Stock Market Prediction of FAANG Stocks Using Twitter

Daniel Forgosh
Computer Science
Hood College
djf7@hood.edu

Ian Wittler
Computer Science
Hood College
imw2@hood.edu

December 18, 2018

Abstract

Twitter is a social media platform which can be used to analyze numerous social patterns [1]. By querying tweets based off of their relationship to companies found in the FAANG (Facebook, Amazon, Apple, Netflix, and Google) stock group, the overall sentiment of these companies through Twitter can be extracted. Then, through the use of Random Forest classification, a model can be derived which predicts whether the price of a stock will rise or fall based off of the Twitter sentiment data.

The main goal of this paper is to determine whether or not Twitter plays a role in predicting a stock's price. By comparing a model with Twitter data to a model without Twitter data, it can be determined whether or not Twitter Sentiment is useful for stock prediction. Mixed results were concluded from this paper and leave room for future work to get a more confident answer.

CCS Concepts

• Computing methodologies Classification and regression trees • Computing methodologies Modeling methodologies

Keywords

Twitter mining; sentiment analysis; random forest; stock price prediction

1 Introduction

In the United States, the stock market is worth north of 30 trillion dollars. In fact, the U.S. stock market exchanges billions of dollars daily. A vast amount of data points can be collected in stock markets, including stock prices (each and every second), overall trade volume on a given day, and many more variables. Naturally, the stock market is a common interest for data scientist and presents a constantly evolving challenge that can be rewarded by financial benefit or punished by incorrect predictions.

There is no precise method to know exactly how many daily trades are controlled by machines, but according to one fund manager, 80% of daily trades are done by machines [2]. This idea invokes the following question: what factors control these machine's algorithms that decide when to buy and sell stocks? Company performance, company sentiment, industry performance, and multiple other economic factors all contribute to what makes a stock rise or fall [3]. In other words, supply and demand are what makes a stock more or less valuable. This paper focuses on how public sentiment affects a companies stock prices.

2 Related Work

2.1 Stock Prediction

People have been trying to predict the stock market since its creation for monetary gain. Predicting the stock market is inherently a difficult task because of the volatility of factors that influence the value of stocks. One approach to predicting the stock market is using data mining techniques.

By using past stock data, a predictive data mining technique has been used multiple times to attempt to predict the stock market [4]. Although this method shows some positive results, it is an unreliable method due to the stock market price fluctuating in what can be considered a random manner. To deal with the randomness in stock market prediction, the randomness can be studied along with past stock market data to help predict the stock market [4].

Since stock prices are based on supply and demand, breaking news can indicate public opinion on specific stocks. A data set of breaking news headlines can be used to help predict how a specific headline will affect the prices of specific stocks [5]. Another approach to predicting the stock market is to analyze internet traffic associated with specific stocks [6]; it is also proposed to analyze the sentiment of the internet traffic associated with specific stocks.

2.2 Twitter Sentiment

Twitter is a large social network platform that allows users to post short messages called "tweets." The sentiment of tweets can be determined through many data science techniques. Using the sentiment of tweets can help with predicting many topics including stock market prediction [7, 8, 9, 10, 11, 12].

When analyzing results, there may be spikes in the data. There could be many reasons for spikes in data [13]. One reason is a real-world event may have triggered emotions within a certain community, state, country, or even world. By looking at the date of specific real-world events, this theory can be tested [14]. For example, December 25 can be analyzed to see if there is a spike in positivity on Twitter since Christmas is thought to be a happy day.

Something that could cause a spike in data, not caused by a real-world event, is a controversial topic. Arguments can be a source of emotions, positivity and negativity. Controversies can be found through sentiment analysis [15]. By using sentiment analysis, not only can positivity and negativity be determined, but it can also be determined if the positivity or negativity is a result of a controversy.

3 Methods

3.1 Data Collection

3.1.1 Stock Data

Alphavantage provides an API that allows people with a developer account to query and download stock information. The stock information can be for a specific company, a subsection of the stock market, such as the technology industry, or the overall stock market. The stock information it provides is comprehensive and contains daily data from the past few months. The specific attributes of stock data that were collected are date, open, high, low, close, and volume. Open is the price of the stock when the stock market opened for the day; high is the highest price of a stock from the day; low is the lowest price of a stock from the day; close is the price of the stock when the stock market closed for the day; volume is the total trade activity of a stock from the day.

3.1.2 Twitter Data

Twitter provides an API that allows people with a developer account to query and download tweets on Twitter [16]; however, the Twitter API does have restrictions. It will only allow people with developer accounts to make 350 calls per hour [17].

Since Twitter restricts the length of posts users are allowed to post by a certain amount of characters, tweets do not provide a complete and reliable data set. Compared to other kinds of data mining, most data sets that are data mined are complete. Having an incomplete data set creates a new research field within data mining [18]. There are different theories on how to data mine Twitter based on social theories from social sciences.

3.2 Sentiment Analysis

Sentiment analysis consists of processing text and categorizing it as positive, negative, or neutral. It is important to consider that because of the complexity of human language and simplicity of three categories, while sentiment analysis can be useful, it is not 100% correct [19]. There are a couple of techniques to determine the sentiment of a text. The first is to consider words individually as positive or negative. This idea generally works; however, it makes the incorrect assumption that words act alone. For instance, "not good" is definitely a negative sentiment, yet if each word is analyzed individually then the result would be neutral. This reason is why most sentiment analysis techniques consider not individual words, but multiple words as vectors in relationship to each other. Many times, how words are related to each other is more important than the word itself [20]. Natural language processing libraries are easy to locate and tons of work has been done to understand the meaning of words. So, while sentiment analysis is not 100% accurate it can achieve relatively close margins.

3.3 Random Forest Classification

Decision trees are a powerful machine learning tool. They can be trained with data to then predict future results. There is a problem with using decision trees to predict complex situations: there are other factors that play a role in the outcome. Random Forest Classification is another machine learning tool. It combines multiple decision trees together to find more accurate results that can incorporate other factors that normal decision trees are not able to [21].

3.4 Model Averaging

When working with a small data set, finding consistent results when predicting can be a difficult task. Since the data set is small, the training and testing data will not have many entries in comparison to large data sets. Since the training and testing data are small sets, there is not a large variety for a model to train and test; this can lead to varied results. A way to deal with the varied results from using a small data set is using model averaging. Model averaging is splitting the data into training and testing sets many different ways. The results of all the models created using these training and testing sets would be averaged to create the model average result. This result is more reliable to represent the model's results because it incorporates all of the varied results into one result.

4 Experiment

4.1 Pre-processing

The pre-processing phase in this experiment combined data from the stock market and data from Twitter. The method in which each set of data was extracted is explained in the proceeding sections. After data from each set was collected, the data was combined into one CSV file which can be found in the Combined_Stock_Twitter_Data folder of this project's GitHub. The two data sets were joined together via the date column of both sets.

Although twitter and stock data was collected from October 22, 2018 through December 2, 2018, since the stock market is not open on the weekends and other holidays, less days could be analyzed.

4.1.1 Stock Data

The alphavantage API was used to collect stock data on Amazon, Apple, Facebook, Google, and Netflix from October 22, 2018 through December 2, 2018; the stock data for the technology industry and overall stock market from October 22, 2018 through December 2, 2018 were also collected using this API. The Python code in getStocks.ipynb on this project's GitHub shows how the stock data was collected.

The attribute "change" can be calculated by determining if a stock gained or lost value from the day. The change attribute was calculated for the previous day instead of the day the data came from so it can add information of the movement of the stock to the previous day.

4.1.2 Tweet Data

Using the Twitter API, tweets that contain "#amazon", "#apple", "#facebook", "#google", and "#netflix" were collected from October 22, 2018 through December 2, 2018. Each company's tweets are saved into separate CSV files. The Python code in getTweets.py on this project's GitHub shows how the tweets were queried and saved into CSV files.

Analysing the tweets showed that preprocessing must be preformed before preforming sentiment analysis on the tweets. There are strange symbols and URLs contained in many of the tweets. The Python code in SentimentGenerator.ipynb on this project's GitHub shows how the tweets were cleaned.

Sentiment analysis can be preformed using the textblob library in Python. This library can analyze the cleaned tweets that are collected and determine if they are positive, negative, or neutral in their sentiment. The Python code in SentimentGenerator.ipynb on this project's GitHub shows how the cleaned tweets were analyzed to determine their sentiment.

To make the sentiment values for each tweet useful to help with the prediction of the stock market, further preprocessing is required. The number of positive, negative, and total tweets were calculated for each company for each day from October 22, 2018 through December 2, 2018. In addition. The Python code in tweetSentimentCounter.py on this project's GitHub shows how the numbers were calculated.

4.2 Results

The following results were collected for each stock model generated: accuracy, model score, and feature importance. The Python code in ClassifierUsingRandomForest.ipynb on this project's GitHub shows how the results were generated. Additionally, in order to answer the question, does Twitter sentiment play a role on stock price, two models were created, one that included Twitter data and one that did not include Twitter data. By comparing the different results from each model, it can be seen whether Twitter data adds value to predicting stock prices.

4.2.1 With Twitter Data

Using Twitter data for each of stock of interest, 75 percent of data was used to train a Random Forest model. The remaining 25 percent was used for testing. One example of results: Google's accuracy was 0.77 and model score was .95 (the model score represents how fitted the model is to the training data). For each stock, a confusion matrix was generated a seen in the below figure.

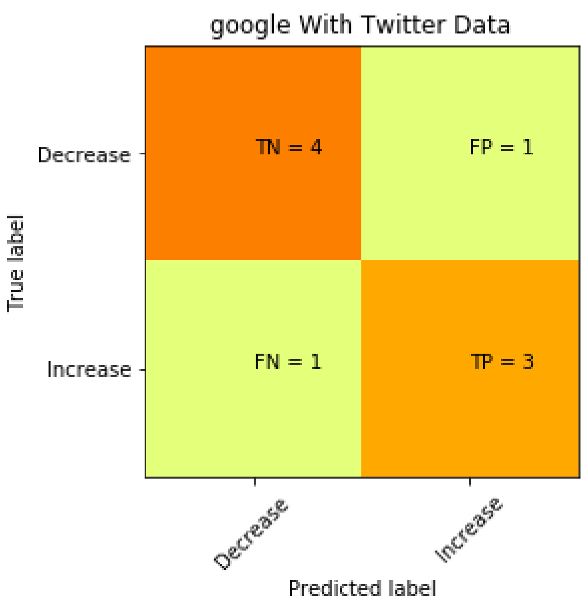

Figure 1: confusion matrix of Google's stock prediction using Twitter data

Also a chart of feature importance on the stock's model was also generated. Google's model's most important feature was the positive number of tweets in a day. Which means that the model is using the Twitter data as a primary decision factor.

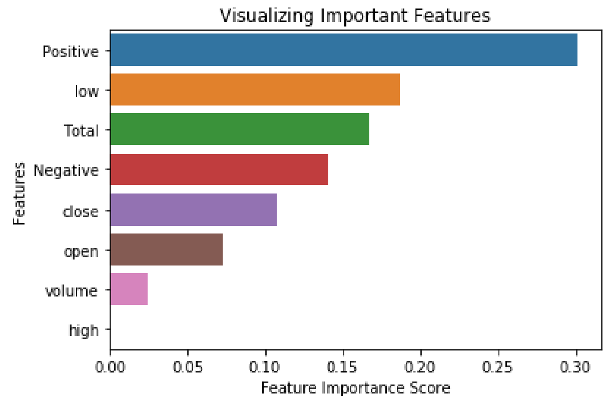

Figure 2: chart of important attributes of Google's stock prediction using Twitter data

4.2.2 Without Twitter Data

The without Twitter data models were generated the exact same way albeit without Twitter data. Continuing with the Google example, accuracy was .44 and model score was .95. Stock volume was the most important feature, where in the counter-part model volume was one of the least important feature. Obviously in this scenario the Twitter data helped to predict the stock value, however there are many things which make generalizing this statement suspect. When comparing with vs. without, not every stock was better with Twitter data, in fact more were better without the data.

4.3 Model Evaluation

Varied results for each stock were found based off of what data was split and how the decision trees were generated. This was likely amplified due to the small data set used in this project. To deal with the varied results due to the random state, model averaging was preformed. The random state was set to 1 through 100. The results for each company was averaged within the two categories of stock data with tweets and only stock data.

The following are the results from the model averaging implemented in the Python code in ModelAveraging.ipynb on this project's GitHub.

```
Amazon With Twitter Accuracy: 0.5022
Apple With Twitter Accuracy: 0.4744
Facebook With Twitter Accuracy: 0.5366
Google With Twitter Accuracy: 0.5000
Netflix With Twitter Accuracy: 0.5277
---------------
Amazon Without Twitter Accuracy: 0.5055
Apple Without Twitter Accuracy: 0.4433
Facebook Without Twitter Accuracy: 0.5367
Google Without Twitter Accuracy: 0.4311
Netflix Without Twitter Accuracy: 0.5456
```

It can be seen through the above results that Twitter data does not guarantee better accuracy. However, it should be noted that in the two models which see a significant jump from Twitter data, Apple and Google, the Twitter data tends to be the most important factors in the model. By collecting more data, the confidence level of these results could potentially be higher.

4.4 Analysis

From the above results and evaluation, it can be seen that Twitter data may or may not be useful and it seems to depend on the stock. It could be possible that some stocks Twitter sentiment more closely follow the stock then others. To truly find this answer out, more data would be needed. It should also be noted that the way in which Twitter data was collected could be improved. Currently only #stock_name was chosen, while a better data collecting method would likely improve the results. With the current method, many tweets are being missed with valuable sentiment.

5 Conclusion and Future Work

Overall more data points and improved methods would give more insight into the question sought after. While mixed results were concluded, the framework into how this experiment should be carried out was laid down. Using techniques such as sentiment analysis and random forest classification helped to connect why a stock changes to the actual results. By continuing to find different factors that affect a stock and creating models, results would improve. Considering other factors such as overall market success or monitoring internal trading would only improve the models that have been created in this paper. Additionally, the data collection method for Twitter data could be improved and stock data could be specified down to the hour, minute, or second. These types of improvements and future works would provide more depth to what makes stock prices move.

Acknowledgements

We would like to show our gratitude to both Twitter and Alphavantage for granting us access to both the Twitter API and alpha_vantage API. These two API's allowed us to gather FAANG related tweets and stock data. The data derived from these tweets and stock data were crucial in implementing the proposed stock prediction model discussed in this paper.

References

[1] Maksym Gabielkov, Ashwin Rao, and Arnaud Legout. 2014. *Studying social networks at scale: macroscopic anatomy of the twitter social graph.* In Proceedings of the 2014 ACM international conference on Measurement and modeling of computer systems (SIGMETRICS '14). ACM, New York, NY, USA, 277-288.

[2] Silvia Amaro. 2018. *Sell-offs could be down to machines that control 80% of the US stock market, fund manager says.* CNBC. CNBC, Englewood Cliffs, NJ, USA.

[3] Ontario Securities Commission. 2018. *Company news and performance.* OSC. OSC, Toronto, Ontario, Canada.

[4] Hyong-Jun Kim, Seong-Min Yoon, and Hwan-Gue Cho. 2009. *Capability and limitation of financial time-series data prediction using symbol string quantization.* In Proceedings of the 2009 International Conference on Hybrid Information Technology (ICHIT '09). ACM, New York, NY, USA, 203-208.

[5] Robert P. Schumaker and Hsinchun Chen. 2009. *Textual analysis of stock market prediction using breaking financial news: The AZFin text system.* In Proceedings of the 2ACM Transactions on Information Systems (TOIS). ACM, New York, NY, USA, Volume 27 Issue 2, February 2009 Article No. 12.

[6] Pierpaolo Dondio. 2012. *Predicting Stock Market Using Online Communities Raw Web Traffic.* In Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence

and Intelligent Agent Technology - Volume 01 (WI-IAT '12). ACM, New York, NY, USA, 230-237.

[7] Arijit Chatterjee and Kendall Nygard. 2017. *Predicting Stock Close Price Using Microsoft Azure.* In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 (ASONAM '17). ACM, New York, NY, USA, 749-757.

[8] Man Li, Chi Yang, Jin Zhang, Deepak Puthal, Yun Luo, and Jianxin Li. 2018. *Stock market analysis using social networks.* In Proceedings of the Australasian Computer Science Week Multiconference (ACSW '18). ACM, New York, NY, USA, Article No. 19.

[9] Andrew Huang. 2016. *Data mining twitter to model stock price movement.* In Proceedings of the 3rd Multidisciplinary International Social Networks Conference on SocialInformatics (MISNC, SI, DS 2016). ACM, New York, NY, USA, Article No. 32.

[10] Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. 2012. *A system for real-time Twitter sentiment analysis of 2012 U.S. presidential election cycle.* In Proceedings of the ACL 2012 System Demonstrations (ACL '12). ACM, New York, NY, USA, 115-120.

[11] Brian Dopson, Cardavian Lowery, and Deepti Joshi. 2014. *Collection and analysis of social media datasets.* Journal of Computing Sciences in Colleges. Volume 30 Issue 2, December 2014, 254-261.

[12] Michael J. Paul, Abeed Sarker, John S. Brownstein, Azadeh Nikfarjam, Matthew Scotch, Karen L. Smith, Graciela Gonzalez. 2011. *Social media mining for public health monitoring and surveillance.* doi:10.1142/9789814749411_0043.

[13] Yuexin Mao, Wei Wei, Bing Wang. 2013. *Twitter volume spikes: analysis and application in stock trading.* In Proceedings of the 7th Workshop on Social Network Mining and Analysis (SNAKDD '13). ACM, New York, NY, USA, Article No. 4.

[14] Ema Kušen, Mark Strembeck, Giuseppe Cascavilla, and Mauro Conti. 2017. *On the influence of emotional valence shifts on the spread of information in social networks.* In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '17). ACM, New York, NY, USA, 321-324.

[15] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2016. *Quantifying controversy in social media.* In Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (WSDM '16). ACM, New York, NY, USA, 33-42.

[16] Adam Marcus, Michael S. Bernstein, Osama Badar, David R. Karger, Samuel Madden, Robert C. Miller. 2011. *Tweets as data: demonstration of TweeQL and Twitinfo.* In Proceedings of the 2011 ACM SIGMOD International Conference on Management of data (SIGMOD '11). ACM, New York, NY, USA, 1259-1262.

[17] Changhyun Byun, Hyeoncheol Lee, and Yanggon Kim. 2012. *Automated Twitter data collecting tool for data mining in social network.* In Proceedings of the 2012 ACM Research in Applied Computation Symposium (RACS '12). ACM, New York, NY, USA, 76-79.

[18] Jiliang Tang, Yi Chang, and Huan Liu. 2014. *Mining social media with social theories: a survey.* ACM SIGKDD Explorations Newsletter. ACM, New York, NY, USA, Volume 15 Issue 2, December 2013 20-29.

[19] Mahnaz Roshanaei and Shivakant Mishra. 2014. *An analysis of positivity and negativity attributes of users in twitter.* In Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (SIGMETRICS '14). ACM, New York, NY, USA, 365-370.

[20] Fotis Aisopos, George Papadakis, and Theodora Varvarigou. 2011. *Sentiment analysis of social media content using N-Gram graphs.* In Proceedings of the 3rd ACM SIGMM international workshop on Social media (WSM '11). ACM, New York, NY, USA, 9-14.

[21] Fatai Anifowose, Jane Labadin, Abdulazeez Abdulraheem. 2013. *Predicting Petroleum Reservoir Properties from Downhole Sensor Data using an Ensemble Model of Neural Networks.* In Proceedings of the Workshop on Machine Learning for Sensory Data Analysis (MLSDA '13). ACM, New York, NY, USA, 27-35.