

CS-579 Online Social Network Analysis

Spring 2018

Project – Milestone

1. **Project Title** - Predicting Fluctuations in Bitcoin with Twitter data

2. **Authors** - Saumil Ajmera, Akash Gairola, FNU Ashish

3. **Problem Overview** -

The problem statement is related to one of the trending topics - crypto currency and how its popularity spread in global world. We would analyze correlations and casualties between Bitcoin value and Tweets which are related to Bitcoin. We would try to determine the relationship between social media and bitcoin returns and would analyze upon what extent, social media can predict future bitcoin prices and trading volume.

4. **Data** -

- We will be using tweets from Twitter through APIs for a specific date range by certain keywords.
- Query parameter would be altered in order to classify positive, negative and fear tweets.
- Moreover Bitcoin prices has been collected within the timeframe from October 1st 2017 until February 28th 2018.
- We will be retrieving time series data of bitcoin values using quandl package provided in python.

5. **Method** -

- First we would collect our data using Twitter API which would contain keyword Bitcoin. Also collecting bitcoin price data from API and storing into csv files for quick access.
- We also need to filter out those tweets which would provide some sentimental information. Tweets which contains Bitcoin and some “emotional” terms such feel, happy, great, love, awesome, lucky, good, sad, bad, upset, unhappy, hope, fear, worry etc. would be fetched.
- Based on that, we would classify these tweets into positive tweets, negative tweets and fear tweets.
- We would also perform some data cleaning process to discard some tweets which do not provide any sentimental information. For example we need to discard tweets like “Good morning bitcoin investors”. It would be classified as positive tweets but in fact this tweet has to be removed.
- Based on the data of the Bitcoin in same time bound as of the above posts, we will train the model to correlate the rise in value with positive values, decrease in value with negative sentiment and consolidated movement with neutral sentiment.
- We would notice the correlation values between classified tweets and bitcoin values to get some insight about their relationship.

6. Intermediate/Preliminary Experiments & Results -

- First we would collect positive tweets based on query 'q': 'Bitcoin' AND 'feel' OR 'happy' OR 'great' OR 'love' OR 'awesome' OR 'lucky' OR 'good'.
- Similarly we would collect negative tweets and fear tweets based on different keywords.
- Using AFINN Lexicon, classification of tweets has been done. It would rate each sentimental terms. However, we would also re-assign weights to some terms according to the results obtained
- Top negative tweets(positive rating, negative rating, text) based on keyword bitcoin are –

```
0 4 RT @Les_G: Bitcoin Mining is not a scam.. Finding a legit broker to mine with is the problem
0 4 RT @piratebeachbum: Every time the market tanks Bitcoin Dominance increases. Current BTC Dominance 44.4%. Oh I know...some
of you shit coin...
0 3 Co-founder of Ethereum lost $ 2 billion in two months due to the fall courses | Bitcoin, cryptocurrency, Ethereum,... http
s://t.co/1RRotb0ZpS
1 2 https://t.co/5VDPDDX2pw https://t.co/8F6LuyCSmx the Solution To Bitcoin's Scaling Problem Could Change Your Networks http
s://t.co/fszuF0HcTj
1 2 https://t.co/A05u7n8TC6 https://t.co/PkSLb5InCJ agree with Buffett: Don't waste your hard-earned cash on bitcoin https://t.
co/TEIAug7Nvh
0 2 https://t.co/KOjoNfk3Wx https://t.co/SXBxZDY09i Price Jumps $1K After Carney Tells G20 Crypto Does Not 'Pose Risk' https://
t.co/SSjIu4KAUJ
0 2 https://t.co/jSEwfSho7i https://t.co/SXBxZDY09i Returns to Above $8K, But Sell-Off Risks Remain https://t.co/92866WTwyu
0 1 RT @LoganPaul: checking my bitcoin investments while training for the @KSI01ajidebt fight, going well so far on both fronts
https://t.co/Nq...
0 1 RT @Remi_Vladuceanu: Demand For Bitcoin Hardware Wallets Rise in South Korea, as Users Develop Awareness https://t.co/HGW3J
4EyGg
#newsofth...
0 1 https://t.co/Q232ucN6Lu https://t.co/YpML0LD3qn states Bitcoin mining no longer profitable https://t.co/SmbYZjlwqF
0 1 RT @LoganPaul: checking my bitcoin investments while training for the @KSI01ajidebt fight, going well so far on both fronts
https://t.co/Nq...
```

Figure 1 Positive Tweets on Bitcoin

- Top positive tweets(positive rating, negative rating, text) based on keyword bitcoin are –

```
5 0 RT @CryptoCobain: Imagine being one of the people that sold bitcoin in the low 7000s yesterday. How do you even pleasure yo
urselves with ha...
4 0 RT @ebitToken: #Giveaway $100 in $ETH!

-Like
-Retweet
-Follow
-Comment down below with your #Ethereum (preferably) ERC20 address.

Like an...
4 0 RT @erkancamli: #Eventum #Bounty

join Eventum Bounty Share and invite to earn free EVT tokens!

signup : https://t.co/RPm4RhV9AX...
3 0 RT @bethereumteam: The whole #betting process in only 5 steps!
Easy, secure and engaging using: https://t.co/dXpy0T6IUp
#socialbetting #onl...
3 0 1/ Just listened to @Murch on the @NodedPodcast. Very excited about the new Coin Selection. It will optimize fees w... http
s://t.co/logNuiFcTL
3 0 RT @DecibelsToken: Are you all ready for the final round of the Decibels Bounty Program?

We are excited for the upcoming contests and can'...
3 0 RT @GymRewards: Submit your ICO Token for listing today and join GYM Rewards Token at GYMBase! https://t.co/Zod85qqwK3 #I
CO #cryptocurren...
```

Figure 2 Negative Tweets on Bitcoin

- Using quandl package, we also retrieved bitcoin price data from October 1st 2017 to February 28th 2018 and stored in excel file with values like open, high, close and volume per day.
- We will now work performing deeper sentiment analysis on specific words by fine tuning the model and with least false positive and false negatives derived from the confusion matrix

- Depending upon the output of positive tweet reflecting increase in price and negative tweet reflecting decrease in price we may require to change some of the keywords and assign weights to the impactful words.



Figure 3 Bitcoin Price for past five months

7. Related 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.
- We will also try to implement granger causality test which will help in forecasting price of bitcoin based on tweets shared by people.

8. Responsibilities and Timeline –

Project Time Line & Responsibilities								
		<i>Color References</i>	<i>Task Status</i>	<i>Color References</i>	<i>Team Member</i>			
			Completed		Saamil			
			In Progress		Akash			
			Remaining		Ashish			
<i>Sr. No.</i>	<i>Task</i>	<i>Social Network Data</i>	<i>Crypto Currency Data</i>	<i>Evaluation and Correlation</i>	<i>Time line</i>	<i>Saamil</i>	<i>Akash</i>	<i>Ashish</i>
1	Exploratory Analysis	Finding triggering words	Understanding the principles of price variation		1/15/18 - 2/6/18			
2	Data Gathering & Cleaning	Twitter API	Quandl API		2/7/18 - 2/25/18			
3	Research Paper Analysis			Understanding research paper's objective and results	2/7/18 - 2/25/18			
5	Sentiment Analysis	Dividing data based on pos, neg and neutal words			2/26/18 - 3/19/18			
6	Statistical Analysis		Plotting charts to know the trends		2/26/18 - 3/19/18			
7	Applying known algorithms by tuning			Correlating social network data to cryptocurrency data	2/26/18 - 3/25/18			
8	Model Building on Training data	Fine tuining weights assigned to word	Forecasting Bitcoin price from the type of tweets	Combining both results and developing a model	3/26/18 - 4/9/18			
9	Evaluating Model and Visualization			Confusion Matrix & Validating against true values	3/26/18 - 4/9/18			
10	Documentation			Report and Presentation	4/10/18 - 4/15/18			

Figure 4 Timeline and Responsibilities

9. References –

Apart from the above reference papers we have also collected few ideas from the following links –

Twitter API - <https://developer.twitter.com/en/docs>

Reference Journal- <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0161197>

Quandl - <https://www.quandl.com/>