Cryptocurrency and the Wisdom of the Crowds

Predicting changes in Bitcoin prices using Public Interest Measures

Group 1: Aashka Trivedi (aht323), Minji Kim (mk7773), and Omkar Darekar (oad245)

Abstract

- This project aims to study the influence of <u>public sentiment & interest</u> on the price of the most popular cryptocurrency, **Bitcoin**

- Study the specific correlation between changes in Bitcoin prices and three social media interest measures

  - **Google Trends**

  - **Twitter Sentiment**

  - **Reddit Networks**

- Train a classification model that predicts the **direction** of changing Bitcoin prices given the sentiment interest scores

- Used datasets are both <u>Daily</u> and and <u>Hourly</u> granularity

Business Understanding
Why Predict Bitcoin Prices?

- Bitcoin is one of the most widely used forms of digital currency

- Cryptocurrency prices are **highly volatile**

- No clear reason for the changing price - Different from stock market

- A lot of research assume the public sentiment mostly affects price

- Various attempts have been made to predict Bitcoin prices using various factors

  - No conclusive result - these attempts were on **single** data sentiment

Prior Works

- Google Trends

  - [1] Weekly Google Search trends are correlated to Bitcoin prices

  - [2] Dissent that Google Trends **alone** are a strong predictor of price changes

- Twitter Sentiment

  - [3] Validates that Twitter feeds being an efficient predictor of a massive conglomerate

  - [4] Proved the efficacy of taking into consideration the tweet volumes of users along with the sentiment as a predictor

  - [5] High degree of correlation between the twitter user sentiment -> Helps in predicting extremely volatile asset like Bitcoin

- Reddit Networks

  - [6] Defines a medium-term positive correlation between price & online activity

  - [6] Reddit has been successfully used as a data source used to model user behavior

  - [6] **Combination** of the features from Reddit along with past price fluctuations gave better forecasting -> Insights to gain **3 different web data**

[1] "Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomenaof the internet era,", L. Kristoufek
[2] "What drives cryptocurrency prices? an investigation of google trends and telegram sentiment,", N. Smuts
[3] "Algorithmic trading of cryptocurrency based on twitter sentiment analysis,", S. Colianni and M. Signorotti
[4] "Cryptocurrency price prediction using tweetvolumes and sentiment analysis,", J. Abraham
[5] "Does twitter predict bitcoin?", A. Urquhart and Wang
[6] "Extracting cryptocurrency price movements from the reddit network sentiment,", S. Wooley, A. Edmonds, A. Bagavathi, and S. Krishnan

Data Understanding and Preparation
Google Trends Data

- Daily Google Search Numbers (PyTrends API)

- Constellation of Keywords related to Bitcoin

- Daily Search Score

  - Average Score

    - Average daily score for all keywords

  - Weighted Score

    - Weighted by how much the term is usually searched for

- Hourly Search Score

  - Replicate the Daily Search Score

Data Understanding and Preparation
Twitter Sentiment

- Scrapped tweets using the hashtag #bitcoin

  - Kaggle Dataset, Twitter API

- VADER Sentiment Analysis

  - Compound Score: Measure of how Positive (+), Negative (-), or Neutral (0)

- Daily and Hourly Sentiment Score

  - Average Compound Sentiment per Day/Hour

```
'$TRB from 37$ to 55%. Sweet little profits in a day. 🚀 \n\n#Bitcoin #BTC #DogecoinRise
tps://t.co/awOIX7XtOn'

{'neg': 0.0, 'neu': 0.699, 'pos': 0.301, 'compound': 0.6815}
```

Data Understanding and Preparation
Reddit Sentiment

- Scrap Reddit comments under subreddit "r/Bitcoin"

  - PushShift API

- VADER Sentiment Analysis

  - Compound Score: Measure of how Positive (+), Negative (-), or Neutral (0)

- Hourly Sentiment Score

  - Average Compound Sentiment per Hour

  - Hour of Comment referred to the original comment, not the thread

	Date	body		sentiment	neg	neu	pos	compound
0	2021-02-12 14:23:46	Bitcoin is the profits	{'neg': 0.0, 'neu': 0.508, 'pos': 0.492, 'comp...	0.000	0.508	0.492	0.4404

Data Understanding and Preparation
Bitcoin Prices

- Daily and Hourly Bitcoin Prices

  - AlphaVantage API and CryptoDataDownload

- Categorization of Price Change

- Daily Price Change (5 Categories)
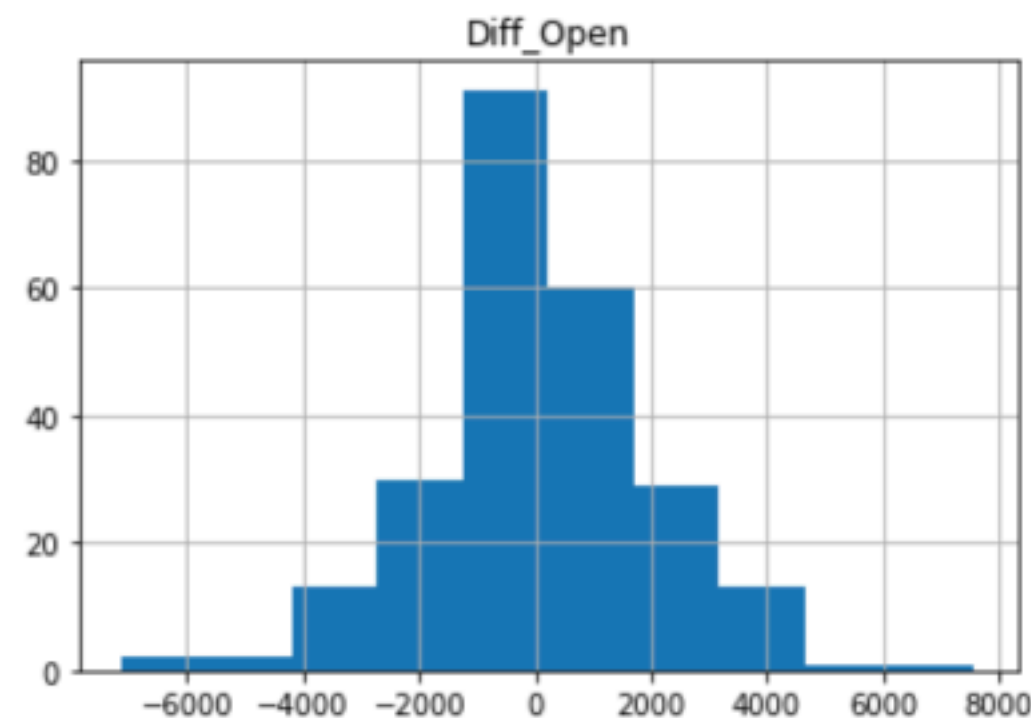
- Hourly Price Change (3 Categories)

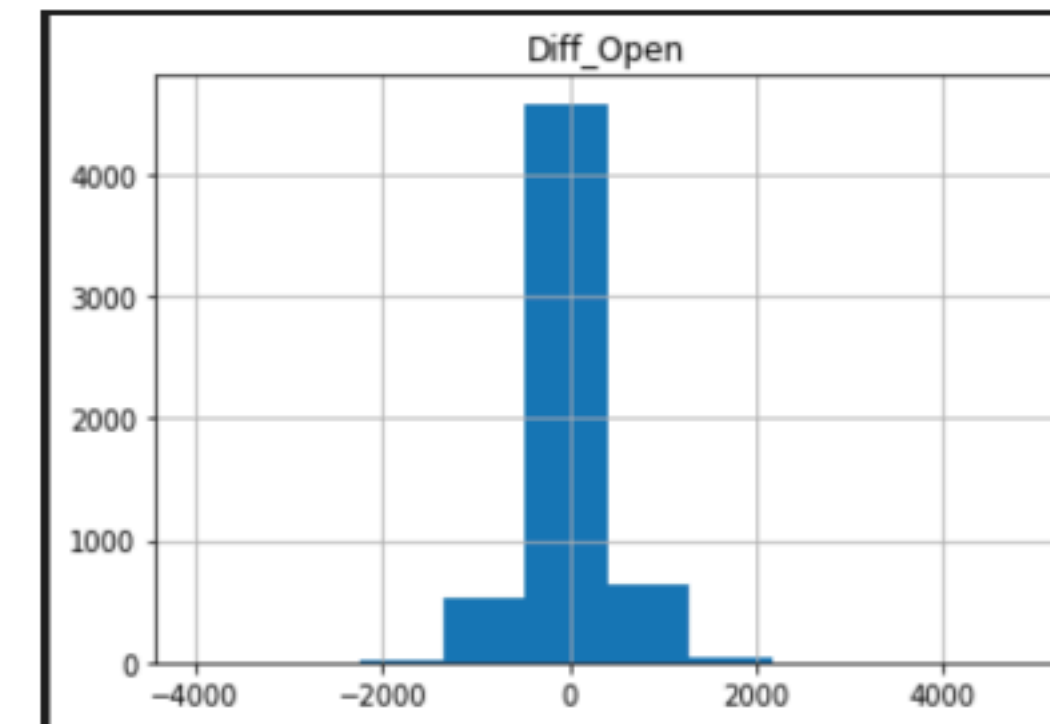Figure 1: The number of days vs the difference between daily Bitcoin values

Figure 2: The number of hours vs the difference between Hourly Bitcoin values

Modeling and Deployment
Granger Causality Analysis

- Granger Causality

  - Chi Square Test

  - Stationary Time Series

  - P Value : 0.05 (Causality Values < 0.05 implies Causation)

- Daily Granularity

- Hourly Granularity

	Daily Bitcoin Open Price
Daily Twitter Sentiment	**0.0011**
Daily Reddit Sentiment	0.2147
Daily Average Google Search Score	**0.0420**
Daily Weighted Google Search Score	0.0502

Table 1: Granger Causality Matrix for Daily Granularity

	Hourly Bitcoin Open Price
Hourly Twitter Sentiment	**0.0010**
Hourly Reddit Sentiment	0.6644
Hourly Average Google Search Score	**0.0075**

Table 2: Granger Causality Matrix for Hourly Granularity

- Twitter Sentiment and Google Trends Granger Cause Bitcoin Prices

Modeling and Deployment
Supervised Learning

- Missing Values: Averaged Sentiment Score

- Random Forest Classifier, Decision Tree Classifier, Support Vector Classifier

Performance Measure	Random Forest Classifier	Decision Tree Classifier	Support Vector Classifier
Accuracy	0.545	0.454	0.272
F1-Score	0.546	0.464	0.116
Precision	0.753	0.772	0.074
Recall	0.545	0.454	0.272

Table 3: Evaluation of Different Algorithms to predict Daily Bitcoin Price Change

Performance Measure	Random Forest Classifier	Decision Tree Classifier	Support Vector Classifier
Accuracy	0.409	0.467	0.500
F1-Score	0.417	0.537	0.374
Precision	0.431	0.493	0.353
Recall	0.409	0.537	0.500

Table 4: Evaluation of Different Algorithms to predict Hourly Bitcoin Price Change

Evaluation
Hyperparameter Tuning using Grid Search

- Daily Granularity

- Hourly Granularity

Performance Measure	Hyperparameter-Optimized Random Forest Classifier
Accuracy	0.566
F1-Score	0.481
Precision	0.589
Recall	0.4137

Table 5: Performance of Random Forest Model after Hyperparameter Optimization

Performance Measure	Hyperparameter-Optimized Decision Tree Classifier
Accuracy	0.588
F1-Score	0.495
Precision	0.433
Recall	0.557

Table 6: Performance of Decision Tree Model after Hyperparameter Optimization

Conclusion
Observations and Future Scope

- Relationship between Public Interest Measures and Bitcoin Prices

- Supervised Learning Models (~60% Accurate)

- Future Scope and Improvements

  - More Data, Different Time Frame

  - Different type of Cryptocurrency

  - Complex Sentiment Score Aggregation Techniques

Questions