

Predicting Stock Performance by Analyzing Emotions of the Public

Yash Jajoo

Thakur College of Engineering and Technology
University of Mumbai, India

Shridhar Kamble

Assistant Professor
Thakur College of Engineering and Technology
University of Mumbai, India

ABSTRACT

Prediction of stock markets has been a significant research area. Especially the study of changes in stock prices due to non-quantifiable factors. Here, the concept of fluctuations in the values of stocks due to people's emotional state is explored. In this approach, sentiment analysis is performed on Twitter data (tweets), the results of which are fed into a prediction algorithm along with stock data from Dow Jones Industrial Average (DJIA). Here, sentiment analysis is based on lexicons as well as heuristics and it determines the tweets' emotional polarity and classifies them as either positive or negative. Results obtained show 100% accuracy in mapping the tweets' sentiments to the change in stock prices and the average deviation between predicted and real stock values is 1.77.

Keywords

Prediction algorithm, Sentiment analysis, Stock prediction.

1. INTRODUCTION

Stock trends are very volatile and are affected by several factors like past performance of the stock, competitors' stock values, overall economic condition, etc. According to the "Efficient Market Hypothesis", prices of stocks are dependent on latest, real-time data and are random [5]. While this information is quantifiable, non-quantifiable factors also have a hand in determining the value of a given company's stock and it is much harder to make sense of such information. That is why, despite the above mentioned hypothesis receiving widespread acceptance, there has been an increase in research on establishing a relationship, with a certain degree of specificity, between non-quantifiable information and the performance of the stock market. One such piece of information is the public opinion and news about companies. Depending on whether or not people are optimistic about a company, there is a possibility that the stock price of that company will increase or decrease.

Multilayer Perceptrons (MLP) are most commonly used for predicting stock values on the basis of emotions but they are not always reliable. Here, an attempt is made to predict the price of a stock using lexicons and heuristics by taking into account the emotions underlying the posts on twitter (tweets) about a company. The tweets are classified on the basis of their "emotional polarity" (positive/negative).

2. RELATED WORK

Our work is most closely related to Bollen et al's strategy [1] in which the researchers showed that the emotional states of people are predictive of changes in stock behaviour. Nagar and Hahsler [2] demonstrated an approach which combined text mining and NLP to collect news articles and create a corpus to be analyzed. They put forward a metric called "Twitter Sentiment" and conclude that the time variation of this metric has a strong correlation with real time stock fluctuations. J. Bean [3] keywords about customer satisfaction

with airlines were tagged to gain insight into any airline's reputation. In [4], researchers show that collectively using multiple categories of tweets leads to better prediction as compared to using a single type of tweets.

3. TECHNICAL APPROACH

The proposed approach includes two steps. First is the sentiment analysis in which the tweets about a given company are classified as either positive or negative depending on their content. Twitter data (raw texts) and stock data (past "Dow Jones Industrial Average" values) are used as inputs for the model which gives the polarity of tweets and its relationship with stocks. Second, results of the analysis of the tweets' sentiments along with the stock of the company to which the tweets refer to are fed into the algorithm. The prediction process is thus dependent on the classification of tweets.

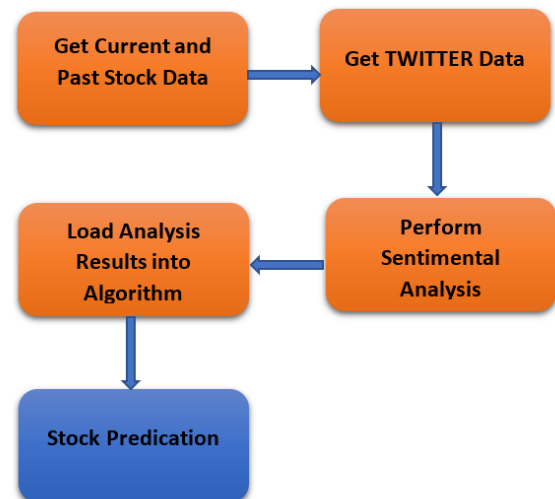


Figure 1: Process flow

3.1 Sentiment Analysis

3.1.1 Apply heuristics

This step is performed to take symbols and emoticons into account as they possess the ability to convey emotions and represent users' moods. Emoticons constitute strong indications of emotions. Regardless of the sentiment showed in the tweet, an emoticon at the last has a stronger effect [6].

3.1.2 Assess hashtags

Hash-tags are used to emphasize a part of the tweet that is of significance. Here, lexicons are matched with a hash-tag's parts. For example, the tag #notatallbad, "not" is taken as a negative lexicon, "at" and "all" are taken as lexicons conveying the strength of the emotion indicated and "bad" is taken as a negative lexicon. Combined, the hash-tag has a positive sentiment [6].

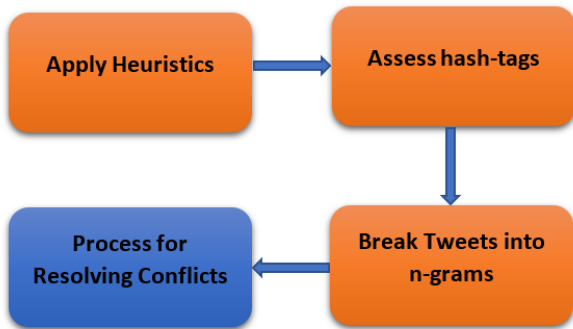


Figure 2: Steps in sentiment analysis

3.1.3 Break tweets into n-grams

The tweet is broken into n-grams (n=1,2,3,4). This is done because some emotion markers are recorded in multi-term expressions. Consider the tweet, "We want good products from this company". In this case, the text is broken into a list including "we, want, good, products, from, this, company, we want, want good, good products, products from, from this, this company, we want good, want good products, good products from, products from this, from this company, we want good products, want good products from, good products from this, products from this company". Here, "we want good products" denotes a negative sentiment whereas "good products" denotes a positive sentiment. All the n-grams are traversed and classified by the model according to the sentiment they represent [6].

3.1.4 Process for resolving conflicts

This step is performed to ensure that a tweet doesn't have opposing sentiments. In the example, "We want good products from this company", there's both positive and negative sentiments. Here, the two classifications are resolved and the tweet is classified as negative [6].

3.2 Algorithm

The sentiment classification for the tweets and the name of the stock are fed into the algorithm. Four local variables are used to store the predicted value of the stock (predict), iteration (iter), change in stock's value (change) and the difference between the no. of positive and negative tweets for the stock (dev). Variables "iter" and "change" are initialized at 0. First, the difference between the no. of positive and negative tweets for the company is stored in "dev". Second, the change in the stock's price is calculated. Since users continue to tweet about the company, the no. of tweets about the company (both positive and negative) changes repeatedly. To account for this, the two aforementioned steps are performed 5 times. Then, the average of the sum of all 5 values of the variable "change" is calculated which is then added to the current price of the company's stock to get its predicted price. The algorithm is given below,

Inputs: stock, emotion

Variables: predict, change, dev, iter

change←0

iter←0

while: iter<5

```

dev ← emotion.getPositive() – emotion.getNegative()
change← change+(dev*100/emotion.getTotalTweets())
iter++
end
change←change/5
predict←stock.currentPrice + change
return predict
  
```

Now, suppose that a company's stock is currently at \$100. The difference between no. of positive and negative tweets for 5 iterations is 100, -115, 90, 130 and 85 respectively. The total tweets for 5 iterations is 1000, 1250, 1890, 2600 and 3400 respectively.

So, following the algorithm, the value of the variable "change" after 5 iterations is $10+(-9.2)+4.76+5+2.5 = 13.06$. Then, taking average of this value, we get, $13.06/5 = 2.61$. The stock's predicted value is $predict = 100+2.61=102.61$

4. RESULTS

The changes in the real values of the stocks are consistent with the sentiment classification. The stocks with a negative classification saw a decline in their value while those with a positive classification saw an increase.

Table 1. Comparison of sentiment and change in value

Stock	Sentiment	Value
Adidas	Negative	Decrease
Cisco Systems	Negative	Decrease
IBM	Positive	Increase
Nike	Negative	Decrease
Tesla, Inc.	Negative	Decrease
Netflix	Positive	Increase
eBay	Positive	Increase
Baidu, Inc.	Negative	Decrease
Microsoft Corp	Negative	Decrease
CalAmp Corp	Negative	Decrease
Qualcomm Inc.	Negative	Decrease
Starbucks Corp	Negative	Decrease
Veeco Instruments	Negative	Decrease
NVIDIA Corp	Negative	Decrease

The real and predicted stock values are given in the table below.

Table 2. Comparison of predicted and real values

Stock	Value Predicted (\$)	Real Value (\$)	Deviation
Adidas	177.5	177.30	0.2
Cisco Systems	32.65	32.00	0.65
IBM	161.69	166.00	4.31
Nike	56	55.90	0.1
Tesla, Inc.	305.52	303.00	2.52
Netflix	139.76	143.50	3.74
eBay	33.86	34.00	0.14
Baidu, Inc.	177.81	177.50	0.31
Microsoft Corp	65.04	65.00	0.04
CalAmp Corp	18.01	16.50	1.51
Qualcomm Inc	52.61	52.50	0.11
Starbucks Corp	59.04	58.50	0.54
Veeco Instruments	29.45	28.00	1.45
NVIDIA Corp	99.68	99.50	0.18

Here, the closing value of the stocks on the following day is considered. The mean deviation is 1.77. Therefore, the approach is able to classify tweets' underlying emotions with 100% accuracy. But, albeit small, there are differences between a stock's predicted and actual value.

5. CONCLUSION

The proposed approach has led to promising results. The accuracy observed in mapping the emotions behind the tweets to the change in actual stock prices is 100%. However, there is a small difference between the values predicted and the actual values. The predicted stock prices were close to the actual ones with mean and highest deviation being 1.77 and 4.31 respectively. Because of heuristics, breaking the tweets into n-grams and taking them into consideration individually, the analysis of sentiments is robust and the method is able to handle varied language constructs.

6. FUTURE WORK

The analysis, however, is limited to English speaking Twitter users and separate models need to be used for other languages. The process here includes stocks listed on DJIA only but those listed on other stock exchanges can also be accommodated as long as their data is used as well.

REFERENCES

- [1] J. Bollen and H. Mao., "Twitter mood as a stock market predictor," *IEEE Computer*, vol. , no. 44(10):91–94
- [2] Anurag Nagar, Michael Hahsler, "Using Text and Data Mining Techniques to extract Stock Market Sentiment from Live News Streams," *IPCSIT* vol., no. XX (2012) IACSIT Press, Singapore
- [3] J. Bean, "R by example: Mining Twitter for consumer attitudes towards airlines," In *Boston Predictive Analytics Meetup Presentation*, Feb 2011
- [4] Yauheniya Shynkevich, "T.M. McGinnity, Sonya Coleman, Ammar Belatreche, Predicting Stock Price Movements Based on Different Categories of News Articles," 2015 *IEEE Symposium Series on Computational Intelligence*
- [5] T. Rao and S. Srivastava, "TweetSmart: Hedging in markets through Twitter," 2012 *Third International Conference on Emerging Applications of Information Technology*, Kolkata, 2012, pp. 193-196. doi: 10.1109/EAIT.2012.6407894.
- [6] Clement Levallois, Umigon. Retrieved from <http://www.clementlevallois.net/download/umigon.pdf>