Automated Stock Price Prediction and Trading Framework for Nifty Intraday Trading

Aparna Anant Bhat
Department of Information Technology
National Institute of Technology, Karnataka
Surathkal, India
aparnabhat89@gmail.com

Sowmya Kamath S.
Department of Information Technology
National Institute of Technology, Karnataka
Surathkal, India
sowmyakamath@ieee.org

***Abstract*—Research on automated systems for Stock price prediction has gained much momentum in recent years owing to its potential to yield profits. In this paper, we present an automatic trading system for Nifty for deciding the buying and selling calls for intra-day trading that combines various methods to improve the quality and precision of the prediction. Historical data has been used to implement the various technical indicators and also to train the Neural Network that predicts movement for intra-day Nifty. Further, Sentiment Analysis techniques are applied to popular blog articles written by domain experts and to user comments to find sentiment orientation, so that analysis can be further improved and better prediction accuracy can be achieved. The system makes a prediction for every trading day with these methods to forecast if next day will be a positive day or negative. Further, buy and sell calls for intra-day trading are also decided by the system thus achieving full automation in stock trading.**

Keywords—Stock price prediction, Sentiment analysis, Neural Network, Technical analysis

I. INTRODUCTION

Stock prediction has been area of intense research for many years now owing to the many challenges in making accurate predictions due to the volatility of the data. Various methods and theories have been proposed for Stock price prediction and each have their own pros and cons. A theory named Random Walk Hypothesis [1] proposes that stock prices follow a random pattern and hence they cant be predicted at all. In the other extreme, there are people who believe market follows a trend which can be observed over time hence is predictable. A lot of research in this area aims at predicting the movement of Stock prices. Historical data, i.e. statistics generated with market transactions like opening closing prices, volume etc has been used for prediction. Various works show the usefulness of technical indicators in identifying trend and prices for near future [2].

In the stock market, there are various ways to invest. Some people go for long term investment where they buy a particular stock and sell it after a few months or years, whereas intraday trading deals with stock transactions for a day. Though both methods have their own merits and demerits, intra-day trading is said to be riskier as the movements of stocks are highly varying. An investor has to devote a lot of time and effort in observing the movements of the prices for long time of the day. Moreover, it is not very easy to predict the status of stock for the immediate next trading day. Even with a bullish trend seen with various technical indicators, there is always a chance that prices may fall.

Some of the standard methods that are time tested and recommended by most investment watches are the Technical Indicators. Technical Analysis deals with the study of charts. There are various powerful technical indicators which show the current trend, trend reversals etc, which can also help in deciding the buy/sell calls. Some of these are Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), Money Flow Index (MFI), Exponential Moving Average (EMA) etc. MACD shows the trend and trend reversals. RSI, MFI etc help in deciding trend reversals. EMA also helps in detecting the direction. Support and resistance level help in identifying high or low prices as well as trend change [2].

The paper is organized as follows. In Section II, we discuss the related work in the area and Section III provides the required background for understanding the various stock price prediction techniques. Section IV presents the proposed system and its various phases. Section V presents a discussion of the experimental results, followed by conclusion and references.

II. RELATED WORK

With the development of technology and mathematics various new technological methods have been proposed for stock price prediction, which are applied along with technical methods. Genetic Algorithm (GA), Support Vector Machine (SVM), or Neural Networks (NNs) etc have been applied to predict Stock prices. Bonde et al. [3] applied genetic algorithms and evolution strategies for stock price prediction. Phua et al. [4] applied Neural Networks to the financial prediction and tested the influence of volume data on Stock price prediction. Khan et al. [5] used the Neural Networks with different number of hidden layers to analyze the prediction of the Stock prices. Khan et al. [6] also used back propagation algorithm with sigmoid function to predict Stock prices. Feng Pai et al. [7] proposed a hybrid approach with SVM and ARIMA (Autoregressive Integrated Moving Average) model and found it gave promising results.

A more recent approach is based on fundamental analysis. Fundamental Analysis deals with real world accountability than mere chart analysis. Various factors affect the stock market like change in government and various other activities across the world or companies accounts. It also depends on the opinion of people. Hence it is important to know the sentiments expressed in news articles and comments on the Web. A. Khan at al. [8] used Sentiment Analysis for finding sentiment orientation of online reviews. Deng et al. [9] used SentiWordNet for stock news analysis and applied Multiple Kernel Learning (MKL) for stock price prediction.

III. Background

A. Technical Indicators

There are strong technical indicators available for technical analysis. They use various statistics generated in the market like closing prices of history, volume traded etc. We have used Moving Average Convergence Divergence (MACD), Exponential Moving Average (EMA) and Relative Strength Index (RSI).

Fig. 1: Moving Average Convergence and Divergence

1) MACD: This has two lines namely MACD line and Signal Line as shown in figure 1, which give us signals of trend changes with cross overs. These two lines also show the movement of Stock with their coming closer to each other(convergence) and departing from each other(divergence).

$$MACD\ Line:\ (12\ day\ EMA\ -\ 26\ day\ EMA) \quad (1)$$

$$Signal\ Line:\ (9\ days\ EMA\ of\ MACD\ Line) \quad (2)$$

2) EMA: While Simple Moving Average gave average of past few days price, EMA gave higher weight to recent prices. A 10 day EMA can be calculated by using equation 5. As seen from Figure 2, when EMA line crosses from below, it shows the beginning of uptrend and when it crosses from above to the trend line, it shows the beginning of down trend.

$$SMA:(\frac{10\ period\ sum}{10}) \quad (3)$$

$$Multiplier:\frac{2}{(Time\ periods+1)}=(\frac{2}{(10+1)})=0.1818(18.18\%) \quad (4)$$

$$EMA:\{Close-EMA(previous\ day)\}\ multiplier$$
$$+EMA(previous\ day) \quad (5)$$

Fig. 2: Exponential Moving Average

3) RSI: indicates the overbought and over sold regions and hence the change in momentum. This oscillates between 0 and 100. Above 70 is marked as oversold region and below 30 is marked to be overbought region. RSI can also be used to see the general trend. Average gain and average loss in equation 6 are simple 14 period averages.

$$RS=(AverageGain/AverageLoss)$$
$$RSI=100-(\frac{100}{(1+RS)}) \quad (6)$$

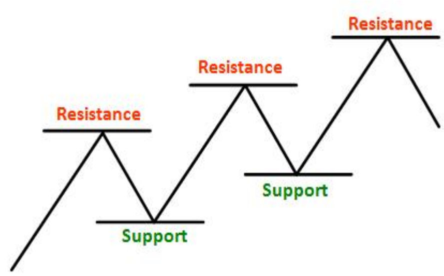

Fig. 3: Support and Resistance

4) Support, Resistance and Pivot Price: These are useful for guessing how high or low can Stock prices go for next day as seen from Figure 3. These values are calculated for everyday. The buy and sell calls are adjusted based on values around these.

$$
\begin{aligned}
&Pivot = (High\ Point + Low\ Point + Close\ Point)/3\\
&Range = High\ Point - Low\ Point\\
&3R\ [3rd\ Resistance] = 1R + Range\\
&2R\ [2nd\ Resistance] = Pivot + Range\\
&1R\ [1st\ Resistance] = Pivot + (Pivot - Low)\\
&1S\ [1st\ Support] = Pivot - (High - Pivot)\\
&2S\ [2nd\ Support] = Pivot - Range\\
&3S\ [3rd\ Support] = 1S - Range
\end{aligned} \tag{7}
$$

The first resistance (support) level is the expected high (low) that is usually seen for a normal non trend day. If this level is crossed early in the day, there are chances that it might reach 2nd Resistance (support) level. And if this is broken it might reach the third resistance (support) as well, that is usually called a trend day when high gain (loss) on Stock price is seen.

B. Backpropagation Neural Network

A Neural networks ability to identify non linear patterns can aid in prediction based on time series data. Error back propagation helps in adjusting the weights in the network for further refinement of results. The network has a number of input neurons, at least one layer of hidden neurons followed by a layer of output neurons. An Activation Function triggers the neurons and along with the weights in the network determines the output. The errors are propagated back to adjust the weights. A trained Neural Network is an expert system that can predict the stock prices with reasonable accuracy [6].

C. Sentiment Analysis

Sentiment Analysis is a Natural Language Processing (NLP) approach that deals with classification of sentiments of written text/articles/blogs etc. Various words occurring in the document are classified as positive, negative or null to identify sentence polarity and thus the text/comments polarity. We use two ways to perform the sentiment classification.

1) SentiWordNet: SentiWordNet (SWN) [10] is an opinion lexicon with a huge list of words where each word is associated with scores between 0 to 1 indicating positive negative and objective sentiment. It is publicly made available for research purposes. It uses a weak supervision, semi supervised learning step and then a random walk step to do the sentiment classification [10]. We use SentiWordNet 3.0 for initial Sentiment Analysis of the web content.

2) Dynamic words list: This is a domain specific list which grows dynamically to improve the performance and time needed for sentiment classification. It contains a list of words, phrases and grows as more documents tested and included.

IV. Proposed System

The proposed system is composed of specific modules as follows -

A. Data Collection

Historical prices and volumes of the Stocks were gathered from National Stock Exchange (NSE) India and Google Finance. Data was gathered for the time period 1990 to 2013 for different time frames. These values are used for both the Technical Analysis component and the Neural Network. For Sentiment Analysis, blog articles and comments were collected with their dates of publication. It was seen that predictions for intra-day were not available for every trading day at times and the frequency of comments and opinions varied for different working days.

B. Technical Analysis

Each of the Technical indicators (MACD, EMA, and RSI) was implemented with the help of historical data collected and predictions were made based on each of them. With the movement and direction of the graph lines, trends were indentified. The trend helped us identify what majority of the next few days are going to be. This usually helps with identifying whether there will be increase or decrease in the prices.

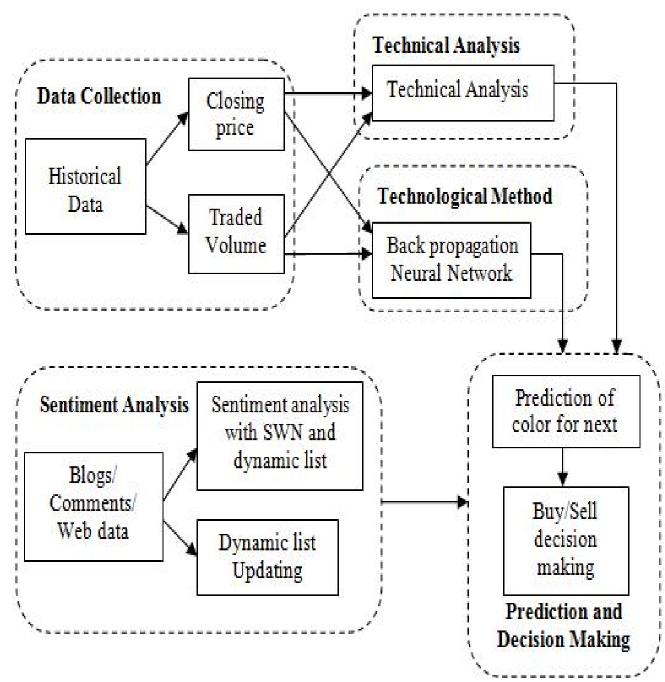

Fig. 4: Modular view of the Proposed System for Intra-day trading

C. Prediction with Neural Network

Neural network model with different sets of input and hidden layer neurons and different activation functions were tested for their ability to predict. Activation functions like sigmoid and hyperbolic tangent function were used to trigger the neurons. Initially weights of paths in the Neural Network were set to be 0.5 and then with Back propagation method they were altered until their optimum values were reached. Training data used was over a period of two years and involved closing prices and volume. Training can be stopped when weight change value of the paths stops changing significantly or the number of given epochs are reached [6]. Time series prediction was also done using the inbuilt functions available in MATLAB. Levenberg Marquardt Backpropagation Algorithm [11] was used for training the Neural Network. Data divisions were 70% 15% 15% for each of training, validation and testing respectively.

While training, the input neurons were given the historical data as input. Closing prices of the past along with the volume of trading is given to input neurons. Output was given to be actual closing price for the next day. Since activation functions used were sigmoid and hyperbolic tangent functions, the inputs were also normalized to fall in the range of 0 to 1. Equation 8 is used for normalization, where Y is the input and Y' is the normalized value.

$$Y' = \frac{(Y - Y_{min})}{(Y_{max} - Y_{min})} \quad (8)$$

D. Sentiment Analysis

Sentiment analysis based approaches deal with identifying the sentiment orientation of public/expert opinion on stock prices. Reviews are collected from various sites where people have discussed stock's fate over the course of the day. Once the reviews are collected they are processed further to derive knowledge that can be used towards improving the accuracy of prediction.

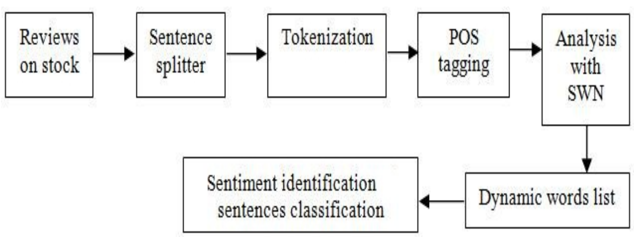

Fig. 5: Steps of Sentiment Analysis

Various phases in analysing the sentiment of Web and blog content are as follows-

1) Sentence splitting: Sentences are assumed to end with a . or ? or ; or !. All the sentences are split at these characters and are stored separately. Each word in the sentence is then Part-of-speech (POS) tagged using Stanford POS tagger [12]. This is necessary because a sentiment word's sentiment score differs when it occurs as a noun or as an adjective. For example, "apple was very good" is more positive than "a good apple". Therefore it is necessary to know a word's form of occurrence. Once sentences are split, words are split and stored for further processing.

2) POS tagging: A Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads text and assigns parts of speech to each word (and other tokens), such as noun, verb, adjective, etc. For example, the result of POS tagging for the input "This is a sample text" is the form of: This_DT is_VBZ a_DT sample_NN text_NN. Where _NN is noun, _VB is verb, _JJ is adjective, _RB adverb. Only these words are used for sentiment identification along with their Part of speech.

3) Sentiment Analysis using SentiWordNet: Figure 6 shows a SentiWordNet fragment. It lists sentiment words and their positive negative scores along with their parts of speech. This is given in .txt file.

The first entry specifies the part of speech, as noun/verb/adverb/adjective. After POS tagging, the sentiment scores of each sentiment word is obtained and overall score of the sentence is calculated. And sentence is classified as positive negative or none of these. The words are stored as a hashmap along with their part of speech. Then they are retrieved by a retrieve function call which takes both word and its part of speech as input and gives back its sentiment score.

Category	WNT Number	pos	neg	Synonyms
A	01123148	0.875	0	good#1
A	00106020	0	0	good#2 full#6
A	01125429	0	0.625	bad#1
A	01510444	0.25	0.25	big#3 bad#2
N	03076708	0	0	trade_good#1 good#4 commodity#1
N	05144079	0	0.875	badness#1 bad#1

Fig. 6: SentiWordNet fragment

4) Dynamic words list: It was observed that SentiWordNet gave results for huge domain of English language and it was seen that results can be improved with additional words often found in the opinions of the Stock Predictions. These words/phrases like "uptrend", "bullish rise", "high", "gain", "bearish trend", "fall", "loss" etc. Phrases like "it may not be a good idea to invest" or "market seems unsteady" caution investors against investing when there is lot of volatility in the market. All such words and phrases are separately listed with positive-negative tags. When such words occur in a sentence, their sentiment is classified as positive or negative in accordance with this list. As more documents are classified, words missed with SentiWordNet are listed. These are added to the dynamic words list with their sentiment tag based on knowledge of English language. Over a period of time, this list covered most of the words that usually appeared in the stock news.

E. Buy Sell Call Decision

1) Deciding the Color for next day: The aim here to indicate if the next trading day will end in gain or loss. In this experiment, after going through each result it was found that EMA was better than SMA for Trend detection. While EMA showed the trend fairly, MACD detected the change in trend faster than EMA. Thus it was found that using MACD results at suitable places like, where the cross over occurs, would allow further refinement of the results. This change in trend was detected by EMA 2-3 days later. Also, we can detect this change in MACD if convergence/divergence is approaching zero and these were found correct at most of the times. Further, along with these results, the results of Neural Network method are taken into account followed by the results of Sentiment Analysis. Highest priority is given to the results of Sentiment Analysis, followed by the other two. This logic was applied for intra-day Nifty prediction. This helped us in predicting whether next day will be a high day or low day, i.e. Nifty index value would go higher or lower compared to today.

However, the main limitation while using technical indicators is that, as stock value depends on multiple, disparate factors, chart analysis alone cannot yield excellent results. Hence, various machine learning techniques are added for technical and fundamental analysis.

2) Buy and Sell Calls: Once we are done with predicting about high day or low day, next step is to calculate support and resistances for next day. This system works as follows.

- On a high day, the system buys stock at open price and sells at high price for the day. This ”high” is set at 90% of the 1st resistance level initially.
- If this first resistance level was seen to be crossed earlier in the day, chances are high that it might reach 2nd resistance level as well. Hence sell signal is called at 80% of 2nd resistance level. These can happen during trend days when Nifty gains/falls sharply. If neither of these levels were reached, sell signal is given at the closing price for the day. Having predicted it to be a high day, we still end up on gaining side. It is seen that price usually doesnt reach the resistance mark, but slightly lesser than that, hence 90% of resistance level is kept for the sell call.
- Similarly for a low day, short calls are made, i.e. sell signal is called at open price while buy at 90% of first support level. If this is broken very early in the day, sell call is not executed on first support but on 80% of second support level. There by increasing the gain. If this mark is not reached buy signal is called at the closing price. Having predicted it as a low day, we would still gain positive amount even at closing price.

Various percentages of support and resistance levels were tested and it was found that highest profits were obtained at 90% of first resistance level, and 80% of second resistance level. But if the color prediction for the day is wrong this algorithm will not give profit for that day. Losses can be reduced by putting appropriate stop losses on such days. Hence aim is to get better and better prediction for-whether next trading day will be a high day or a low day.

V. Experimental Results

The proposed approach combines various methods of stock prediction and consists of two modules, one predicts the color of next day as green if positive day and red if it is a negative day. The second module decides where the buy and sell calls should get executed. Technical analysis was done for a period of 100 days and results with each of Technical Indicators were noted. The results obtained from technical analysis were not very attractive; Neural Network improved the results compared to Technical analysis. Fundamental analysis with the help of Sentiment analysis further improved the results.

For first module i.e. color prediction, results obtained with different methods are shown in figure 7. While Technical indicators gave upto 54% accuracy, Prediction with Neural Network varied with the Different time frame and different sets of inputs and number of hidden neurons. It gave accuracy of 61%. Prediction with Sentiment Analysis largely depended on availability of correct prediction data and user comments. It was tested over a period of one month. With the addition of domain specific dynamic words list the sentiment classification accuracy sharply increased and correctly classified documents as positive or negative.

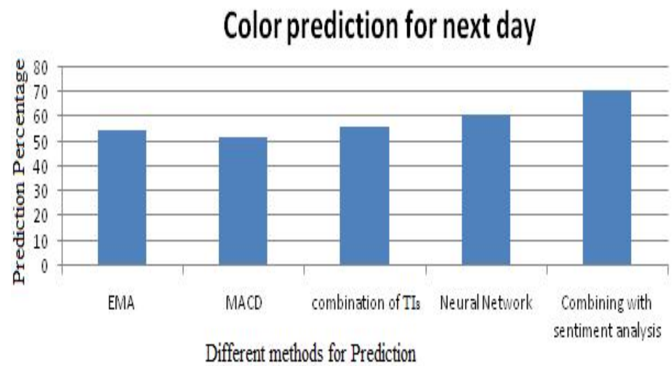

Fig. 7: High/Low prediction with various methods

The accuracy of color prediction was found to be 61% on the lower side and up to 71% on the upper side with the tested Nifty data. The results may vary with the availability of user/expert comments for Sentiment Analysis in future.

The system was tested for a period of one year, total trading days being 249. The second module as we know decided the buy and sell calls. Assuming the correct color predictions i.e. 100% correct predictions of color for next day, algorithm gave profit for 86% of total tested days with buy sell decisions. When tested with 5000 rupees as a virtual investment, the system gave profit up to 129% on Nifty as whole. But given that system achieved 70% of accuracy with color prediction, overall system gave a profit of 75% with Intra-day Nifty.

VI. Conclusion

In this paper, we combined three methods for stock price prediction and proposed to build an automatic trading system which decides when to execute the buy and sell calls for next trading day. Since intra-day trading requires user to continuously monitor the positions of stock, an automated trading system would help user to make profits and save time. Though combining of Technical Analysis with technological methods like Neural Network gave better performance, addition of Sentiment Analysis further increased the results as stock prices highly depended on day to day activities of the market and fundamental analysis. System first predicted the status for next day i.e. whether there will be a rise in price or fall, and then it predicted where should the buy and sell calls be placed.

References

[1] M. D. Godfrey, C. W. Granger, and O. Morgenstern, “The random-walk hypothesis of stock market behaviora,” *Kyklos*, vol. 17, no. 1, pp. 1–30, 1964.

[2] J. Murphy, “Technical analysis of the financial markets, prentice hall, london,” 1998.

[3] G. Bonde and R. Khaled, “Stock price prediction using genetic algorithms and evolution strategies.”

[4] X. Wang, P. K. H. Phua, and W. Lin, "Stock market prediction using neural networks: does trading volume help in short-term prediction?" in *Neural Networks, 2003. Proceedings of the International Joint Conference on*, vol. 4. IEEE, 2003, pp. 2438–2442.

[5] A. U. Khan *et al.*, "Stock rate prediction using back propagation algorithm: Analyzing the prediction accuracy with different number of hidden layers," *Glow gift, Bhopal*, 2005.

[6] A. U. Khan, T. Bandopadhyaya, and S. Sharma, "Comparisons of stock rates prediction accuracy using different technical indicators with backpropagation neural network and genetic algorithm based backpropagation neural network," in *Emerging Trends in Engineering and Technology, 2008. ICETET'08. First International Conference on*. IEEE, 2008, pp. 575–580.

[7] P.-F. Pai and C.-S. Lin, "A hybrid arima and support vector machines model in stock price forecasting," *Omega*, vol. 33, no. 6, pp. 497–505, 2005.

[8] A. Khan, B. Baharudin, and K. Khan, "Sentiment classification using sentence-level lexical based semantic orientation of online reviews," *vol*, vol. 6, pp. 1141–1157, 2011.

[9] S. Deng, T. Mitsubuchi, K. Shioda, T. Shimada, and A. Sakurai, "Combining technical analysis with sentiment analysis for stock price prediction," in *Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on*. IEEE, 2011, pp. 800–807.

[10] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *Proceedings of the 7th conference on International Language Resources and Evaluation (LREC10), Valletta, Malta, May*, 2010.

[11] J. J. Moré, "The levenberg-marquardt algorithm: implementation and theory," in *Numerical analysis*. Springer, 1978, pp. 105–116.

[12] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. Association for Computational Linguistics, 2003, pp. 173–180.