



Stock Price Prediction Using Hybrid Approach of Rule Based Algorithm and Financial News

Jageshwer Shriwas

*M.Tech Scholar, Department of IT ,NRI Institute of Science and Technology,
Bhopal(MP). jagesh.sh@gmail.com*

Dr Samidha Dwivedi Sharma

*Professor and HOD, Department of IT ,NRI Institute of Science and
Technology, Bhopal(MP). samidhad2000@gmail.com*

ABSTRACT

This research examines and analyzes the use of neural networks as a prediction tool. Specifically a neural network's ability to predict future trends of Stock Market Indices is tested. Accuracy is compared against a traditional forecasting method. While only briefly discussing neural network theory, this research determines the feasibility and practicality of using data mining as a forecasting tool for the individual investor. Data mining algorithms have a great capability of finding hidden patterns and trends, if they are provided with a reasonable amount of input data and desired output. As the number of input values increases, the quality of prediction increases as well. Thus for a better indexes predictor, you would like to use more parameters than just the prime interest rate and indexes historical data. Prediction of stock market trends has been an area of great interest both to researchers attempting to uncover the information hidden in the stock market data and for those who wish to profit by trading stocks. The extremely nonlinear nature of the stock market data makes it very difficult to design a system that can predict the future direction of the stock market with sufficient accuracy. Thus in this paper, we introduced to the financial predictor based upon neural networks.

INTRODUCTION

Each year, the field of computer science becomes more sophisticated as new types of technologies hit the market. Despite that, the problem of developing intelligent agents that will precisely simulate human brain activity is still unsolved. Data mining is well founded on the theory that the historic data holds the essential memory for predicting the future direction. Mining

news articles and the time series data concurrently, for predicting the stock market prices is an emerging topic in data mining and text mining communities. Information in the form of quotes and financial news is released into the market all the time. While quotes data is structured and can be directly used, the challenge is to manage the large amounts of textual information. We can employ techniques to parse the news articles and identify the key features most likely to have an impact on the stock market. By automating this process, machines can take advantage of arbitrage opportunities faster than human counterparts by repeatedly forecasting price fluctuations and executing immediate trades to make profits. In this paper, we describe such an application driven data mining system for stock market prediction. Financial news implicitly affects financial markets of investors, stock-price, political, economic and financial attributes.

REVIEW OF LITERATURE

M.Suresh Babu [1] primarily aims at designing an algorithm based on pattern matching technique which makes stock market forecast using historical data, and constructing formulas which are used to calculate future index points. The secondary objective is to analyze potential relationship between stock index and some attributes of the index. David Marten [2] having two aims. Firstly they provide an overview of previous ant-based approaches to the classification task and compare them with state-of-the-art classification techniques, such as C4.5, RIPPER, and support vector machines. Secondly a new ant-based classification technique is proposed named AntMiner+. The key differences between the proposed AntMiner+ and previous AntMiner versions are the usage of the better performing MAX-MIN ant system. The proposed AntMiner+

accuracy that is superior to that obtained by the other AntMiner versions, and competitive or better than the results achieved by the compared classification techniques. Schumaker, R. P.[3] examines a predictive machine learning approach for financial news articles analysis using several different textual representations: bag of words, noun phrases, and named entities. The proposed model called AzfinTS based on support vector machine approach. It calculate the stock price after 20 minute to released news. The model containing both article terms and stock price at the time of article release had the best performance in closeness to the actual future stock price, the same direction of price movement as the future price and the highest return using a simulated trading engine. Proper Noun scheme performs better than the de facto standard of Bag of Words in all three metrics. Manisha V. Pinto [5] provides a framework for predicting stock magnitude and trend for making trading decisions by making use of a combination of Data Mining and Text Mining methods. The prediction model predicts the stock market closing price for a given trading day 'D', by analyzing the information rich unstructured news articles along with the historical stock quotes. In particular, we investigate the immediate impact of the news articles on the time series based on Efficient Market Hypothesis (EMH). Key phrases provide semantic metadata that summarize and characterize documents. This framework incorporates Kea, an algorithm for automatically extracting key phrases from news articles. The prediction power of the Neural Network is used for predicting the closing price for a given trading day. The Neural Network is trained on the extracted key phrases and the stock quotes using the Back propagation Algorithm. Anton Nakov and Galo Nuño [6] explore how replacing the rational expectations hypothesis (REH) with "learning from experience" modifies the results of a simple general equilibrium model of the stock market. They were interested in the dynamics of heterogeneous beliefs and in the feedback loop that arises when individuals learn about variables that are the result of their collective decisions given their beliefs, a type of self referentiality emphasized.

Binoy B. Nair [10] presents a data mining based stock market trend prediction system, which produces highly accurate stock market forecasts. The proposed system is a genetic algorithm optimized decision tree-support vector machine (SVM) hybrid, which can predict one-day-ahead trends in stock markets. The uniqueness of the proposed system lies in the use of the hybrid system which can adapt itself to the changing market conditions and in the fact that while most of the attempts at stock market trend prediction have approached it as a regression problem, present

study converts the trend prediction task into a classification problem, thus improving the prediction accuracy significantly. Performance of the proposed hybrid system is validated on the historical time series data from the Bombay stock exchange sensitive index (BSE-Sensex). The system performance is then compared to that of an artificial neural network (ANN) based system and a naïve Bayes based system. It is found that the trend prediction accuracy is highest for the hybrid system and the genetic algorithm optimized decision tree-SVM hybrid system outperforms both the artificial neural network and the naïve bayes based trend prediction systems. Luna C. Tjung [12] presents a Business Intelligence (BI) approach to forecast daily changes in seven financial stocks' prices. The purpose of our paper is to compare the performance of ordinary least squares model and neural network model to see which model does a better job to predict the changes in the stock prices and identify critical predictors to forecast stock prices to increase forecasting accuracy. The BI approach uses a financial data mining technique to assess the feasibility of financial forecasting compared to a regression model using an ordinary least squares estimation method. We used eight indicators such as macroeconomic indicators, microeconomic indicators, market indicators, market sentiment, institutional investor, politics indicators, business cycles, and calendar anomalies to predict changes in financial stock prices. We found that NN provided superior performance with up to 96% forecasting accuracy compared with OLS model with only 68%.

National Stock Exchange of India (NSE)

NSE was incorporated in November 1992, and received recognition as a stock exchange under the Securities Contracts (Regulation) Act, 1956 in April 1993. Since its inception in 1992, NSE of India has been at the vanguard of change in the Indian securities market. This period has seen remarkable changes in markets, from how capital is raised and traded, to how transactions are cleared and settled. The market has grown in scope and scale in a way that could not have been imagined at that time. Average daily trading volumes have jumped from Rs. 17 crore in 1994-95 when NSE started its Cash Market segment to Rs.16, 959 crore in 2009-10. Similarly, market capitalization of listed companies went up from Rs.363, 350 crore at the end of March 1995 to Rs.36, 834,930 crore at end March 2011. Indian equity markets are today among the most deep and vibrant markets in the world. NSE offers a wide range of products for multiple markets, including equity shares, Exchange Traded Funds (ETF), Mutual Funds, Debt instruments, Index futures and options, Stock futures and options, Currency futures and Interest

rate futures. Our Exchange has more than 1,400 companies listed in the Capital Market and more than 92% of these companies are actively traded. The debt market has 4,140 securities available for trading. Index futures and options trade on four different indices and on 223 stocks in stock futures and options as on 31st March, 2010. Currency futures contracts are traded in four currency pairs. Interest Rate Futures (IRF) contracts based on 10 year 7% Notional GOI Bond is also available for trading. The role of trading members at NSE is to the extent of providing only trading services to the investors; the Exchange involves trading members in the process of consultation and participation in vital inputs towards decision making.

Stock Market

A stock market index is a method of measuring a stock market as a whole. The most important type of market index is the broad-market index, consisting of the large, liquid stocks of the country. In most countries, a single major index dominates benchmarking, index funds, index derivatives and research applications. In addition, more specialized indices often find interesting applications. In India, we have seen situations where a dedicated industry fund uses an industry index as a benchmark. In India, where clear categories of ownership groups exist, it becomes interesting to examine the performance of classes of companies sorted by ownership group.

Stock Classification

Stocks are often classified based on the type of company it is, the company's value, or in some cases the level of return that is expected from the company. Below is a list of classifications which are generally known to us Growth Stocks, Value Stocks, Large Cap Stocks, Mid Cap Stocks, and Small Cap Stocks. Stocks are usually classified according to their characteristics. Some are classified according to their growth potential in the long run and the others as per their current valuations. Similarly, stocks can also be classified according to their market capitalization. S&P CNX NIFTY has NIFTY (50), Junior NIFTY (50), CNX IT (20), Bank NIFTY (12), NIFTY Midcap50, CNX Realty (10) and CNX Infra (25). The sectoral distribution of NSE are Financial services or banks, Energy, Information Technology, Metals, Automobile, FMCG, Construction, Media & Entertainment, Pharma, Industrial Manufacturing, Cement, Fertilizers & Pesticides, Textiles, Power and Telecom. Two ways of analyzing stock prices namely fundamental analysis and technical analysis are described in the next section.

Fundamental Analysis

Fundamental analysis involves analysis of a company's performance and profitability to determine its share price. By studying the overall economic conditions, the company's competition, and other factors, it is possible to determine expected returns and the intrinsic value of shares. This type of analysis assumes that a share's current (and future) price depends on its intrinsic value and anticipated return on investment. As new information is released pertaining to the company's status, the expected return on the company's shares will change, which affects the stock price. So the advantages of fundamental analysis are its ability to predict changes before they show up on the charts. Growth prospects are related to the current economic environment.

Technical Analysis

Technical analysis is a method of evaluating securities by analyzing the statistics generated by market activity, such as past prices and volume. To predict the future movement and identify patterns of a stock, technical analysts use tools and charts ignoring the fundamental value. Technical analysis studies the historical data relevant to price and volume movements of the stock by using charts as a primary tool to forecast possible price movements. According to early research, future and past stock prices were deemed as irrelevant. As a result, it was believed that using past data to predict the future stock price was impossible, and that it would only have abnormal profits. However, recent findings have proven that there was, indeed, a relationship between the past and future return rates. Furthermore, arguments have been made that by using past return rates, future return rates could also be forecasted. There are various kinds of technical indicators used in futures market as well. There are 26 technical indicators which can be primarily used to analyze the stock prices. Based upon the analysis, stock trend either up or down can be predicted by the investor.

STATEMENT OF THE PROBLEM

Financial investors who invest in stock market are usually not aware about the stock market behavior. They might face the problem of stock trading as they do not know which stock to buy and which to sell in order to gain more profit. All these users don't know that the progress of stock market depends a lot on relevant news. Any financial news make the big or small impact on all the sector, but some news make the big impact on the specific sectors as shown in Table 1.

News	Sector
Petrol price hike by 5%.	Petroleum
Cylinder gas is again costlier by Rs. 200.	Petroleum
Announces Q3 results	IT
TCS add another company to their belt	IT
Detroit Auto Show's newest hotties	Automobile
U.S. Treasury to sell some Ally Financial shares	Automobile

TABLE 1: NEWS AND IMPACT OF THAT ON SECTOR

Business / financial news stories can contain information (words, sequence of words or overall story opinion) which has an economic value which is not reflected in the current price of a specific share or market index. Business/ financial news therefore can be used as a basis of a successful trading strategy. In summary, the underlying hypothesis for the work of this paper is that it is possible to improve an existing trading strategy by adding information from a financial news text mining system.

PROPOSED WORK

The objective of this research is to develop a model that would accurately predict the future closing price of the BSE. Once this was accomplished, the probability of an accurate forecast would be calculated. Given the accuracy of the forecast, the benefits of the network to the investor would be determined. Financial time series consists of multidimensional and complex nonlinear data that result in of pitfalls and difficulties for accurate prediction. We believe that a successful financial time series prediction system should be hybrid and custom made, which means it requires a perfect combination of sophistication in the prediction system design and insight in the input data preprocessing. Both areas have been extensively researched. Data mining people have developed some very useful techniques of dealing with data, such as missing values, data multiplication. Traders and technical analysts also have done lots of work. They derive enhanced features of interest from the raw time series data, such as the variety of indicators including Moving Average Convergence Divergence (MACD), Price by Volume, Relative Strength Index (RSI) etc. some of which have been confirmed to pertaining useful information. Figure 1 shows the stock price prediction process model.

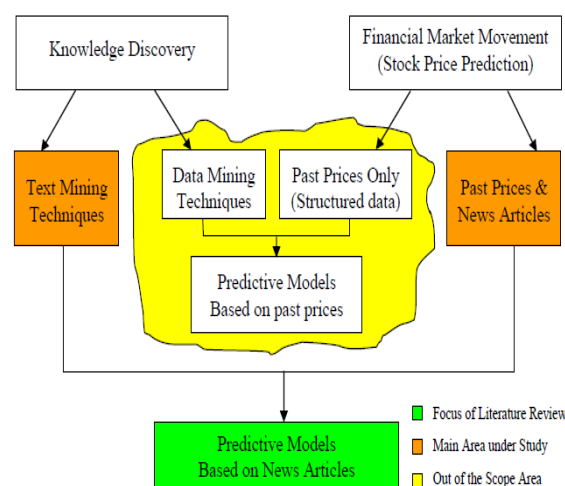


Figure 1: Stock Price Prediction Process Model

There are three major difficulties about accurate forecast of financial time series. First, the patterns of well-known financial time series are dynamic, i.e. there are no single fixed models that work all the time. Second, it is hard to balance between long-term trend and short-term sideways movements during the training process. In other words, an efficient system must be able to adjusting its sensitivity as time goes by. Third, it is usually hard to determine the usefulness of information. Misleading information must be identified and eliminated.

Rosennean Complexity

Rosennean complexity suggests that if a single dynamical description is capable to describe a system successfully, then the behaviors of that system will always be predicted correctly. This kind of system does not have any complexity. On the other hand, if multiple partial dynamical descriptions are insufficient to describe a system successfully, it is complex. Financial markets are certainly complex systems; no one could correctly, successfully describe the system. Hence, no one could predict the financial market behavior, or else there would be no financial market. Some people believe that prices and indices movement obey the random walk theory, which says prices move independently in any point of time, implying that historical data cannot be used to predict future movement. This is, indeed, obvious, because a financial markets responses to many factors every day.

Method

Pattern matching using Ant Algorithm is employed to generalize similar segments. This helps us predict the future movement direction of a target segment. After determining the future movement direction, several formulas are applied

to calculate the index point for the next trade date. The logical flow of the whole financial forecast process is shown in Figure 2.

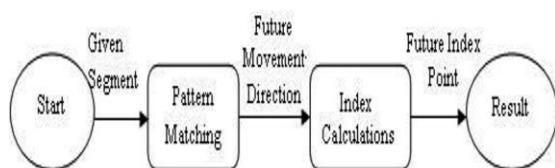


Figure 2: Logical flow of the financial forecast process

Pattern Matching

Generalization of segments is done according to the some of the segment attributes stated in Table 2. Segments which are too large or too small cannot be generalized well. In order words, the segment size affects prediction accuracy. As the ants learn how segments should be generalized, the optimal segment size could be determined by picking the size that the best prediction accuracy was achieved with. The time-series will be chopped into segments. Each segment is associated with a number of attributes, which are sets of values for each index point in the segments.

Segment attribute	Description
Absolute index	The exact index point
Movement magnitude	The change in index values
Movement percentage	The percentage change in index value
Movement direction	A rise, a falls or zero index movement
Percentage difference between index and averaged index	The percentage difference between index and averaged index

TABLE 2: SEGMENT ATTRIBUTES

Ant Algorithm

For the application of Ant algorithm to our pattern matching problem, some of the listed segment attributes are used to find out the matched segment. Different ants give different weights to these segment attributes, and they also give a weight to the other ants' choices, which simulates the effect of pheromone. A segment and a sub time series to be evaluated will be given, and different ants will start from different locations in the sub-time series, which are randomized. An ant moves along the sub-time-series, inspects its left hand side and right hand side, and stop at the point that it

believes that the matched pattern was found on each side. It then chooses the one with the lowest difference among both sides. The track it went through will attract other ants' choices, and the degree will be influenced by the weight given by the other ants. During the learning process, each ant find out the segment which it thinks is matched. After all ants made their decisions, ants' choices are compared to the target segment. A choice is considered to be incorrect, if the next movement direction of the chosen segment is different from that of the target segment. In this case, the neighbors' of the chosen segment will be inspected to determine how the ant should adjust its weights.

Index Calculation

A number of formulas are used to calculate the predicted index point for the next trade date after determined the movement direction, and their performances are being assessed by comparing to the differences to the actual index point. Table 3 lists these formulas. The predicted index point is basically calculated by the last index point plus a movement. The movement direction predicted determines the sign of the movement.

No	Formula	Description
1	$y_{n+1} = y_n \pm (\sum_{i=1}^n v_i) / n$	Take the averaged movement of the segment as the next movement
2	$y_{n+1} = y_n \pm \sum_{i=1}^n \{i / (\sum_{j=1}^n j)\} v_i$	Take the weighted averaged movement of the segment as the next movement, while the more recent the index point, the higher the weight applied
3	$y_{n+1} = y_n \pm y_n * \{(\sum_{i=1}^n p_i) / n\}$	Take the averaged movement percentage of the segment as the next movement percentage
4	$y_{n+1} = y_n \pm y_n * \{(\sum_{i=1}^n \{i / (\sum_{j=1}^n j)\} p_i)\}$	Take the weighted averaged movement percentage of the segment as the next movement percentage, while the more recent the index point, the higher the weight applied

TABLE 3: FORMULAS USED FOR INDEX CALCULATION

Multilayer Perceptron

For the task of predicting the indexes, we'll be using the so called multilayer feed forward network which is the best choice for this type of application. In a feed forward neural network, neurons are only connected forward. Each layer of the neural network contains connections to the next layer, but there are no connections back.

Typically, the network consists of a set of sensory units (source nodes) that constitute the input layer, one or more hidden layers of computation nodes, and an output layer of

computation nodes. In its common use, most neural networks will have one hidden layer, and it's very rare for a neural network to have more than two hidden layers. The input signal propagates through the network in a forward direction, on a layer by layer basis. These neural networks are commonly referred as multilayer perceptrons (MLPs).

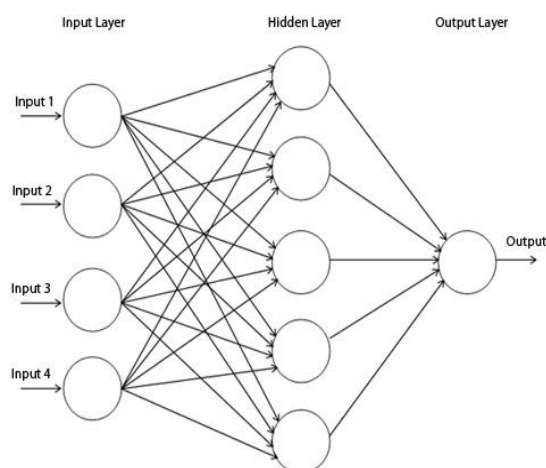


Figure 3: A simple MLP with 4 inputs, 1 output, and 1 hidden layer.

Algorithm Snapshot

```
PredictionStockMarket(){
    If(data != null){
        for (int I=0; i<data.count-1;i++){
//absolute index of data
            absIndex [i]= double(close_price)[i]
            absIndex1 [i]= double(close_price)[i+1]
// sum of absolute index
            sumofAbsIndex = sumofAbsIndex + absIndex[i]
//movement magnitude of absolute index
            movmentMagnitude[i] = absIndex[i] - absIndex[i+1]
// sum of movement magnitude of absolute index
            movSum = movSum + movmentMagnitude[i]
// percentage of movment magnitude of absolute index
            movPercentage[i] = (movmentMagnitude[i] / absIndex[i])*100
// average of absolute index
            avgIndex = (sumofAbsIndex / absIndex.length )/absIndex[i]
// percentage difference between average index and absolute index
            perDiffIndexAvgIndex = (absIndex[i] / avgIndex) *100
// percentage index
            predictedIndex = absIndex + movSum
            if(predictedIndex > absIndex[i]){
                err = ((predictedIndex - absIndex[i]) /
absIndex[i]) * 100
                sp = (predictedIndex + err)
            }Else if(predictedIndex < absIndex[i]){
                err = ((absIndex[i] - predictedIndex) /
absIndex[i]) * 100;
                sp = (predictedIndex + err);}}
            Else{
                Message("Insufficient Data");}
        }
    }
```

RESULT

Stock market samples were collected from [12-15]. For the 20 success prediction, we compared the performance of each index calculation formulas.

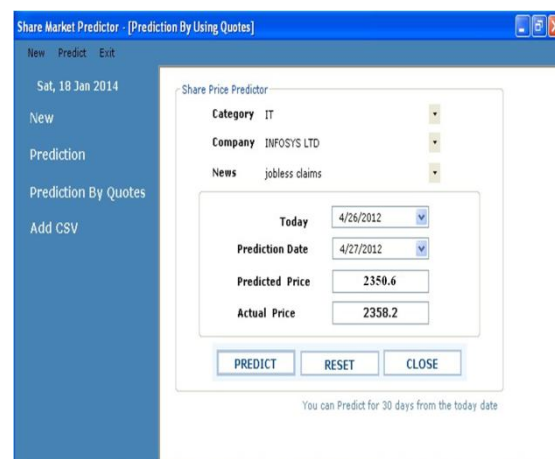


Figure 4: Stock Price Prediction with News Sample

The result is presented in Table 3. The system predicted the Predicted index movement direction successfully for 16 out of 20 samples, the accuracy was 80%.

N o	Predicted on	Predicted for	Actual index move ment direction	Predicted index move ment direction	correct
1	2/9/2013	3/9/2013	2986	2960	✓
2	5/9/2013	6/9/2013	2961	2897	✓
3	9/9/2013	10/9/2013	3084	2944	✓
4	16/9/2013	17/9/2013	3007	2989	✓
5	27/9/2013	28/9/2013	3100	3206	✓
6	2/10/2013	3/10/2013	3105	3102	✓
7	13/10/2013	14/10/2013	3023	2703	×
8	22/10/2013	23/10/2013	3014	2719	×
9	31/10/2013	1/11/2013	3015	2783	×
10	5/11/2013	6/11/2013	3032	2755	×
11	11/11/2013	12/11/2013	3348	3332	✓
12	25/11/2013	26/11/2013	3334	3339	✓
13	1/12/2013	2/12/2013	3264	3260	✓

14	3/12/2013	4/12/2013	3305	3208	✓
15	10/12/2013	11/12/2013	3339	3331	✓
16	16/12/2013	17/12/2013	3443	3389	✓
17	30/12/2013	31/12/2013	3348	3205	✓
18	1/1/2014	2/1/2014	3365	3242	✓
19	3/1/2014	4/1/2014	3126	3118	✓
20	10/1/2014	11/1/2014	3373	3369	✓

TABLE 3: FORECAST RESULT WITH CORRECTNESS

CONCLUSION

In this article, the topic of neural networks and their prediction capabilities have been analyzed. Feed forward neural networks proved to be a reliable solution for applications that need to predict something. Generally speaking, function interpolation is one of the major fields of study in stock market environment. A strategy based upon technical indicators, can really help you in achieving good trading results. Of course, the application that is presented in this article cannot be used in a real world environment, because normally you would need not only an almost precise prediction, but also a program that will perform the market analysis in short bursts (each 15-30 seconds), opposite to the values predicted in this application (closing stock value). In order to achieve better results, you would rather want to combine classical trading strategy with one based upon real-time technical indicators. As for the studying purposes, the main objective has been achieved. It can be concluded that neural networks do have the capability to forecast financial markets and, if properly trained, the individual investor could benefit from the use of this forecasting tool.

BIBLIOGRAPHY

- [1] M.Suresh Babu, N.Geethanjali and B.Satyanarayana, Forecasting of Indian Stock Market Index Using Data Mining & Artificial Neural Network, International Journal of Advanced Engineering & Application, June 2011.
- [2] David Martens, Manu De Backer, Raf Haesen, Student Member, IEEE, Jan Vanthienen, Monique Snoeck, and Bart Baesens Classification With Ant Colony Optimization,

IEEE Transaction on evolutionary computation, vol.11, No. 5, October 2007.

- [3] Schumaker, R. P. and Chen, Textual analysis of stock market prediction using breaking financial news: The AZFin text system. ACM Trans. Inform. Syst. 27, 2, Article 12, February 2009.
- [4] Azadeh Nikfarjam ,Ehsan Emadzadeh, Saravanan Muthaiyah, Text mining approaches for stock market prediction, IEEE, 2010.
- [5] Manisha V. Pinto and Kavita Asnani, Stock Price Prediction Using Quotes and Financial News, International Journal of Soft Computing and Engineering (IJSCE) Volume-1, Issue-5, November 2011.
- [6] Anton Nakov and Galo Nuño, Learning From Experience in the Stock Market, Working Paper Series, No 1396, November 2011
- [7] K.K.Sureshkumar and N.M.Elango, An Efficient Approach to Forecast Indian Stock Market Price and their Performance Analysis, International Journal of Computer Applications, Volume 34– No.5, November 2011
- [8] Johan Bollena, Huina Mao, Xiaojun Zeng, Twitter mood predicts the stock market, Journal of Computational Science 2, 2011.
- [9] Holger Breinlich, Heterogeneous Firm-Level Responses to Trade Liberalisation: A Test Using Stock Price Reactions, CEP Discussion Paper No 1085, November 2011
- [10] Binoy B. Nair, V.P Mohandas and N.R. Sakthivel, A Genetic Algorithm Optimized Decision Tree-SVM based Stock Market Trend Prediction System, International Journal on Computer Science and Engineering, Vol. 02, September 2010.
- [11] Binoy.B.Nair, V.P Mohandas and N. R. Sakthivel, A Decision Tree- Rough Set Hybrid System for Stock Market Trend Prediction, International Journal of Computer Applications, Volume 6, September 2010.
- [12] Luna C. Tjung, Ojoung Kwon, K. C. Tseng and Jill Bradley-Geist, Forecasting Financial Stocks using Data Mining, Global Economy and Finance Journal, September 2010.
- [13] Wilton.W.T. Fok, Iaeng, Vincent.W.L. Tam, Hon Ng, Computational Neural Network for Global Stock Indexes Prediction, World Congress on Engineering 2008 Vol II, July, 2008.
- [14] Nikhil Bakshi, Stock Market Prediction Using Online Data: Fundamental and Technical Approaches, August 2008.
- [15] www.moneycontrol.com
- [16] www.indiaonline.com
- [17] www.bseindia.com
- [18] www.nseindia.com