

OUTLINE

TOPICS TO BE COVERED

- 1/ Introduction
- 2/ Method
- 3/ Conclusion
- 4/ Next Steps

1/ INTRODUCTION

1/ INTRODUCTION

PROBLEM STATEMENT

- Is there a correlation between Twitter sentiment and BTC price fluctuation?
- Can a naive prediction model based on sentiment changes yield better than random accuracy?

PURPOSE

- Identify drivers of bitcoin price fluctuation.
- Build and optimize a prediction model given sentiment analysis.



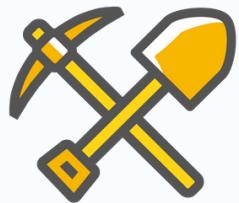
2/ METHOD

I. DATA COLLECTION



TWEETS

- Retrieved Tweets by using Twitterscraper
- Twitterscraper: An open source framework written in Python to retrieve the tweeter and use BeautifulSoup4 to parse the retrieved content.
- Filtered based on: hashtags --bitcoin; English; 2014-2018.



MINER'S PERFORMANCE

- Hash Rate:measure of miner's performance.
- The higher hash rate means increased opportunity for cryptocurrency mining and receiving block reward.

GOOGLE TREND



- BTC price has historically correlated with search interest.
- It shows how interested the public is in the use of Bitcoin, which may suggest that at present there is a high possibility of attracting new investors.
- Retrieved by using Pytrends written in Python.



BITCOIN PRICE

- Target variable
- Historical daily average of BTC market price ((UTC))

II SENTIMENT ANALYSIS

REDUCING NOISE

- scrubbed from any non-alphabetic symbols (excluding !@#\$+%^*:()'')
- Used Spacy to remove stop words and punctuation

INDIVIDUAL TWEET SENTIMENT ANALYSIS

- VADER is used to derive a sentiment score from each tweet.
- The result of this process is that each tweet row in the dataset is appended with it's individual sentiment score.

CALCULATED SENTIMENT SCORES FOR 20 TWEETS WITH TOP RETWEETS

```
print_sentiment_scores(" I bought bitcoin today but I realized it would go down soon").
```

```
I bought bitcoin today but I realized it would go down soon {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

```
print_sentiment_scores(" I bought bitcoin today but I realized it would go down soon :(").
```

```
I bought bitcoin today but I realized it would go down soon :(' {'neg': 0.278, 'neu': 0.722, 'pos': 0.0, 'compound': -0.59271}
```

PREPROCESSING:TWITTER

TOP 20 RETWEETED TWEETS PER DAY

date	retweet	text
2017-05-02	405	online casino reviews & gambling games. Poker sp
2017-05-02	210	Mechs stalking humans in SpacePirate: Age of Rus
2017-05-02	41	The latest Bitcoin Price Index is 1,444.02 USD http:
2017-05-02	28	I will be keynoting http://www.theblockchain.nz/
2017-05-02	13	#Bitcoin skyrockets to fresh record high on a spike
2017-05-02	6	Bitcoin soars above \$1,400 to all-time high Article
2017-05-02	6	??Are we in another crypto bubble??
2017-05-02	5	...
2017-05-02	3	Alt Coin Mining Rigs https://youtu.be/d2xWzPDre

.....

date	retweet	text
2018-04-30	264	Gold was officially made the standard of value in Engl
2018-04-30	148	#HOLDVEST Launch Universal Crypto Trading Portal
2018-04-30	124	Cryptfunder is a funding source of the community and
2018-04-30	114	I will choose 20 random followers to win 0.1 ETH
2018-04-30	39	A big Brand you can Trust is Offering the Premium, Hy
2018-04-30
2018-04-30	24	Japanese @sbigroup to launch SBI Virtual Currencies
2018-04-30	23	Bitcoin has jumped 36%, the biggest monthly gain since

GET SENTIMENT SCORE

date	ct1	rt1	...	ct20	rt20
4/24/2018	Airdrop	36		Some simp 1	
4/26/2018	whole	1565		er meets 8	
4/28/2018	Neither Co	12		I'm about t 0	
4/30/2018	Gold was o	264		@Fantasy_ 9	

ADD WEIGHTS TO SCORE

score_x_ret1	...	score_x_ret20
0	...	4.9644
-37.7325	...	0
87.72	...	-21.5712
23.46	...	24.897

III MODEL

MODELS RESULTS

- Our different models and their performances

1 year dataset	
model	accuracy
Logistic Regression	55.15%
Xgboost Regression	57.02%
Neural Network	58.86%

4 years dataset	
model	accuracy*
Logistic Regression	47.45%
Xgboost Regression	50.00%
Neural Network	51.65%

* this considers just twitter data

3/CONCLUSIONS



3/ CONCLUSIONS

WHAT WE LEARNED

Bitcoin price is dependent on many factors, and we believe a model taking into accounts more sources of informations might perform well.



4/ NEXT STEPS



4/ NEXT STEPS

POTENTIAL IMPROVEMENTS :

1. Combining NLP techniques with time-series analysis could help predict trend more efficiently

2. Run the model using the change of sentiment to predict the change of price

3. Include other sources of information as:
<https://cryptopanic.com/>

THANK
YOU !