ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

Survey Paper on Stock Market Prediction

Milind Kolambe, Prof. Aparna Junnarkar

Postgraduate Research Scholar, PESMCOE, Pune, India

Department of Computer Engineering, PESMCOE, Pune, India

ABSTRACT: Stock market prediction involves predicting future value of company stock or other financial instrument traded on an exchange. Various types of trading can be done in stock market. It could be short term trading or even long term trading but if someone can predict the value or class of that entity, it can yield very good return for the investment done. Prior to evolution of digital world, predictors continued to use paper work methods like fundamental and technical analysis. Various useful technical indicators like SMA, EMA, MACD found to be very usefulbut with the advent of computer technologies and algorithms, prediction moved into technological realm. Analysts started building prediction system using Neural Network, Support Vector Machine, Decision Trees, Hidden Markov Model. Prediction accuracy really improved using algorithmic approach.This surveycovers various traditional as well asevolutionary data mining techniques used for stock market prediction.

KEYWORDS: Stock trading, data mining, support vector machine, neural network,hidden markovmodel, decision trees, technical indicators

I. INTRODUCTION

Goal behind making any financial investment is to achieve above average return for invested money while maintaining certain level of involved risks [1] but as the stock market is a very complex, volatile and non-linear dynamical system, stock market prediction has become a tough challenge for researchers and investors.

Traditionally majority analysts depend on Fundamental analysis methods [3].Fundamental analysis is all about using concrete information about a company's business to try to find the real value of a stock. It is the investigation of the forces that affect the well being of the economy, industry groups, and companies. As with most analysis, the goal is to derive a forecast and profit from future price movements.

For day trading or short term trading Technical analysis [4],[5]found to be very effective. Itnot only enables the trader to define a concrete opinionon a particular stock or index but also helps to definethe trade keeping in mind the entry, exit and risk perspective.Technical analysis involves use offunctions, formulas such asindicators and oscillators derived by time series, and heuristic rules able to revealsignals of change in the market trends. Popular examplesof methods are Moving Average Convergence Divergence(MACD), Relative Strength Index (RSI), and stochastic oscillator [3] Result of technology evolution invented machine learning methods for stock price prediction. Various new data mining methods and algorithms are proposed like Genetic Algorithm (GA), Support Vector Machine(SVM), Neural Networks (NNs).

This survey paper has been organized as follows. In Section II we discussvarious proposed methods used for stock market prediction. Basicsof fundamental analysis used in trading stocks,most popular technical indicators being used for stock market prediction and various machine learning methods, algorithms that can used in this prediction followed by conclusion and references.

II. RELATED WORK

A. *FUNDAMENTAL ANALYSIS*

At the company level fundamental analysis includesanalysis of financial data, management reports, business concepts and competition. It also explore the relation between financial statement information and fundamental attributes such as revenue rate growth, price to book ratio etc [6],[7]. At the industry level, there might be an examination of supply and demand forces for the products offered. At economy level fundamental analysis might target economic data to assess the present and future growth of the economy. To forecast future stock prices, fundamental

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

analysis combines economic, industry, and company analysis [8],[9],[10] to derive a stock's current fair value and forecast future value. If fair value is not equal to the current stock price, fundamental analysts believe that the stock is either over or under valued and the market price will ultimately gravitate towards fair value. Fundamentalists do not pay attention to the advice of the random opinions and believe that markets are weak-form efficient. By believing that prices do not accurately reflect all available information, fundamental analysts look to capitalize on perceived price discrepancies. Various aspects regarding fundamental analysis are examining business plan, management, financial analysis etc.Advantage of fundamental analysis is to find out long term trend. It also helps to uncover companies with valuable assets, a strong balance sheet, stable earning etc.

Yuh-Jen Chen and Yuh-Min Chen proposed [11] afundamental analysis-based method for stock market forecasting by calculating the weight of financial indicators, evaluating and selecting individual stocks, selecting financial news features, determining stock trading signals based on financial news.ChingHsue Cheng, You-shyang Chen proposed[12] fundamental analysis of stock trading system using classification techniques.

B. ***TECHNICAL ANALYSIS (INDICATORS)***

Largenumber of technical indicators isavailable for technical analysis. They use various statistics generated in the market like closing prices of history, volume traded etc. Earlier in 1960s and 1970s severalresearchers studied trading rules based on TIs. Though they did notfind them much profitable [13],[14] recent studies [15],[16]show that they are very useful.Commonly used TIs aresimple moving averages (SMA), exponential moving averages (EMA), moving average convergence divergence (MACD), exponential moving average (EMA) and relative strength index (RSI).

1. Simple Moving Averages (SMA)

Moving averages give smooth price data to form a trend following indicator. Though they cannot predict price direction, but rather give some ideal about the current direction with a lag. Moving averages include lag because they are computedusing past prices. Despite this lag, moving averages gives smooth price action and filter out the noise. They are also useful as the building blocks for many other technical indicators and overlays, such as bollinger bands, MACD and the McClellan Oscillator.

A simple moving average is computedas the average price of a security over a specific number of periods. Most moving averages are calculated using closing prices. A 5-day simple moving average is the five day sum of closing prices divided by five. As its name indicates, a moving average is an average that moves. Past data is averaged as new data becomes available. This causes the average to move along the time scale. Below is an example of a 5-day moving average evolving over three days.

Daily closing prices = 6010, 6020, 6030, 6040, 6050, 6060, 6070
First day of 5-Days SMA = (6010 + 6020 + 6030 + 6040 + 6050) / 5 = 6030
Second day of 5-Days SMA= (6020 + 6030 + 6040 + 6050 +6060) / 5 = 6040
Third day of 5-Days SMA= (6030 + 6040 + 6050 +6060+6070) / 5 = 6050

Yu-Feng Lin, Chien-Feng Huang, Vincent S. Tseng used Simple moving Averages along with a technique of episode mining [17].

2. Exponential Moving Average (EMA)

The lag in SMA can be reduced by applying more weight to recent prices. EMA is the extension of SMA. The weighting applied to the most recent price depends on the number of periods in the moving average. There are three steps to calculating an exponential moving average. In first step simple moving average is caluclated. An exponential moving average (EMA) has to start somewhere so a simple moving average is used as the previous period's EMA in the first calculation. In second step, the multiplier (weighing multiplier) is computed. Finally, the exponential moving average can be computed using the formula.A 12-day EMA can be computed as follows.

SMA: 12 period sum / 12
Multiplier: (2 / (Time periods + 1)) = (2 / (12 + 1)) = 0.1538 (15.38%)
EMA: {Close - EMA (previous day)} x multiplier + EMA (previous day)

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

A 12-period exponential moving average applies 15.38% weighting to the most recent price. A 12-period EMA can also be called an15.38% EMA. A 20-period EMA applies a 9.52% weighing to the most recent price (2/(20+1) = .0952). Notice that the weighing multiplier for the shorter time period is greater than the weighing multiplier for the longer time period. In fact, the weighting drops by half every time the moving average period doubles.YauheniyaShynkevich, T.M. McGinnity, Sonya Coleman, Yuhua Li, AmmarBelatreche [18] used EMA to predict the directions of the future price movements.

3. Relative Strength Index (RSI)

RSI was developed by J. Welles Wilder. It is a momentum oscillator [19] that measures the speed and change of price movements. RSI oscillates between zero and hundred. Traditionally according to Wilder's calculation, RSI is said to be overbought when above 70 and said to be oversold when below 30. Signals can also be generated by looking for divergences, failure swings and centerline crossovers. RSI is useful for identifying general trend too.

RSI is one the most popular momentum indicators that has been featured in a number of articles, interviews and books over the years. Technical Analysis for the Trading Professional, the book byConstance Brown, features the concept ofbull market and bear market ranges for RSI. Andrew Cardwell, Brown's RSI mentor, implemented positive and negative reversals for RSI. In addition, Cardwell brought the concept of divergence, literally and figuratively, on its head.

$$RSI = 100 - (1/ (1 + RS))$$

$$RS = \text{Average Gain} / \text{Average Loss}$$

Jianxue Chen built svm based application [20] of financial time series forecasting using some empirical technical indicators like RSI.

4. Moving Average Convergence/Divergence Oscillator (MACD)

MACD was introduced by Gerald Appel in the seventies, the Moving Average Convergence/Divergence oscillator (MACD) is one of the simplest and most effective momentum indicators available. MACD tracks the changes in strength, direction, momentumand direction of a trend [21], [22]. The MACD turns two trend-following indicators, moving averages, into a momentum oscillator by subtracting the moving average for longer period from the moving average for shorter period. As a result, the MACD offers the best of both worlds: trend following and momentum. The MACD goes up and down with respect to the line called as zero line as the moving averages converge, cross and diverge. Someone canwatch for signal line crossovers, centerline crossovers and divergences to generate signals. Because the MACD is not bounded, it is not particularly useful for finding overbought and oversold status.

MACD Line: (EMA of 12 days- EMA of 26 days)
Signal Line: EMA of 9 days of MACD Line
MACD Histogram: MACD Line - Signal Line

The MACD Line is the 12-day Exponential Moving Average (EMA) subtracted by the 26-day EMA. Closing prices are used for these moving averages. A 9-day EMA of the MACD Line is plotted with the indicator to act as a signal line and identify turns. The MACD Histogram shows the difference between MACD and its 9-day EMA, the Signal line. The histogram is positive when the MACD Line is above its Signal line and negative when the MACD Line is below its Signal line.12, 26 and 9 are the most popularly used values used with the MACD; however other values can be substituted depending on your trading style and goals.

5. Resistance and Support

As the name suggests, resistance is something which stops the price from rising further. The resistancelevel is a price point on the chart where traders expect maximum supply (for selling)for the stock/index. The resistance level is always above the current market price.The likely hood of the price rising up to the resistance level, consolidating, absorbing all the supply,and then declining is high. The resistance is one of the critical technical analysis tool which market participants look at in a rising market. The resistance often acts as a trigger to sell.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

Figure 1: Support and Resistance

Understanding the support level should be quite simple and intuitive.As the name suggests, the support is something that prevents the price from falling further.Trader expects maximum demand (for buying) coming into the stock/index at this price point in the chart that is support level. Whenever the price falls to the support line, it islikely to bounce back. The support level is always belowthe current market price.There is a maximum likelihood that the price could fall till the supportconsolidate, absorb allthe demand, and then start to move up direction. The support is one of the critical technical levelmarket participants look for in a falling market. The support often acts as a trigger to buy.Aparna Bhatt and SowmyaKamath [23] used support and resistance to decide buy sell decision for a particular stock.

C. *Machine Learning Methods*

1. Artificial Neural Network (ANN)

An ANN has several advantages but one of the most recognized of these is the fact that it can actually learn from observing data sets. In this way, Artificial Neural Network is known for as a random function approximation tool. These types of tools help estimate the most cost-effective and ideal methods for arriving at solutions while defining computing functions or distributions. ANN use data samples instead ofcomplete data sets to arrive at solutions, which saves both time and money. ANNs are considered fairly simple mathematical models to improveeffectiveness of the available the data analysis technologies. ANNs includes three layers. These layers are connected to each other. The first layer consists of input neurons. Those neurons send data on to thesecond layer, which in turn sends the output neurons to the third layer. Training an artificial neural network involves choosing from allowed models for which there are several associated algorithms.

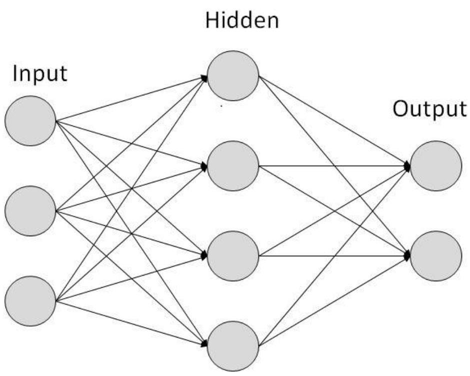

Figure 2: Artificial Neural Network

Phua et al. [24] applied Neural Networks to the financial prediction. He tested the influence of volume data on Stock price prediction. Khan et al. [25] applied the Neural Networks with different number of hidden layers to analyze the prediction of the Stock prices.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

2. Support Vector Machine (SVM)

Although SVM can be applied to various optimization problems such as regression, the typical problem is to classifythe data. The basic idea is shown in figure. The data points are identified as being positive or negative, and the problem is to find a hyper-plane. This plane separates the points (data) by a maximal margin.

"Support Vector Machine" (SVM) falls under a category of supervised machine learning algorithms which can be used for both classification and regression challenges.However researchersmostly use it for classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot).

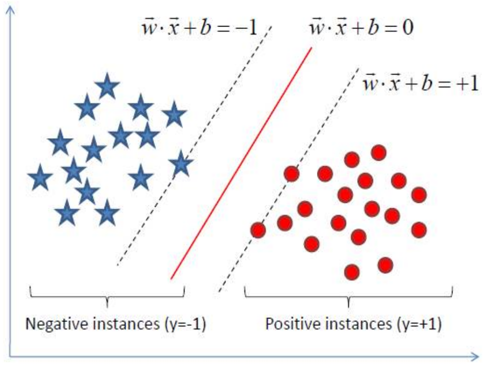

Figure 3: Support Vector Machine

Above figure only shows the 2-dimensional case where the data points are linearly separable.SVM approachto linear regression amounts to (simultaneous) minimization of ε-insensitive loss and minimization of the norm of linear parameters. This can be formally described by introducing (non-negative) slack variables,to measure the deviation of training samples outside ε-insensitive zone. Pai et al. [26] proposed a hybrid approach with SVM and ARIMA (Autoregressive Integrated Moving Average) model and found it gave promising results.

3. Hidden Markov Models

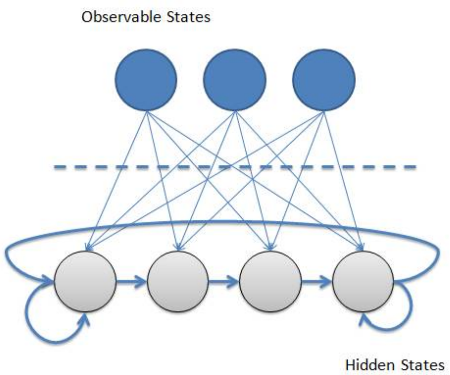

Figure 4: Hidden Markov Model

A Hidden Markov Model (HMM) is a finite statemachine. This has some fixed number of states. It gives a probabilistic framework for modeling a timeseries of multivariate observations. Hidden Markovmodels were introduced in the beginning of the 1970's. It is used asa tool in speech recognition. This model which is based onstatistical methods has become increasingly popular in thelast several years due to its strong mathematical

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

structureand theoretical basis as it is used a wide range ofapplications. Recently researchers proposed HMMas a classifier or predictor for speech signal recognition, DNA sequence analysis, handwrittencharacters recognition, natural language domains etc.It shows that HMM is a very powerful tool for variousapplications.The advantage of HMM can be summarizedas:

- HMM has strong statistical foundation
- It is able to handle new data robustly
- Computationally efficient to develop and evaluate(due to the existence of established trainingalgorithms).
- It is able to predict similar patterns efficiently

Luigi Troiano and PraveshKriplani applied [27] HMM for Predicting Trend in the Next-Day Market.

4. Decision Trees

Decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into small and then even smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches. Leaf node represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.

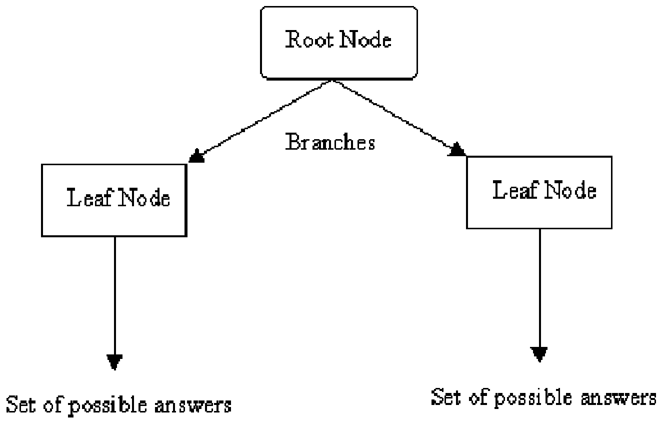

Figure 5: Decision Tree

A decision tree is built top-down from a root node and involves partitioning the data into subsets that contain instances with similar values (homogenous). ID3 algorithm uses entropy to calculate the homogeneity of a sample. If the sample is completely homogeneous the entropy is zero and if the sample is an equally divided it has entropy of one. The information gain is based on the decrease in entropy when a dataset is split on an attribute. Constructing a decision tree is all about finding attribute that returns the highest information gain (i.e., the most homogeneous branches).

In 2010 a study by Nair B. B. [28] et al proposed a system based on a genetic algorithm optimized decision tree-support vector machine hybrid, which can predict one-day-ahead trends in stock markets

III. CONCLUSION AND FUTURE SCOPE

This survey paper concludes that though various approaches and techniques are available to increase profit in stock market investment, every methods has its advantages and limitations. Fundamental analysis really helps to find a stock's intrinsic value but it is not much profitable for short term trading. Technical indicators look to predict the future price levels by looking at past patterns and hence useful for long term trading as well as short term trading.SMA smoothen the price movement thus eliminating most fake outs but it also cause a lag in buying and selling signals. EMA reduces the lag by applying more weight to recent prices hence better than SMA in terms of recent movements in the market. Benefit of using RSI is that it immediately indicates the overbought and oversold levels to traders but since the indicator is showing momentum, as long as momentum remains strong (up or down) the indicator can stay in overbought or oversold territory for long periods of time. Therefore, price analysis or some other confirmation is still needed for reversals.The MACD indicator is a so called trend following indicator. With the moving average, a trend can be found, with the MACD, the strength of the trend and the possible turning points can be determined but sometimes it is difficult to find reversal in the market using MACD. Machine learning methods have also their advantages and limitations. Neural network is really an adaptive learning method having well self-organized structure but it sometimes it converges on local minima in optimization problem. Overfitting is another issue with Neural network.Overfitting occurs in complex decision trees too. Hard concept learning could be difficult in decision trees. Learning model parameter is another constraint in decision trees. HMM uses large number of parameters resulting into large amount of data needed to train it. Although SVMs have good generalization performance, they can be abysmally slow in test phase. Though having limitations with every non algorithmic techniques if properly applied we can predict stock market prices at some extents but use of machine learning algorithm have shown better results.We can predict value as well as trend effectively.

ACKNOWLEDGEMENT

A special thanks to Prof. Mrs. AparnaJunnarkar for providing invaluable guidance.

REFERENCES

1. Yu-Feng Lin, Chien-Feng Huang, Vincent S. Tseng, “A Novel Methodology for Stock Investment using Episode Mining and TechnicalIndicators”, 2012 Conference on Technologies and Applications of Artificial Intelligence
2. Abu-Mostafa YS, Atiya AF, “Introduction to financial forecasting”, Applied Intelligence 1996; 6:205–13.
3. S. Nison, “Japanese Candlestick Charting Techniques:A Contemporary Guide to the Ancient InvestmentTechnique of the Far East”, New York Institute ofFinance, 1991
4. Fernández-Blanco, P., Bodas-Sagi, D. J., Soltero, F. J., and Hidalgo, J. I.(2008) Technical market indicators optimization using evolutionaryalgorithms. In: Keijzer (Ed.) Proceedings of the Genetic andEvolutionary Computation Conference (GECCO’2008). New York2008. pp. 1851-1858.
5. Bodas-Sagi, D. J., Fernández, P., Hidalgo, J. I., Soltero, F. J., Risco-Martín, J. L. (2009) Multi-objective optimization of technical marketindicators. In: Proceedings of the 11th Annual Conference on Geneticand Evolutionary Computation. GECCO’09. pp. 1999–2004.
6. John Sneed “Estimating earning forecasting model using fundamental analysis: Controlling for difference across industries” American business review, pp.17-24, 1999
7. Vanstone B., Finnie G. and Tan C., Applyng fundamental analysis and neural network in the Australina stock market” Proceding of international conference on Artificial intelligence in Science and Technology, Hobart, Tasmania, pp.21-25, November 2004.
8. E.I. Altman, “Financial ratios, discriminant analysis and the predictionof corporate bankruptcy,” Journal of Finance, vol. 23, 1968, p 589 609.
9. S.P. Kothari, “Capital markets research in accounting,” Journal ofAccounting and Economics, vol. 31, 2001, pp. 105-231.
10. D. Nissim and S.H. Penman, “Ratio analysis and equity valuation: fromresearch to practice,” Review Of Accounting Studies, vol.6, pp.109-154.
11. Yuh-Jen Chen and Yuh-Min Chen, “A Fundamental Analysis-based Method for Stock Market Forecasting”, 2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP)June 9 – 11, 2013, Beijing, China
12. ChingHsue Cheng, You-shyang Chen, “Fundamental analysis of stock trading system using classification techniques” Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 19-22 August 2007
13. Alexander S. (1961), Price movements in speculative markets: trends orrandom walks, Industrial Management Review, (2), pp. 7–26.
14. Fama E. F. and Blume M. E. (1966) Filter Rules and Stock-MarketTrading, Journal of Business, 39 (1), Part2: Supplement on Security Pricing, pp. 226–241.
15. Fernández-Blanco, P., Bodas-Sagi, D. J., Soltero, F. J., and Hidalgo, J. I.(2008) Technical market indicators optimization using evolutionaryalgorithms. In: Keijzer (Ed.) Proceedings of the Genetic andEvolutionary Computation Conference (GECCO’2008). New York2008. pp. 1851-1858.
16. Bodas-Sagi, D. J., Fernández, P., Hidalgo, J. I., Soltero, F. J., Risco-Martín, J. L. (2009) Multi-objective optimization of technical marketindicators. In: Proceedings of the 11th Annual Conference on Geneticand Evolutionary Computation. GECCO’09. pp. 1999–2004.
17. Yu-Feng Lin, Chien-Feng Huang, Vincent S. Tseng, “A Novel Methodology for Stock Investment using Episode Mining and Technical” 2012 Conference on Technologies and Applications of Artificial Intelligence.
18. YauheniyaShynkevich, T.M. McGinnity, Sonya Coleman, Yuhua Li, AmmarBelatreche, “Forecasting stock price directional movements usingtechnical indicators: investigating window size effects on one-step-ahead forecasting”
19. J. Welles Wilder, “New Concepts in Technical Trading Systems”, ISBN0-89459-027-8, 1978.
20. Jianxue Chen, “SVM application of financial time series forecasting using empirical technical indicators”, 201O International Conference on Information, Networking and Automation (ICINA)
21. D. Moldovan, M. Moca, and S. I. Nitchi, “A stock trading algorithm model proposal, based on technical indicators signals,” InformaticaEconomica, vol. 15, no. 1, pp. 183–188, 2011.
22. W. Brock, J. Lakonishok, and B. LeBaron, “Simple technical trading rules and the stochastic properties of stock returns,” The Journal of Finance, vol. 47, no. 5, pp. 1731–1764, 1992.
23. Aparna Bhatt, SowmyaKamath S., “Automated Stock Price Prediction and TradingFramework for Nifty Intraday Trading”, IEEE - 31661

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

24. X. Wang, P. K. H. Phua, and W. Lin, "Stock market prediction usingneural networks: does trading volume help in short-term prediction" in Neural Networks, 2003. Proceedings of the International JointConference on, vol. 4. IEEE, 2003, pp. 2438–2442.
25. A. U. Khan et al., "Stock rate prediction using back propagationalgorithm: Analyzing the prediction accuracy with different number ofhidden layers," Glow gift, Bhopal, 2005.
26. P.-F. Pai and C.-S. Lin, "A hybrid arima and support vector machinesmodel in stock price forecasting," Omega, vol. 33, no. 6, pp. 497 505,2005.
27. Luigi Troiano, PraveshKriplani, "Predicting Trend in the Next-Day Market by Hierarchical Hidden Markov Model", 978-1-4244-7818-7/10/$26.00_c 2010 IEEE
28. Binoy B. Nair et. aI., 2010, "A Genetic Algorithm Optimized DecisionTree- SVM based Stock Market Trend Prediction System" International Journal on Computer Science and Engineering Vol. 02, No. 09, 2010, 2981-2988

BIOGRAPHY

Milind Kolambe is a Research Scholar in Computer Department, PESMCOE, Pune. He is also working as an assistant professor for department of information technology, Cummins college of engineering for women, pune since 2008. His research interests are Data mining,Software Engineering etc.