A Study of Soft Computing Models for Stock Market Forecasting

Synopsis of the Thesis submitted to **Madurai Kamaraj University**

in partial fulfilment of the requirements for the award of the Degree of

DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE

Submitted by

J SHARMILA VAIZ

(Registration No. **P4234**)

Under the Guidance of

Dr. M.RAMASWAMI

Professor

Department of Computer Applications

School of Information Technology

Madurai Kamaraj University

DEPARTMENT OF COMPUTER APPLICATIONS

MADURAI KAMARAJ UNIVERSITY

Madurai - 625 021

Tamilnadu, India

D E C E M B E R – 2019

A Study of Soft Computing Models for Stock Market Forecasting

1. Introduction

"I will tell you how to become rich. Close the doors. Be fearful when others are greedy. Be greedy when others are fearful." –Warren Buffet.

The words of successful traders and investors not only give us inspiration but also guide us on how to be on the right path and avoid mistakes. Quotes of Warren Buffet have always been golden words for investors.

Over the last few decades, the average person's interest in the Stock Market has grown exponentially. Stock investing became all the rage during the late 1990s. The stock market plays a pivotal role in the growth of the industry and commerce and it affects the economy of the country to a greater extent. Shares are the most glamorous investment option over the long term and they offer the highest returns. There is risk involved in stock trading, but no other investment such as savings account, bond holds as much as potential as stocks over long run. Good stocks may be discovered that meet all these test and it can be added to portfolio to get good yield.

Stock market analysis is the evaluation of a particular trading instrument in an investment sector or the market as a whole. Stock analysis can be lumped into two major categories first one is fundamental analysis and the other one is technical analysis. Fundamental Analysis of a company is a study of basic fundamentals of a company's performance that can decide its future as well as growth rate. Factors that are considered in fundamental analysis are earnings per share, price-earnings ratio, dividend yield, etc. Technical analysis is a methodology for forecasting the stock trend based on the past market data, primarily price and volume. Stock trend prediction is one of the most widely studied and challenging problems, attracting researchers from many fields including economics, history, finance, mathematics, and computer science. Ups and downs in the market is compared with two big animals the ***bull*** and the ***bear***. Trend lines are used to identify trends which connect a series of

highs or lows. Technical indicators are powerful tools that are used in technical analysis that turn patterns into actionable trading plans. In this study, stock trend was forecasted with different machine learning algorithms such as ID3, C50 and RPART Decision tree algorithms, Naïve Bayesian, K Nearest Neighbor, Support Vector Machine, Neural Networks, Generalized Linear Model and ensemble learning algorithms bagging, boosting and stacking.

2. Literature Review

Senthamarai et al., (2010) [1] used five methods Bollinger Bands(BBands), Relative Strength Index(RSI), Chaikin Money Flow(CMF), Stochastic Momentum Index(SMI) and Moving Average(MA) to predict the stock direction for a sample of 400 signals. Their experimental study reveals that MA produced 60% profitable signals, CMI produces 51.45% and the profit percentage of RSI is 56.04%. They concluded that SMI and BBands produce more profitable signals than the other three indicators.

Chitra(2011) [2] in her study analyzed the price movements of 10 companies of NSE using certain computational measures like Beta calculation, Relative Strength Index(RSI) and Simple Moving average(SMA).

Vasantha et al.,(2012) [3] used four enriched technical indicators such as Relative Strength Index(RSI), Bollinger Bands(BBands), Moving Average Convergence Divergence(MACD) and Simple Moving Average(SMA) to make a decision on whether to buy or sell the stocks of the IT sector stocks such as Tata Consultancy Services (TCS), Hindustan Computers Limited Technologies (HCL), Infosys, Wipro and Polaris.

Karuna Dhutti(2014) [4] in his study discussed about the role of the technical indicators like Relative Stock Index(RSI), coefficient of variation, beta and chart patterns elaborately to take investment decisions to buy or sell the stocks of five IT Companies.

Boobalan(2014) [5] in his study used weekly share price movements of the companies 1) WIPRO 2) SBIN 3) GAIL 4) ONGC 5) ITC to predict the future trend of the companies. The major tools and techniques used in this study are Stock (Candlestick) chart, Exponential moving average (EMA), Moving average convergence divergence (MACD) and Relative strength index (RSI).

Pushpa et al.,(2017) [6] in their study used technical analysis to analyze the stock movements of nine companies of NSE 50 index from different sectors such as Adani ports, Ambuja Cements, ITC, Aurobindo Pharma, BPCL, HeroMotorCorporatition, TCS, HDFC, Zee Entertainment. Moving average(MA), Relative Strength Index(RSI), Bollinger Bands(BBands) and Moving Average Convergence/Divergence(MACD) were used as technical tools.

Qasem et al., (2013) [7] studied the effect of decision tree based classification model of **ID3 and C4.5** with six response variables attributes viz; Previous Close, Open, Min, Max, Last and a predictor variable Action(Buy/Sell). Their findings showed the classifier accuracy of these models were in between 44% - 54%.

Sadegh et al., (2014) [8] attempted to develop three forecasting models **Decision tree, Random forest and Naïve Bayesian Classifier** and compared their performance in predicting stock price movement direction in Tehran Stock Exchange (TSE) Index with ten microeconomic 10-day SMA, 10-day WMA, momentum, stochastic %K, stochastic %D, RSI, MACD, WPR, A/D Oscillator and CCI and three macroeconomic variables oil, gold and USD/INR. The experimental results showed that Decision tree model (80.08%) surpass other two models Random Forest (78.8%) and Naïve Bayesian Classifier (73.8%).

Lamartine et al(2010) [9] proposed a method for automatic stock trading which combines technical analysis and **nearest neighbor** classification algorithm. Stop loss, stop gain and RSI filter were used in technical analysis tool. The proposed method gives higher profitability and reduces the risk of market exposure.

Saahil Madge (2015) [10] in his study used **Support Vector Machine** to predict the stock trend of 34 technology stocks. Price volatility and momentum for individual stocks and for the overall sector was derived from closing prices of stocks. These variables were used as parameters to the SVM model. His model predicted whether a stock price sometime in the future will be higher or lower than it is on a given day. `

Xinjie Di(2014) [11] forecasted stock trend of 3 companies Apple, Amazon and Microsoft with technical indicators WPR, ROCR, MOM, RSI, CCI, ADX, TRIX, MACD, OBV, TSF, ATR, MFI. Extremely Randomized tree was used for feature selection and subsequently a model is fitted with **SVM using Radial Bias Function (RBF) kernel** and gained predictive accuracy of above 70% for these 3 stocks.

Vishal and Parth(2015) [12] have attempted to develop models for predicting the stock trend using **naïve bayes and random forest classification** technique for BSE-200 top 10 gainers and BSE-200 top 10 losers and they used technical indicators RSI, MACD, Stochastic, WPR, Bollinger Bands, MFI, CCI, PROC, OBV as features.

Jacinta et al. (2016) [13] enhanced profitability of trading rules by the use of **neural networks** on the Kuala Lumpur Composite Index (KLCI) with actual close (Ct) and previous close (Ct-1), MA, RSI, Momentum (M) and Stochastic %K and %D as features with the output of the next predicted close. The profitable returns on KLCI from January 2008 to December 2014 with technical trading rules using neural networks out-perform the buy-and-hold threshold benchmark.

Khan et al. (2016) [14] predicted stock trend using nine attributes Date, Open, High, Low, Close, Volume, Trend, Sentiment and Future Trend Value with three machine learning algorithms **KNN, SVM and NAÏVE BAYES** and for the exchanges Karachi, New York and London Stock exchanges before and after feature selection using Principal Component Analysis (PCA). Their experimental study showed that KNN yields an accuracy of 67% before PCA and 68% after PCA in

predicting LSE, SVM provided 55% accuracy for NSE after PCA and an accuracy of 45% was observed using NB when predicting LSE.

Shubharthi et al.(2016) [15] applied feature reduction with PCA and **Extreme Gradient Boosting** is used to predict the stock trend with technical indicators RSI, Stochastic, WPR, MACD, PROC and OBV values which were derived from exponentially smoothed historical data.

3. Objectives of the Research

Based on the survey of the research work carried out by various authors, it has been observed that limited technical indicators alone were used in their research study and the role of other reliable stock market technical indicators are not discussed so for. In order to enhance visibility of research study twenty two technical indicators were induced and its efficacy was analyzed with different machine learning algorithms. It also aims to construct a stock trend forecasting model that generates trading signals.

The study aims,

- To get a clear picture about the role of technical indicators in predicting the stock trend.
- To analyse the efficacy of other reliable technical indicators in predicting the stock trend which were not used by the majority of the researchers
- Whether feature selection techniques can be used to decrease the complexity of learning and increase the optimality of the classification algorithm accuracy.
- Different machine learning algorithms such as Iterative Dichotomiser, Recursive Partitioning, C50, Naïve Bayesian, Support Vector Machine, K nearest neighbour, Generalized Linear Model, Neural Networks were analysed in predicting the stock trend.
- To enhance the predictability, the use of the significance of an ensemble learning methods such as bagging, boosting and stacking

were also experimented with three different algorithms such as Random Forest, Gradient Boosting Machine and Stacking. The fruitfulness of ensemble learning methods were analysed and evaluated with several evaluation measures.

- To construct a stock trend forecasting model with ensemble stack method to predict the direction of the stock movement.

The overall objective of this research is to suggest the investors in making trading decision (buy/sell) and thereby they can increase their investment returns. Investors can make use of ensemble stock trend forecasting model to get suggestions about the trading signal.

4. Structure of the Research

The research work is organized as nine chapters

Chapter 2 give the gist of the research work carried out by several researchers in predicting the stock direction. An overview of the technical indicators used by various researchers and the types of learning algorithms used by them are discussed. Several researchers used 4 to 14 technical indicators in their study to predict the stock trend using algorithms such as Decision Tree, Neural Network, Support Vector Machine, K Nearest Neighbor etc.,

Chapter 3 elaborates the characteristics of data set used in this study and it is organized as three sections. In the first section the role of technical indicators in stock trend prediction are observed with the help of chart patterns, in the second section machine learning algorithms which were used in this study are discussed and in the last section the evaluation measures which were used in the study to check the optimality of the prediction models are expounded.

In Chapter 4 machine learning process with basic prediction models such as Decision Trees, Naïve Bayesian, K Nearest Neighbor, Support Vector Machine, Artificial Neural Network and Generalized Linear Model has been explored in predicting the stock movement. Stock trend was predicted for six companies of NSE

and the efficacy of each of the algorithms was observed with the help of evaluation measures Accuracy, F1-measure, Kappa, ROC and AUC.

Chapter 5 summarizes the list of feature selection techniques used in the research and the impact of it was flashed using several evaluation measures. An attempt was made to find relevant features using twelve feature selection techniques such as chi-square, info gain, gain ratio, forward search, backward search, best first search, hill climbing search, linear correlation, rank correlation, oneR, random forest importance, relief and a comparative study was made to find the impact of feature selection techniques using classification algorithms. Principal Component Analysis feature selection was also experimented and an experimental study was carried out with relevant features by using machine learning algorithms in predicting the direction of the market. A comparative study was also made to check the efficacy of the feature selection techniques using evaluation metrics.

Chapter 6 investigates deep learning using Deep Neural Networks with six different activation functions in predicting the stock trend. Efficacy of the activation functions in increasing the optimality of the classification algorithm was observed on experimentation while predicting the stock trend of six companies of NSE.

Chapter 7 analyzes hybrid model that combines SVM and NNET to predict the stock trend. SVM was used to retrieve relevant features and NNET was used as classification algorithm to predict the stock trend. Efficacy of feature selection was analyzed by comparing the prediction accuracy of NNET before and after feature selection.

Chapter 8 discusses the three types of ensemble learning algorithms bagging, boosting and stacking. Random forest was used for bagging algorithm and Gradient Boosting Machine was chosen for implementing boosting algorithm. A comparative study was made with basic prediction models, bagging and boosting. Finally, an ensemble stock trend forecasting model (ESFTM) was constructed with a method called stacking to predict the stock trend. The optimality of the ensemble learning model was observed with the help of evaluation metrics. A comparative

study was made by combining all the individual classification algorithms with stacking to highlight the efficacy of the stacking method.

Chapter 9 describes flow of the research from the beginning to the end and discusses about the results arrived at various stages of the research.

Chapter 10 gives the conclusion of the result, limitation and the future modification that can be added to the current research.

5. Stock Trend Forecasting using Soft Computing Models

To analyze the stock movement six stocks from different sectors were chosen. Stocks that were chosen for analysis are listed below. Stock trend forecasting of six different companies of NSE has been carried out in phases.

i) Tata Consultancy service (TCS) - IT Sector
ii) Reliance industries (RIL) – Communication Sector
iii) Housing Development Finance Corporation (HDFC) – Banking Sector
iv) Hindustan Unilever Limited (HUL) – Personal Care
v) Sun Pharmaceutical Limited (SPIL) – Pharmaceutical Sector
vi) Imperial Tobacco Company Limited(ITCL) – Cigarette Sector

5.1 Analysis of Basic Classifier Models in Forecasting Stock Trend

ID3, C50, RPART, Naïve Bayesian, K Nearest Neighbor, Support Vector Machine, Neural Networks, Generalized Linear models were used to forecast the stock trend of six companies of NSE and a comparative analysis of the evaluation measure accuracy for TCS is depicted in the Figure. 1.

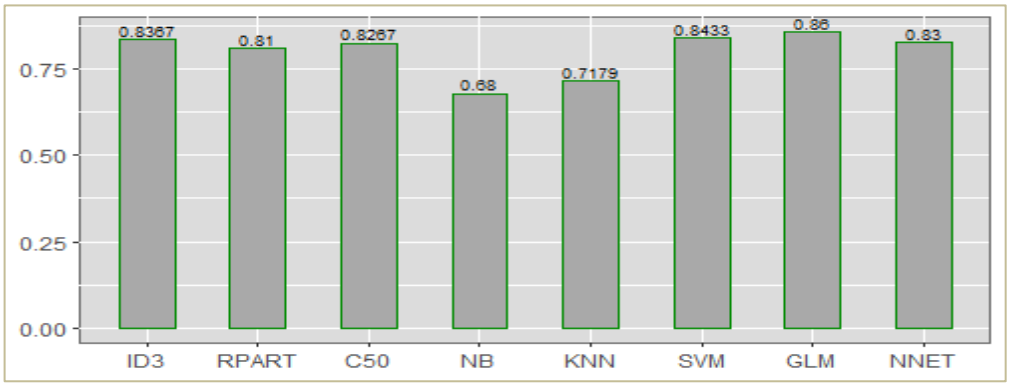

Figure 1: Prediction models accuracy for TCS

5.2 Analysis of Feature Selection Techniques

To heighten the prediction accuracy an analysis was carried out with feature selection techniques. The study was initiated with twelve feature selection techniques such as Chi-square, Information Gain, Forward Search, Best First, Hill Climbing, Linear Correlation, Rank Correlation, OneR, Random forest importance, RRelief filter with the help of FSELECTOR package in R. With each of the feature selection techniques different feature set was selected based on its principle and the stock trend prediction process was iterated with the relevant attributes selected. The performance of NB classification was lesser than other classification models such as decision trees, SVM, etc. Hence, we reiterate the process of stock trend forecast with NB classification for the reduced dataset derived from the twelve feature selection techniques. For the stock HDFC a 7% hike in accuracy was observed using Forward Search and RIL increases by 3% using NB classification. With other features selection techniques no variation or increase in performance was observed. The process with ID3 classification was reiterated with reduced feature set and an increase of 5% prediction accuracy was observed for HDFC. Linear correlation feature selection technique decreases the performance of the model by 21% for RIL when trained with ID3 classification model. As a second step in feature selection techniques principal component analysis was implemented and observed the performance of the models with relevant attributes retrieved from PCA for TCS. Bi-plot of 2 PCA components is depicted in Figure 2. Experimental study revealed that there was a drastic decrease in accuracy with PCA selected feature set. ID3 accuracy was decreased from 82% to 42% and NB classification accuracy was reduced from 72% to 47%.

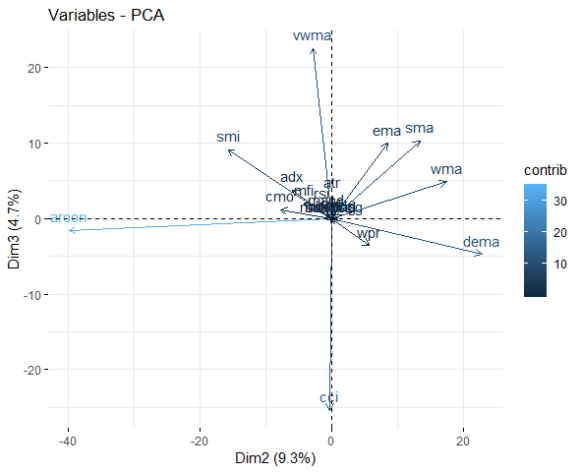

Figure 2: Bi plot of PCA2 and PCA3

5.3 Analysis of Deep Learning with Different Activation Functions in Forecasting Stock Trend

In the second phase, experimental analysis was implemented with DNN using six different activation functions such as TAN, TANWD, MAX, MAXWD, REC and RECWD for the six stocks of NSE to predict the stock trend. The evaluation measures of six different activation functions for the company TCS is shown in Table 1. Among the six activation functions REC gives an accuracy of 87.16% for TCS. F1 measure for TCS stock prediction is high with 0.90 using REC function and AUC values using MAX activation function was high with the value of 0.9353.

Table 1: Evaluation measures of activation functions

Symbol	Activation	Accuracy	F-Measure	AUC
TCS	TAN	0.8328	0.8747	0.9140
	TANWD	0.8000	0.8553	0.8272
	MAX	0.8597	0.8834	**0.9353**
	MAXWD	0.8507	0.8809	0.9156
	REC	**0.8716**	**0.9005**	0.9252
	RECWD	0.8388	0.8726	0.9137

5.4 Analysis of Hybrid Models in Stock Trend Forecasting

A hybrid stock trend forecasting model was constructed using SVM and NNET to analyze stock trend prediction performance. SVM was used to retrieve relevant attributes from the dataset and NNET was used for classify the stock trend from the reduced data set. Initially, the model is trained with NNET classification and stock trend was predicted for the test instances and then NNET classification was performed with the reduced dataset retrieved from SVM feature selection. The classification accuracy of NNET before and after feature selection is shown in Table 2 for six companies of NSE.

Table 2: NNET and SVM-NNET model evaluation measures

Symbol	NNET			Hybrid(SVM & NNET)		
	Accuracy	F-measure	AUC	Accuracy	F-measure	AUC
TCS	0.8167	0.7909	0.9274	0.8600	0.8250	0.9416
RIL	0.8833	0.8679	0.959	0.9000	0.8888	0.9638
HDFC	0.8300	0.8061	0.9245	0.8667	0.8571	0.9499
HUL	0.7800	0.7626	0.8878	0.8533	0.8321	0.9360
SPIL	0.8333	0.7899	0.8829	0.8333	0.7965	0.9323
ITC	0.8667	0.8601	0.9413	0.8663	0.8530	0.9528

Experimental study reveals that there is 4% increase for TCS, 2% for RIL, 3% for HDFC, 7% for HUL using hybrid classification models. Two stocks SPIL and TCS showed no variation in applying hybrid model learning. However, no drastic decrease in performance was observed using SVM feature selection techniques.

5.5 Comparative analysis of basic prediction models and ensemble models

Upon experimental observations of basic prediction models, deep learning models and hybrid models the study moved on to ensemble learning methods bagging, boosting and stacking. Random Forest algorithm and Gradient Boosting was used to implement bagging and boosting. Experimental study was carried out with random forest and gradient boosting to predict stock

trend and the optimality of the algorithms were revealed with evaluation metrics. A comparative plot of basic prediction models and ensemble prediction models is shown in Figure 3.

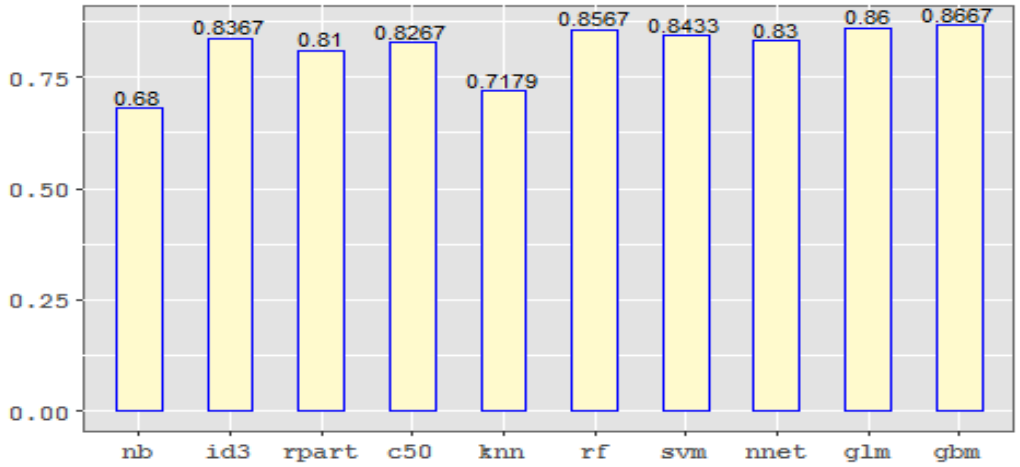

Figure 3: Comparative plot for different classification algorithm for TCS

5.6 Construction of Ensemble Stock Trend Forecasting Model

Steps involved in construction of ESTFM

- Data preparation
- Preprocess the Data
- Build ESTFM model with individual classification algorithms
- Remove relevant features using SVM Feature Selection Techniques
- Build the model with relevant features using individual classification algorithms
- Stack the predictions of individual classification algorithms
- Build the model with meta-learning algorithms
- Predict the stock trend using meta-learning algorithm
- Reiterate meta-learning algorithms
- Generate buy and sell signals using ESTFM

The experimental study reveals that stacked ensemble model yields nearer or above 90% prediction accuracy. The optimality of the classification is more when compared to individual classification model. Investors can increase their portfolio with stacked ensemble model.

6 Conclusion

From the research study, we found that ensemble learning methods like bagging, boosting can be used in stock trend prediction and can gain above 80% accuracy. Before training the model with learning algorithms, features can be extracted using feature selection techniques to reduce the time consumption and the clumsiness of the model. Stacked ensemble method can be used with high confidence which aggregates all the classification algorithms predictions. In this study we analyzed two methods, in the base learner phase the model was trained with ten different algorithms, in the meta learning phase the model was built with the prediction outcome of the all the algorithms from base learner phase and stock trend was predicted for test instances. Meta learning process was also reiterated with ten different algorithms. In the second method, major voting method was used to select trend based on the majority of the predictions from the ten different classification functions. A sample of buy/sell signals were also generated based on major voting method. Investors can frame set of rules to exit based on the percentage of increase or decrease when entered into buy or sell action. As the experimental observations of the present study yielded about 90% of prediction accuracy in predicting the type of the market using stacked ensemble method, the investors can make use of stacking method with high confidence to swell their investments.

References

1) Senthamarai Kannan, Sailapathi, Mohammed and Arumugam. (2010). "Financial Stock Market Forecast using Data Mining Techniques", 2010. *Proceedings of the international multiconference of engineers and computer scientists.*

2) Chitra R. (2011). "Technical Analysis on Selected Stocks of Energy Sector". *International Journal of Management &Business Studies(ijmbs), Volume 1*(Issue 1).

3) Vasantha S, Dhanraj V and Varadharajan R. (2012). "Stock price movement through technical analysis: Empirical evidence from the I.T sector". *Indian Journal of Finance , Volume 6* (No 10), pp 4-17.

4) Karuna Dhutti. (2014). "Stock Price Movement of Information Technology Sector through Technical Analysis". *Journal of Academia and Industrial Research (JAIR) , Volume 2* (Issue 9).

5) Boobalan C. (2014). "Technical Analysis In Select Stocks Of Indian Companies". *International Journal of Business and Administration Research Review , Volume 2 (Issue 4).*

6) Pushpa B, Sumithra C G and Madhuri Hegde. (2017). " Investment Decision Making Using Technical Analysis: A Study on Selected Stocks in Indian Stock Market",. *IOSR Journal of Business and Management (IOSR-JBM), e-ISSN: 2278-487X, p-ISSN: 2319-7668. Volume 19, Issue 9. Ver. VI. (September. 2017), PP 24-33 .*

7) Qasem A. Al-Radaideh, Adel Abu Assaf and Eman Alnagi. (2013). " Predicting Stock Prices Using Data Mining Techniques". *The International Arab Conference On Information Technology (ACIT'2013).*

8) Sadegh Bafandeh Imandoust and Mohammed Bolandraftar. (2014). " Forecasting The Direction Of Stock Market Index Movement Using Three Data Mining Techniques: The Case Of Tehran Stock Exchange". *Int. Journal Of Engineering Research And Applications, ISSN : 2248-9622, Vol. 4, Issue 6(Version 2), June 2014, pp.106-117.*

9) Lamartine Almeida Teixeira A and Adriano Lorena Inacio De Oliveira. (2010). " Method For Automatic Stock Trading Combining Technical Analysis And Nearest Neighbor Classification". *Expert Systems With Applications 37 (2010) 6885–6890 .*

10) Saahil Madge. (2015). "Predicting Stock Price Direction Using Support Vector Machines" . *Independent Work Report Spring 2015*.

11) Xinjie D. (2014). "Stock Trend Prediction with Technical Indicators using SVM". *Stanford University, 2014* .

12) Vishal Parikh and Parth Shah. (2015). "Stock Prediction and Automated Trading System", *IJCSC 0973-7391, Vol- 6 • Issue-1 Sep - Mar 2015 pp.104-111, www.csjournalss.com*.

13) Jacinta Chan Phooi M'ng and Azmin Azliza Aziz. (2016). "Using Neural Networks To Enhance Technical Trading Rule Returns": *A Case With KLCI, Athens Journal Of Business And Economics.*

14) Khan Ghazanfar. (2016). " Predicting Trend In Stock Market Exchange Using Machine Learning Classifiers", *Sci.Int.(Lahore),28(2),1363-1367,2016, ISSN 1013-5316; CODEN: SINTE 8.*

15) Shubharthi Dey, Yash Kumar, Saha S and Basak S. "Forecasting To Classification: Predicting The Direction Of Stock Market Price Using Xtreme Gradient Boosting". *Working Paper. DOI: 10.13140/RG.2.2.15294.48968.*

List of Publications

Sharmila Vaiz J and Ramaswami M (2016), Forecasting Stock Trend Using Technical Indicators with R, International Journal of Computational Intelligence and Informatics, Vol. 6: No. 3, December 2016, ISSN: 2349-6363

Sharmila Vaiz J and Ramaswami M (2016), "A Study on Technical Indicators in Stock Price Movement Prediction Using Decision Tree Algorithms, American Journal of Engineering Research (AJER)", UGC Approved Journal - e-ISSN: 2320-087, pISSN No: 2320-0936.

Sharmila Vaiz J and Ramaswami M (2017), A Hybrid Model of Stock Trend Forecasting Using Support Vector Machine and Neural Networks, International Journal of Engineering Research and Development, UGC Approved Journal - e-ISSN : 2278-067X, p-ISSN : 2278-800X.

Sharmila Vaiz J, Dhamayanthi B and Ramaswami M (2017). "A Study of Deep Neural Networks in Stock Trend Prediction using Different Activation Functions", International Conference on Recent Trends in Engineering, Computers, Information Technology and Applications (ICRTECITA-2017), PSNA College of Engineering and Technology, Dindigul

Sharmila Vaiz J and Ramaswami M (2018), "Stacked Ensemble Model in Predicting Stock Trend of Three Major Capitalization Companies of NSE**",** Journal of Emerging Technologies and Innovative Research , (An International Open Access Journal & UGC and ISSN Approved, Peer-reviewed, Refereed Journals) © 2018 JETIR May 2018, Volume 5, Issue 5, (ISSN-2349-5162)