Can Twitter sentiment predict Bitcoin performance?

Devon Ankar – MS in Data Science, Wayne State University (DSE 6300 Term Project) 4/26/2019

1 Abstract

Data analytics, quantitative finance, and algorithmic trading have brought the benefits of automation to the fingertips of everyday investors. An individual investor, such as a day trader, might want to use each day's Twitter sentiment to decide whether to buy, sell, or hold Bitcoin shares on that day itself. A day trader needs to know how to react to what has happened since the last day's close. Bitcoin data was collected from Yahoo! Finance. Twitter data was collected from a random sample of 296,117 Tweet IDs which were "hydrated" to get full Tweet information. Sentiment analysis was performed on the Tweets using TextBlob. Models were developed using logistic regression, support vector machines, and random forest. All found that daily Twitter sentiment can be used to predict whether to buy, sell, or hold Bitcoin shares on a given day with an accuracy of over 85%. This result is mostly consistent with the literature. The author intends to further develop this model into an algorithm that can be used for real-world algorithmic trading in the quantitative finance space.

2 Introduction

2.1 Overview

The financial industry relies heavily on big data to develop models that find correlations among disparate datasets. Likewise, the social media industry relies heavily on data such as engagement rates and text mining procedures such as sentiment analysis. Big data is critical in both industries to gain a competitive advantage.

Moreover, as data analytics and quantitative finance become more popular, individual investors such as day traders have taken to algorithmic trading. When certain criteria defined by the investor are met, a certain trade (buy or sell) is performed. If the criteria are not met, then no trade is made (a hold). The benefits of algorithmic trading are twofold. Firstly, the investor does not need to monitor criteria in real-time and does not need to complete trades manually. Secondly, the trade can happen as soon as the criteria are met with no lag introduced by a human investor. Thus, algorithmic trading brings the benefits of automation to everyday investors.

A few online platforms, such as Alpaca Markets, let you build algorithms in the language of your choice and trade with real-time market data for free. Other similar platforms charge a nominal fee. As real-time trading is now highly accessible to individual investors, the author chose to approach this course project from the perspective of an individual investor (day trader) who wants to use each day's Twitter sentiment to decide whether to buy, sell, or hold Bitcoin shares on that day itself.

2.2 PRIOR RESEARCH

There have been several prior studies investigating Twitter sentiment alone, correlation between Twitter sentiment and stock market prices, and correlation between Twitter sentiment and Bitcoin price (or other cryptocurrency prices). For this course project, the focus must be only on investigating the correlation, if any, between Twitter sentiment and Bitcoin price.

The most relevant study found is “Twitter mood predicts the stock market” by Johan Bollen, Huina Mao, and Xiao-Jun Zeng published in 2011. They found that Twitter sentiment has a positive correlation with stock market prices. While it does not focus specifically on currency prices or Bitcoin prices especially, they found 88% accuracy in predicting the daily up and down changes of the DJIA based on Twitter sentiment. Because they were looking at daily up and down changes, they took the perspective of a day trader, similar to the one the author of this paper intends to take here. However, Bollen and coinvestigators only looked at the period Feb. 28, 2008 to Nov. 3, 2008, which is a relatively short time period.

Another study, “Stock Prediction Using Twitter Sentiment Analysis”, by Anshul Mittal and Arpit Goel of Stanford University, used a daily sentiment score to find that average daily sentiment in terms of calmness and happiness was predictive of the DJIA. Again, this is stock prediction, not Bitcoin, but it offers a similar model to the one that will be used here.

The 2018 study “Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis” by Jethin Abraham, Daniel Higdon, John Nelson, and Juan Ibarra of Southern Methodist University found that "Sentiment of tweets was determined to not be a reliable indicator when cryptocurrency prices were falling […] Both Google Trends and tweet [a ‘tweet’ is a message posted to Twitter] volume were highly correlated with price." (Abraham, et al, pg. 16) This suggests that perhaps Tweet volume should be considered as another feature if sentiment alone does not provide enough predictive value. The authors also considered that perhaps most of the Tweets mentioning “Bitcoin” by name were factual messages, which tend to be of neutral sentiment, rather than subjective expressions that could be of positive or negative sentiment. Perhaps, then, all Tweets should be considered in the analysis and not just those Tweets mentioning “Bitcoin”.

The study “Bitcoin Price Forecasting using Web Search and Social Media Data” by Rishanki Jain, Rosie Nguyen, Linyi Tang, Travis Miller, and Dr. Venu Gopal Lolla of Oklahoma State University found that “Both Google Trends and Twitter volume present correlation with Bitcoin’s price fluctuation. Specifically, Twitter volume and Bitcoin price shows a strong positive correlation.” (Jain et al, pg. 5). This is closely related to our project, but not an exact match; the authors report that the volume of Tweets correlates with Bitcoin’s price fluctuation, but they did not investigate the question of Twitter sentiment’s impact, if any, on Bitcoin price.

3 METHODOLOGY

3.1 DAY TRADING FOCUS

As mentioned in the introduction, the author of this paper chose to approach this course project from the perspective of a day trader who wants to use each day's Twitter sentiment to decide whether to buy, sell, or hold Bitcoin shares on that day itself.

For example, the paper by Abraham et al did not look at day-to-day variation in Bitcoin price; it only looked at the correlation between Twitter sentiment and Bitcoin price over the long-term from 2014 to 2018 and only looked at short-term trends going back 15 days. A day trader still needs to look at long-term trends because currency markets, like stock markets, operate over decades-long time horizons. However, what needs to be evaluated over the long-term are hourly trends, not weekly or monthly trends. A day trader needs to know how to react to what has happened since the last day's close.

3.2 DATA COLLECTION

For this class project, it was recommended that we collect only recent streaming data for training our models. Unfortunately, this limitation did not align with the real-world goal of predicting daily fluctuations within a market trend that occurs over many years. It is possible to get "enough" data – as in, a sufficient number of data points – from collecting only recent streaming data; however, the data will not be representative. When attempting to predict a trend that occurs over years, if not decades, one must use years of training data.

Ideally, the author wanted to collect 10 years of relevant data with at least hourly granularity. That is, the author wanted 10 years of historical hourly share price data for Bitcoin-USD conversion, and 10 years of Tweets about all topics (a random sample of such Tweets; not all of them!).

Unfortunately, this data was not publicly available. Historical Bitcoin data is only available with daily granularity, so daily close prices were used. Daily data for Bitcoin, as well as other currencies, stocks, and funds, are freely available from Yahoo! Finance[1].

Twitter's API allows developers to find very limited historical Tweets. It is not possible to retrieve more than a few thousand Tweets this way. Thus, collecting historical Twitter data proved to be more complicated.

As Justin Littman of George Washington University writes:

> "Twitter's Developer Policy (which you agree to when you get keys for the Twitter API) places limits on the sharing of datasets. If you are sharing datasets of tweets, you can only publicly share the ids of the tweets, not the tweets themselves. Another party that wants to use the dataset has to retrieve the complete tweet

[1] Bitcoin USD (BTC-USD) Prices via Yahoo! Finance: https://finance.yahoo.com/quote/BTC-USD/history?period1=1388552400&period2=1554436800&interval=1d&filter=history&frequency=1d

> from the Twitter API based on the tweet id ("hydrating"). Any tweets which have been deleted or become protected will not be available."[2]

The author of this paper obtained a dataset of only Tweet IDs via George Washington University's libraries[3]. The dataset contained the ids of almost 40 million Tweets collected from the accounts of 4,500 global news outlets between August 4, 2016 and July 20, 2018 by the researchers are GWU. A random sample of 296,117 Tweet IDs were then "hydrated" by the author using a script. The fields for `created_at`, `text`, `retweets`, and `likes` were obtained.

A RANDOM SAMPLE OF 296,117 TWEET IDS WERE THEN "HYDRATED" BY THE AUTHOR USING A SCRIPT

Twitter's API allows developers to make up to 900 requests in a 15-minute window, which translates to 1 request per second. To rehydrate almost 300,000 Tweets, the author set the script to run every 15 minutes for a few days.

3.3 SENTIMENT ANALYSIS

Prior to sentiment analysis, each Tweet's text was cleaned using regular expressions. URLs and user mentions (beginning with "`@`") were removed. Hashtags were left untouched because they can potentially indicate a positive or negative sentiment.

The author used a pre-trained sentiment analysis algorithm for sentiment analysis. The algorithm is called TextBlob. Prior to use, it was tested against Stanford's "Sentiment140"[4] labeled dataset. When the threshold was set at 0.7, i.e. sentiment polarity >= 0.7 is positive and <= -0.7 is negative, TextBlob performed with over 98% accuracy. When the threshold was set at 0.5, the accuracy deteriorated to 95%. 0.5 was used for this investigation because the number of Tweets classified as positive or negative at 0.7 is too small.

4 ANALYSIS

4.1 DATA PROCESSING: BITCOIN

The following operations were performed on the Bitcoin data:

- Select columns for date and close price for each day
- Convert date string to datetime object so it can be combined with Twitter data later

---

[2] Littman, Justin. "Where to get Twitter data for academic research" via GWU Libraries. Available at: https://gwu-libraries.github.io/sfm-ui/posts/2017-09-14-twitter-data

[3] Littman, Justin; Wrubel, Laura; Kerchner, Daniel; Bromberg Gaber, Yonah, 2017, "News Outlet Tweet Ids", https://doi.org/10.7910/DVN/2FIFLH, Harvard Dataverse, V3, UNF:6:I38WJ5vqwDky1fkEOeexvQ== [fileUNF]. Available at: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/2FIFLH
Found via DocNow.io's Tweet Id Datasets Catalog: https://www.docnow.io/catalog/

[4] Stanford's "Sentiment140" labeled dataset. Information at: http://help.sentiment140.com/for-students/. Download available at: http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip

- Detrend by differencing: subtract yesterday's close price from today's, for each day
  - Plot to make sure detrending is complete – that there are not still trends
- Create field for buy/sell/hold determination based on criteria. If difference in close price is in:
  - 93^{rd} percentile and higher: buy
  - 7^{th} percentile and lower: sell
  - else: hold
- Make sure the # of buys and sells aren't too small; buy: 135 days; hold: 1650 days: sell: 135 days

4.2 DATA PROCESSING: TWITTER

The following operations were performed on the Twitter data:

Time Operations

- Twitter data was at UTC, Bitcoin data was at UTC-5 (for NYSE's time) – Twitter data was adjusted by -5 hours to put them in the same time zone
- For each Tweet, the hour, day of week, month of year, and day of month were extracted into separate fields to look for potential cyclical trends later
- Since the recommendation needs to be received early enough in the day for a day trader to act on it, a cutoff of 1400 (UTC-5) was chosen. Only Tweets made before this time each day were considered for determining that day's sentiment. Tweets made after this time were dropped.

Text Operations

- Create a new field for the length of each Tweet
- Create sentiment fields for sentiment found by TextBlob, classification of that sentiment as positive, classification of that sentiment as negative

Aggregation by Day: Tweets were then aggregated by date (incl. month, weekday, day of month) on:

- sentiment – average daily sentiment
- length – average daily tweet length
- retweets – average daily retweets (# of retweets per tweet)
- likes – average daily likes
- positive/negative sentiment tweets – count
- tweet volume – # of tweets each day
  - volume was then detrended by differencing to account for Twitter's growth across time

Count of positive sentiment tweets was then divided by count of negative sentiment tweets for each day to get each day's sentiment ratio of positive tweets to negative tweets. The threshold was 25, meaning that a ratio of 25 positive tweets to 1 negative tweet was the minimum requirement for a positive sentiment day. There were 268 days with a positive sentiment ratio (i.e. "good days" on Twitter) and 1218 days with a neutral or negative sentiment ratio.

4.3 DATA PROCESSING: BOTH

The Bitcoin and Twitter data were then merged on date and plotted to make sure all series were fully detrended. Dates with very low tweet volume were then dropped (<= 10 tweets).

4.4 MODEL

The author developed a predictive model to answer the question, “Can Twitter sentiment predict Bitcoin performance?” The ultimate goal of the predictive model is to offer a recommendation for each day: buy, sell, or hold (do nothing), with the highest predictive accuracy possible.

In addition to average daily sentiment, the following features were considered as potential predictors:

- month of year, day of week, day of month
- sentiment ratio
- “good day” or not – this is highly colinear with sentiment ratio; it’s sentiment ratio binarized
- length, retweets, likes – retweets and likes are also fairly colinear with each other
- volume, volume detrended

Of these, only “good day” was found to be a strong predictor. Sentiment ratio was also a good predictor, but not as good as “good day”. When combined with other features, the overall predictive accuracy did not improve, so “good day” was taken alone.

The combined data (Bitcoin and Twitter) was then split into two datasets for training and test. The standard of 80% training and 20% test was used. The recommended allotment for the course project was 2/3 training and 1/3 test, but it is usually recommended to use at least 70% of the data for training. Using 70/30 and 75/25 splits did not significantly change the result.

The data was then visualized to look at the shapes of the distributions and for outliers. Data was found to be non-normal for several features that had long right tails such as retweets and likes. Thus, IQR was chosen for scaling instead of min/max; a scaler was chosen that is robust to outliers because it uses IQR.

The model was then fit with three different algorithms: logistic regression, support vector machines (SVM), and random forest.

5 RESULTS

Each model was run 100 times from the train/test split stage with the following results:

MODEL ALGORITHM	PREDICTIVE ACCURACY – MEAN ± STANDARD DEV.
LOGISTIC REGRESSION	0.855 ± 0.016
SUPPORT VECTOR MACHINES	0.854 ± 0.021
RANDOM FOREST	0.855 ± 0.018

Logistic regression had the highest performance with the highest predictive accuracy and lowest standard deviation. However, there was not much variance in the result between the three algorithms, and the small differences were not statistically significant. Random forest took the longest to run.

The logistic regression model seems to offer quite good predictive accuracy at 85.5%. It can predict, and thus recommend to a day trader, to sell, buy, or hold on any given day with over 85% accuracy.

Logistic regression is used when the dependent (Y) variable is categorical in nature, in this case a classification of buy, hold, or sell, based on the independent (X) variables (features). This problem requires use of a multinomial logistic regression because the Y variable is not binary (e.g. only buy or sell); there are three possible classifications (buy, sell, or hold).

However, thinking back to the Bitcoin data processing, the baseline is about 85% accuracy. Remember that there were 135 buy days, 135 sell days, and 1650 hold days. Thus, if the model simply predicted “hold” on all days, it would be accurate 85.9% of the time! (1650 / (135*2 + 1650) = 0.859)

Overfitting was determined not to be an issue because the model has the same predictive accuracy of 85.5% even on the training data. Thus, it can be concluded that the model is not sticking too tightly to the training data.

TWITTER SENTIMENT CAN BE USED TO PREDICT WHETHER TO BUY, SELL, OR HOLD BITCOIN SHARES WITH AN ACCURACY OF OVER 85%.

Undersampling of “hold” days was also attempted. “Buy” and “sell” days are smaller fractions, due to imbalanced data. However, even when “hold” days only comprised 50% of days, the predictive accuracy did not rise above the baseline (50% in that case).

Lastly, lowering the threshold for a “buy” or “sell” day also lowered the baseline, but the predictive accuracy was still not able to rise above the baseline.

6 DISCUSSION

6.1 RESULTS VS. LITERATURE

These findings are consistent with the literature in that a correlation was observed between Bitcoin price fluctuations and Twitter sentiment. More specifically, Twitter sentiment can be used to predict whether to buy, sell, or hold Bitcoin shares with an accuracy of over 85%. However, the literature also found a correlation between Tweet volume and Bitcoin share price, which was not found here.

6.2 TWITTER DATA

All Tweets in the sample were used because selecting only Tweets about Bitcoin made for a very small Tweet volume. Perhaps ideally, this study would be conducted using only Tweets about Bitcoin. However, that would require a much larger number of Tweets in total. If one were to “hydrate” all 40 million Tweets in the original Tweet IDs dataset, such research might be possible. Within the timeframe of this course project, it was not realistic.

Additionally, because the Tweets were collected only from the accounts of news outlets, they are less likely to reflect the sentiments of individuals and more likely to reflect the characteristics of actual world events. For example, Tweets about a natural disaster are likely to be of negative sentiment not because journalists or social media account managers are feeling bad (although they likely are), but because the Tweets are covering an inherently negative event. Likewise, as positively as journalists may feel about a possible cure for HIV, Tweets about such a cure would be positive because they cover a positive event. What has been found here is more so a correlation between the characteristics of world events and Bitcoin price.

6.3 REAL-WORLD APPLICATIONS

The author intends to further develop this model into an algorithm that can be used for real-world algorithmic trading in the quantitative finance space. To do this, several improvements will need to be made.

Firstly, stock market data (such as the S&P 500 index) will need to be obtained instead of Bitcoin data – this can be done easily from Yahoo! Finance, especially using the share price of an index fund. However, Yahoo! Finance and similar websites only have daily data and the author has not yet been able to find a free source of historical hourly stock market data. Ideally, hourly data would be used so that the buy/sell/hold recommendation can be provided hourly instead of daily. Hourly data would also help with predicting at what time the daily low and high occur, so that an algorithm could buy on the low and sell on the high each day.

Secondly, the imbalanced data issues will need to be further investigated and addressed. It is possible that stock market data won't share the same imbalance as the Bitcoin data to begin with. Moreover, when using hourly stock market data instead of daily data, perhaps hours categorized as "buy" and "sell" will be a larger proportion of all data points and the data will no longer be imbalanced as a result of that.

Thirdly, as discussed earlier, the Twitter dataset used for this project contains only Tweets from the accounts of news outlets. Ideally, a dataset could be obtained containing a wider cross-section of Tweets or Tweet IDs in order to capture the wider public sentiment that could affect consumer confidence, in turn affecting the stock market.

In addition to Twitter, text data such as comments can also be collected from other social media platforms such as Reddit. It is possible that sentiment of certain subreddits may better predict the stock market than sentiment of other subreddits. This is another area to investigate prior to developing a real-world implementation.

Fourthly, the model's predictive accuracy will need to be improved to over 90% using more features and feature engineering, if it is not already over 90% after the above improvements.

Lastly, a real-time recommendation engine needs to be developed. The engine would batch Twitter data hourly, perform sentiment analysis, and use the hourly average sentiment or sentiment ratio to give a buy/sell/hold recommendation for that hour. Since it is possible to use an online platform that lets users trade algorithmically with real-time market data, it would be optimal if the recommendation engine completed the trade automatically rather than simply providing a recommendation to a human trader. In order to do this effectively, the trading algorithm would need to include stop-loss conditions and other safeguards against extreme or unpredictable occurrences. In addition, the engine would need to be tested against real-time data for at least a few months before being allowed to run with real money.

7 CONCLUSION

Models developed using logistic regression, support vector machines, and random forest found that daily Twitter sentiment can be used to predict whether to buy, sell, or hold Bitcoin shares on a given day with an accuracy of over 85%. While this is a good result, it needs much improvement before it can be implemented as a real-world trading algorithm. The author intends to further develop this model into an algorithm that can be used for real-world algorithmic trading in the quantitative finance space. The ultimate goal is to develop a recommendation engine with appropriate safeguards that can be used in the real world.

8 REFERENCES

1. Bollen, Johan; Mao, Huina; and Zeng, Xiao-Jun (2011). “Twitter mood predicts the stock market”. Journal of Computational Science: Vol. 2: No. 1, Pages 1-8. Available at: https://doi.org/10.1016/j.jocs.2010.12.007

2. Mittal, Anshul, and Goel, Arpit (2012). “Stock Prediction Using Twitter Sentiment Analysis”. Stanford University. Available at: http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf

3. Abraham, Jethin; Higdon, Daniel; Nelson, John; and Ibarra, Juan (2018). “Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis”. SMU Data Science Review: Vol. 1: No. 3, Article 1. Available at: https://scholar.smu.edu/datasciencereview/vol1/iss3/1

4. Jain, Rishanki, Rosie Nguyen, Linyi Tang, Travis Miller, and Dr. Venu Gopal Lolla (2018). “Bitcoin Price Forecasting using Web Search and Social Media Data”. Oklahoma State University: Paper 3601-2018. Available at: https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/3601-2018.pdf