Does Twitter Influence Stock Prices?

Steve Bogdan

Abstract

Social media is extremely pervasive in society today. Its impact should be studied to determine how much it influences our day to day activities, beliefs, decision making, and spending habits. Can its influence be too far reaching, thus allowing the companies that control social media to use these platforms to sway populations to think and act according to their wishes?

Motivation

As an investor for many years, there are some adages about investing that I've heard repeatedly with "Buy on the rumor, sell on the news" being one of the most prevalent. With the speed which information is disseminated on social media, could investors be choosing information from this source over traditional sources of investment information? If investors choose information from social media rather than from traditional sources, could they be manipulated into making poor investment decisions?

Dataset(s)

Data required for this project was a list of stocks and their closing prices over a period of a week. To obtain this data I first downloaded all securities traded on the NASDAQ exchange from: ftp://ftp.nasdaqtrader.com/symboldirectory/nasdaqlisted.txt

After cleaning up the list eliminating all securities that are not stocks, the resulting list of NASDAQ stocks was loaded into http://finance.jasonstrimpel.com/ to get historical stock closing prices for desired time period.

The Twitter API was used to download tweets about the companies on a target list of companies that had the largest positive and negative price changes over the timeframe of the study.

Data Preparation and Cleaning

3 sources of data were used for this project. The first is a download of all the securities that are traded on the NASDAQ exchange. This dataset was used to generate a list of stock tickers that would be used to lookup stock closing prices over a period of a week. All non-stock records needed to be removed and the dataset reduced to just the column of stock ticker symbols.

The second dataset is the closing prices of all the stocks generated from the first dataset. This dataset was prepped by removing all stocks with a price below $25 and then the dataset was transposed so that each date in the data was a column so that each stock record was a row.

The third data set was tweets about the companies selected for analysis based on having a large change in price over the timeframe of the study. Many times a group of stocks are mentioned in a tweet, so the tweets were filtered so that only tweets about companies on the target list and were it was the only company mentioned in the tweet was selected.

Research Question(s)

Can Twitter be used to predict stock price increases or declines?

Methods

After the stock closing price data was collected the data was sorted according to price change over the timeframe of the study. The top 30 biggest gainers and top 30 biggest decliners were chosen for analysis. Tweets for the target 60 companies were downloaded, with tweets for the 30 gainers designated as having positive sentiment and the 30 losers designated as having negative sentiment for purposes of training.

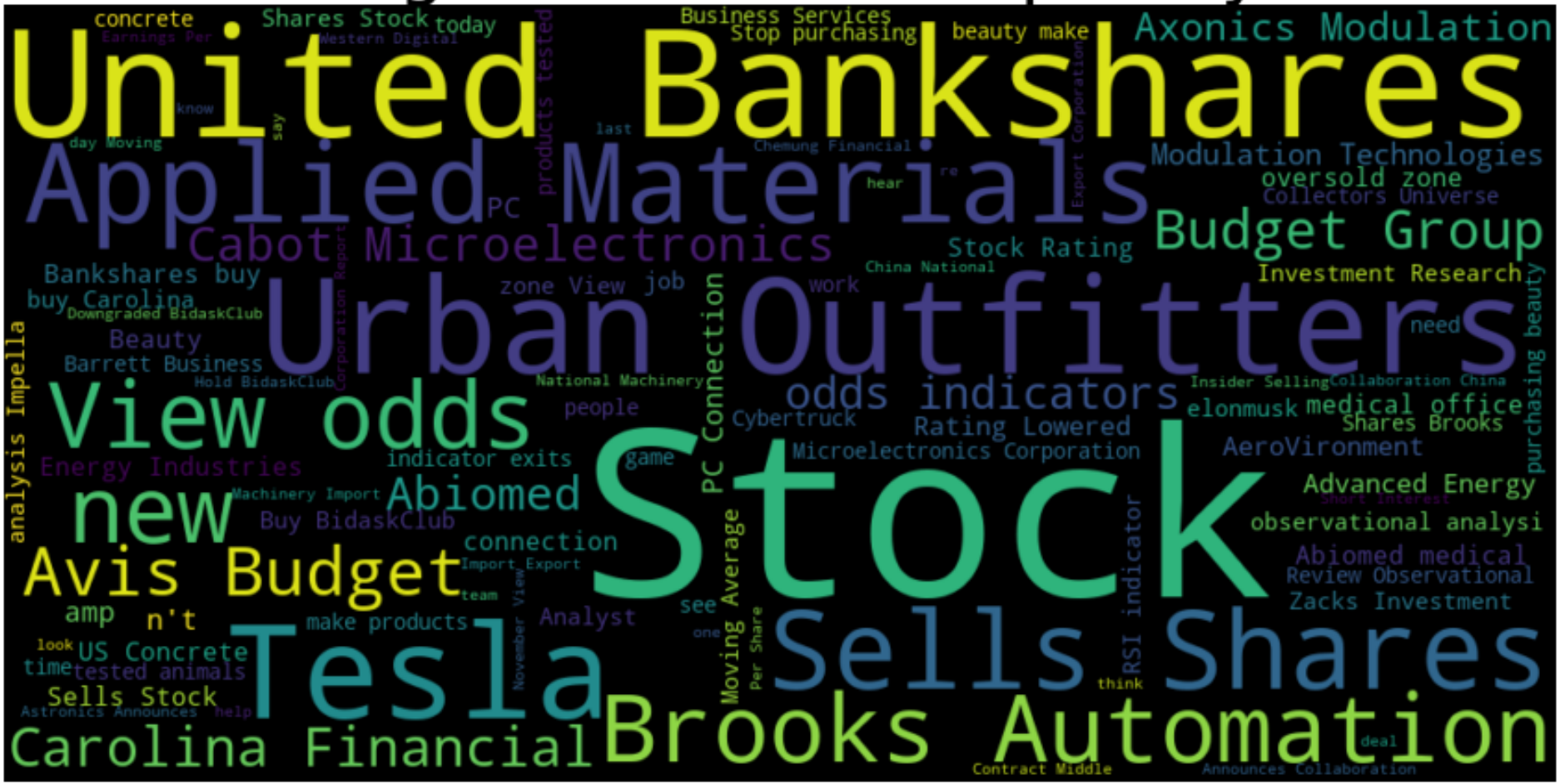
Once the tweets were divided into positive and negative training and testing sets, a NaiveBayesClassifier was trained and tested on the respective datasets.

Findings

The model created by training a NaiveBayesClassifier on tweets about the target companies scored 98.27% accuracy when evaluating the training dataset and scored 88.38% accuracy when evaluating the test dataset. The accuracy of 88.38% is on the high side for a sentiment analysis, indicating that this result supports the hypothesis that Twitter sentiment correlates to stock price movement. The 2 plots shown represent the top 100 words by frequency in the positive and negative tweets.

Positive Word Frequency

Negative Word Frequency



Limitations

The stock closing prices and the tweets collected were for the same week. This limited data set was due to the short time span given for this project and that only tweets for the prior 7 days can be accessed.

Even though there is a correlation between the Twitter sentiment and stock price change as shown by analysis, it may be that the sentiment reflects the change of the stock prices versus the hypothesis that Twitter sentiment causes stock price changes.

The order of tweets was collected was by company, so the order of the tweets was randomized before being analyzed so that each company would have tweets in both the training and testing dataset. The analysis was run multiple times with the tweet lists being randomized each time. The accuracy of the analysis on the testing set varied between 83% and 94%. This result was probably due to the small sample size of tweets.

Limitations (Continued)

A further study should be done with a larger sample of both company stocks and tweets over a longer period time to get a more consistent analysis result and to determine whether tweets influence stock prices or vice versa.

Conclusions

This study shows that there is a strong correlation between stock price changes and Twitter sentiment. But, as noted in the Limitations section, using only one week of stock prices and tweets is not a sufficient amount of data to determine if Twitter sentiment leads to stock price changes or vice versa.

Acknowledgements

All data used in this study was collected by myself and there was no collaboration or feedback sought from anyone else.

References

No references were used in conducting this study. All work conducted for this study was solely my own.