

AUTOMATED STOCK TRADING USING MACHINE LEARNING

Anup Munde

Pune Institute Of Computer technology,
Pune, India.

Harshal Kaute

Pune Institute Of Computer technology,
Pune, India

Ashwin Jadhav

Pune Institute Of Computer technology,
Pune, India.

Ajay Gosavi

Pune Institute Of Computer technology,
Pune, India.

Abstract: Prediction of stock market movements is a well-known problem of interest. Today's social media are fully representing the public sentiment and opinion about current events. In particular, Twitter has gained a lot of attention from researchers for studying public sentiment. Stock market prediction based on public sentiments expressed on Twitter has been an intriguing area of research. The approach, through sentimental analysis, is to see how well stock prices change i.e. the rise and fall in the opinions of the people they express on Twitter is correlated. Sentimental analysis helps in analyzing the public sentiments on Twitter, this approach is our approach through using make of sentimental analysis. Another approach in the same subject of our project uses technical analysis. We model the movement of the stock price based on these input characteristics and solve it as a regression problem in a multi-kernel learning regression framework. Machine learning coupled with fundamental and / or technical analysis also gives satisfactory market share forecast results. We evaluated the end-of-day buy-sell decision-making model, also known as intraday trading.

Keywords—Machine learning, Stock price, Stock market, Supportvector machine, Natural language processing

I. INTRODUCTION

Correct forecasting of stocks can to huge profits Seller and broker. Often it is highlighted that the prediction is chaotic rather than random, which means that it can be predicted by carefully analysing the history of the associated stock market. Machine learning is an efficient way to represent such processes. It predicts a market value closer to the tangible value, thus increasing the accuracy. The introduction of machine learning in the field of inventory forecasting has called for much research because of its efficient and accurate measurement. In this project, supervised machine learning is used on a dataset obtained from Yahoo Finance.

II. METHODOLOGY

1. Output of sentimental analysis can't be the only deciding factor in predicting process. Use of technical analysis to

check the actual movement of the market is essential. Many times market movement can't be predicted due to pseudo force and it can't sustain more or less than the upper and lower circuit. So here investors should be aware of retracement of market movement at any time.

2. In the proposed model, we extract features from sentimental analysis results mining from Twitter and from time series data source (price, volume) .

3. Using the historical data of a stock to analyse how the movement of the stock has been in the last one year it becomes easy to predict the future movement of the stock.

III. ALGORITHM AND METHODOLOGIES

1. NATURAL LANGUAGE PROCESSING

Natural language processing (NLP) may be a subfield of linguistics, computer science, information engineering, and AI, related to the interaction between computer and human (natural) languages, in particular the processing and analysis of large amounts of linguistic data. The way to program a computer to do. The challenges of language processing often include speech recognition, language comprehension, and language training.

2. NAIYE BAYES ALGORITHM

The Naive Bayes classifier uses applied mathematics to classify data. Nave's basic classification algorithms use Bayes' theorem. The main insight into Bayes' theorem is that the probability of an event can be adjusted as new data is introduced. The one that classifies a naive Bayes is the naive belief that each of the properties of a point of knowledge is independent of each other. A grader sorting fruit into apples and oranges would know that apples are red, round, and a certain size, but will not outshine all of them.

Chat at times. Oranges are round too, after all

3. TIME SERIES PREDICTION

Time series analysis involves methods of statistical analysis. Data to extract meaningful statistics and other information characteristics Time series predict that

Using a Model to Predict Previously Supported Future Values

Observed the values. While multivariate analysis is typically used in such a way that the current values of one or more independent data on the test theories again affect the current value of the series, this type of study of your time series is called "time series analysis. called" no, "which focuses on comparing the values of a multiple statistical or dependent time series at different times. Constrained statistical analysis is the analysis of interference on one static.

4. AUTO REGRESSION

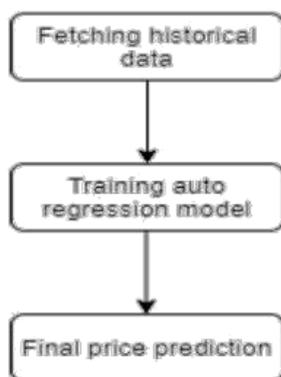
Autoregression is essentially a time series model that predicts that subsequent time steps are worth as input of past time steps or a regression of x over y as input of past data This is a simple concept that can result in an accurate prediction of how far your time series issues will spread, which is a great way to estimate the future price of a company's shares.

5. ARIMA FROM STATS MODEL

Auto regressive integrated moving average (ARIMA) is the statistical model for time series prediction. The standard notation for the model is ARIMA (p, d, q) where p is the total observation of the delay, also called the order of delay, d occurs so many times that the raw observations are differentiated, also called degrees of deference, and q is the size of the moving average window.

IV. IMPLEMENTATION

1. TECHNICAL ANALYSIS



2. SENTIMENTAL ANALYSIS

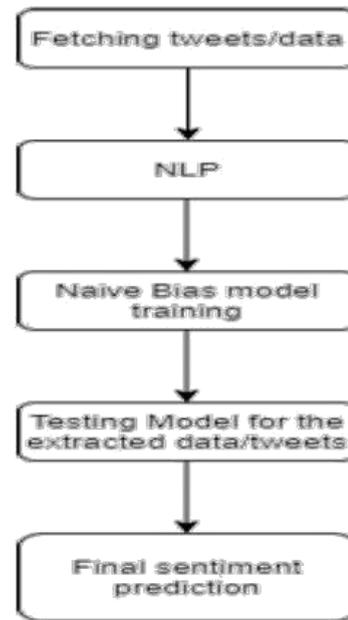


Fig 2: Sentimental analysis

V. IMPLEMENTATION TECHNIQUES

1.Raw data pre-processing component:

It takes content of news and comments as input, and also takes historic price and volume data as raw data for further process the content of tweets require cleaning which helps us in creating the sentiment analysis.

2. Feature extraction component:

It breaks down each and every word in the fetched tweets and compares it with bag of words, if that word exists in bag of words then it assigns it labels from that respective bag of words.

3. Sentiment analysis component:

It analyzes the overall sentiments (objective, subjective, positive, negative), a threshold value for overall sentiments is set to classify each sentiment. The output of process will be the actual percentage of positive and negative sentiment from the fetched data.

4. Technical analysis component:

It formulates different indicators based on price and volume. These indicators generate signals movement of market. Then we combine outputs of each signal using simple combination or using Neural Network combination. The trend of market is predicted as output of technical analysis.

VI. CONCLUSION

The proposed system will help beginner traders as a decision support tool and help them take decisions accordingly. Collective analysis of news regarding market have been made easy using machine learning algorithms.

VII. REFERENCES

- [1] "Approaches, Tools and Applications for Sentiment Analysis Implementation", Alessia D'Andrea, Fernando Ferri, Patrizia Grifoni, Tiziana Guzzo, International Journal of Computer Applications (0975 – 8887) Volume 125 – No.3, September 2015.
- [2] "Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers", Ghaith Abdulsattar A.Jabbar Alkubaisi, Siti Sakira Kamaruddin & Husniza Husni, Computer and Information Science; Vol. 11, No. 1; 2018, ISSN 1913-8989 E ISSN 1913-8997
- [3] "Stocks Market Prediction Using Support Vector Machine", Zhen Hu, Jie Zhu, and Ken Tse, 2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering
- [4] "Sentiment Analysis in Twitter using Machine Learning Techniques", Neethu M S, Rajasree R, 4th ICCCNT 2013, July 4 - 6, 2013, Tiruchengode, India.
- [5] "The Integration of Fundamental and Technical Analysis in Predicting the Stock Price", Isnaini Nuzula Agustin, Volume 18, Nomor 2, Mei 2019, pp 93-102. Studi Manajemen, Fakultas Ekonomi, Universitas Kristen Maranatha. ISSN 1411-9293 | e-ISSN 2579-4094.
- [6] "STOCK MARKET FORECASTING TECHNIQUES: LITERATURE SURVEY", Vivek Rajput, Sarika Bobde, International Journal of Computer Science and Mobile Computing, Vol.5 Issue.6, June- 2016.

- [7] "Quantitative Analysis of Stock Market Prediction for Accurate Investment Decisions in Future", Surbhi Sharma, Baijnath Kaushik, Journal of Artificial Intelligence, ISSN 1994-5450, 2018. Baijnath Kaushik, Journal of Artificial Intelligence, ISSN 1994-5450, 2018. Baijnath Kaushik, Journal of Artificial Intelligence, ISSN 1994-5450, 2018.