

Cryptocurrency Price Prediction with Social Media

...

By Derek Plemons

Table of Contents

1. Business Problem
2. Datasets
3. Features
4. Bitcoin Comments Wordcloud
5. Exploratory Analysis Findings
6. Modeling
7. Model Metrics
8. Conclusion
9. Next Steps

Business Problem

- Using the cryptocurrency sentiment analysis dataset of the r/Bitcoin subreddit and this historical bitcoin price, can we predict whether the price of bitcoin will decrease or increase for the following day?
 - Applicable to cryptocurrency traders

Datasets

1. Reddit Comments from r/Bitcoin subreddit
 - a. Using Pushshift api
 - b. Average of 2k comments per day
 - c. Date range of 1/1/19 to present
2. Bitcoin Historical Price Data
 - a. Using Historical-Crypto
 - b. Data from 2014 to present

Features

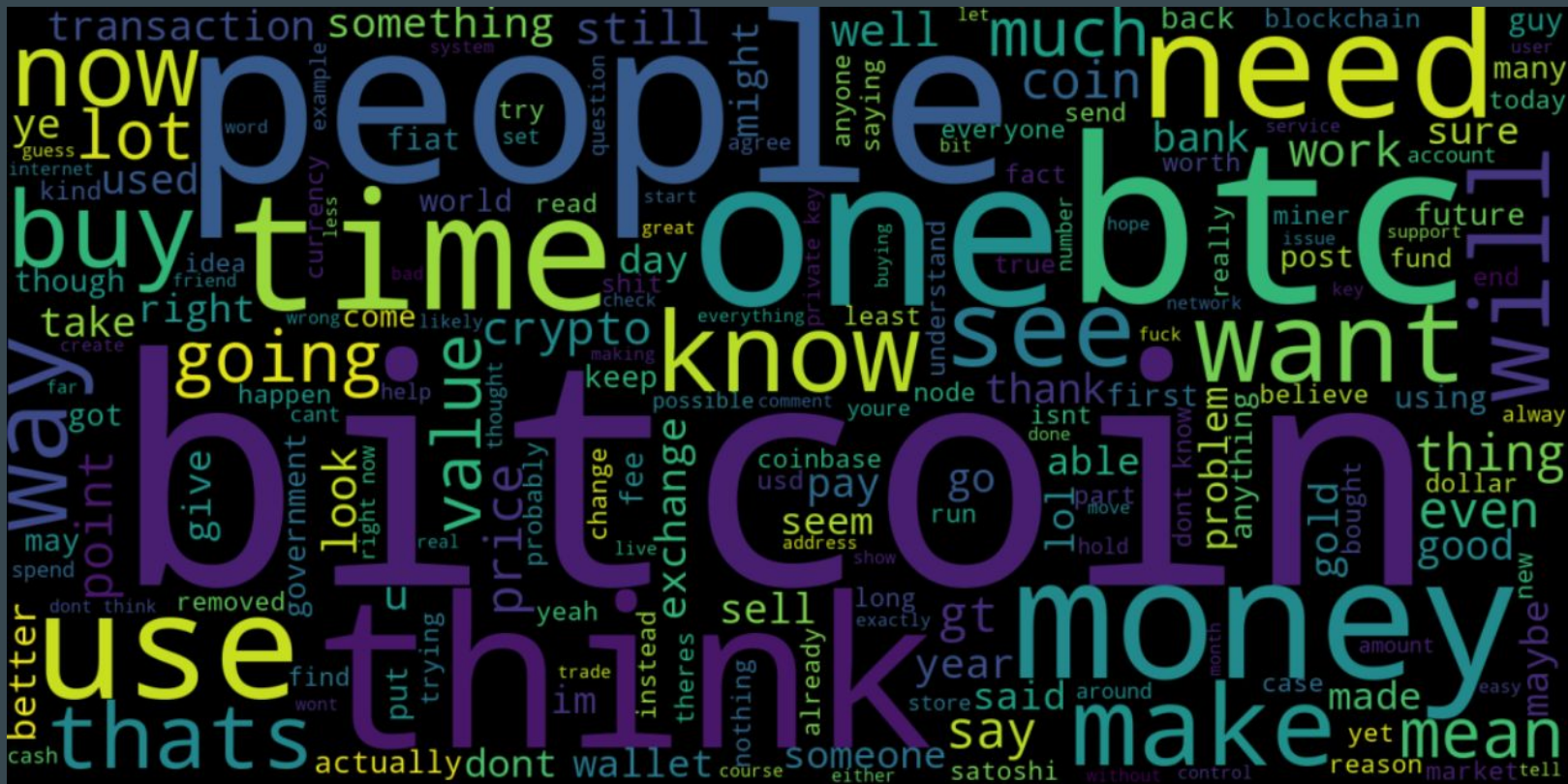
r/Bitcoin Subreddit Comments

- 1,565,474 comments from 01/01/19 to present
- time, avg_neutral, avg_negative, avg_compund

Historical Bitcoin Price

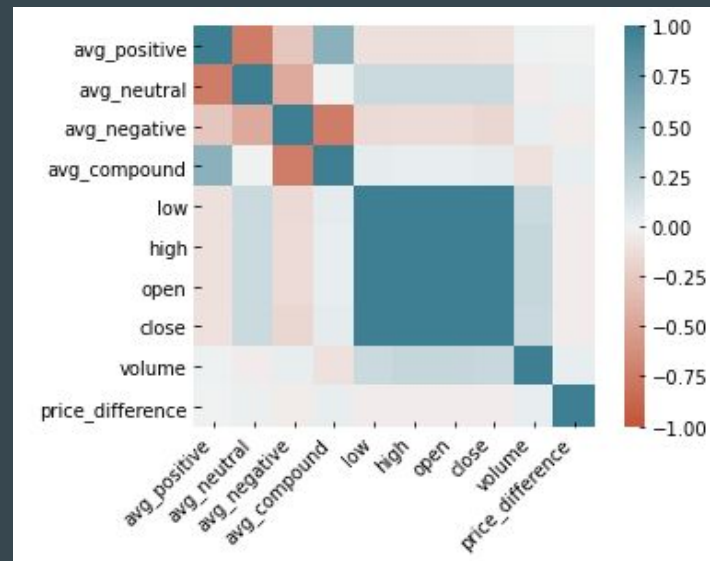
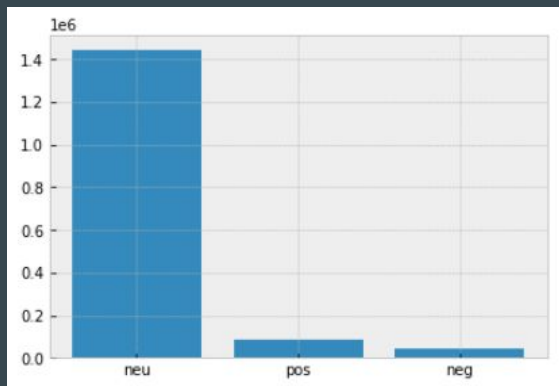
- Low, high, open, close, volume

Bitcoin Comments Wordcloud



Exploratory Analysis Findings

- Majority of comments are neutral
- There does not appear to be a significant correlation between average sentiment and the price difference of bitcoin.



Modeling

Logistic Regression

K Nearest Neighbors

Support Vector Machine

Random Forest

Gradient Boosting

Model Metrics

Model	Recall	Precision	Accuracy	F1
Logistic Regression	0.97	0.49	0.5	0.51
K Nearest Neighbors	0.53	0.46	0.46	0.47
Support Vector Machine	1.0	0.49	0.48	0.5
Random Forest	0.53	0.46	0.46	0.46
Gradient Boosting	0.65	0.53	0.55	0.56

Conclusion

Using the cryptocurrency sentiment analysis dataset of the r/Bitcoin subreddit and historical bitcoin price data from January 1st 2019 to present, there was not an accurate way to predict if the price of bitcoin would go up or down the following day.

Next Steps

- Need more data. Currently, only able to gather ~5k comments per day using pushshift from the r/bitcoin subreddit.
- Perform bag of words/ CountVectorization on dataset
- Use Twitter and Google trends data to calculate sentiment and bag of words
- Use XGboost and LightGBM