \\ \xi_i, \xi_i^* \ge 0, i = 1, \ldots, n \end{cases} \tag{5}$$

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

$$f(x) = \sum_{i=1}^{n_{sv}} (\propto_i - \propto_i^*) K(X_i, X) \tag{6}$$

Where n_{sv}, the number of support vectors and the kernel function is is

$$K(X_i, X) = \varphi(X_i) T \varphi(X) \tag{7}$$

It is well known that SVM generalization performance (estimation accuracy) depends on a good setting of meta-parameters c, Ɛ and the kernel parameters. Selecting a particular kernel type and kernel function parameters is usually based on application-domain knowledge and also should reflect distribution of input (x) values of the training data. Parameter c determines the trade-off between the model and the degree to which deviations larger than Ɛ are tolerated in optimization formulation [9].

3.2 Evaluation Parameters

Generally, the performance of the prediction techniques is evaluated using predictor accuracy. Predictor accuracy measures how far off the predicted value is from the actual known value. The loss function measures the error between actual value y_i and the predicted value y'. The average loss over the set of test cases determines the generalization error of the model and the generalization error can be computed using the following basic parameter.

MEAN ABSOLUTE ERROR

$$\frac{\sum_{i=1}^{d} |y_i - y_i'|}{d} \tag{8}$$

RELATIVE ABSOLUTE ERROR

$$\frac{\sum_{i=1}^{d}|y_i - y_i'|}{\sum_{i=1}^{d}|y_i - \bar{Y}|} \tag{9}$$

MEAN SQUARD ERROR

$$\frac{\sum_{i=1}^{d}(y_i - y_i')}{d} \tag{10}$$

RELATIVE SQUARED ERROR

$$\frac{\sum_{i=1}^{d}|y_i - y_i'|}{\sum_{i=1}^{d}|y_i - \bar{Y}|} \tag{11}$$

The mean squared-error exaggerates the presence of outliers in the training data and commonly used as a measure of performance evaluation. Another important measure is correlation coefficient, which measures the degree of the correlation between the actual and the predicted values. The higher correlation coefficient value is better in the model.

4 Exprements and Results

The data analysis and the prediction of closing price of stock have been carried out using LIBSVM tool for support vector regression. LIBSVM is simple, easy to use and efficient software for classification and regression [10]. The experiment is also carried out in WEKA environment for training the dataset using linear regression. The weka open source, portable, GUI-based workbench is a collection of state-of-the-art machine learning algorithms and data pre-processing tools [11]. To facilitate training, the historical data are collected from Aditya Birla Money limited, Erode. The training dataset consisting of 970 instances with 7 attributes is prepared and converted into the format as required by LIBSVM and WEKA.

Optimization problem of support vector regression with linear, polynomial, RBF kernels is implemented using LIBSVM with parameter c- regulation parameter, d- degree of polynomial and g- gamma. The support vector regression is trained with ‘d’ as 2 in case of polynomial SVR, ‘g’ as 2 in case of SVR with RBF kernel and ‘ε’ as 3 in case of epsilon parameter. The performance of the trained models are evaluated using 10 fold cross validation for its prediction accuracy based on mean squared error and correlation coefficient with the threshold 2.

The performances of the support vector regression and linear regression were evaluated based on the two estimators, the mean squared error and correlation coefficient. The mean squared error is generalization error between actual value and predicted value. The correlation coefficient measures the degree of the correlation between the actual and the estimated values. The result of the performance analysis is summarized in Table 1, Figure 2, Figure 3 and Figure 4.

Table 1. Performance of the models

Evaluation Criteria	SUPPORT VECTOR REGRESSION			Linear Regression
	Linear	Polynomial	RBF	
Mean Squared Error	1.41789	2.57416	0.55132	0.6604
Correlation Coefficient	0.86417	0.83416	0.98599	0.93499
Accuracy	89%	86%	94%	92%

Figure 2 shows the comparative analysis based on mean squared error of the models and it is observed that prediction model based on support vector regression with RBF kernel has lower mean squared error than statistical model.

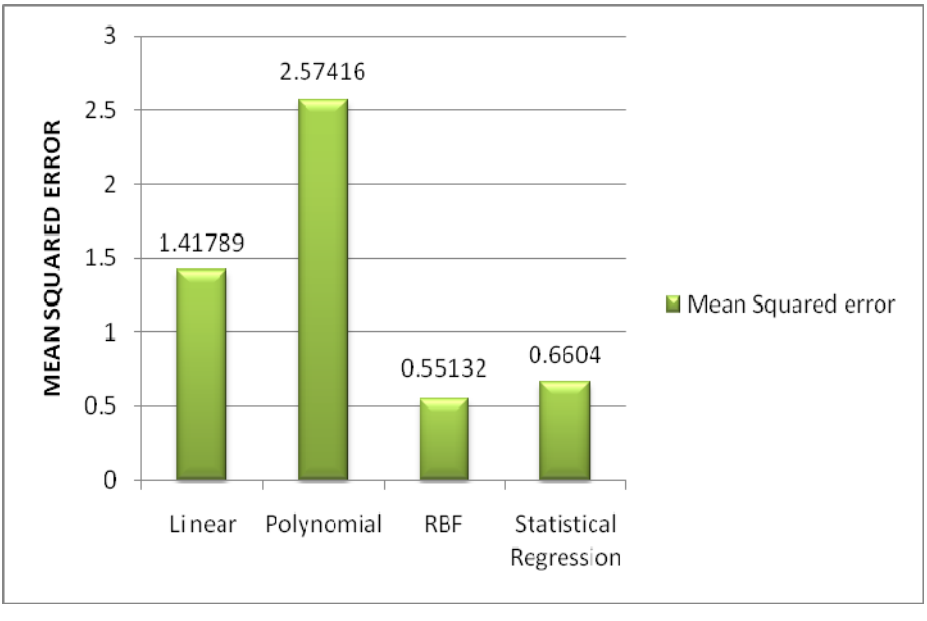

Fig. 2. Comparison based on Mean Squared Error

Figure 3 shows the performance analysis of the models based on correlation coefficient between the actual and the predicted values of the closing price.

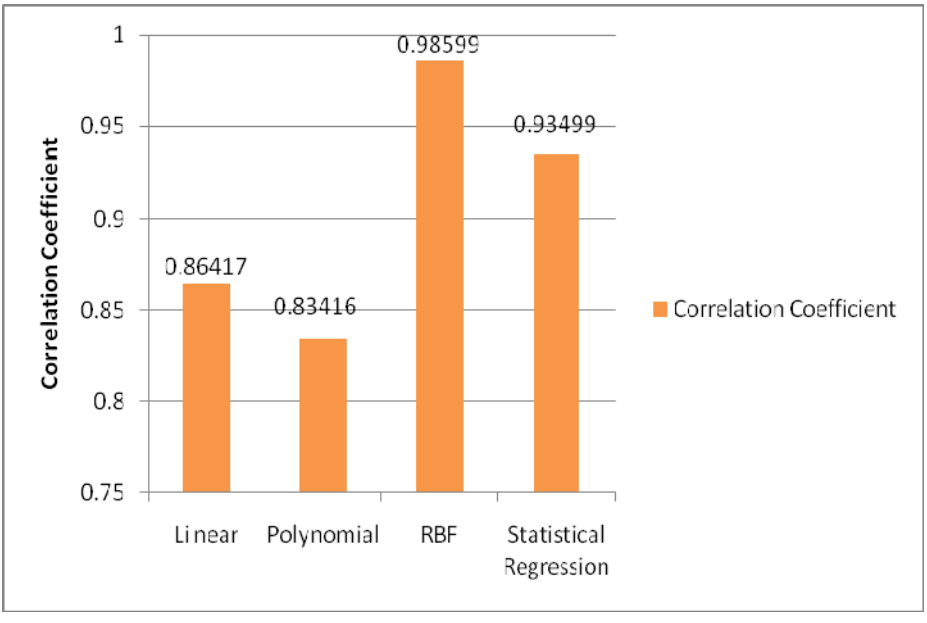

Fig. 3. Comparison based on Correlation coefficient

The comparative results show that prediction model based on support vector regression with RBF kernel has higher correlation coefficient than the statistical model.

From the above results, it is concluded that support vector regression with RBF kernel significantly reduce the mean squared error and improves the correlation

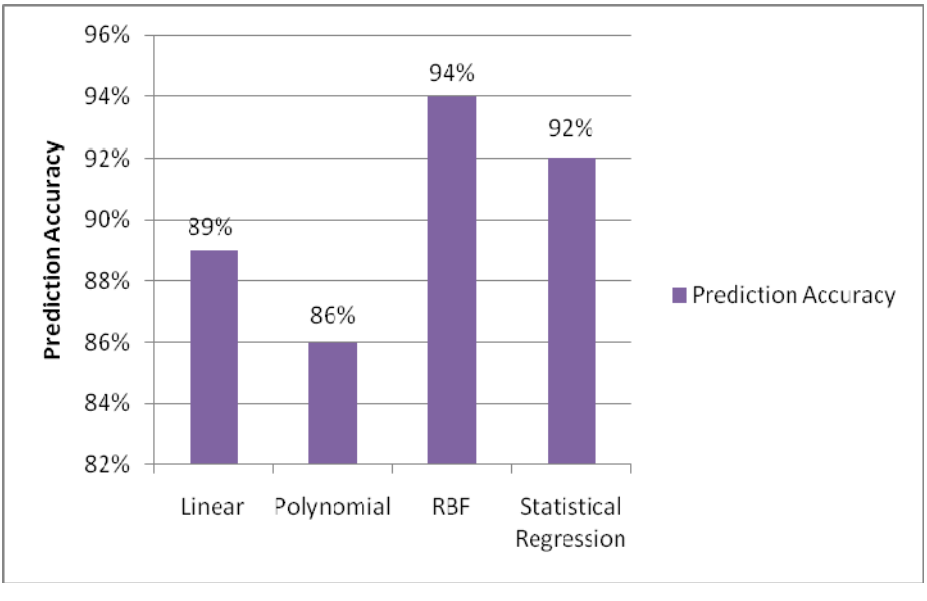

Fig. 4. Accuracy Prediction

coefficient. Figure 4 depicts the comparison of the models based on the prediction accuracy. The prediction accuracy is computed as the ratio of number of correctly predicted instances in the test dataset and the total number of test cases. The total number of cases that have approximately same stock price between actual and predicted values with *threshold T=2* is taken as the number of correctly predicted instances.

From the results it has been observed that stock prediction model based on SVR with RBF kernel produce higher accuracy than statistical regression model.

5 Conclusion

This paper demonstrates the applicability of a supervised learning algorithm namely support vector regression, for forecasting the stock price. The SVR model and linear regression model have been implemented using the dataset collected from a financial services company. The experiments have been carried on the historical data with 970 instances under various parameter settings. The performance of the models has been analysed based on the evaluation criteria such as predictive accuracy, mean squared error and correlation coefficient. It has been observed that the stock prediction model based on SVR with *RBF kernel* shows more efficiency in forecasting the stock value. Finally it is concluded that support vector regression can be well adopted for modelling the stock value prediction.

References

[1] Senthamarai Kannan, K., Sailapathi Sekar, P.: 'Financial Stock Market Forecasting using Data Mining Techniques. In: IMECS (2010)
[2] Huang, W., Nakamori, Y.: Forecasting stock market movement direction with support vector machine. Elsevier (2004)
[3] Watanabe, H., Chakra borty, B., Chakra borty, G.: Soft computing Approach for Stock Price Trend Forecasting from Multivariate Time Series. IEEE (2007)
[4] Shah, V.H.: Machine-learning techniques for stock prediction (spring 2007)
[5] Choudhry, R., Grag, K.: A Hybrid Machine Learning System for Stock Market Forecasting. World Academy of Science, Engineering and Technology (2008)
[6] Witten, L.H., Frank, E.: Data Mining – Practical machine Learning Tools and Techniques, 2nd edn. Elsevier (2005)
[7] Cristianiniand, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines and other Kernel-based learning methods. Cambridge University Press (2000)
[8] Gunn, S.R.: Support Vector Machines for Classifications and Regression Technical Report, University of Southampton (1997)
[9] Soman, K.P., Loganathan, R., Ajay, V.: Machine Learning with SVM and other Kernel Methods. PHI, India (2009)
[10] LIBSVM software, `http://www.csie.ntn.edu.tw/~cjlin/libsvm`
[11] Witten, I.H., Frank, E., Trigg, L., Hall, M., Holmes, G., Cunningham, S.J.: Weka: Practical Machine Learning Tools and Techniques with java implementations. weka software, `http://www.cs.waikato.ac.nz/ml/weka`
[12] Han, J., Kambar, M.: Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers (2001)