# Project Report

# Predicting Fluctuations in Bitcoin With Twitter Data

Spring 2018
CS-579 Online Social Network Analysis
Illinois Institute of Technology
Department of Computer Science

Saumil Ajmera
A20397303
sajmera4@hawk.iit.edu

Akash Gairola
A20395744
agairola@hawk.iit.edu

FNU Ashish
A20403925
fashish@hawk.iit.edu

## Abstract

The problem statement is related to one of the trending topics - crypto currency and how its popularity spread in global world. The sudden volatility in its price very well captures the correlation between the sentiment of people and its effect over the price. This paper analyzes user comments in online crypto currency communities to predict fluctuations in the prices of crypto currencies through the classification model. We gathered the bitcoin related tweets for nine months from twitter scraper and bitcoin's price (Kraken Exchange) variation in US $ from quandl API. We were able to establish a moderate accuracy of 59% for train data (seven months) and average accuracy of 26% for test data (two months). We considered price fluctuation (closing price – opening price) above or below 200$ per day (positive or negative) as deviation point and any movement in range of 200 $ was considered as neutral. We standardized the difference of opening and closing prices by standardization formula. Now each tweet is given a polarity score based on TextBlob module provided in python according to overall sentiment of sentence. We applied supervised algorithm by iterating multiple iterations of standardized bitcoin value to classify tweets into positive, negative and neutral based from an average score of polarity of tweet (20 per day). The model was trained to decide a threshold for tweet polarity values from bitcoin's standardized value. The threshold values obtained was given to twitter test data to directly classify and its output is matched against bitcoin's classified value. We prepared a confusion matrix table for ease of inference of actual and predicted classes. The model further can be trained with more data and machine learning techniques to obtain higher accuracy.

# 1. Introduction

Crypto currency isn't so much a revolution as it is an evolution in the way humans transfer and store their valued assets. John McAfee, Founder of McAfee Security and the foremost expert on cyber-security says: *"You can't stop things like Bitcoin. It will be everywhere and the world will have to readjust. World governments will have to readjust."* Also the following attributes makes it attractive for commoner – Global Accessibility, Immutability, Security, Speed and Price, Transparency and the most important feature is Decentralized – no common center of power.

Prior to social media development, people often find hard to trust any products and also were less aware about global news. With the penetration of social media in the last decade, gives you an awesomely efficient, cheap, and effective way to build that trust provided, of course, that you're a good egg to begin with. The truth is, social media when used strategically over time is the most powerful form of marketing and market research the world has ever seen.

So keeping in mind the above both factors we decided to come with idea that how we channelize the power of social media platform – Twitter into monetary gains obtained by providing trading ideas of Bit coin. This project deals with algorithmic trading of bitcoin, it is different from stock market where many other factors (company's quarterly performance, overall economic growth, company's share holding pattern etc.) are affecting the stock price. However, in case of bitcoin's price is irrelevant to above mentioned factors rather it is primarily dependent on demand supply and sentiments. So we have used sentiments as our predicting factor for matching the actual price fluctuations.

# 2. Background Work

1) Nowcasting the Bitcoin Market with Twitter Signals:
https://arxiv.org/pdf/1406.7577v3.pdf

This paper analyzes how correlated are the bitcoin market indicators with the sentiments of people on twitter regarding bitcoin. Here the twitter posts related to bitcoin are collected and analyzed for positive, negative and uncertainty related words. The analyzed data shows that a higher Bitcoin trading volume Granger causes more signals of uncertainty within a 24 to 72- hour timeframe.

2) From Bitcoin to Big Coin: The Impacts of Social Media on Bitcoin Performance:
http://www.fmaconferences.org/Orlando/Papers/Bitcoin_FMA.pdf

This paper examines predictive relationships between social media and bitcoin returns by considering the relative effect of different social media platforms (Internet forum vs. microblogging) and the dynamics of the resulting relationships using vector autoregressive and vector error correction models.

3) Rapid Prototyping of a Text Mining Application for Crypto currency Market Intelligence:
https://arxiv.org/ftp/arxiv/papers/1611/1611.00315.pdf

This paper uses block chain technology which is used to establish a shared and immutable version of truth between networks of users. Machine learning and Natural language processing is used for prototyping the applications for the analysis of trends that occur during emergence of block chain technology.

4) Predicting fluctuation in crypto currency transactions based on user comments and replies:
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0161197

This paper proposes a prediction model for prediction the fluctuations in crypto currency transactions based on the user comments in online crypto currency communities.

5) Algorithmic Trading of Crypto currency Based on Twitter Sentiment Analysis:
http://cs229.stanford.edu/proj2015/029_report.pdf
This paper uses logistic regression, naive bayes and support vector machines for cleaning and analyzing data and the prediction accuracy is increased by 90%

Our project is very much related to all these papers and we are using some the techniques from these papers for analyzing positive and negative words to reflect the overall sentiment of the people regarding bitcoin.

## 3. Approach

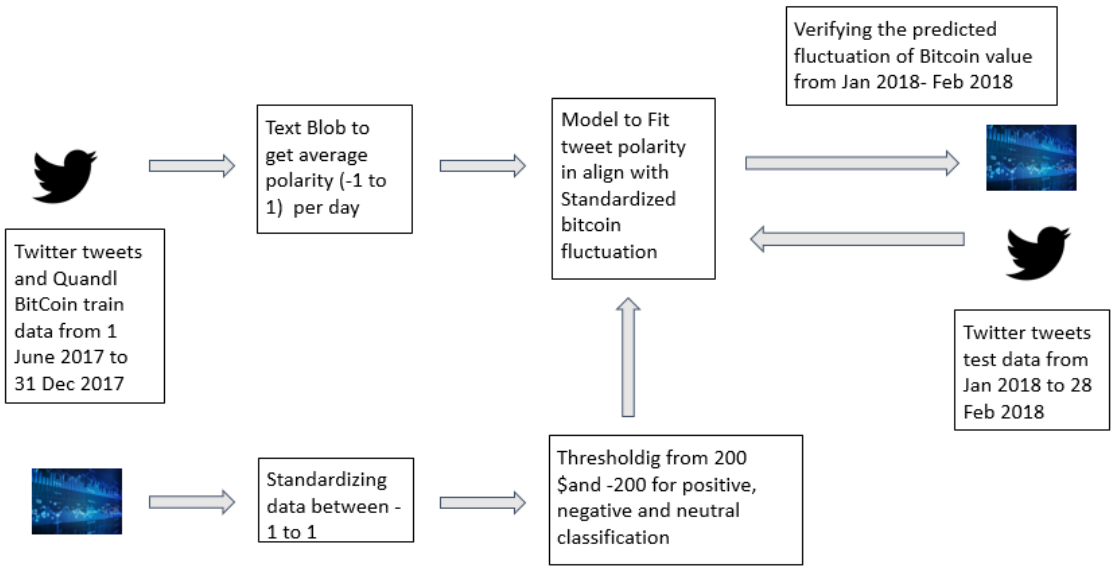


*Figure 1: Flow Diagram*

I. First Twitter and Quandl valid authentication is done for data collecting purpose from their provided API
II. We added three columns bitcoin data set - *Diff* to know the closing and opening price difference, *Change* to set into positive, negative and neutral depending upon *Diff* column variation above 200, less than 200 and in between -200 to 200, *Score* to set into -1 to 1 as standardization of *Diff* column by formula ( 2*(value-min)/(max-min)-1)
III. We obtained JSON file by running the following script in downloaded twitterscraper folder for each month's twitter data for querying with keyword *Bitcoin*, date range and English language.

twitterscraper Bitcoin -l 20 –p 15 -–lang=en -bd 2017-06-01 -ed 2017-06-30 -o tweetsJun.json

l => limit of tweets per day
p => pool of 15 tweets
lang => language english
-bd => begin date
-ed => end date
-o => output JSON file name

IV. Now we converted JSON file into CSV file by removing special characters and links from tweets and scoring tweet polarity using Text blob module of python
V. We merged all such CSV file into one file with each date representing average score as twitter sentiment related to Bitcoin
VI. Now we decided threshold for tweet score to segregate tweets into positive, negative and neutral by running multiple iterations against bitcoin segregated data obtained least error and maximum accuracy from confusion matrix.
VII. Tweets having polarity score less than -0.05 are labeled as negative, greater than 0.13 are labeled as positive and in between -0.05 and 0.13 are labeled as neutral.
VIII. Repeat the above steps except VI step for Jan and Feb month data of twitter against bitcoin to correctly predict fluctuation.

# 4. Experimental Result

I. Here is the sample of Bitcoin CSV file with newly added columns

| Date | Open | High | Low | Close | Volume (B | Volume (C | Weighted | Change | Diff | Score |
|---|---|---|---|---|---|---|---|---|---|---|
| 1/1/2018 | 13971.1 | 14442.9 | 13050 | 13506.1 | 2226.373 | 30453735 | 13678.63 | Negative | -465 | -0.15267 |
| 1/2/2018 | 13506.1 | 15300 | 13088.8 | 14882 | 3228.214 | 45657585 | 14143.3 | Positive | 1375.9 | 0.816966 |
| 1/3/2018 | 14882 | 15599.7 | 14701.9 | 15185.7 | 2740.653 | 41376944 | 15097.48 | Positive | 303.7 | 0.252219 |
| 1/4/2018 | 15186.7 | 15500.1 | 14417.1 | 15266.6 | 3497.9 | 52267644 | 14942.58 | Neutral | 79.9 | 0.134339 |
| 1/5/2018 | 15266.6 | 17000 | 14961 | 16990 | 3728.59 | 59665096 | 16002.05 | Positive | 1723.4 | 1 |
| 1/6/2018 | 16988.3 | 17245 | 16167.5 | 17088 | 2126.638 | 35480125 | 16683.67 | Neutral | 99.7 | 0.144768 |

*Figure 2: Bitcoin Data*

II. Here is the sample of Tweets CSV file with polarity score

| | | |
|---|---|---|
| 0.4 | 1/27/2018 | Bitcoin is not controlled by any person company or government Its run by the community of its users Read that twice because its important |
| 0.1875 | 1/27/2018 | DSCR Company is excited to announce that the Company is pivoting its commercial operations into the mining of Bitcoin Gold coins |
| 1 | 1/27/2018 | A Bitcoin Embassy With Legendary Economist Jeffrey Tucker Greg Photon Decentralized Storage |
| -0.05 | 1/27/2018 | Why does bitcoin mining take so long |
| 0 | 1/27/2018 | Bitcoin investment trust gbtc bitcoin will change how you think about money smarter analyst |
| -0.15556 | 1/27/2018 | Consider Bitcoin Price goes down to 3000 what will be the effect on miners will they stop mining and the affect on difficulty to mine a Bitcoin |

*Figure 3: Twitter Data*

III. Here are the plots for variation of Bitcoin standardized value and Twitter sentiment variation for Train Data (Jun-Dec 2017)

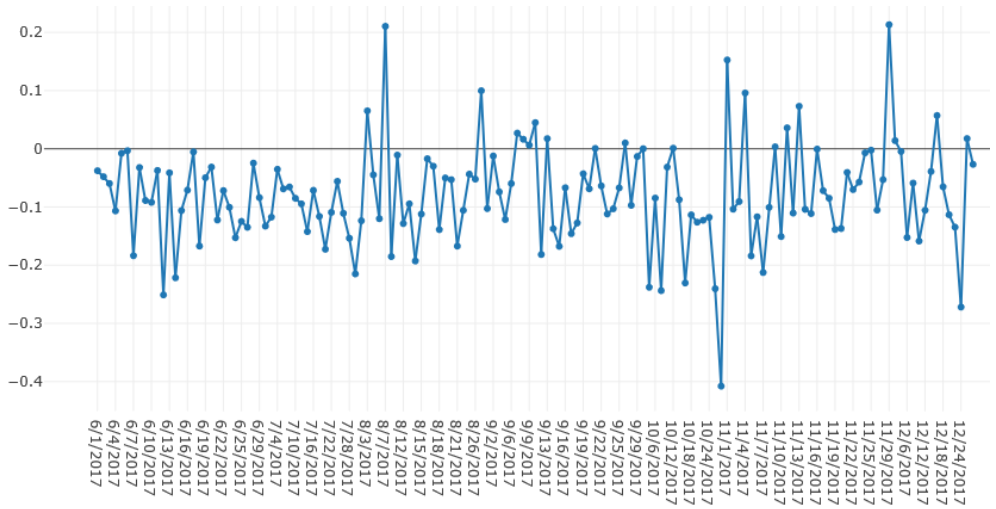


*Figure 4 Bitcoin Price Variation*

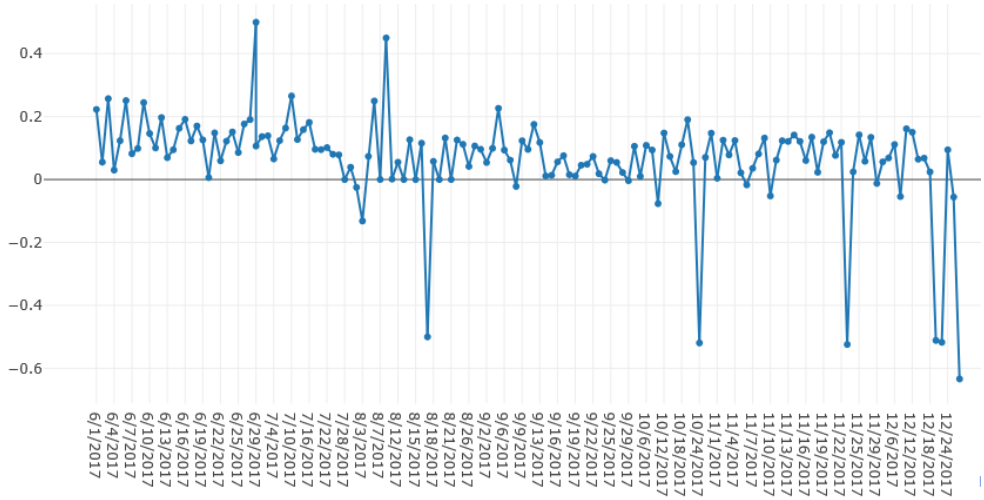


*Figure 5: Tweets Sentiment*

IV. Confusion Matrix of Train Data

| | Bitcoin Positve | Bitcoin Neutral | Bitcoin Negative |
|---|---|---|---|
| Twitter Positive | 1 | 34 | 2 |
| Twitter Neutral | 13 | 85 | 7 |
| Twitter Negative | 2 | 2 | 1 |

Accuracy = 1+85+1 / (147) = 59.2 % (Total Correct classified / Total Values)
Error = 60/147 = 40.8% (Total Misclassified / Total Values)

V. Now following are plots for variation of Bitcoin standardized value and Twitter sentiment variation for Test Data (Jan-Feb 2018)

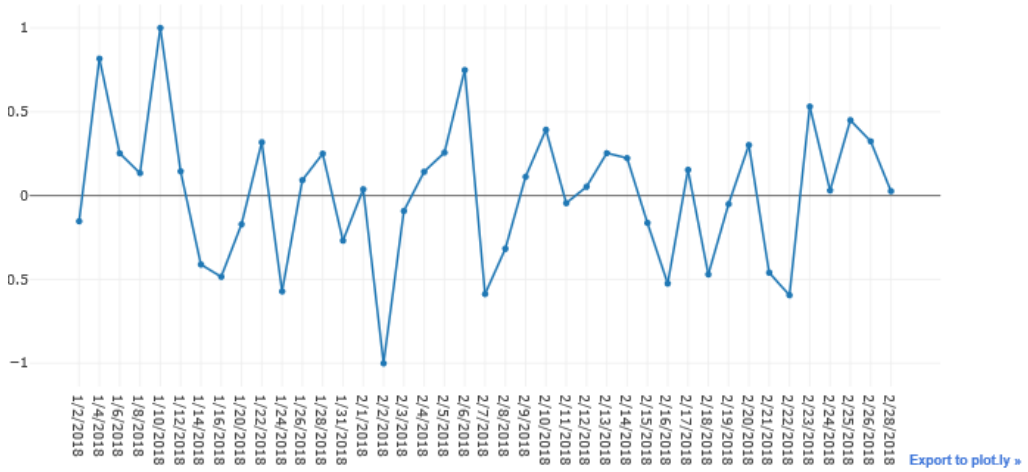


*Figure 6 Bitcoin Price Variation*

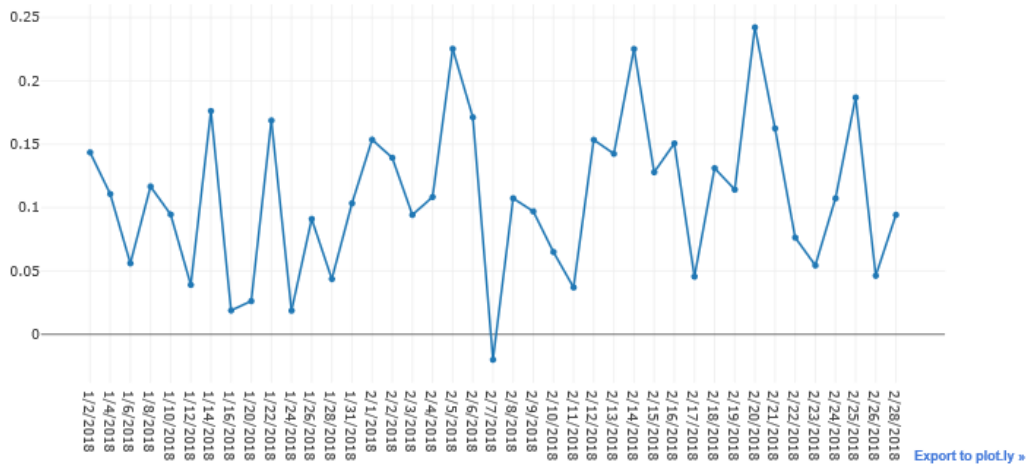


*Figure 7 Tweets Sentiment*

VI. Confusion Matrix of Test Data

| | Bitcoin Positve | Bitcoin Neutral | Bitcoin Negative |
|---|---|---|---|
| Twitter Positive | 4 | 3 | 8 |
| Twitter Neutral | 10 | 7 | 9 |
| Twitter Negative | 0 | 0 | 0 |

Accuracy = 11 / (41) = 26.8 % (Total Correct classified / Total Values)
Error = 30/41 = 73.1% (Total Misclassified / Total Values)

## 5. Conclusion & Future Scope

We were able to infer twitter tweets for determining right time to initiate a buy or sell call of bitcoin and how algorithmic trading works and how real time sentiments drives the equity market or crypto currency. Although it will always remain mystery to predict the correct price of stock or bitcoin for a particular time, however we can achieve a close result that can help us to achieve maximum profitability with minimum risk factor.

It shows how much important factor social media plays for decision making and how it has penetrated into routine life of people.

Future Scope –

I. Also we can achieve high accuracy by getting twitter data of Bitcoin exchange provider's twitter account like Kraken account.
II. We can apply sentiment analysis with some keywords and check whether it has strong predictive power for Bitcoin. Our manual cut off point evaluation point can be automated using Machine learning which helps in identifying hidden signals.
III. We can also set up a regression model like linear or polynomial to find the best correlation between bitcoin price and tweets polarity which can be used to predict the future outcome.

## 6. References:

1) http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0161197
2) https://www.quandl.com/
3) https://cryptoanswers.net/what-is-cryptocurrency/