

Forecast on Close Stock Market Prediction using Support Vector Machine (SVM)

Elijah Joseph¹

Dept. of Computer Science
IBB University, Lapai, Nigeria

Amit Mishra²

Dept. of Computer Science,
Baze University, Abuja

Idris Rabi³

Dept. of Computer Science
IBB University, Lapai, Nigeria

Abstract:-The use of support vector machine (SVM) technique to improve the performance of quadratic, cubic, linear and fine Gaussian (SVM) for forecasting stock price prediction was developed in this paper. State of stock market price 170 days was divided into 119 data and 51 data and the first 119 data was used for training and second 51 data was used for testing to predict the close stock price. The four model's prediction result were compared with the actual value of stock market price to predict the future stock prices. The system was implemented using the support vector machine (SVM) and machine learning tool boxes of MATLAB 2015(a). The performance of the system was evaluated using Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) and Mean Squared Error (MSE) and compared with the models. The performance evaluation result shows that the linear (SVM) gives RMSE of 0.124 and MAPE of 97% and with less error in the quadratic (SVM) which gives an RMSE of 0.097 and MAPE of 98% and cubic (SVM) gives an RMSE of 0.10 and MAPE 98% with existing of fine Gaussian has more less error gives an RMSE of 0.009 and MAPE of 98.6%. This shows that, the developed model performs better than the linear and quadratic (SVM) because of the fine Gaussian which performs better than the Linear and cubic in support vector machine (SVM).

Keywords: ANN, Forecasting, Close Stock Market, SVM

INTRODUCTION

Stock, is a share in a company; it represents claims on a company's assets and earnings (Investopedia, 2010; Kiplinger, 2005). The price of the stock goes up or down, depending on the present and perceived future performance of the company. Stock market is a position where different stock sellers and stock buyers meet for the aim of transaction (or switch over) (Perwej & Perwej, 2012). The major movement in a typical stock market is the trading of shares among traders or stock brokers. In Nigeria, there are more than 261 different stock owners operating as entities in the Nigerian Stock Market (Saha, 2013).

Recently, United States of America was estimated to be contributing 34% of the total 55 trillion US dollars (US\$55 Trillion) of the world's total stock market price, with United Kingdom (UK) having 6% and Japan contributing 7% of the total stock price. Stock brokers sometimes get confused on the trend of the closing prices, hence, find it difficult to

predict tomorrow's closing price in order to prepare for it. This made forecasting of stock price an important mission for productivity and to help in decision making. Forecasting the future trend of stock price can be used to form an effective and guided trading plan, which may lead to better profit margins (Dase & Pawar, 2010).

There exist different soft computing methods aimed at forecasting future closing stock prices of a particular stock market; they include the use Artificial Neural Network (ANN), K Nearest Neighbours (KNN), Hidden Markov Model (HMM), Linear Programming (LP) and Support Vector Machine (SVM) among others.

The support vector machine (SVM) is a data classification technique that has been recently proven to perform better than other machine learning techniques especially in stock market prediction (Zhang, 2004). SVM try to build a model using a set of training examples given to it. Each training data instance is marked as belonging to one of two categories. The SVM will attempt to classify the data instances into those two categories. The trained SVM model can then be tested with new data instances to predict which category they belong to base on the training performance.

There are two classes of SVM, linear and nonlinear. Linear SVMs are fast to train and execute, but they tend to underperform on complex datasets with many training input and not too many features. Nonlinear SVMs can be more consistent in performance across different problems, and are the preferred option in many applications, however losing instructive power. (Huerta, Corbacho & Elkan, 2013). In this work, a nonlinear SVM is proposed to forecast the future trend of stock market closing prices. This work aimed at forecasting closing stock market price index using Support vector machine; to formulate different support vector machine models for stock data; to apply the formulated SVM models in forecasting closing stock market price index and evaluate the performance of the models in terms of Mean Absolute Percentage Error (MAPE) and correlation coefficient between the actual and predicted stock prices. The remainder of this paper is organized as follows: Section 2, a brief Literature review, while in section 3, the methodology adopted in this work is presented, performance analysis and results obtained is contained in section 4. The paper ends with conclusion in section 5.

LITERATURE REVIEW

Primary market deals with the new issues of securities and also securities are brought directly from the companies. But the secondary market, securities are bought and sold among investors. Secondary market deals with excellent securities. This market is made of organized exchanges and it has trading floor, where orders are transmitted for exchange. All the trading of stocks is maintained and guided by the exchanges. The rules and regulations are set down by the exchanges. For investors, indices give the direction of the entire market. They use indices to track the performance of the stock market. If possible, a change in the price of an index represents an exactly proportional change in the stocks included in the index. The ASI is one of the principal stock indices of the CSE and it measures the movement of share prices of all listed companies based on market capitalization.

Overview of Stock Market Concept.

The most basic purpose is to provide a measure to understand the direction or the movements of the market as a whole. An increase in the index indicates a rising market and decrease indicates a falling market. Market indices enable us to calculate market return. It represents the rate of return earned by investing in a group that impersonates the market range. Market return and risk are typically used as primary benchmarks to judge investment performance of a group.

Technical analysts try to predict future price movements by looking at the behavior of past price trends. Market indices also enable us to study factors that influence aggregate security price movements. Security analysts, selection managers and academics investigate these factors that impact the performance of the market. Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit (Wikipedia 2015).

Overview of Support Vector Machine

The support vector machine (SVM) is a data classification technique that has been recently shown to outperform other machine learning techniques when applied to stock market forecasting. In a possible particular state observation or outcome can be generated which is associated symbol of observation of probability distribution. It is only the outcome, not the state that is visible to an external observer and therefore states are provided to the outside; hence the name support vector machine.

Similarly, to HMMs, given a set of training input, SVMs will try to build a model. Each training data instance is marked as belonging to one of two categories. The SVM will attempt to separate the data instances into two categories with a $p-1$ -dimensional hyperplane, where p is the size of each data instance. This model can be used on a new data instance to predict which category it falls into (Rao & Hong, 2010).

Components of SVM and The Functions.

Support Vector Machines SVMs are Linear Learning Machines represented in a dual fashion Data appear only

within dot products (indecision purpose and in training algorithm) Linear classifiers cannot deal with Non-linear separable data Noisy data this formulation only deals with factorial data One solution creating a net of simple linear classifiers (neurons) a Neural Network (problems: local minima; many parameters; heuristics needed to train; etc.) Other solution map data into a richer feature space including nonlinear features, then use a linear classifier to map data into a feature space where they are linearly separated (Saahil, 2015).

How SVM Works.

There are two general classes of machine learning techniques. The first is supervised learning, in which the training data is the target where each example is collections of features that determine with the correct output corresponding to the feature set. This means that the algorithm is given features and outputs for a particular dataset (training data), and must apply what it "learns" from this dataset to predict the outputs (target) for another dataset (test data). Unsupervised learning, on the other hand, consists of examples where the feature set is untargeted. Supervised learning can be further broken down into classification and degeneration problems.

In classification problems there are a set number of outputs that a feature set can be target as, whereas the output can take on continuous values in degeneration problems. In this I am try to treat the problem of close stock price market forecasting as a classification problem. The feature set of a stock's recent price volatility and momentum, along with the index's recent volatility and momentum, are used to predict whether or not the close stock's price in this days in the future will be higher (+1) or lower (-1) than the current day's price (Saahil Madge 2015)

Strengths and Weakness

The support vector machine (SVM) which was first suggested by Vapnik, has recently been used in a range of applications, including financial stock market prediction. As Chen and Shih improved, the SVM technique, in general, is widely regarded as the state of art classifier. Previous researches indicated that SVM prediction approaches are superior to neural networks approaches.

Training is relatively easy no local optimal, unlike in neural networks. It scales relatively well to high dimensional data Tradeoff between classifier complexity and error can be controlled explicitly Non-traditional data like strings and trees can be used as input to SVM, instead of feature vectors.

Stock Market

The stock market is the market in which shares of publicly held companies are issued and traded either through exchanges or over-the counter market. Also known as the equity market the stock market is one of the most vital components of a free-market economy, as it provides companies with access to capital in exchange for giving investors a slice of ownership in the company. The stock market makes it possible to grow small initial sums of money into large ones, and to become wealthy without taking the risk of starting a business or making the sacrifices

that often accompany a high-paying career (Investopedia 2016).

Review of Related Literature (Work)

Forecasting of stock market price index has become an important field of research due to the impact of stock market price index on the economic development. In most cases the researchers had attempted to establish a linear relationship between the input macroeconomic variables and the stock market price index. But with the discovery of nonlinearity in the stock market price index returns there has been a great shift in the focus of the researchers towards the nonlinear prediction of the stock market price index (Abhyankar, 2007).

Although, there after many literatures have come up in nonlinear statistical modeling of the stock market price index, most of them required that the nonlinear model be specified before the estimation is done, (Wu, 2003). Not much has been done on SVM for Stock Prediction. Here is some literature review:

In literature, different sets of input variables are used to predict stock returns. In fact, different input variables are used to predict the same set of stock market price index data. Some researchers used input data from a single time series where others considered the inclusion of heterogeneous market information and macroeconomic variables. Some researchers even preprocessed these input data sets before feeding it to the ANN for forecasting (Abhyankar, 2007).

Dase (2010) predicated the stock rate because it is a challenging and daunting task to find out which is more effective and accurate method so that a buy or sell signal can be generated for given stocks. Predicting stock index with traditional time series analysis has proven to be difficult an artificial neural network may be suitable for the task. Neural network has the ability to extract useful information from large set of data. In the research, the author also presented a literature review on application of artificial neural network in stock market Index prediction.

Halbert (2012) reported some results of an on-going project using neural network modeling and learning techniques to search for and decode nonlinear regularities in asset price movements. The researcher focuses on case of IBM common stock daily returns. Having to deal with the salient features of economic data highlights the role to be played by statistical inference and requires modifications to standard learning techniques that may prove useful in other contexts. Abdulsalam (2010) used the moving average [MA] method to uncover the patterns, relationship and to extract values of variables from the database to predict the future values of other variables through the use of time series data. The advantage of the MA method is a device for reducing fluctuations and obtaining trends with a fair degree of accuracy. This technique proven numeric forecasting method using regression analysis with the input of financial information obtained from the daily activity equities published by Nigerian stock exchange.

METHODOLOGY

The Model Architecture

The architecture of the SVM model for stock closing price prediction is shown in Figure 1. The description of each block in the architecture and how it relates to other blocks is provided in the subsequent sections.

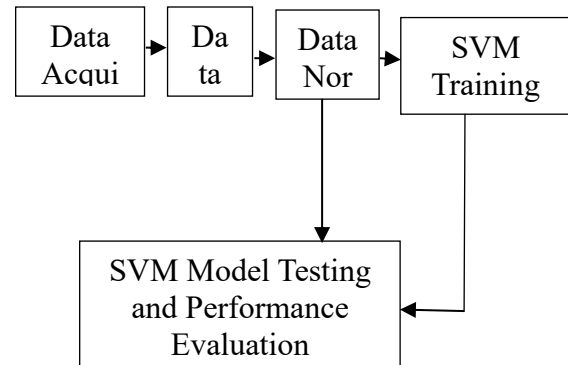


Figure 1: The model architecture for stock closing price prediction

Data Acquisition and Description

The data used for training and testing of the model was collected from yahoofinance.com. It is the stock price data for S&P 500. The data is a daily trading from 23/02/1950 to 11/03/2016 and it comprised of the opening price, high price, low price, closing price, volume and adjacent close of each trading day as shown in Table 1.

Data Selection

The data selected from the acquired data is from 10th July, 2015 to 11th March, 2016 making a total of 170 trading days. The selected data was divided into 70% for training and 30% for testing. Table 1 shows a sample of the data used. Figure 2 shows the pattern of the selected closing stock price and their corresponding trading dates.

Table 1: Sample of S&P 500 stock data used

Date	Open	High	Low	Close	Volume	Adj Close
3/11/2016	1994.71	2022.37	1994.71	2022.19	4.08E+09	2022.19
3/10/2016	1990.97	2005.08	1969.25	1989.57	4.38E+09	1989.57
3/9/2016	1981.44	1992.69	1979.84	1989.26	4.04E+09	1989.26
3/8/2016	1996.88	1996.88	1977.43	1979.26	4.64E+09	1979.26
3/7/2016	1996.11	2006.12	1989.38	2001.76	4.97E+09	2001.76
3/4/2016	1994.01	2009.13	1986.77	1999.99	6.05E+09	1999.99
3/3/2016	1985.6	1993.69	1977.37	1993.4	5.08E+09	1993.4
3/2/2016	1976.6	1986.51	1968.8	1986.45	4.67E+09	1986.45
3/1/2016	1937.09	1978.35	1937.09	1978.35	4.82E+09	1978.35
2/29/2016	1947.13	1958.27	1931.81	1932.23	4.59E+09	1932.23
2/26/2016	1954.95	1962.96	1945.78	1948.05	4.35E+09	1948.05
2/25/2016	1931.87	1951.83	1925.41	1951.7	4.12E+09	1951.7
2/24/2016	1917.56	1932.08	1891	1929.8	4.32E+09	1929.8
2/23/2016	1942.38	1942.38	1919.44	1921.27	3.89E+09	1921.27
2/22/2016	1924.44	1946.7	1924.44	1945.5	4.05E+09	1945.5
2/21/2016	1916.74	1918.78	1902.17	1917.78	4.14E+09	1917.78

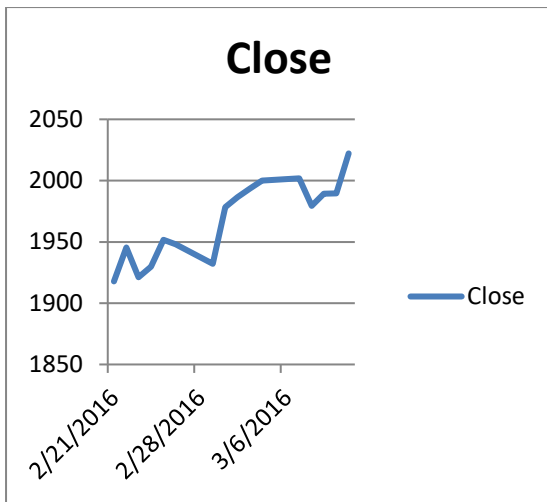


Figure 2: Plot of the closing stock price for S&P 500.

Data Normalization

Due to the inconsistency in the used data and to avoid the classifier from been biased towards larger values, the selected data were normalized to the range of 0 and 1. This was done to improve the performance of the models during testing.

Model Training

The selected period from the data was divided into training and testing data, with training data taking 70% (119 trading days) of the selected data. The SVM model was trained using the opening, high and low prices as inputs to predict the closing price as target or.

Model Testing

After training, each model was saved and tested externally with the 30% (51 trading days) of selected data to ascertain its performance. The performance of each model was recorded and analyzed.

Model Performance Evaluation

The forecasting model was evaluated using Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). The equations for evaluating the two metrics are shown in equations 1 and 2 respectively. The two equations determine the error resulting from the prediction by the proposed model.

$$MAPE = \frac{1}{N} \sum \left| \frac{d(k)-y(k)}{d(k)} \right| \times 100 \dots\dots(1)$$

$$RMSE =$$

$$\sqrt{\frac{1}{N} \sum E^2 \dots\dots\dots(2)}$$

Where:

- N = Number of test samples
- d(k) = Actual closing stock price
- y(k) = Predicted closing stock price
- E = Error

RESULTS

This Section presents the results obtained from the various stages of the development of the closing stock price

forecasting model using support vector machine (SVM). As shown in the hybrid model diagram described in Figure 2 the closing stock Price Index was selected to train and test the performance of the models and determine the effect of the optimization of SVM models. The models were trained using 170 trading days because it is sufficiently large enough for the model to understand the pattern and tested with 51 days because it was desired to forecast the closing stock for all ahead. The graph of the original stock prices, actual and observations. This is shown in order to visualize the original market trend before prediction.

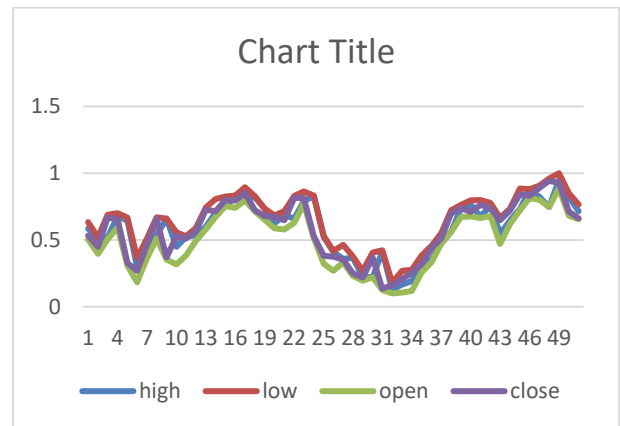


Figure 3: Close Stock Price, High, Low and Open

The graph in Figure 3 shows that the stock closing prices are higher in the day of 6 days and 35 days and lowest in 6 to 33 days (black line) while their low corresponding (dark red line) goes high moderate and the open fluctuates between up, down and same as shown in the red line.

Training Results

The training of 119 days' were used for the four models (Linear SVM, Quadratic SVM, Cubic SVM and Fine Gaussian SVM) described in previous sections produced the results of root mean square error (RMSE) and mean absolute percentage error (MAPE) the actual values for each model for the training and testing data results. The results were used to calculate the RMSE and MAPE the high, low and open stock price are used to forecast the closing price market stock.

Parameter optimization results

Table 2 shows the state closing price, open price, low price and high stock market price for the 119 for training data and 51 for testing data that were trained as explained in previous section. The table show the initial value for the difference models been trained.

Models Value Results

The model value of the model trained and testing was calculated as explained in previous section, different numbers of stock market price were used to determine the model that predict best the SVM model (Linear, Quadratic, Cubic and Fine Gaussian) actual closing market price and the predicted value of each model. The results are shown in table 3.

Table 3: Training Data Before Normalize

DATE	OPEN	HIGH	LOW	CLOSE
31/12/2015	7938.45	7972.55	7909.8	7963.2
30/12/2015	7897.8	7821.4	7891.15	7946.35
29/12/2015	7938.6	7825.4	7889.85	7896.25
28/12/2015	7929.2	7771.25	7902.75	7928.95
27/12/2015	7863.2	7702.85	7863	7925.15
26/12/2015	7888.75	7691.95	7835.5	7861.05
25/12/2015	7830.45	7703.05	7826.1	7865.95
24/12/2015	7829.4	7663.95	7776.85	7786.1

Table 4: Testing Result

Actual closing	Cubic	Quadratic	Linear	Fine Gaussian
0.687068966	0.681883	0.681883	0.576005	0.713678547
0.742834052	0.743578	0.743578	0.681883	0.753624748
0.981654095	0.898884	0.898884	0.888756	0.898883658
1	0.75879	0.812011	0.681883	0.890087424
0.939789871	0.838077	0.838077	0.681883	0.871203766

Table 5: show the actual close market price and predicted value of the four (SVM) model.
 The full table is in appendix.

From this table 4 it observed that, the model's gives difference predicted values from the actual closing market price it observed that each model predict difference with one another its only quadratic and cubic (SVM) predict the same value that both have the same understanding for the S&P data I used for these researches but linear and cubic predict difference value not the same with any of them.

Testing Result

Fifty-one trading days' data were used as the test data whose closing stock market prices were predicted and compared with the actual close stock price. Table 4 shows the testing result for the four models Linear (SVM), Quadratic (SVM), Cubic (SVM) and Fine Gaussian (SVM) whose results were being compared. The result contains four difference models their predicted close stock prices. Other items on the table include the actual closing stock price and predicted value for each model.

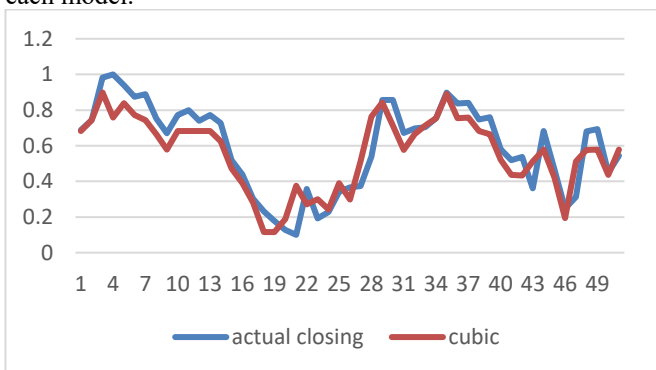


Figure 5: Actual close stock market price vs. Quadratic and Cubic predicted. A

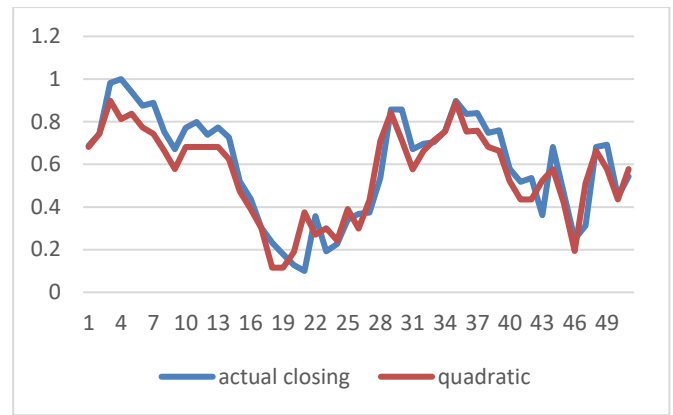


Figure 6: Comparison of actual stock price versus predicted on quadratic and cubic.

From this graph the blue line indicates the actual close stock price and red line indicate predicted close stock price from the observed the space between the two line is not much this shows that the error in quadratic and cubic prediction is minimum because is prediction was at 100%.

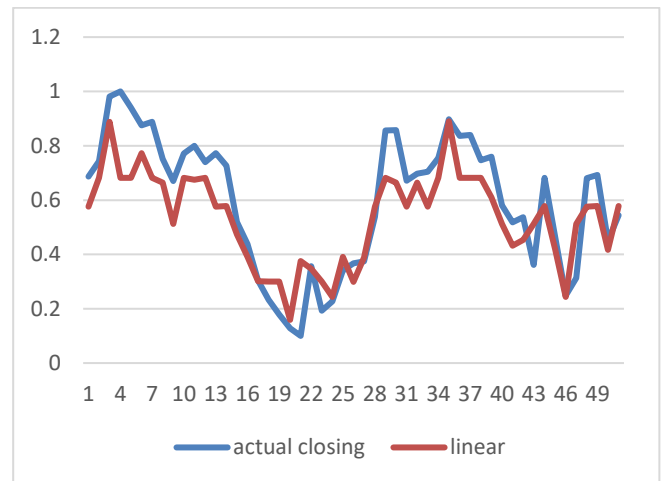


Figure 7: Actual Close Stock market price vs. Linear predicted.

From this graph the blue line indicates the actual close stock price and red line indicate predicted close stock price from the observed the space between the two lines is much this shows that there is a lot of error on linear prediction because it prediction out of 100% the linear (SVM) predict its own at 26.9%.

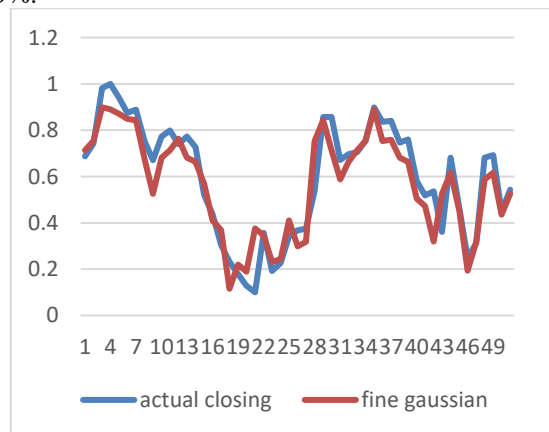


Figure 8: comparison of actual stock price versus predicted on Fine Gaussian.

From this graph the blue line indicates the actual close stock price and red line indicate predicted close stock price from the observed the space between the two line is not much this shows that the error in fine Gaussian prediction is minimum because is prediction was at 100%.

Performance Evaluation Results

The performance of the models was measured using Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE) calculated as explained in chapter three under performance evaluation. Table 4.3 shows the results of MAPE and RMSE for the four models Linear, quadratic, cubic and fine Gaussian (SVM).

Table 5: Performance Evaluation Results (Linear SVM)

Actual close price	predicted close price (linear SVM)	Error	Error sqr
0.687068966	0.57600538	0.111064	0.012335
0.742834052	0.681882986	0.060951	0.003715
0.981654095	0.888755884	0.092898	0.00863
1	0.681882986	0.318117	0.101198
0.939789871	0.681882986	0.257907	0.066516
0.875134698	0.772427707	0.102707	0.010549

Table 6: Comparative Value for Linear SVM

TOTAL OF ERROR	0.784473
MSE	0.0153
RMSE	0.124
MAPE	97%

Table 7: Performance evaluation results (Quadratic SVM)

Actual close price	predicted close price quadratic	Error	Error Square
0.687068966	0.681882986	0.005186	2.69E-05
0.742834052	0.743577673	-0.00074	5.53E-07
0.981654095	0.898883658	0.08277	0.006851
1	0.81201076	0.187989	0.03534
0.939789871	0.838076664	0.101713	0.010346

Table 8: Comparative Value for Quadratic SVM

TOTAL OF ERROR	0.481558
MSE	0.0097
RMSE	0.097
MAPE	98%

Table 9: Performance evaluation results (Cubic SVM)

Actual close price	predicted close price cubic	Error	Error Square
0.687068966	0.681882986	0.005186	2.69E-05
0.742834052	0.743577673	-0.00074	5.53E-07
0.981654095	0.898883658	0.08277	0.006851
1	0.758789509	0.24121	0.058183
0.939789871	0.838076664	0.101713	0.010346

Table 10: Comparative Value for Cubic SVM

TOTAL OF ERROR	0.548954
MSE	0.01
RMSE	0.10
MAPE	98%

Table 11: Performance evaluation results (Fine Gaussian SVM)

Actual close price	predicted close price fine Gaussian	Error	Error sqr
0.687068966	0.713678547	-0.02661	0.000708
0.742834052	0.753624748	-0.01079	0.000116
0.981654095	0.898883658	0.08277	0.006851
1	0.890087424	0.109913	0.012081
0.939789871	0.871203766	0.068586	0.004704

Table 12: Comparative Value for Fine Gaussian SVM

TOTAL OF ERROR	0.405983
MSE	0.009
RMSE	0.089
MAPE	98.6%

The performance evaluation result shows that the developed model predicted the stock prices with difference kind of error the linear (SVM) gives more error than others gives an RMSE of 0.124 and MAPE of 97% and with less error in the quadratic (SVM) gives an RMSE of 0.097 and MAPE of 98% and cubic (SVM) gives an RMSE of 0.10 and MAPE 98% with existing of fine Gaussian has more less error gives an RMSE of 0.009 and MAPE of 98.6%. This shows that, the developed model performs better than the linear and quadratic (SVM) because of the fine Gaussian which performs better than the Linear and cubic in support vector machine (SVM). The performance evaluation result and errors of each model is show in the figure 9.

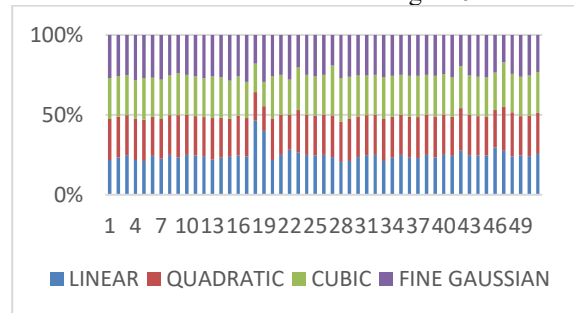


Figure 9: performance and errors of the models

From these graph show that quadratic SVM perform better than the Linear SVM and the Fine Gaussian performance better than the three models because it has less error than others.

CONCLUSION:

The use of support vector machine (SVM) technique to improve the performance of quadratic, cubic, linear and fine Gaussian (SVM) for forecasting stock price prediction was applied. State of stock market price 170 days was divided into 119 data and 51 data and the first 119 data was used for training and second 51 data was used for testing to predict

the close stock price. The four model's prediction result were compared with the actual value of stock market price to predict the future stock prices.

The system was implemented using the support vector machine (SVM) and machine learning tool boxes of MATLAB 2015(a). The performance of the system was evaluated using Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) and Mean Squared Error (MSE) and compared with the models. The result showed that the developed Fine Gaussian model have less prediction errors than the other three models.

REFERENCES:

- [1] Dase, R. K., & Pawar, D. D. (2010), "Application of Artificial Neural Network for stock market predictions: A review of literature", International Journal of Machine Intelligence, Vol. 2, No. 2, Pp14 – 17.
- [2] Huerta R., Corbacho F. & Elkan C., (2013), "Nonlinear support vector machines can systematically identify stocks with high and low future returns", Algorithmic Finance, Vol. 2, Pp45–58
- [3] Investopedia, (2010), "Stock Basics", A tutorial on stocks downloaded on 12/03/2016 from http://i.investopedia.com/inv/pdf/tutorials/stock_basics.pdf
- [4] Perwej, Y., & Perwej, A. (2012), "Prediction of the Bombay Stock Exchange (BSE) Market Returns Using Artificial Neural Network and Genetic Algorithm", Journal of Intelligent Learning Systems and Applications, Vol. 4, Pp108 – 119
- [5] Saha, S. (2013), "Comparison of Performance Analysis using Different Neural Network and Fuzzy Logic Models for Prediction of Stock Price", Unpublished M.Tech Thesis, Department of Computer and Engineering, National Institute of Technology Rourkela, Odisha, India.
- [6] Zhang Y. (2004), "Prediction of Financial Time Series with Hidden Markov Models", Simon Fraser University, May 2004.
- [7] Saahil Madge. (2015), Predicting Stock Price Direction using Support Vector Machines, Independent Work Report Spring 2015.
- [8] Shom Prasad das. (2012), Support Vector Machines for Prediction of Futures Prices in Indian Stock Market. Department of computer science \$ Engineering National Institute of science \$ Technology Palur Hills, Odisha 761008, India.
- [9] Sudarsan Padhy. (2012), support vector machines for prediction of futures prices in Indian Stock Market Department of computer science \$ Engineering International Institute of Technology Bhubaneswar Odisha- 751003, India
- [10] Martin Law. (2015), Lecture for CSE 802 Department of Computer Science and Engineering