International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2020 IJSRCSEIT | Volume 6 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT206290

Stock Market Prediction using Data Mining Techniques

Karunesh Makker, Prince Patel, Hrishikesh Roy, Sonali Borse
Computer Engineering Department, NMIMS MPSTME, Shirpur, Maharashtra, India

ABSTRACT

Stock market is a very volatile in-deterministic system with vast number of factors influencing the direction of trend on varying scales and multiple layers. Efficient Market Hypothesis (EMH) states that the market is unbeatable. This makes predicting the uptrend or downtrend a very challenging task. This research aims to combine multiple existing techniques into a much more robust prediction model which can handle various scenarios in which investment can be beneficial. Existing techniques like sentiment analysis or neural network techniques can be too narrow in their approach and can lead to erroneous outcomes for varying scenarios. By combing both techniques, this prediction model can provide more accurate and flexible recommendations. Embedding Technical indicators will guide the investor to minimize the risk and reap better returns.

Keywords : Machine Learning, Sentiment Analysis, Stock Market

I. INTRODUCTION

This section describes the limitations of traditional approach in Stock Market analysis and lists the benefits of using machine learning and sentiment analysis.

Stock market is a very volatile in-deterministic system with vast number of factors influencing the direction of trend on varying scales and multiple layers. Efficient Market Hypothesis (EMH) states that the market is self-correcting i.e. current stock price reflects the most relevant cumulative price which is nether undervalued nor overvalued and any new information is instantly depicted by the price change. In layman's term "The market is unbeatable ", as you cannot gain any advantage over the market but existing research proves otherwise. It is possible to predict the market trends by analyzing the patterns of stock movement. Traditional approach applies the following models for this.

Fundamental analysis

This approach focuses mainly on a company's past performance and credibility. Performance measures like P/E ratios are utilized to filter stock, which may incline towards a positive price surge. This approach is based on theory that profitable companies will continue to be so because of uptrend influenced by rewarding nature of the market.

Technical analysis

This approach is based on predicting the future prices by applying time series analysis on previous trends. Statistical techniques such as Bollinger Bands, Simple moving averages etc. are applied to predict the successive trends.

Modern approach to Stock market analysis

Computer science provides us with innovative tools for Machine learning like SVM and EML, which can analyze and perform knowledge discovery at large scales in short amount of time. Two approaches for

prediction of stock market are proposed in this research.

Qualitative Analysis

News feeds regarding stock market highly affect the market trend and thus forms a downhill movement in case of a negative news. Thus, the media / social network and stock market data are highly coupled and make the system more unpredictable. Existing research points out that in case of crisis, stocks mimic each other and lead to market crashes [1]. Nowadays, twitter has come forth as the most reliable and fastest way of consuming media. With combined resources of news feed and twitter feed, general population sentiment about a company can be highlighted. Text mining and sentiment analysis are useful tools for such a high scale analysis.

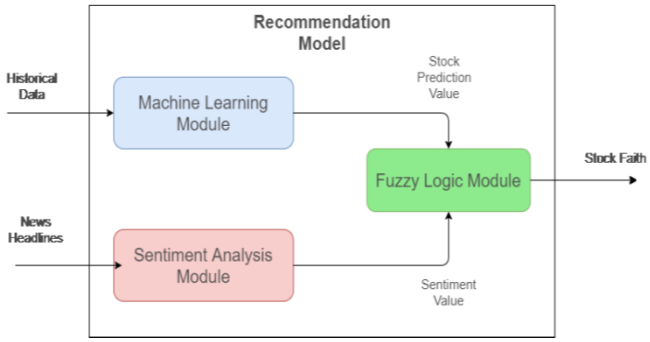
Quantitative Analysis

Historical data is now readily available for most markets. Using this dataset, we can apply multiple machine learning models to give accurate results for future investments. These models can be trained for individual stocks with adjusted bias for most reflective features. These models can also be trained to work in different scenarios and overall market movement.

Traditional approach focuses on fundamental analysis and technical analysis to predict the market at a large scale which rarely translates to low level individual stock prediction but it can be clearly observed that individual stocks contribute to whole market movement rather than the other way around. With technology advancing at such a rapid pace and abundance of computing power we can now easily strive towards a comprehensive system to accurately predict the market trend and reap beneficial financial returns. Existing research proves that modern approach outperforms traditional approach and can output the most accurate results

II. METHODS AND MATERIAL

- The aim of this project is to build an application, which outputs accurate recommendations in a quantifiable manner. For this purpose, 3 modules are implemented which are as follows:
- Machine Learning module
- Sentiment Analysis module
- Fuzzy logic Module



III. MACHINE LEARNING MODULE

The purpose of this module is to output Stock Prediction value. Stock Prediction value is the strength of difference in opening price and closing price. For this, we need to predict the closing price of the stock. This is achieved by applying Machine Learning on Historical data of the stock. Research in [3] affirms that maximum number of features required to accurately predict a stock's closing price for a specific day are given as follows.

1. Opening price of prediction day
2. Lowest and highest prices of the prediction day
3. Simple Moving Average
4. Exponential moving average of opening and closing prices of the prediction day
5. Exponential moving average of lowest and highest prices of the prediction day
6. Bollinger Bands of opening and closing prices of the prediction day
7. Bollinger Bands of lowest and highest prices of the prediction day

The training data is then fitted by a machine learning module and is used to predict the closing price of testing data through supervised learning. There are many regressors available **scikit learn** library. Their accuracy was measured in terms of percentage error rate with accuracy calculated as shown Equation (1).

$$\begin{cases} \dfrac{|\text{Predicted Closing Price} - \text{Actual Closing Price}|}{\text{Actual Closing Price}} \times 100 < \alpha & \text{Accurate} \\ \text{else} & \text{Inaccurate} \end{cases}$$

Where α is the acceptable error rate.

On Finding accuracy for α=2 and 5, the accuracies observed are illustrated in Table 1 and Table 2

Table 1. Accuracy table for Closing price prediction (Error rate less than 2 %)

Classifier	Accuracy
Lasso	40.79 %
LassoLars	51.61 %
Elastic Net	40.79 %
Ridge Regressor	85.4 %
SVR (kernel = linear)	0.97 %
SVR (kernel = RBF)	0.97 %
Random forest	15.44 %
Ada boost	3.99 %
Decision Tree	3.67 %

Table 2. Accuracy table for Closing price prediction (Error rate less than 5 %)

Classifier	Accuracy
Lasso	64.03 %
LassoLars	72.49 %
Elastic Net	64.03 %
Ridge Regressor	94.2 %
SVR (kernel = linear)	2.37%
SVR (kernel = RBF)	2.37 %
Random forest	29.49 %
Ada boost	7.49 %
Decision Tree	9.29 %

As it is obvious that Ridge Regressors give most accurate outcome for our dataset, it was selected to be used as the regressor for Machine Learning module to provide the Stock Prediction value. The formula in Equation(2) gives the Stock Prediction value.

$$\left(\frac{\text{Actual opening price} - \text{Predicted closing price}}{\text{Actual opening price}}\right) \times 100 + 50$$

IV. LITERATURE SURVEY

Mehak Usmani et al in [2] proposed an intuitive idea of combining results from historical data, news and twitter feed sentiment analysis. This dual approach predicts the stock market trend with high accuracy. It uses technical analysis like ARIMA and SMA to get an idea of the market trend.. This research considers other factors like depreciation and exchange rates. This research utilizes technical analysis for prediction, which has been proven inferior to machine learning in terms of accuracy. Machine learning can handle noise and lack of information more efficiently. This approach has chances of inaccuracy for market scenarios not covered in training data.

The work proposed in [3] by Rodolfo C. Cavalcante et al improves upon previously existing trading rules and produces results better than research proposed before. This research uses multiple proven market strategies to stimulate a real time autonomous trader. This research focuses on short-term gains, which is excellent for hands off trading. Their model accumulates lot of revenue by trading in small time frames (minutes) Improvements can be made on choosing more features and making it more flexible

In [4], Paul D. Yoo et al investigate the success of machine learning models and event driven models like sentiment analysis in predicting the stock market trends. It also illuminates the fact that macro-economic conditions like International and political events affect market trends and need to be taken into consideration Alexander Porshnev et al in [5] states that addition of twitter sentiment analysis doesn't add any valuable information to the prediction model and doesn't increase the accuracy. Thus, this research takes news feeds into consideration to add credibility to sentiment analysis.

The research done by Dongning Rao et al in [6] provides great insight into proper implementation of

sentiment analysis. They propose increasing the size of corpus (training data) with each test. This is done by adding non-polarizing words found in the test data not present in the corpus. Thus, making the training data more efficient with each successive testing.

V. CONCLUSION

This study presents a proposal to use the decision tree classifier on the historical prices of the stocks to create decision rules that give buy or sell recommendations in the stock market. Such proposed model can be a helpful tool for the investors to take the right decision regarding their stocks based on the analysis of the historical prices of stocks in order to extract any predictive information from that historical data. The results for the proposed model were not perfect because many factors including but not limited to political events, general economic conditions, and investors' expectations influence stock market.

As for the future work, there is still big room for testing and improving the proposed model by evaluating the model over the whole companies listed in the stock market. Also, the evaluation of a larger collection of learning techniques such as neural networks, genetic algorithms, and association rules can represent a rich area for future investigation. Finally, reconsidering the factors affecting the behavior of the stock markets, such as trading volume, news and financial reports which might impact stock price can be another rich field for future studying.

VI. REFERENCES

[1]. Savinderjit Kaur and Veenu Mangat (2012) Applications of Data Mining in Stock Market. Journal of Information and Operations Management ISSN: 0976–7754 & E-ISSN: 0976–7762, Volume 3, Issue 1, pp-66-68.

[2]. Al-Radaideh, Qasem & Abu Assaf, Adel & Alnagi, Eman. (2013). Predicting Stock Prices Using Data Mining Techniques.

[3]. K. Senthamarai Kannan, P. Sailapathi Sekar, M.Mohamed Sathikand P. Arumugam, Financial Stock Market Forecast using Data Mining Techniques.

[4]. Chin-Yin Huang and Philip K.P. Lin,Application of integrated data mining techniques in stock market forecasting

[5]. Stock Market Prediction: Using Historical Data Analysis ,Vivek Kanade ,Bhausaheb Devikar, Sayali Phadatare ,Pranali Munde and Shubhangi SononeI.

[6]. Chouhan, Lokesh & Agarwal, Navanshu & Parmar, Ishita & Saxena, Sheirsh & Arora, Ridam & Gupta, Shikhin & Dhiman, Himanshu. (2018). Stock Market Prediction Using Machine Learning. 10.1109/ICSCCC.2018.8703332.

[7]. Shetty, Nisha & Pathak, Ashish. (2017). Indian Stock Market Prediction Using Machine Learning and Sentiment Analysis. 10.1007/978-981-10-8055-5_53.

Cite this article as :

Karunesh Makker, Prince Patel, Hrishikesh Roy, Sonali Borse, "Stock Market Prediction using Data Mining Techniques", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 2, pp. 310-313, March-April 2020. Available at doi : https://doi.org/10.32628/CSEIT206290
Journal URL : http://ijsrcseit.com/CSEIT206290