Predicting the uptrend and downtrend in stock markets for intraday trading using machine learning algorithms

[1]Ishan Bhatt

[1]Computer science and engineering, Amity School of Engineering & Technology, Amity University, Noida, Uttar Pradesh, India (201303)
Email: ishanbhatt.korba@gmail.com

[2]Kartik Ramaswamy

[2]MBA Finance, Amity Business School, Amity University, Noida, Uttar Pradesh, India (201303)
Email: kartik.ramaswamy96@gmail.com

[3]Nidhi Vij

[1]Computer science and engineering, Amity School of Engineering & Technology, Amity University, Noida, Uttar Pradesh, India (201303)
Email: ernidhivij@gmail.com

Abstract

A country's economy is dependent on several parameters among these parameters stock markets plays a very important role. There are typically two sorts of risks in regard with the security exchange which are systematic risk and unsystematic risks and this is the reason why stock market is stochastic in nature. From years, scholars are trying to find a definitive solution for better decision making in market to generate more returns and reduce risk. There are many ratios, formulas and theorems which attempts to predict the stock market but in reality these theorems are made on countless assumptions. With the new age technology and fast computing, we can now solve this problem by advanced algorithms and machine learning. We will take help of probability to solve problems generating because of stochastic nature of Stock market. Computing series of probability at different scenarios and parameters of stock market by using machine learning.

Keywords:- Stock market, Random forest regression, technical analysis, security analysis, Portfolio management

I. INTRODUCTION

Countries economy stock markets have a dominant role. Stock markets are very large and basically operate on the principle of demand and supply. Stock market has stochastic nature and it is very difficult to predict trends of securities. There are typically two sorts of risks in regard with security exchange:

1. Systematic risk: is the innate risk of the security exchanges which arises due to the nation's economy and international rules, strategies and policies. These types of nests are uncontrollable and affect every security.
2. Unsystematic risk: are associated with company or sector. They are usually controllable and arise due to malfunctioning of a sector or a company. For example, increased debt to equity ratio, labour strike, etc.

Until now prediction in stock market using fundamental analysis and technical analysis. These analyses have separate tools to analyse a security of a stock market but they failed miserably because these tools are built on countless assumptions and it is not necessary that the stock market principal every time.

With the advance and cutting-edge technology, we can now predict the movement of securities and stock market efficiently. Random forest regressionis capable of finding the probability that market will go up or down. Predicting the upward or downward trend of the market will ensure an efficient decision making for buying and selling securities.

II. RELATED WORK DONE

Economic growth and stock development are interlinked. Understanding the relationship between them might help the investors, commercial banks, fund managers take crucial decisions in stock market [1]. In [2] the author predicts the performance of the stock using logistic regression. [4] The research paper displays a positive correlation between DJIA values and individual behavioural. The author [7] uses ANN to predict the BSE, NSE BSVP, FTSE, MCX etc indexes. In [9] the author takes help of RNN(recurrent Neural Network) in order to predict the closing price in BSE, NSE BSVP, FTSE, MCX etc. In [6] the author has used convolution neural networks which help to improve upon the 4 out of 7 tasks.

The research paper [3] codes and classifies previously published research papers to summarize and bridges the gaps in the analysis of security exchanges, multi commodity exchange etc. [8] The research paper used clustering algorithm and correlation coefficient of financial time series, where a map is assigned to each company and then using the mentioned techniques, the robustness between maps is determined. [10] The paper gives a detailed study of various algorithms used for stock prediction and gives a brief of areas that should be focused upon for getting optimal results. [11] The extensive use of machine learning techniques like sentiment analysis, supervised, unsupervised learning and other hybrid techniques are mentioned for stock market prediction. [12] The paper promotes usage of technical analysis to increase investment consciousness and to rely on proficient data and optimal analysis. In [5] the paper overviews the relation between Indian security exchanges, multi commodity exchange etc with other Asian stock markets.

In [13] the economic time series has been analysed and it results in the similarity between a genuine time series and a series where one of the systematic elements is weak. In [15] the need for investors to predict the stock market is mentioned, and how this prediction helps retain investors' attention for stocks [14] the paper suggests that an amalgamation of artificial neural networks and any other statistical tool or machine learning algorithm provides better results for financial time series predictions. In [16] the research paper keeps into account factors like pricing, social behaviour, regulations and how it affects the stock market. The paper aims at promoting pragmatic policy orientation. In [18] the paper measures the

volatility between different stock markets around Asia and displays a high correlation between them and in [17] the correlation analysis between BSE and other stock markets such as BSVP, FTSE and MCX show a high positive correlation.

III. METHODOLOGY

We will be using random forest regression which is a part of machine learning data analysis and is known to give results with highest accuracy. We will be using these parameters for analysis: -

1) Risks – Risk can be defined as the probability of a future loss. Long term investments, poor decisions may increase the probability of risks.

   Systematic risk are those risks which effect the whole industry and not any specific business. These occur due to external factors of a business such as market risks, purchasing power risk and interest risk. Systematic risks are usually uncontrollable and unavoidable by a business while Unsystematic risks occur due to the internal factors and it affects a particular business not the whole. These risks occur due to incapable operational activity and due to its inability to maintain competitive advantage. These risks are usually financial or business specific risks. Financial risks occur due to ineffectual capital structure which results in financial instability. Business specific risks includes credit, currency, country and liquidity risks. Unsystematic risks are avoidable and can be reduced through diversification.

2) Beta of stock market – Beta is the measure of the systematic risk of a business in comparison to unsystematic risk of the entire market or the volatility. It represents the fluctuation of a stock to changes in the overall stock market. Beta coefficient calculation is measured through regression. If the beta value of a stock is 1.2, then the stock is said to be 20% more volatile than the other stocks in the stock market. If beta is 1.2 and the stock is expected to move by 10% then the stock should move up by 12% (1.2*10). Beta measure helps the investor to decide whether he should invest with a less volatile stock (beta less than 1) or a riskier stock with high correlation value (beta more than 1). Higher beta stocks have frequent and wider price changes which could increase the chance of investor losing the money while lower beta stocks have opposite effect. Beta is a useful statistical tool to have in calculating company's cost of equal component in the cost of capital.

3) William percentage range moves between O to 100. It is a momentum indicator that measures oversold and overbought levels. It is used to find entry and exit levels of stock market. When the indicator is between -20 and 0 it represents that the stock is overbought i.e. the price of stock is close to the higher end of the recent price range of the stock. When the indicator is between -80 and -100 it represents that the stock is oversold i.e. the price of stock is closer to lowerend of recent price range of the stock. It is used to generate trade signals. It can be calculated as highest price in the lookback subtracted by the most recent closing price divided by highest price in the lookback subtracted by lowest price in the lookback. Since, the

indicator is looking at only the last 14 values it may be too responsive that means it may give false signals.

4) Working of stock market – Stock markets enables sellers and buyer to negotiate prices and trades. The demand and supply influence the pricing of BSE, NSE BSVP, FTSE, MCX etc. Supply alludes to total number of investors who are willing to share their stock at any price. Demand alludes to the aggregate sum of stock potential purchasers who would purchase at any cost. As the price of stock increases, the individuals ready to purchase the share decrease. In the event that the economy is performing better than normal, it makes more demand for stocks. All the potential purchaser and dealers exchange until there is nobody left who concurs on that value, this is where demand and supply meet and is called as market equilibrium. Issuing new shares, public offerings can increase the rate of supply since the supply of stocks tend to change at a slower pace as compared to demand.

IV. IMPLEMENTATION

Data mining is a process of discovering predictive information by analysing large data bases. Database for stock market can be generated from nifty India's website and for are analysis we will be analysing nifty 50 companies. we will be analysing past 2 years of data because within the span of 2 years stock will be affected by many factors such as bonus, trade was, reception, split, mergers and acquisitions, infringement, change in the upper management, etc. Taking such a large database will ensure a robust proof-of-concept and in data mining the larger the database results a better predictive analysis.

1. At first we will be extracting information from nifty website and then we will keep relevant information and delete an irrelevant attribute.
2. Since nifty previous volume database does not provide results and risk calculated so we will first calculate risk and returns as well as beta of the market.
3. The same process will be done for all the companies and then database will be further filtered and arranged for random forest classifier
4. We will be analysing first 14 days. Random forest regression will find the predicted %William of uptrend and downtrend considering previous records of risk, returns and beta ratios.
5. The calculated %william will be then matched with the actual 15th day percentage William ratio. If percentage William suggests that on the 15th day stock was overbought that means there is an uptrend and the same trend will be predicted by random forest classifier.
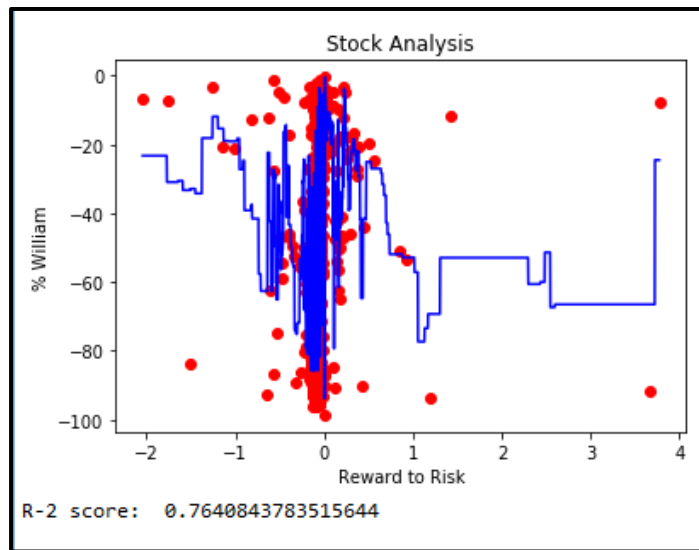6. At last we will find out R-squared value.

V. RESULTS

Figure 1 Random Forest Regression

VI. CONCLUSION AND COMPARISON

The R-squared value is 0.7640 but in the field of stock markets it is very accurate. Also, the comparatively lower R-square value is because of movement large quantity of stocks. The selected stocks were traded averagely 1,00,000 each day. That is the reason why R-squared value is affected by 1,00,000 securities. Given the highly stochastic nature of NSE and BSE these results are satisfactory to predict trend.

We can conclude this study in two ways. Firstly we can compare the traditional used methods by investors for intraday trading. These customary ways are:
1. Fundamental analysis: This uses company's performance such as top level management people, debt on company, P/E ratio, etc.
2. Technical analysis: This uses technical factors such as beta ratios, Elliot wave equation and Dow Theory.

Random forest somehow takes use of tools technical analysis and combining with stochastic nature of stock market and calculates the probability of stock to be over sold or over bought. Random forest regression will create a new way of predicting movement of securities by analysing a span of 14 days. Our results show that Random forest regression is capable of finding probability.

1) Fundamental Analysis
   a. Advantages:
      i. Not at all for short term analysis
      ii. Easily understandable
      iii. Influences global pricing
   b. Disadvantages:
      i. It is limited to linear relations only.
      ii. Times consuming analysis
      iii. Too many assumptions are taken
2) Technical analysis
   a. Advantage:
      i. Very efficient for intraday trading and short selling

ii. Provide efficient entry and exit levels
iii. There are many theories dependent on this analysis technique such as Dow theory, Elliot wave pattern, etc.
   b. Disadvantage:
      i. Does not incorporate probabilistic nature of stock market
      ii. Difficult to understand
3) Random forest regression
   a. Advantage:
      i. incorporate probabilistic nature of stock market
      iii. provide mostly accurate result
      iv. easy for investors and venture capitalist to predict movement
      v. we cannot completely depend on results this is the reason why it is open for further analysis
      vi. calculated by advance algorithms and cutting edge technology
   b. Disadvantages:
      i. No definitive or discrete solution is given by algorithms.
      ii. Still decision is dependent on investor
      iii. Attributes must be independent

References

[1] Paramati, S., & Gupta, R. (2013). An Empirical Analysis of Stock Market Performance and Economic Growth: Evidence from India. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2335996

[2] Imran, K. (2018). Prediction of stock performance by using logistic regression model: evidence from Pakistan Stock Exchange (PSX). *Asian Journal Of Empirical Research*, *8*(7), 247-258. doi: 10.18488/journal.1007/2018.8.7/1007.7.247.258

[3] Farias Nazário, R., e Silva, J., Sobreiro, V., & Kimura, H. (2017). A literature review of technical analysis on stock markets. *The Quarterly Review Of Economics And Finance*, *66*, 115-126. doi: 10.1016/j.qref.2017.01.014

[4] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal Of Computational Science*, *2*(1), 1-8. doi: 10.1016/j.jocs.2010.12.007

[5] Bose, S., & Mukherjee, P. (2006). A Study of Interlingages Between the Indian Stock Market and Some Other Emerging and Developed Markets. *SSRN Electronic Journal*. doi: 10.2139/ssrn.876397

[6] Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *Proceedings Of The 2014 Conference On Empirical Methods In Natural Language Processing (EMNLP)*. doi: 10.3115/v1/d14-1181

[7] Guresen, E., Kayakutlu, G., & Daim, T. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems With Applications*, *38*(8), 10389-10397. doi: 10.1016/j.eswa.2011.02.068

[8] Basalto, N., Bellotti, R., De Carlo, F., Facchi, P., & Pascazio, S. (2005). Clustering stock market companies via chaotic map synchronization. *Physica A: Statistical*

Mechanics And Its Applications, *345*(1-2), 196-206. doi: 10.1016/j.physa.2004.07.034

[9] Lin, X., Yang, Z., & Song, Y. (2009). Short-term stock price prediction based on echo state networks. *Expert Systems With Applications*, *36*(3), 7313-7317. doi: 10.1016/j.eswa.2008.09.049

[10] Shah, D., Isah, H., & Zulkernine, F. (2019). Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. *International Journal Of Financial Studies*, *7*(2), 26. doi: 10.3390/ijfs7020026

[11] Mitra, S. (2011). How rewarding is technical analysis in the Indian stock market?. *Quantitative Finance*, *11*(2), 287-297. doi: 10.1080/14697680903493581

[12] Masry, M. (2017). The Impact of Technical Analysis on Stock Returns in an Emerging Capital Markets (ECM's) Country: Theoretical and Empirical Study. *International Journal Of Economics And Finance*, *9*(3), 91. doi: 10.5539/ijef.v9n3p91

[13] Kendall, M., & Hill, A. (1953). The Analysis of Economic Time-Series-Part I: Prices. *Journal Of The Royal Statistical Society. Series A (General)*, *116*(1), 11. doi: 10.2307/2980947

[14] Ican, Ö., & Çelik, T. B. (2017). Stock Market Prediction Performance of Neural Networks: A Literature Review. International Journal of Economics and Finance, 9(11), 100. https://doi.org/10.5539/ijef.v9n11p100

[15] Yang, Z., Liu, J., Yu, C., & Han, J. (2017). Quantifying the effect of investors' attention on stock market. *PLOS ONE*, *12*(5), e0176836. doi: 10.1371/journal.pone.0176836

[16] Haslam, J., Ji, J., & Sun, H. (2018). Towards a well-functioning stock market in context. *Journal Of Capital Markets Studies*, *2*(1), 21-36. doi: 10.1108/jcms-01-2018-0003

[17] Patel, R. (2017). Co-Movement and Integration Among Stock Markets : A Study of 14 Countries. *Indian Journal Of Finance*, *11*(9), 53. doi: 10.17010/ijf/2017/v11i9/118089

[18] Malepati, V. (2016). A Study on Volatility in Indian Stock Market. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2890890