



## Hybrid Model based on unification of Technical Analysis and Sentiment Analysis for Stock Price Prediction

S. B. Wankhade, Divyesh Surana, Neel Mansatta, Karan Shah

Asst. Professor, Computer Engineering Department, Rajiv Gandhi Institute of Technology, Mumbai, India

[sunil.wankhade9@yahoo.com](mailto:sunil.wankhade9@yahoo.com)

Computer Engineering Department, Rajiv Gandhi Institute of Technology, Mumbai, India

[divyesh\\_surana@yahoo.com](mailto:divyesh_surana@yahoo.com)

Computer Engineering Department, Rajiv Gandhi Institute of Technology, Mumbai, India

[neelmansata@gmail.com](mailto:neelmansata@gmail.com)

Shah and Anchor Kuttchi College of Engineering, Information Technology Department, Mumbai, India

[karan.pace@gmail.com](mailto:karan.pace@gmail.com)

### ABSTRACT

Stock price forecasting phenomenon has been majorly made on the basis of quantitative information. Over the time, with the advent of technology, stock forecasting used technical analysis to get more accurate predictions. Until recently, studies have demonstrated that sentiment information hidden in corporate reports can be effectively incorporated to predict short-run stock price returns. Soft computing methods, like neural networks, fuzzy models and support vector regression, have shown great results in the forecasting of stock price due to their ability to model complex non-linear systems.

In this paper we propose a hybrid method for stock price prediction, which is combinational feature from technical analysis and sentiment analysis (SA). The features of sentiment analysis are based on a Point wise Mutual Information (PMI) and we apply neural network and  $\epsilon$ -support vector regression models to predict the yearly change in the stock price.

### Indexing terms/Keywords

Stock price; forecasting; prediction; sentiment analysis; corporate report; neural networks;  $\epsilon$ -support vector regression; Pointwise Mutual Information; Technical Analysis.

### Academic Discipline And Sub-Disciplines

Computer Science, Artificial Neural Networks

### TYPE (METHOD/APPROACH)

Theoretical Research and Hypothesis

---

# Council for Innovative Research

Peer Review Research Publishing System

Journal: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Vol 11, No. 9

[editor@cirworld.com](mailto:editor@cirworld.com)

[www.cirworld.com](http://www.cirworld.com), [member.cirworld.com](http://member.cirworld.com)



## INTRODUCTION

The ability to predict stock market behavior has always had a certain appeal to researchers. Even though numerous attempts have been made, the difficulty has been the inability to capitalize on the behaviors of human traders. Behavioral patterns have not been fully defined and are constantly changing, thus making accurate predictions quite difficult. Previous literature has shown that the problem of stock price forecasting has to be taken as complex since stock price changes in time are highly nonlinear with a changing volatility and many micro and macroeconomic determinants.

With the advent of cheap computing and the ease of gathering information, the role of computers in stock prediction has increased dramatically. In general, Technical

Analysis (TA) is based on historical developed regularities in the stock exchange with an assumption that the same result will repeat in the future. There are many influential indicators and trading rules based on them. Technical indicators might provide advice to traders on whether a trend will continue, such as MACD, or whether a stock is oversold or overbought, such as BIAS. One of the important issues for forecasting market trend is to know sentiment of stock news, that it's good or bad trend, when the financial stock prices go through the up/down cycle. The Sentimental Analysis (SA) can be applied to make trading decisions where some of potentially important information affecting investor's decisions is the news. Sentimental Analysis can use text mining technique to find best information. SA has been developed and good performance in many researches. The sentimental analysis is a different way to mining stock information compares to uses the trading information to predict future stock trends. In addition, many researchers using textual information to improved prediction performance. However, these approaches have a problem that textual data is highly complex information representation, whether using dictionary or manual lexicon by analyzer may miss many of distinctive features. Therefore, the feature extraction can capture more effective variables as information to improved classification or predication problem, such as using SentiWordNet, Association Rule Mining (ARM), Pointwise Mutual Information (PMI) and Mutual Information (MI).

In this paper we will demonstrate that the long run behavior of stock price can be effectively predicted employing NNs and  $\epsilon$ -SVRs. To overcome the complexity of information representation, we employ PMI model in the sentimental analysis part of the prediction. Therefore, we develop a hybrid model that combines quantitative input variables (mostly fundamental analysis indicators) with qualitative sentiment from corporate reports using PMI model. Then, NNs and  $\epsilon$ -SVRs are used to perform a one year ahead stock return forecast.

This paper is arranged as follows:

Section 2 provides an overview of literature concerning stock market prediction, textual representations and sentiment analysis techniques.

Sections 3 and 4 describe our proposed approaches, where we first show a macroscopic view of our proposed model and then further expands the components of the model to explain the sequence of execution. With the methodology of PMI followed by NNs and  $\epsilon$ -SVRs; explained precisely with the block diagram for the model.

Section 5 provides an overview of our experimental design.

Section 6 delivers our conclusions and a brief discourse on future research directions.

## LITERATURE REVIEW

Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on a financial exchange. The successful prediction of a stock's future price could yield significant profit. The ability to predict stock market behavior has always had a certain appeal to researchers [1]. Even though numerous attempts have been made, the difficulty has been the inability to capitalize on the behaviors of human traders. Behavioral patterns have not been fully defined and are constantly changing, thus making accurate predictions quite difficult.

Sentiment analysis has several subtasks, all of them concern with tagging a given text according to expressed opinion. Work has been done in the field of sentiment analysis by professionals as well as students in order to understand emotions of people, to know about movie or any product review etc. They have mainly been implemented using either WordNet or SentiWordNet lexicon [2]. Lexicon is essentially a catalogue of a given language's words and grammar, a system of rules which allows for the combination of those words into meaningful sentences. Several research teams in universities around the world have focused on understanding the dynamics of sentiment in e-communities through sentiment analysis.

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process [3]. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyse. This expert can then be used to provide projections given new situations of interest and answer "what if" questions.



The PMI of a pair of outcomes  $x$  and  $y$  belonging to discrete random variables  $X$  and  $Y$  quantifies the discrepancy between the probability of their coincidence given their joint distribution and their individual distributions, assuming independence. Mathematically:

$$\text{pmi}(x; y) \equiv \log \frac{p(x, y)}{p(x)p(y)} = \log \frac{p(x|y)}{p(x)} = \log \frac{p(y|x)}{p(y)}.$$

The mutual information (MI) of the random variables  $X$  and  $Y$  is the expected value of the PMI over all possible outcomes (w.r.t. the joint distribution  $p(x, y)$ ).

The measure is symmetric ( $\text{pmi}(x; y) = \text{pmi}(y; x)$ ). It can take positive or negative values, but is zero if  $X$  and  $Y$  are independent. PMI maximizes when  $X$  and  $Y$  are perfectly associated, yielding the following bounds:

$$-\infty \leq \text{pmi}(x; y) \leq \min [-\log p(x), -\log p(y)].$$

Finally,  $\text{pmi}(x; y)$  will increase if  $p(x|y)$  is fixed but  $p(x)$  decreases.

PMI is a measure of association between a feature (in your case a word) and a class (category), not between a document (tweet) and a category [34]. In that formula,  $X$  is the random variable that models the occurrence of a word, and  $Y$  models the occurrence of a class. For a given word  $x$  and a given class  $y$ , you can use PMI to decide if a feature is informative or not, and you can do feature selection on that basis. Having less features often improves the performance of your classification algorithm and speeds it up considerably. The classification step, however, is separate- PMI only helps you select better features to feed into your learning algorithm.

A Support Vector Machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks [12]. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space. To keep the computational load reasonable, the mappings used by SVM schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function  $K(x, y)$  selected to suit the problem. The hyperplanes in the higher-dimensional space are defined as the set of points whose dot product with a vector in that space is constant. The vectors defining the hyperplanes can be chosen to be linear combinations with parameters  $\alpha_i$  of images of feature vectors that occur in the data base. With this choice of a hyperplane, the points  $x$  in the feature space that are mapped into the hyperplane are defined by the relation:  $\sum_i \alpha_i K(x_i, x) = \text{constant}$ . Note that if  $K(x, y)$  becomes small as  $y$  grows further away from  $x$ , each term in the sum measures the degree of closeness of the test point  $x$  to the corresponding data base point  $x_i$ . In this way, the sum of kernels above can be used to measure the relative nearness of each test point to the data points originating in one or the other of the sets to be discriminated. Note the fact that the set of points  $x$  mapped into any hyperplane can be quite convoluted as a result, allowing much more complex discrimination between sets which are not convex at all in the original space.

Technical Analysis (TA) is based on historical developed regularities in the stock exchange with an assumption that the same result will repeat in the future. There are many influential indicators and trading rules based on them. Technical indicators might provide advice to traders on whether a trend will continue, such as MACD, or whether a stock is oversold or overbought, such as BIAS. One of the important issues for forecasting market trend is to know sentiment of stock news, that it's good or bad trend, when the financial stock prices go through the up/down cycle.

Several authors have attempted to analyze the stock market. They used quantitative and qualitative information on the net for predicting the movement of the stock.

Vivek John George et. al [1] suggested a new method for automatically predicting the stock price. They have shown that stock prices predicted from historical prices and sentiments are significantly correlated with actual stock prices of a particular company.

Caslav Bozic and Detlef Seese [2] proposed a system for quantifying text sentiment based on Neural Networks predictor. By using the methodology from empirical finance, they proved statistically significant relation between text sentiment of published news and future daily returns.

Liang, X. [3] work used only volume of posted internet stock news to train neural network and predict changes in stock prices.

Liang and Chen [4] employed natural language processing techniques and handcrafted dictionary to predict stock returns. They used feed forward neural network with five neurons in the input layer, 27 in the hidden layer, and one output neuron. Since only 500 news items was used for the analysis, no statistical significance of the results could be found.



Vivek Sehgal and Charles Song [5] proposed a system learns the correlation between the sentiments and the stock values. The learned model can then be used to make future predictions about stock values. They showed that their method is able to predict the sentiment with high precision and also showed that the stock performance and its recent web sentiments are also closely correlated.

Khurshid Ahmad and Yousif Almas [6] developed a system called E-Analyst which collects two types of data, the financial time series and time stamp news stories. It generates trend from time series and align them with relevant news stories and build language models for trend type. In their work they treated the news articles as bag of words.

Shangkun Deng [7] proposed a stock prediction model which includes Raw Data Preprocessing (RDP) component, Sentiment Analysis (SA) component, Feature Extraction (FE) component, Price Prediction and Evaluation (P&E) component. First, the RDP component downloads data of time series, news, and comments; second, the SA component analyze the overall sentiment for every comments and news; third, the FE component extracts features from the three different kinds of data sources; then, the P&E component predicts the stock price based on a MKL regression framework and evaluates the prediction results based on some magnitude evaluation measures.

Yoon and Swales [5] demonstrated the capabilities of forecasting performance of multilayer perceptron NNs (MLPs) compared to multivariate statistical methods.

In a similar manner, MLPs have been employed to predict short-term stock prices or indexes on various stock markets, see e.g. [6,7,8].

Except for MLPs, other NNs' architectures have successfully been applied to stockprice forecasting such as generalized regression NNs [9], radial basis function (RBF) NNs [10, 11] and related SVR [12].

The non-linear character of stock price data have further been examined using other soft-computing and AI methods such as chaostheory [13,14], multi-agent systems [15,16], or fuzzy rule-based systems [17]. The advantages of individual soft computing methods have been combined in hybrid systems [18,19,20].

Fuzzy rule based systems [21] and NNs [22] have been also successfully applied stock market trend where the hit ratio of correctly predicted trends is used as a measure of forecasting performance.

The problem of stock price forecasting becomes even more complex when performing long-run forecasts. Short-run forecasts are mainly based on technical indicators whilst long-run forecasts are performed using fundamental analysis.

Campbell and Ammer [23] report that long-run stock returns of US companies are driven largely by news about future excess stock returns and inflation, respectively.

Campbell and Shiller [24] demonstrated that price earnings ratios and dividend price ratios are important drivers of future stock price changes.

Previous returns seem to affect future stock price returns, too (a long memory property of stock market) [25].

However, large variations in stock prices have not been explained adequately so far. Bak et al. [26] argue that the large variations may be due to a crowd effect (with agents imitating each other's behaviour). The variations were explained by the interplay between "rational traders" and "noisetraders". The rational traders' behaviour is based on fundamental analysis, whereas the noise traders make decisions based on the behaviour of other traders. Then, fundamental analysis can be used to forecast future stock returns effectively only when the number of rational traders (arbitrageurs) is larger.

Researchers in behavioural finance have been working with two basic assumptions[27]:

investors are subject to sentiment; and betting against sentimental investors is costly and risky.

(1) Investor sentiment is measured either bottom-up (investors under react or overreact to past returns or fundamentals)

(2) Top-down (the effect of aggregate sentiment on individual stocks).

Recently, the effect of market sentiment on stock market behaviour has been investigated in agent-based simulators [28].

According to [28], a high sensitivity to aggregate investor sentiment is associated with low capitalization, younger, unprofitable, high volatility, non-dividend paying, growth companies, or stocks of firms in financial distress.

Bollen et al. [29] showed that the aggregate sentiment can be extracted from the text messages on the Twitter. They analyzed the text content of daily Twitter feeds by measuring (1) positive vs. negative mood, and (2) mood in terms of 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). The accuracy of DJIA (Dow Jones Industrial Average) daily predictions were significantly improved by the inclusion of specific public mood dimensions.

Tetlock [30] finds that sentiment in news stories determines both stock price return and volatility. Specifically, high media pessimism predicted downward pressure on market prices followed by a reversion to fundamentals. In addition, unusually high or low pessimism predicted high market trading volumes. These findings conform to noise traders' models.

Demers and Vega [31] investigated the effect of sentiment in earnings announcements.

They conclude that (1) unanticipated net optimism in managers' language predicts abnormal stock returns, and (2) the level of uncertainty in the text is associated with idiosyncratic volatility and predicts future idiosyncratic volatility.

Statistical approaches such as Naïve Bayes classifier, vector distance classifier, discriminant-based classifier, and adjective-adverb phrase classifier were used by [32] to analyze the sentiment of stock message boards. The sentiment analysis proves to be a significant determinant of stock index levels, trading volumes and volatility.

Annual reports are an important vehicle for organizations to communicate with their stakeholders. In addition to quantitative data (accounting and financial data drawn from financial statements), annual reports contains narrative texts, i.e. qualitative data. Besides other things, annual reports describe company’s managerial priorities. Kohut and Segars [33] noticed that communication strategies in annual reports differ in terms of the subjects emphasized when the company’s performance worsens.

Sentiment analysis of text documents is carried out using either word categorization (bag of words) method or statistical methods. The former method requires available dictionary of terms and their categorization according to their sentiment.

However, such a dictionary is context sensitive.

## PROPOSED MODEL

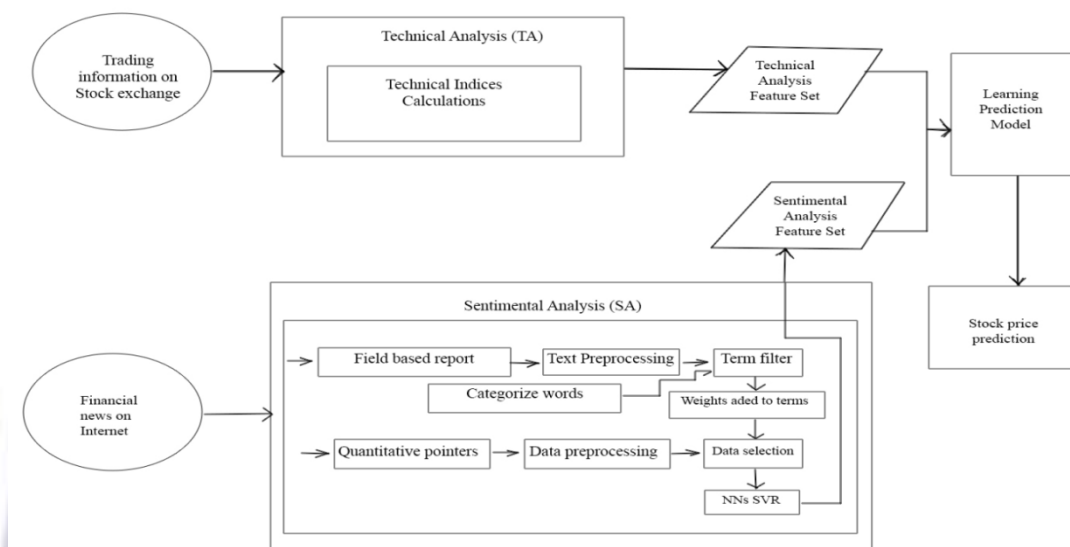


Fig 1: Hybrid Stock Price Prediction Model

## SENTIMENT ANALYSIS BASED ON STOCK NEWS

### Data pre-processing and detecting the seed word set

The textual data collect from online stock news and use part-of-speech method to tagging each word from CKIP system (word segmentation on English word). We select important word according to POS tagging including verb, noun and adjective and then generate multidimensional seed word seta according multidimensional considerations i.e. economy, technology. Each seed word set is detecting by specific field expert because of the reason that human can identify sensitively the seed word according to their background in the field expert.

### Extracting sentiment features and its weighting by PMI-based

In features extraction, we want to use PMI method [9] to analyze word association among word and seed word set. Each word has PMI value from seed word set from last step. The value of  $PMI(word, sword)$  which is the word with the sword seed word calculates follows as:

$$PMI(word, sword) = \log_2 \left( \frac{count(word, sword)}{count(word)count(sword)} \right) \tag{1}$$

where  $count(word, sword)$  is count the co-occur frequency between words.

From the PMI value we calculate the strength of semantic association between word and seed word set of Class (i.e. positive, negative). The word strength is following as:

$$strength(word) = \frac{1}{n} \sum_{sword \in Class} PMI(word, sword_i) \tag{2}$$

where if the value in this class1 more than another class2 is belongs to the class1, contrariwise belongs to class2. Also the value of  $strength(word)$  is the feature weight of word in its class. Therefore, we could know the word that how many

similarities with seed word set of class and each word has strength value. In addition, we can repeat the feature extraction processes from each seed word sets.

### Calculating sentiment intensity for each news

In this step, we provide a function to calculate sentimental intensity which is total stock information in a stock news document. The value of *Intensity* is a balance among positive and negative. The *p\_strength* is positive feature detecting information volume of positive and the *n\_strength* is negative feature detecting

information volume of negative

$$Intensity(k) = p\_strength(k) - n\_strength(k) \tag{3}$$

$$p\_strength(k) = \frac{\sum_{i=1, j \in P, j \in D_k}^n strength(word_i)}{n} \tag{4}$$

$$n\_strength(k) = \frac{\sum_{j=1, j \in N, j \in D_k}^m strength(w_j)}{m} \tag{5}$$

Where the *word<sub>i</sub>* is the word strength of *i*th feature in positive feature set *P* exist document *D<sub>k</sub>*. *word<sub>j</sub>* is the strength of *j*th feature in negative feature set *N* exist document *D<sub>k</sub>*. The *n* and *m* are total number of positive and negative feature appear in stock news *D<sub>k</sub>*. According to the intensity of stock information determination the new affect degree by our proposed sentimental analysis, if the *Intensity* value more than zero then it has positive sentiment otherwise is negative sentiment.

### Sentiment Analysis

Sentiment analysis represents a complex problem due to the ambiguity in word categorization. The ambiguity can be resolved using context knowledge, for example from financial domain. The correct categorization of terms into the bags of words (positive, negative, etc.) is difficult because words may have different meanings and tones in individual domains. Therefore, there have been attempts to propose a domain-specific word categorization recently. In this study we used the word categorization from the financial dictionary proposed by [38] with the following categories of terms:

- Negative (e.g. abandon, abolish, abuse, annoy, annul, assault, bad, loss, bankruptcy. Barrier, calamity, cancel, close, corrupt, critical, crucial, danger decline, default, depress, diminish, disagree, imbalance, improper, problem, suffer, unable, weak), *wf*= 2349,
- Positive (e.g. able, accomplish, achieve, advance, assure, boost, collaborate, compliment, creative, delight, easy, enable, effective, enjoying, excellent, gain, progress, strong, succeed), *wf*=354,
- Uncertainty (e.g. ambiguity, assume, depend, crossroad, deviate, fluctuate, may, maybe, inexact, probably, random, reconsider, risk, unknown, variable), *wf*= 291,
- Litigious (e.g. allege, amend, appeal, arbitrate, attest, attorney, bail, codified, constitution, contract, crime, court, defeasance, delegable, indict, judicial, legal, sue), *wf*= 871,
- Modal strong (e.g. always, best, clearly, definitely, highest, must, never, strongly, undoubtedly), *wf* = 19,
- Modal weak (e.g. almost, appeared, could, depend, might, nearly, possible, seldom, sometimes, suggest), *wf* = 27,

Where *wf* is the frequency of terms in the word categories listed in the financial dictionary. The frequency of net positive words was determined as the positive term count minus the count for negation (positive terms can be easily qualified or compromised).

The most common *tf.idf* (term frequency-inverse document frequency) term weighting scheme was used in this study. The weights can be defined as follows in the *tf.idf*.

$$w_{i,j} = \begin{cases} \frac{(1 + \log(tf_{i,j}))}{(1 + \log(a))} \log \frac{N}{df_i} & \text{if } tf_{i,j} \geq 1, \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Where *N* denotes the total number of documents in the sample, *df<sub>i</sub>* stands for the number of documents with at least one occurrence of the *i*-th term, *tf<sub>i,j</sub>* is the frequency of the *i*-th term in the *j*-th document, and *a* denotes the average term count in the document.

### Technical Analysis based on Stock Price and Volume

Investment managers calculate different indicators from available data and plot them as charts. Observations of price, direction, and volume on the charts assist managers in making decisions on their investment portfolios According to

references that their researches shown the technical indices with stock price have high correlation coefficient. There are many kind of technical index in the stock market for investor decision which considers three major kinds of technical indices including moving average (MA), bias (BIAS) and relative strength index (RSI).

### Learning Predication Model based on Technical Indices and Sentiment Intensity

In this section, we combine two feature set from sentiment analysis and technical analysis for prediction model. The sentimental analysis component can analyse the sentimental intensity of news in a day. The technical analysis component can generate technical indices in a day. The combinational feature set as input data is generated from SA and TA for learn the predication model. The target output is future stock price. Support vector regression will be applied as a machine learning model which can extract the hidden knowledge according to SA and TA. On the kernel function selection, we try to use RBF functions to generate better performance in SVR model.

### Predicting the Daily Future Stock Price

In this part, we will calculate average sentimental intensity of stock news of each dimension of each day and combining the technical indices to stock price predication model. In this paper, we propose predict daily stock price based on our proposed predication mode.

### METHODOLOGY MODEL

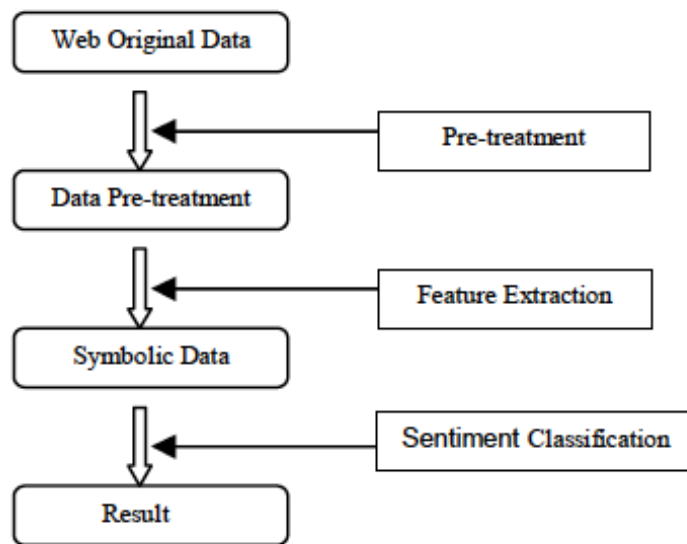


Fig 2: Sentiment Analysis and Evaluation Mechanism

### OVERVIEW OF PROPOSED MODEL

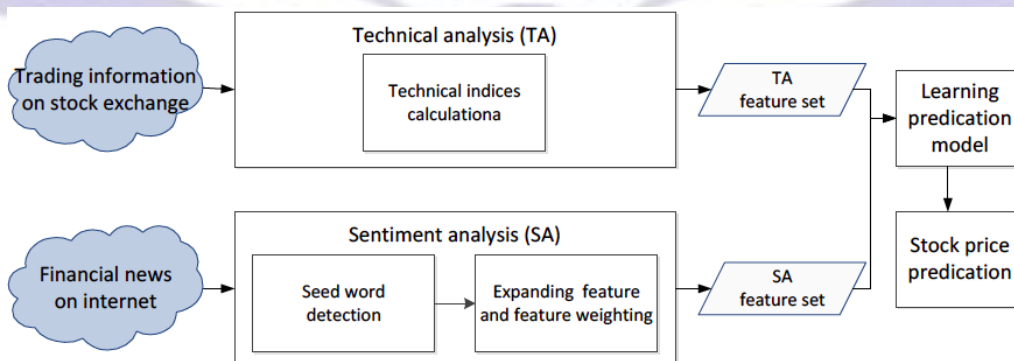


Fig 3: Macroscopic view of Price Prediction Model



## CONCLUSION AND FUTURE RESEARCH DIRECTIONS

Thus we have put forward a hypothesis to formulate a hybrid method for stock price prediction using sentimental analysis. Our future aims and research directions would be to test this hypothesis and compose a working model to implement this hypothesis and to calculate future stock price using technical analysis with the presence of hybrid sentimental analysis method in the base.

## REFERENCES

- [1] Vivek John George, Darshan M.S., Sneha Pricilla, Arun S., Ch. Vanipriya, "Indian Stock Market Predictor System", in International Journal of Soft Computing and Artificial Intelligence, ISSN: 2321-404X Volume- 1, Issue- 1, pg. 39-44.
- [2] Caslav Bozic and Detlef Seese, "Neural Networks for Sentiment Detection in Financial", JEL Codes: C45, D83, and G17, Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology (KIT)
- [3] Liang, X. 2005, "Impacts of Internet Stock News on Stock Markets Based on Neural Networks", : ISNN 2005, LNCS 3497, pp. 897–903, 2005. © Springer-Verlag Berlin Heidelberg 2005
- [4] Liang, X. and Chen, R.-C. 2005, "Mining Stock News in Cyberworld Based on Natural Language Processing and Neural Networks", 13-15 Oct. 2005, Neural Networks and Brain, 2005. ICNN&B '05. International Conference
- [5] Vivek Sehgal and Charles Song, "SOPS: Stock Prediction using Web Sentiment", 2007 IEEE DOI 10.1109.
- [6] Khurshid Ahmad and Yousif Almas, "Visualising Sentiments in Financial Texts
- [7] Shangkun DENG, Takashi MITSUBUCHI, Kei SHIODA, Tatsuro SHIMADA, and Akito SAKURAI Faculty of Science and Technology, Keio University 3-14-1, Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan {dsk8672, mitsubuchi, shioda, shimada, IEEE, International Conference on Dependable, Autonomic and Secure Computing
- [8] Y. Yoon, J. Swales, Prediction Stock Price Performance: A Neural Network Approach, In: Proc. of the 24th Annual Hawaii Int. Conf. on System Science, 1991, pp.156–162.
- [9] J. Yao, H.L. Poh, Forecasting the KLSE Index using Neural Networks, In: Proc. of the IEEE Int. Conf. on Neural Networks, Vol.2, 1995, pp. 1012–1017.
- [10] N. Baba, M. Kozaki, An Intelligent Forecasting System of Stock Price using Neural Networks, In: Proc. of the Int. Joint Conf. on Neural Networks, Vol.1, 1992, pp. 371–377.
- [11] J.H. Choi, M.K. Lee, M.W. Rhee, Trading S&P 500 Stock Index Futures using a Neural Network, In: Proc. of the Annual Int. Conf. on Artificial Intelligence Applications on Wall Street, New York, 1995, pp. 63–72.
- [12] D. Enke, S. Thawornwong, The Use of Data Mining and Neural Networks for Forecasting Stock Market Returns, Expert Systems with Applications, Vol.29, No.4, 2005, pp. 927–940.
- [13] X.B. Yan, Z. Wang, S. Yu, Y. Li, Time Series Forecasting with RBF Neural Network, In: Proc. of the 4th Int. Conf. on Machine Learning and Cybernetics, Guangzhou, Vol.8, 2005, pp.4680–4683.
- [14] K. Chen, H. Lin, T. Huang, The Prediction of Taiwan 10-year Government Bond Yield, WSEAS Transactions on Systems, Vol.8, No.9, 2009, pp. 1051–1060.
- [15] L.J. Cao , F.E.H. Tay, Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting, IEEE Transactions on Neural Networks, Vol.14, No.6, 2003, pp. 1506–1518.
- [16] Ch.K. Volos, I.M. Kyprianidis, S.G. Stavrinos, I.N. Stouboulos, L. Magafas, A.N. Anagnostopoulos, Nonlinear Financial Dynamics from an Engineer's Point of View, Journal of Engineering Science and Technology Review, Vol.4, 2011, pp. 281–285.
- [17] Ch.K. Volos, I.M. Kyprianidis, S.G. Stavrinos, I.N. Stouboulos, Synchronization Phenomena in Coupled Nonlinear Systems Applied in Economic Cycles, WSEAS Transactions on Systems, Vol.11, No.12, 2012, pp. 681–690.
- [18] F. Neri, Learning Predictive Models for Financial Time Series by Using Agent Based Simulations, Lecture Notes in Computer Science, Vol.7190, 2012, pp. 202–221.
- [19] F. Neri, Agent based Modeling under Partial and Full Knowledge Learning Settings to Simulate Financial Markets, AI Communications, Issue 4, Vol.25, 2012, pp.295–304.
- [20] P.Ch. Chang, Ch.H. Liu, A TSK Type Fuzzy Rule based System for Stock Price Prediction, Expert Systems with Applications, Vol.34, No.1, 2008, pp. 135–144.
- [21] R. Tsaih, Y. Hsu, C.C. Lai, Forecasting S&P 500 Stock Index Futures with a Hybrid AI System, Decision Support Systems, Vol.23, No.2, 1998, pp. 161–174.
- [22] G. Armano, M. Marchesi, A. Murru, A Hybrid Genetic-Neural Architecture for Stock Indexes Forecasting, Information Science, Vol.170, No.1, 2005, pp. 3–33.





- [23] J.Y. Campbell, J. Ammer, What Moves the Stock and Bond Markets? A Variance Decomposition for Long-Term Asset Returns, *The Journal of Finance*, Vol.48, No.1, 1993, pp. 3–37.
- [24] J.Y. Campbell, R.J. Shiller, Valuation Ratios and the Long-run Stock Market Outlook: An Update. No. w8221, National Bureau of Economic Research, 2001.
- [25] Z. Ding, C.W.J. Granger, R.F. Engle, A Long Memory Property of Stock Market Returns and a New Model, *Journal of Empirical Finance*, Vol.1, No.1, 1993, pp. 83–106.
- [26] P. Bak, M. Paczuski, M. Shubik, Price Variations in a Stock Market with Many Agents, *Physica A: Statistical Mechanics and its Applications*, Vol.246, No.3, 1997, pp. 430–453.
- [27] F. Neri, Quantitative Estimation of Market Sentiment: A Discussion of Two Alternatives, *WSEAS Transactions on Systems*, Vol.11, No.12, 2012, pp. 691–702. *WSEAS TRANSACTIONS on BUSINESS and ECONOMICS* Petr Hájek, Vladimír Olej, Renáta Myšková E-ISSN: 2224-2899 304 Issue 4, Volume 10, October 2013.
- [28] N. Barberis, A. Shleifer, R. Vishny, A Model of Investor Sentiment, *Journal of Financial Economics*, Vol.49, No.3, 1998, pp. 307–343.
- [29] J. Bollen, H. Mao, X. Zeng, Twitter Mood Predicts the Stock Market, *Journal of Computational Science*, Vol.2, No.1, 2011, pp. 1–8.
- [30] P.C. Tetlock, Giving Content to Investor Sentiment: The Role of Media in the Stock Market, *Journal of Finance*, Vol.62, 2007, pp.1139–1168.
- [31] E.A. Demers, C. Vega, Soft Information in Earnings Announcements: News or Noise? Working Paper, INSEAD, 2010.
- [32] S. Das, M. Chen, Yahoo! for Amazon: Opinion Extraction from Small Talk on the Web. Working Paper, Santa Clara University, 2001.
- [33] G.F. Kohut, A.H. Segars, The President's Letter to Stockholders: An Examination of Corporate Communication Strategy, *Journal of Business Communication*, Vol.29, No.1, 1992, pp. 7–21.
- [34] J.-L. Wu, H.-S. Chu, L.-C. Yu and P.-C. Chang. Sentiment analysis of stock news using a PMI-based term expansion method. *ICIC Express Letters*. 2012, 6 (2): 491 496.
- [35] Pang, Bo; Lee, Lillian (2008) 4.1.2 Subjectivity Detection and Opinion Identification. *Opinion Mining and Sentiment Analysis*. Now Publishers Inc.
- [36] Bo Pang; Lillian Lee and Shivakumar Vaithyanathan (2002): Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

### Author's biography

**S. B. Wankhade** is currently Assistant Professor with Rajiv Gandhi Institute of Technology, Mumbai. He received his Master and Bachelor of computer Engineering degree from Amravati University. His research interests are in the fields of Mobile Ad Hoc Network, Distributed Computing and Neural Network.

**Divyesh Surana** is pursuing B.E. in Computer Science at Rajiv Gandhi Institute of Technology, Mumbai, India.

**Neel Mansatta** is pursuing B.E. in Computer Science at Rajiv Gandhi Institute of Technology, Mumbai, India.

**Karan Shah** is pursuing B.E. in Information Technology at Shah and Anchor Kuttchi College of Engineering, Mumbai, India.