International Journal of Research in Advent Technology, Special Issue, ICATESM 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

Integrating Stock Twits and News Feed with Stock Data for better Stock Market Prediction

Prof. A.G.Sawant , Aditya Dhawane , Gopal Ghate , Piyush Lohana , Umed Kishan
Computer Department , PVG's COET , Pune University
Email: adityadhawane17@gmail.com ,piyushlohana19@gmail.com

Abstract-Stock Market is very dynamic and volatile field , because of ever advancing technology the scope of stock market is rising. Opinions of people which is sentiment can cause the change. In this paper, we propose the prediction of stock prices, rising or falling ,using sentiments. Also combining traditional Predicting method helps to figure whether to buy/hold/sell the stock. For prediction SVC and Naive Bayes is used providing an accuracy of 70-75% .

Keywords-Stock market, sentiment analysis,News analysis, Opinion Mining, machine learning, prediction.

1. INTRODUCTION

Stock prediction is one of the most famous machine learning problem. Stock Market is considered as the most unstable because of its uncertainty in nature. It's because of its uncertainty people tend to avoid stock trading. But research throughout the years shows that its not fully unpredictable. Using certain Machine Learning Algorithms we can predict the stock prices by the recognizing the patterns in its behaviour. With further research in this field [1] it's been proved that the market sentiments do have a certain impact on a stock's price. The recent example being the Tesla incident where a single tweet by Elon Musk to take Tesla private resulted in huge drop in its share prices. This is a prominent example showing that the sentiments do have a significant impact on the stock prices.

Sentiment Analysis is the technique of extracting the opinion of an individual behind the message. The sentiments derived from a message are categorised into 3 types: positive, negative, neutral. Natural language processing is used to get sentiments from a tweet. Earlier stock predictions have been made[2]using Sentiment analysis on Twitter data but the problem with Twitter data is the data can be fake as people who want to make a certain share go up can makeup fake tweets on that stock. So, for our project we have used Python's news API and Stocktwits data.

Stock twits is a platform solely made for Traders who discuss the fate of different stocks on a daily basis. On the other hand Python's news API gives us data from various renowned sources like "The New York Times", "CNBC", "Bloomberg" and many more. We need to download the news API python library and the maximum limit for requests to be made is 1000/day which is more than enough. For getting the data through Stocktwits we need to make a developer's account after which you are given the URL to their various methods. We have used the Symbols method under the Stream API.

After getting the data, the sentiment score must be calculated from the tweets which will be used for prediction as textual data can't be used for prediction. Now, for Sentiment Analysis many Machine Learning models can be used like Naive Bayes, SVM. We have used Python library called Text Blob from python's NLTK package. The Sentiment score ranges between -1 to 1.

The Stocktwits API gives the latest 30 tweets of a single day and the Python News API gives us the past 7 days data from the specified sources.

2. HISTORY & BACKGROUND

Though the idea of ML in Finance has been in the market for more than a couple of years but using external parameters into stock prediction is something which distinguishes from the traditional techniques. The NSE website shows us the techniques used in NSE(national stock exchange).The sentimental analysis is still not a part of the prediction mechanism. Though big names like JP Morgan have considered ML into finance by launching CO-in which is a Contract Intelligence platform. Though companies have started considering use of AI in Finance but they are limited to Security & chatbots. A very few companies have invested in AI for trading purpose.

- *ADVANTAGES:*

1. Increases the Speed
2. Reduces the Labour Cost
3. Automation

- *DISADVANTAGES:*

1. Very few investments made in Trading part of Finance which is a major part of Finance.
2. Though the companies using Algorithmic trading make use of only Internal Parameters

Opinion Mining online social media has attracted significant research interest across many different disciplines such as political science, finance, business, health care, etc. Much progress has been made on sentiment analysis of a variety of text contents found in open forums & microblogs. Recently there have been many efforts to investigate the relationship

between stock market behaviour and public sentiment on social media. It has been shown in many studies that post sentiments on social media are correlated with market movement and can be used to improve the accuracy of stock price prediction. So, we focus on using social media platforms for collecting data & using it in our prediction model. The social media data is openly made available through the company's respective APIs. The opinions will be categorised into 3 parts: Negative Neutral Positive The paper[1] focuses on only sentiment analysis for prediction but we propose the Use of Sentiment along with Traditional Indicators. On Observing the current market trends we have reached the conclusion that our model will make use of real time market data from websites like Yahoo finance, Google Finance. For sentiments, we will be using Python News API & Stocktwits data.

3. DESIGN ISSUES

- To get All Day Stock data:

The Stock data available from the API are only for those days for which the market is open whereas the Sentiments for that stock are available throughout the week. So to handle this we calculated the missing dates from the Stock date and the missing values are calculated by using mean of other values. For example the missing value for closing price of Sunday would be the mean of other days closing price values.

- Converting JSON response to Python format:

The response from the respective APIs was in JSON format which was a list of dictionaries. So, the challenge was here to extract the required information from the JSON response like "Published At" value which was the date on which the tweet was posted. The "Description" value which was the body of the message.

- To get Sentiment data:

The Sentiment data was collected from Stock twits and Python News API Sources. The body of the message consisted of stop words like "#,$,%" which needed to be removed .So, in first step all the stop words were removed. In the second step the Sentiment score for each tweet was calculated using the Text-Blob library of Python.

- Combining Stock & Sentiment data for analysis:

Here the features from each file were extracted and combined into one single file like "Date, Open, Close, Sentiment Score". The problem here was that the Sentiment data was available for only about a month which we have collected so far whereas the stock data is available for the past 9-10 years.

4. METHODOLOGY

The idea upon which we have made this project is that there exists a correlation between the market sentiments and the stock prices given the same timeline. We have mathematically proven this by plotting bar and line charts using our collected data from respective APIs. (bar chart fig.)

Following are the steps involved:

1. Stock Market Data Fetching & Pre-processing
The stock data collected was missing for the days when the market was closed. So, to fill up those dates We used Holidays API of python for finding out the public holidays on which the market was closed. For Sundays and Saturdays we used "weekdays" method of Python's Datetime module. The stock data for each stock was stored in a CSV file with features being "Date", "Open", "Close", "High", "Low".

2. Sentiment Data Pre processing
The sentiment data was collected from Stock Twits and Python News API. The sentiment score was calculated using Text Blob library for all the tweets made on a particular day. So, we calculated the aggregated sentiment score for a single day. The Sentiment data was stored for each stock in a CSV file with features being "Date", "Sentiment".

3. Merging of sentiment and stock price data
The stock data and sentiment data was merged according to the respective dates. The columns in the final CSV file were "Date", "Open", "Close", "High", "Low", "Sentiment".

4. SVM & Gaussian NB Model for Stock prediction
We used SVM model to suggest whether a person should buy or sell a share. For training the model we used the merged data from step 3. In the data a new attribute of 'decision' was added that is prediction variable. The value of this attribute for training set was calculated as follows: if the stock and sentiment both are positive the decision is buy otherwise sell. Whole data was split into training and test set and then model was trained with training set and evaluated by calculating prediction accuracy of test set.

5. RESULT & ANALYSIS

The first observation we found is the correlation between the stock prices and the market sentiments in a given timeline.

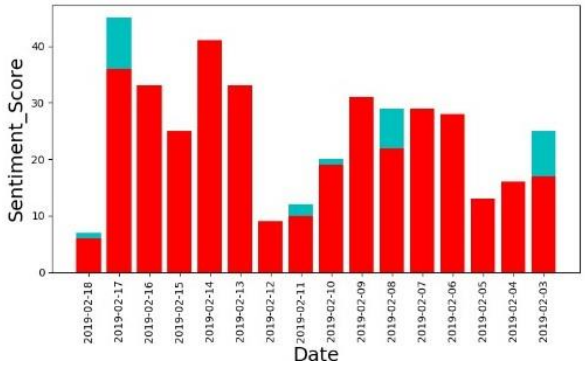

Figure 1.Sentiments Regarding Stock Price .

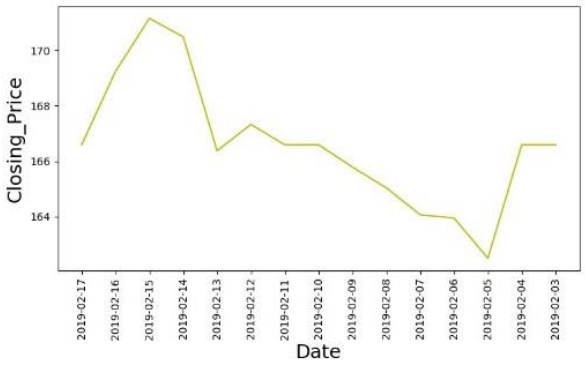

Figure 2.Closing Price of Same Stock.

The Figure 3. shows the predictions we made using the SVM model

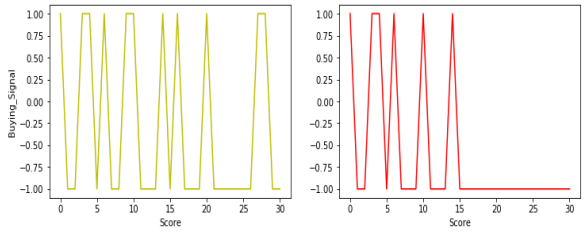

Figure 3. SVM Prediction Model

The Figure 4. shows the predictions we made using the Gaussian NB model

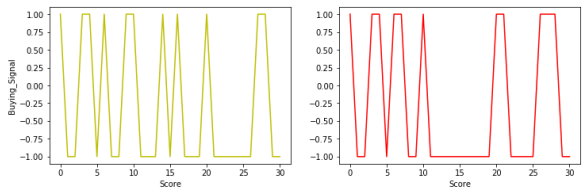

Figure 4. Gaussian NB Prediction Model.

6. CONCLUSION

Early research on Stock Market prediction were totally based on random walks and numerical prediction but with the introduction of behavioural finance, the people's belief and mood were also considered while predicted about stock movement. Making it more efficient we used the idea of sentiment analysis of Stock Twits &; News Feed through machine learning models .We implemented the idea by collecting sentiment data and stock price market data and built SVM & Naïve Bayes models for prediction and we have achieved 72.22% 84.3% training accuracy respectively. The accuracy can be improved by increasing the dataset size. In our case the data available was only for past 30 days.

7. LIMITATIONS & FUTURE WORK

1. *Limitations*
   The Sentiment data was obtained from above mentioned APIs and was available for only 30 days. So, the Dataset size can be increased using other sources.

2. *Future Work*
   The Stocktwits data is available is only available for only 1 day. But with their partner level access(paid) the historic data can be obtained from Stock twits. The accuracy can be increased by increasing the size of the dataset. Also, tweets from important dignitaries of the respective companies can be extracted and more importance can be given to their tweets as they sure will have a significant impact on tomorrow's prices.

REFERENCES

[1] Wang, Yaojun, and Yaoqing Wang. "Using social media mining technology to assist in price prediction of stock market." 2016 IEEE International Conference on Big Data Analysis (ICBDA). IEEE, 2016.

[2] Bollen, Johan, Huina Mao, and Xiaojun Zeng. "Twitter mood predicts the stock market." Journal of computational science 2.1 (2011): 1-8.