International Journal of Current Trends in Engineering & Technology
Volume: 02, Issue: 01 (JAN-FAB, 2016)

Stock Market Prediction Using Support Vector Machine

Mr. Sachin Sampat Patil, Prof. Kailash Patidar, Assistant Prof. Megha Jain
SSSIST, Sehore, Madhya Pradesh, India

Abstract—a lot of studies provide strong evidence that traditional predictive regression models face significant challenges in out-of sample predictability tests due to model uncertainty and parameter instability. Recent studies introduce particular strategies that overcome these problems. Support Vector Machine (SVM) is a relatively new learning algorithm that has the desirable characteristics of the control of the decision function, the use of the kernel method, and the sparsely of the solution. In this paper, we present a theoretical and empirical framework to apply the Support Vector Machines strategy to predict the stock market. Firstly, four company-specific and six macroeconomic factors that may influence the stock trend are selected for further stock multivariate analysis. Secondly, Support Vector Machine is used in analyzing the relationship of these factors and predicting the stock performance. Our results suggest that SVM is a powerful predictive tool for stock predictions in the financial market.

Keywords: - ***Stock Classification; Data Mining; SVM; Forecasting***

1. INTRODUCTION

The macroeconomic environment and the financial market are complex, evolutionary, and non-linear dynamical systems. Before we study the historic volatile days of the ten years, let us first know what are:

a) Stock Markets
b) Stock exchanges

a) Stock Markets:

Stock Market is a market where the trading of company stock, both listed securities and unlisted takes place. It is different from stock exchange because it includes all the national stock exchanges of the country. For example, we use the term, "the stock market was up today" or "the stock market bubble."

b) Stock Exchanges

Stock Exchanges are an organized marketplace, either corporation or mutual organization, where members of the organization gather to trade company stocks or other securities. The members may act either as agents for their customers, or as principals for their own accounts. Stock exchanges also facilitates for the issue and redemption of securities and other financial instruments including the payment of income and dividends. The record keeping is central but trade is linked to such physical place because modern markets are computerized. The trade on an exchange is only by members and stock broker do have a seat on the exchange [15].

2. IMPORTANCE OF STOCK MARKET WITH FUNCTION AND PURPOSE:

The stock market is one of the most important for companies to raise money, along with debt markets which are generally more imposing but do not trade publicly. This allows businesses to be publicly traded, and raise additional financial capital for expansion by selling shares of ownership of the company in a public market. The liquidity that an exchange affords the investors enables their holders to quickly and easily sell securities. This is an attractive feature of investing in stocks, compared to other less liquid investments such as property and other immoveable assets. Some companies actively increase liquidity by trading in their own shares. History has shown that the price of stocks and other assets is an important part of the dynamics of economic activity, and can influence or be an indicator of social mood. An economy where the stock market is on the rise is considered to be an up-and-coming economy. In fact, the stock market is often considered the primary indicator of a country's economic strength and development.

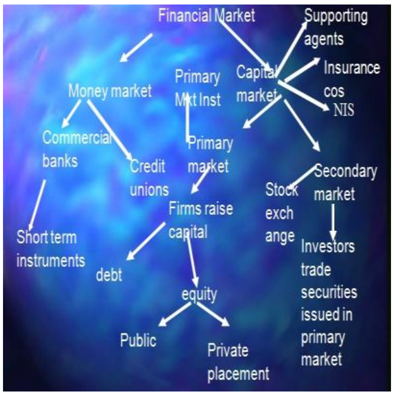


Figure 1 Hierarchical Structure; Financial Market

Rising share prices, for instance, tend to be associated with increased business investment and vice versa. Share prices also affect the wealth of households and their consumption. Therefore, central banks tend to keep an eye on the control and behavior of the stock market and, in general, on the smooth operation of financial system functions. Financial stability is the raison d'être of central banks. Exchanges also act as the clearinghouse for each transaction, meaning that they collect and deliver the shares, and guarantee payment to the seller of a security. This eliminates the risk to an individual buyer or seller that the counterparty could default on the transaction. The smooth functioning of all these activities facilitates economic growth in that lower costs and enterprise risks promote the production of goods and services as well as possibly employment. In this way the financial system is assumed to contribute to increased prosperity, although some controversy exists as to whether the optimal financial system is bank-based or market-based Recent events such as the Global Financial Crisis have prompted a heightened degree of scrutiny of the impact of the structure of stock markets (called market microstructure), in particular to the stability of the financial system and the transmission of systemic risk. The field of financial forecasting is characterized by data intensity, noise, non-stationary, unstructured nature, and hidden relationships. Predicting financial indicators is therefore a difficult task. However, forecasting is important in the sense that it provides concrete data for investment decisions. How can we predict whether the price of a particular stock will go up or down in the upcoming year? In the modern techniques, one way is to develop a predictor based on the information in the historical data. First of all, we should select some major factors that may influence the performance of the stocks; we can further discover an interesting model from our dataset to predict the future performance of any stocks. That is to say, we need to learn a model that can map those factors into the class attribute which indicates the whole performance of stocks. Support vector machine (SVM) is a machine learning technique that can be used for this purpose of classification. Established on the unique theory of the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization error, SVM is shown to be very resistant to the over-fitting problem, eventually achieving a high generalization performance. Another key property of SVM is that training SVM is equivalent to solving a linearly constrained quadratic programming problem so that the solution of SVM is always unique and globally optimal, unlike neural networks training, which requires nonlinear optimization with the danger of getting stuck at local minima. Although SVM approach has been widely applied in financial forecasting, little analysis is extended into the stock market of China, which is the second biggest economic entity and also one of the most important emerging markets in the world. Additionally, while indices, the main indicators of countries' economic condition; have been the efficient instrument for both hedgers and speculators in traditional and derivative market, less work is performed in this area than the analysis on individual stocks. Besides, model inputs in prior research typically involve price and volume data, and may also include a selection of well-known technical indicators but few papers apply fundamental indicators in the model. Compared to the technical analysis, which is the study of collective market sentiment mainly reflected in the price and volume, fundamental\ analysis focuses more on the intuitive physical interpretation and attempts to find the intrinsic value of the assets. Fundamental variables selected and included in the model generally have intuitive justification and have certain connection with the target; while it may be difficult to explain a technical analytical model.

2.2 Support Vector Machine (SVM)

The support vector machine (SVM) is a training algorithm for learning classification and regression rules from data, for example the SVM can be used to learn polynomial, radial basis function (RBF) and multi-layer perception (MLP) classifiers. SVMs were first suggested by Vapnik in the 1960s for classification and have recently become an area of intense research owing to developments in the techniques and theory coupled with extensions to regression and density estimation. SVMs arose from statistical learning theory; the aim being to solve only the problem of interest without solving a more difficult problem as an intermediate step. SVMs are based on the structural risk minimization principle, closely related to regularization theory. This principle incorporates capacity control to prevent over-fitting and thus is a partial solution to the bias-variance trade-off dilemma. Two key elements in the implementation of SVM are the techniques of mathematical programming and kernel functions. The parameters are found by solving a quadratic programming problem with linear equality and inequality constraints; rather than by solving a non-convex, unconstrained optimization problem. SVM algorithm developed by Vapnik is based on statistical learning theory. SVM can be used for both classification and regression task. In classification case we try to find an optimal hyper plane that separates two classes. In order to find an optimal hyper plane, we need to minimize the norm of the vector *w*, which defines the separating hyper plane. This is equivalent to maximizing the margin between two classes. Mathematically, we will obtain a quadratic programming problem where the number of variables is equal to the number of observations. Consider the problem of separating the set of training vector belonging to two separate classes,

$G=\{(x_i, y_i)\}$ Where i=1, 2, 3, 4,……N With a hyper plane $W^T \varphi(x) + b=0$ ($xi \in R^n$) Is the i-th input vector, y_i {1,-1} is known binary target, the original SVM classifier satisfies the following conditions:

$W^T \varphi(x) + b >= 1$, if $y_i=1$ ----------------- (1)

$W^T \varphi(x) + b <= -1$, if $y_i= -1$ -------------- (2)

Or equivalently

$y_i [W^T \varphi(x) + b] 1>=1$ ----------------- (3)

Where i=1, 2, 3 ……N

Where $\varphi: R^n \rightarrow R^m$ is the feature map mapping the input space to a usually high dimensional feature space where the data points become linearly separable. We need to find the hyper plane that optimally separates the data by solving the optimization problem: min ϕ (W) =0.5 $\| W \|^2$ under constrains of Equation (3). The solution to the above optimization problem is given by the saddle point of the Lagrange function:

$$Lp_1 = \frac{1}{2}\|W\|^2 - \sum_{i=1}^{N} \alpha i\, [yi\, (W^T \phi(x_i) + b) - 1] \text{ -- (4)}$$

Under constraints of Equation (3), where α_i are the nonnegative Lagrange multipliers. To generalize the problem to the non-separable case, slack variable ξ_i is introduced such that.

$$y_i [W^T \varphi(x) + b]2 >= 1 - \xi_i, \quad \xi_i >= 0 \text{ ------------------------ (5)}$$

Where, I = 1,2,…..N

Thus, for an error to occur the corresponding ξ_i must exceed unity, So, $\sum_{i=1}^{N} \xi_i$ Is an upper bound on the number o training error. Therefore, we rewrite the objective function to

$$\min \varphi(W, \xi) = \frac{1}{2}\|W\|^2 + C \sum_{i=1}^{N} \xi i \qquad \text{---------(6)}$$

under the constraints of Equation (5), where *C* is a positive constant parameter used to control the tradeoff between the training error and the margin. Similarly, solve the optimal problem by minimizing its Lagrange function.

$$Lp_2 = \frac{1}{2}\|W\|^2 + C \sum_{i=1}^{N} \xi i - \sum_{i=1}^{N} \alpha i\, [yi\, (W^T \phi(x_i) + b) - 1 + \xi_i] - \sum_{i=1}^{N} \mu_i \xi_i \text{ -------------------------------------(7)}$$

under the constraints of Equation (5), where α_i, μ_i are the nonnegative Lagrange multipliers. The Karush-Kuhn-Tucker (KKT) conditions for the primal problem are :

$$\frac{\partial Lp2}{\partial W} = W - \sum_{i=1}^{N} \alpha_i\, y_i\, \phi(x_i) = 0 \qquad \text{------------------- (8)}$$

$$\frac{\partial Lp2}{\partial b} = -\sum_{i=1}^{N} \alpha_i\, y_i = 0 \qquad \text{------- --------------(9)}$$

$$\frac{\partial Lp2}{\partial b} = C - \alpha_i - \xi_i = 0 \qquad \text{------------------------------ (10)}$$

$$y_i [W^T \phi(x_i) + b]2 >= 1 - \xi_i \qquad \text{---------------- (11)}$$

$$\xi_i >= 0 \qquad \text{---------------- (12)}$$

$$\alpha_i >= 0 \qquad \text{------------------(13)}$$

$$\mu_i >= 0 \qquad \text{----------------- (14)}$$

$$\alpha_i\, [y_i [W^T \phi(x_i) + b - 1 + \xi_i] = 0 \qquad \text{------------------(15)}$$

$$\mu_i \xi_i = 0 \qquad \text{---------------- (16)}$$

$$W = \sum_{i=1}^{N} \alpha_i\, y_i\, \phi(x_i) \qquad \text{------------- - (17)}$$

We can use the KKT complementarily conditions, Equations (15) and (16), to determine *b*. Note that Equation (10) combined with Equation (16) shows that $\xi_i = 0$ if $\alpha_i < C$ Thus we can simply take any training data for which $0 < \alpha_i < C$ to use Equation (15) (with $\xi_i = 0$) to compute *b*.

$$b = y_i - W^T \varphi(x) \qquad \text{----------------- (18)}$$

It is numerically reasonable to take the mean value of all *b* resulting from such computing. Hence,

$$b = \frac{1}{Ns} \sum_{0<\alpha i<C} [yi - W^T \phi(x_i)) \qquad \text{------------ (19)}$$

Where, *Ns* is the number of the support vectors. For a new data *x*, the classification function is then given by

$$f(x) = \mathrm{sign}(W^T \varphi(x) + b) \qquad \text{------------------ (20)}$$

Substituting Equations (17) and (19) into Equation (20), we get the final classification function

$$f(x) = \mathrm{sign}\, (\sum_{i=1}^{N} \alpha i\, yi\, \phi(xi) T \phi(xi))$$

$$+ \frac{1}{Ns} \sum_{0<\alpha i<C} [yi - \sum_{i=1}^{N} \alpha i\, yi\, \phi(x_i) T \phi(x_j)] \text{ ----(21)}$$

If there is a kernel function such that $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)^T$, it is usually unnecessary to explicitly know what $\varphi(x)$ is, and we only need to work with a kernel

Function in the training algorithm. Therefore, the non-linear classification function is f(x) = sign $(\sum_{i=1}^{N} \alpha i\ yi\ K(xi, x)\ + \frac{1}{Ns} \sum_{0<\alpha i<C} [\, yi$

$\sum_{i=1}^{N} \alpha i\ yi\ K(x_i, x_j)\,]\,)$ -------------------- (22)

3. PREDICTIONS OF STOCK MARKET BY USING SVM.

Following steps for prediction of stock market:

Step 1: This step is important for the download data from the net. We are predicting the financial market value of any stock. So that the share value up to the closing date are download from the site.

Step 2: In the next step the data value of any stock that can be converted into the CSV file (Comma Separate Value) so that it will easily load into the algorithm.

Step 3: In the next step in which GUI is open and when we click on the SVM button it will show the window from which we select the stock dataset value file.

Step 4: After selecting the stock dataset file from the folder it will show graph Stock before mapping and stock after mapping.

Step 5: The next step algorithm calculated the $\log_2 c$ and $\log_2 g$ value for minimizing error. So it will predict the graph for the dataset value efficiently.

Step 6: In final step algorithm display the predicted value graph of select stock which shows the original value and predicted value of the stock.

4. RESULT ANALYSIS

The data we use in this analysis comes from the Allegiant Travel Company(ALGT), Alliance Fiber Optic Products,(AFOP), AT & T Inc. (T), Bank of New York Mellon Corpora(BK), eBay, Inc(EBAY), EXCO TECH(XTC.TO), Facebook, Inc(FB), FORD, Inc(FORD), IBM, Inc(IBM), Kofax Limited(KFX), Old Second Bancorp, Inc(OSBC), SLM Corporation(SLM), Xilinx, Inc(XLNX). Also, we label each stock in a specific year in our data set as a good or a poor investment. Although there is no definitive method for defining a market investment as "good" or "poor", we use a method that is simple and objective: if the price of a company's stock over a given year rose, it is classified as a good investment; otherwise it is classified as a poor investment. Our training sample was based on a random selection of 13 companies, for all years from 2004-2015, where data was provided in their annual report.

Table 1: Classification of test sample by SVM classifier

Ob	Stock year 2004-15	Symbol	Dataset Value	Actual Group	Predicted Group
1	Allegiant Travel Company	ALGT	2103	Good	Poor
2	Alliance Fiber Optic Products	(AFOP)	2843	Good	Good
3	AT & T Inc.	T	2843	Good	Good
4	Bank of New York Mellon Corpora	BK	2843	Good	Good
5	EXCO TECH	XTC.TO	2839	Good	Poor
6	IBM, Inc	IBM	2843	Good	Good
7	FORD Inc.	FORD	2843	Poor	Good
8	Facebook, Inc.	FB	733	Good	Poor
9	eBay, Inc.	EBAY	2843	Good	Good
10	Old Second Bancorp, Inc.	OSBC	2843	Good	Good
11	SLM Corporation	SLM	2843	Poor	Good
12	Xilinx, Inc.	XLNX	2843	Good	Good
13	Kofex Limited	KFX	344	Good	Good

As the purpose of the classification model should be used to predict how a stock will perform in the upcoming year, when the result is unknown, it is necessary to test the SVM model on a test sample. For the test sample in our case, the SVM model correctly classified 10 of the 13 stocks as a good or poor investment. Table 4.4.1 shows how the SVM model classified the 13 stocks in the "unknown" test samples. The error rate of 23% means that this model correctly predicts whether the value of a particular stock will rise or fall over the next year.

4.1 Result Analysis for SVM

For result analysis we take the global financial data of company name as IBM Inc. from the year 2014 to 2015 only one year data. The data can get with the help of from http://in.finance.yahoo.com. The code written for download the dataset of stock using MATLAB

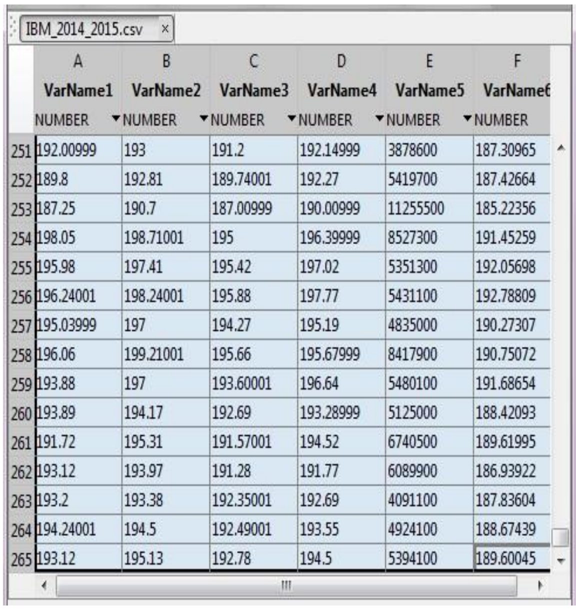
c = yahoo;

Yahoo_HSI_Data = fetch(c,'IBM',{'Open', 'High', 'Low', 'Close', 'Volume', 'Adj Close'},'2014-01-01','2015-04-20','d');save Yahoo_HSI_Data.mat

Now it will generate the day wise stock value of the IBM Incorporation the dataset can get the 265 dataset values. This is a very small amount of data. But we use this dataset for the result analysis. The CSV file of this dataset can load into the algorithm then out of 256 data set value only 70 % data can giving to the training SVM classifier and remaining 30% are giving input to the testing phase

Training Phase = $\frac{265}{100}$ x 70 = 185.5

Testing Phase = $\frac{265}{100}$ x 30 =79.5So that approximate 186 dataset value is used for the traing phase and 79 dataset value is used for the testing phase. And apply these input to the SVM algorithm ithen it will preprosessing and then it will produsce tesult in the form of graph. The result as shown in below


IBM_2014_2015.csv

	A	B	C	D	E	F
	VarName1	VarName2	VarName3	VarName4	VarName5	VarName6
	NUMBER	NUMBER	NUMBER	NUMBER	NUMBER	NUMBER
251	192.00999	193	191.2	192.14999	3878600	187.30965
252	189.8	192.81	189.74001	192.27	5419700	187.42664
253	187.25	190.7	187.00999	190.00999	11255500	185.22356
254	198.05	198.71001	195	196.39999	8527300	191.45259
255	195.98	197.41	195.42	197.02	5351300	192.05698
256	196.24001	198.24001	195.88	197.77	5431100	192.78809
257	195.03999	197	194.27	195.19	4835000	190.27307
258	196.06	199.21001	195.66	195.67999	8417900	190.75072
259	193.88	197	193.60001	196.64	5480100	191.68654
260	193.89	194.17	192.69	193.28999	5125000	188.42093
261	191.72	195.31	191.57001	194.52	6740500	189.61995
262	193.12	193.97	191.28	191.77	6089900	186.93922
263	193.2	193.38	192.35001	192.69	4091100	187.83604
264	194.24001	194.5	192.49001	193.55	4924100	188.67439
265	193.12	195.13	192.78	194.5	5394100	189.60045

Figure 2: Stock dataset for IBM Inc. in CSV file

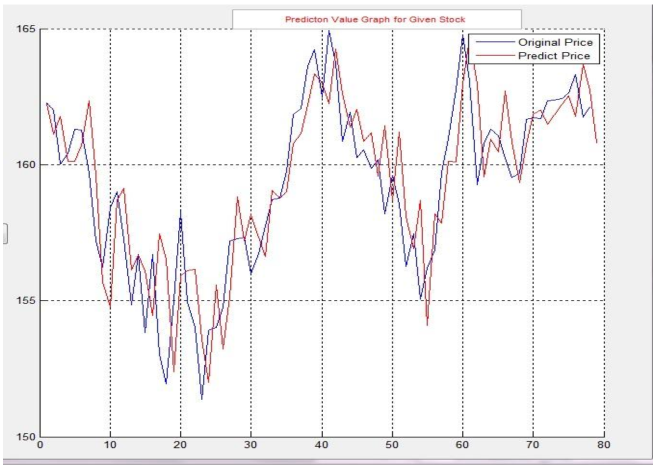


Figure 3: Predicted output by SVM for IBM Inc.

To calculate the efficiency of the predicted output we required the value which good for prediction also poor for prediction. If we are analyzing the above graph some value predicting with large error. So that we can remove that error value for calculating the efficiency of algorithm. In given example we test 79 dataset out of which only 7 value giving the error value so that, only 72 dataset value giving correct prediction. SVM Algorithm Efficiency Rate for Testing Phase = $\frac{72}{79}$ x100 = 91.13%

And the algorithm train 100% to predict the value so that the SVM Algorithm Efficiency Rate for Training Phase is 100%.

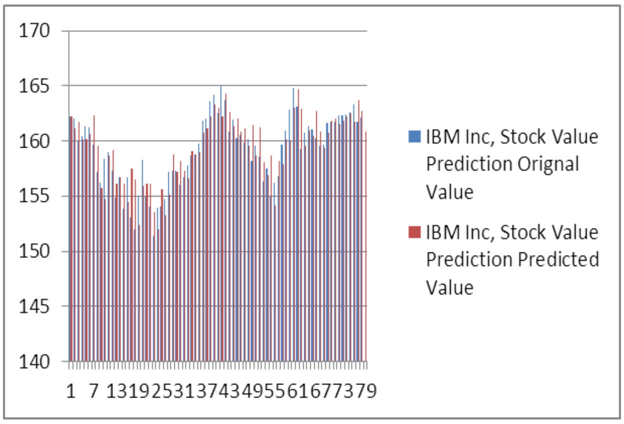


Figure 4 is a bar graph shows the result for the predicted stock value for IBM Inc. in year 2014 to 2015.

CONCLUSION

In the project, we proposed the use of data collected from different global financial markets with machine learning algorithms to predict the stock index movements. Our conclusion can be summarized into following aspects: SVM algorithm work on the large dataset value which collected from different global financial markets. Also SVM does not give a problem of over fitting. Correlation analysis indicates strong interconnection between the Market stock index and global markets that close right before or at the very beginning of trading time. Various machine learning based models are proposed for predicting daily trend of Market stocks. Numerical results suggest high efficiency. A practical trading model is built upon our well trained predictor. The model generates higher profit compared to selected benchmarks.

REFERENCES

[1]. Zhen Hu, Jibe Zhu , and Ken Tse "Stocks Market Prediction Using Support Vector Machine", 6th International Conference on Information Management, Innovation Management and Industrial Engineering, 2013.

[2]. Wei Huang, Yoshiteru Nakamori, Shou-Yang Wang, "Forecasting stock market movement direction with support vector machine", Computers & Operations Research, Volume 32, Issue 10, October 2005, Pages 2513–2522.

[3]. C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, Volume 2, pp. 1-43, 1998.

[4]. C. Cortes and V. Vapnik, "Support Vector Networks", Machine Learning, 20, 273-297, 1995.

[5]. M. Pontil and A. Verri, "Properties of Support Vector Machines", Technical Report, Massachusetts Institute of Technology, 1997.

[6]. E.E. Osuna, R. Freund and F. Girosi, "Support Vector Machines: Training and Applications", Technical Report, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, AI Memo No. 1602, 1997.

[7]. N. Ancona, Classification Properties of Support Vector Machines for Regression, Technical Report, RIIESI/CNR-Nr. 02/99.

[8]. Jianwen Xie, Jianhua Wu, and Qingquan Qian: Feature Selection Algorithm Based on Association Rules Mining Method. ACIS-ICIS 2009: 357-362.

[9]. T. Joachim's, Making Large-Scale SVM Learning practical, Technical, Report, L 8-24, Computer Science Department, University of Dortmund, 1998.

[10]. A.J. Smola and B. Scholkopf, a Tutorial on Support Vector Regression, NEUROCOLT2 Technical Report Series, NC2-TR- 1998-030, 1998.

[11]. Debashish Das and Mohammad shorif uddin data mining and neural network techniques in stock market prediction: a methodological review, international journal of artificial intelligence & applications, vol.4, no.1, January 2013

[12]. Mehpare Timor, Hasan Dincer and Şenol Emir doing a work about the "Performance comparison of artificial neural network (ANN) and support vector machines (SVM) models for the stock selection problem: An application on the Istanbul Stock Exchange (ISE)" African Journal of Business Management Vol. 6(3), pp. 1191-1198, 25 January, 2012.

[13]. Atanu Pal, Diptarka Chakraborty Prediction of Stock Exchange Share Price using ANN and PSO in International journal 2014.

[14]. Sudarsan Padhy and Shom Prasad Das Support Vector Machine for prediction of future international Journal of Computer Applications (0975 – 8887) Volume 41– No.3, March 2012.

[15]. http://en.wikipedia.org/wiki/Stock_market

[16]. http://biz.yahoo.com/p/

[17]. http://en.wikipedia.org/wiki/MATLAB