Impact of Social Consensus on Bitcoin Price Action

Cihan EVREN

Supervised by Professor Patrick GLAUNER

MOTIVATION

- The motivation behind our project stems from the realization that Bitcoin price predictions rely on numerous metrics such as; **Moving Averages**, **Relative Strength Index (RSI)**, **Trading Volume**, **Trading Patterns**. Upon existing approaches, we believe that **Social Sentiment** also holds an immense potential as an additional influential factor.
- By exploring the **relationship between social sentiment and the price action of Bitcoin**, we aim to uncover valuable insights that can enhance the accuracy of price predictions.
- Through understanding how public sentiment impacts Bitcoin’s value, we can contribute to a more comprehensive understanding of the cryptocurrency market and **empower investors with actionable insights** for informed decision-making.

OBJECTIVES

- Investigating the correlation between the **volume of YouTube comments and market conditions**.

- Studying the correlation between the **quantity of YouTube videos and the price of Bitcoin**.

- Exploring the relationship between the **sentiment score of YouTube comments and the price movement of Bitcoin**.

WHAT’S BITCOIN?

- Bitcoin is a **<u>digital asset</u>** that operates on a **<u>decentralized network</u>**, enabling individuals to **<u>store and transfer value</u>** without the need for **<u>intermediaries</u>** such as traditional banks.
- Bitcoin (AKA Digital Gold) **<u>can be bought, sold and stored in digital wallets</u>**. Also, its value can fluctuate based on market demand and other factors.
- As of today, Bitcoin’s market cap is at **<u>$521.23B</u>**
- The total cryptocurrency market cap is at around **<u>$1.13T</u>**
- For comparison; Gold’s market cap is at **<u>$13.3T</u>**

WHY YOUTUBE AS A SOURCE?

- Three primary social media platforms contribute to the expansion of the crypto community: **Twitter**, **YouTube**, **Reddit**.
- Based on a literature review, it is evident that numerous sentiment analysis studies **primarily utilize Twitter as the main data source**. This preference is attributed to the presence of **hashtags** on Twitter. The use of hashtags makes data collection and preprocessing for sentiment analysis relatively easier compared to other sources.
- Due to the challenge of retrieving relevant data without the use of hasthags, many researchers tend to avoid using YouTube as a source for sentiment analysis. However, **we strongly believe that YouTube holds significant value as a source for studying the impact of social sentiment on bitcoin's price action**.

RESEARCH QUESTIONS

- Does a correlation exist between the number of comments made in YouTube and the price movements of Bitcoin ?

- To what extent can social sentiment on YouTube be used as a predictive indicator for Bitcoin’s price action?

RELATED WORK

- Effect of Sentiment on Bitcoin Price Formation by Brian Perry-Carrera (2018)
- The study period is December 1st 2017 and December 31st 2017
- Performing sentiment analysis over **500,000 tweets**
- Uses a lexicon and rule-based sentiment tool that is specifically attuned to sentiments expressed in social media, namely **VADER** (Valence Aware Dictionary and sEntiment Reasoner)
- Additionally studies the correlation between **Gold futures** and **Bitcoin** price action
- Concludes that while the **gold futures is negatively related to the price of bitcoin**, there is a **positive relationship between the sentiment and the bitcoin price**.

RELATED WORK

- The Interplay between Social Media Sentiment and the Cryptocurrency Market by Lukas Eisner (2019)
- Running a Multi-Class Sentiment Analysis (positive, neutral and negative) on Tweets that is scraped for **20 different cryptocurrencies** (BTC, ETH, XRP etc.)
- The models used for sentiment analysis; **Naïve Bayes**, **Logistic Regression Classification Tree**, **Random Forest**, **Gradient Boosting**, **Support Vector Machine** and **Convolutional Neural Network**
- Concludes that through a **lead-lag analysis** there is a **cross-correlation** between the sentiment score and the market value. It is however a **complex interaction** with a shifting lead-lag structure.

CONTRIBUTION

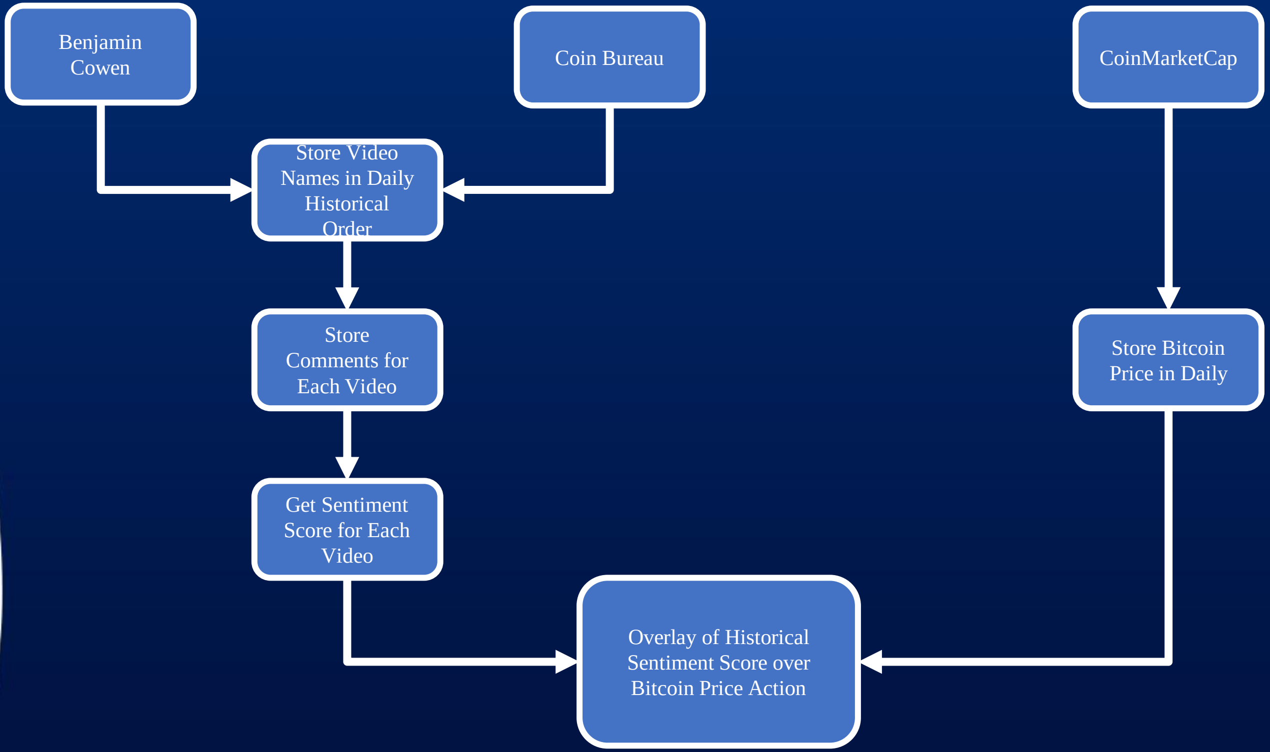
- While previous researches have primarily focused on analyzing social sentiment through **Twitter data**, our study takes a different approach by utilizing **YouTube comments** as the primary source of data.
- In contrast to the common practice of conducting studies **over monthly or yearly periods**, we have opted for a **4-year study period** to gain a comprehensive understanding of the volatility.
- While sentiment analysis is typically conducted using common tools such as **Support Vector Machines**, **Random Forest**, **Logistic Regression**, **CNN**, or lexicon-based tools like **VADER**, our approach will be centered around leveraging advanced tools such as **GPT-3** or **RoBERTA** for more accurate sentiment analysis.

NUMBERS

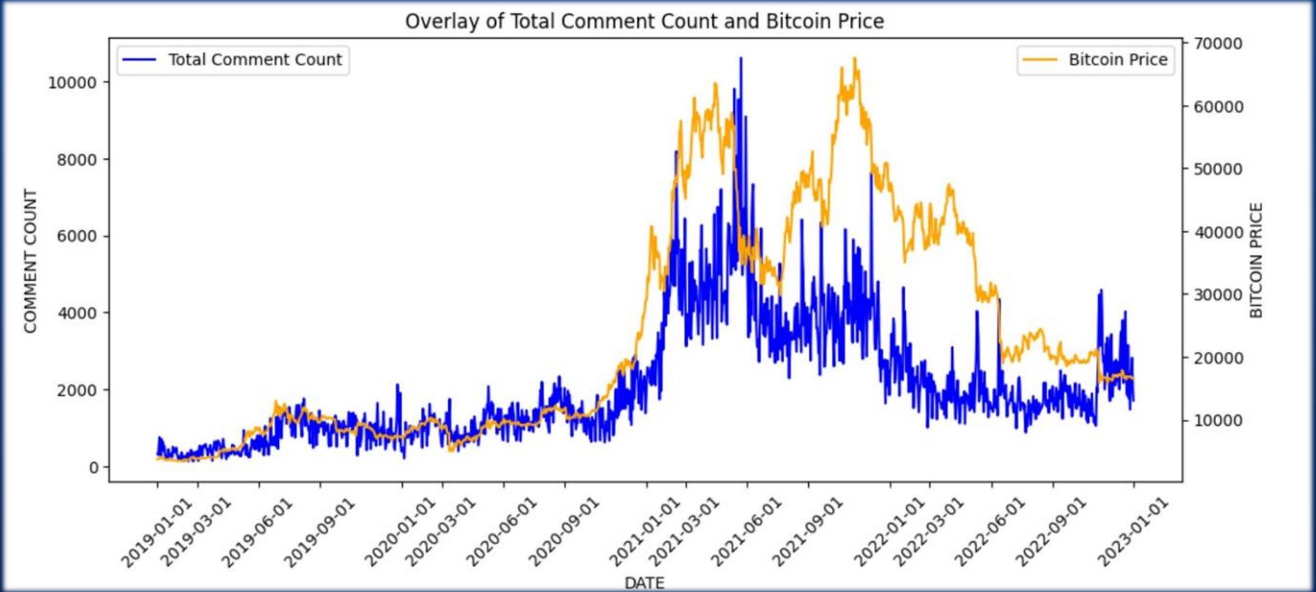
- Research Period : **01.01.2019 – 01.01.2023**
- Channel Count : **13**
- Video Count : **16845**
- Comment Count : **2.992.699**


PEER TO PEER
BTC

WORKFLOW

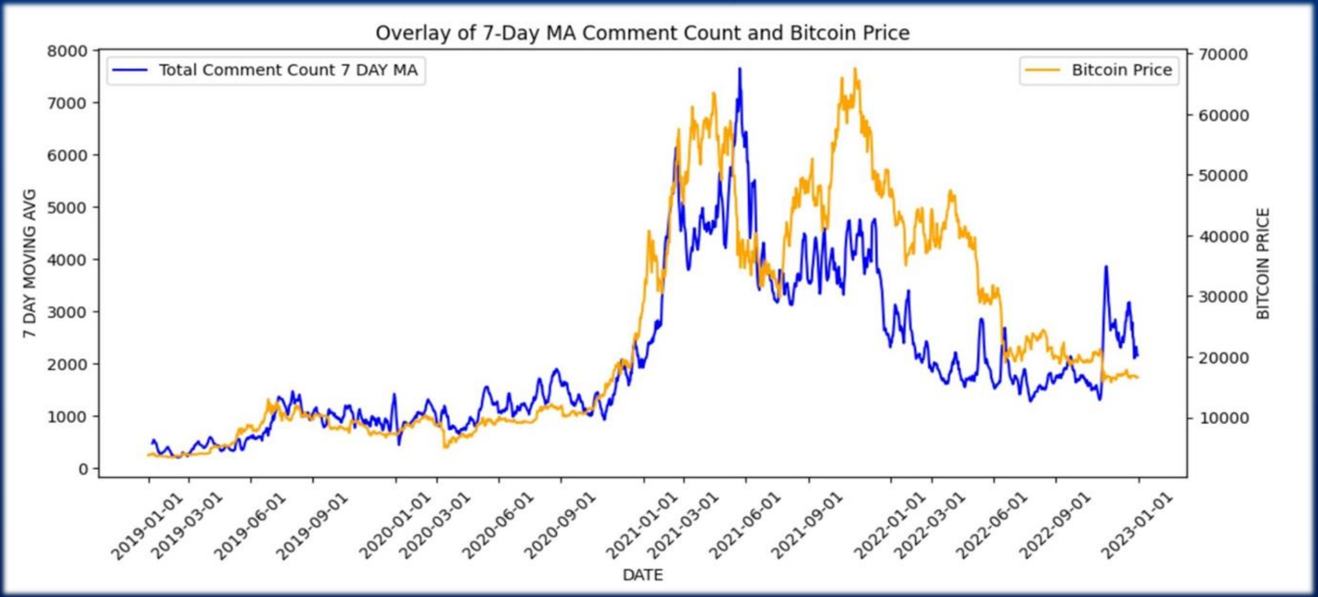

Benjamin Cowen
Coin Bureau
CoinMarketCap
Store Video Names in Daily Historical Order
Store Comments for Each Video
Get Sentiment Score for Each Video
Store Bitcoin Price in Daily
Overlay of Historical Sentiment Score over Bitcoin Price Action

GRAPH 1



- The **orange line** represents the **daily opening** price of Bitcoin from 2019 to 2023.
- The **blue line** represents the **count of YouTube comments** over the same 4-year period.
- Increasing comment count corresponds to rising Bitcoin price, indicating a **potential correlation**.
- To enhance the visualization, we can apply a 7-day moving average to the comment counts.

GRAPH 2



- The **orange line** represents the **daily opening** price of Bitcoin from 2019 to 2023.
- The **blue line** represents the **7-Day moving average of YouTube comments** over the same 4-year period.
- The observed correlation is evident, but to establish its statistical significance, we need to conduct further tests.

GRANGER CAUSALITY TEST

- It is a **statistical hypothesis test** for determining whether one time series is a factor and offer useful information in **forecasting another time series**.
- Granger Causality Test assumes that the both timeseries are **stationary** as it ensures that the relationships between variables are consistent and reliable
- To verify the stationarity of our timeseries, we can perform an **ADF (Augmented Dickey-Fuller) test**. This test helps determine if the time series data is stationary, which is a key assumption for reliable analysis.

ADF TEST RESULTS

- *Null Hypothesis*: **<u>If failed to be rejected</u>**, It suggests the time series is **<u>not stationary.</u>**
- *Alternative Hypothesis*: **<u>The null hypothesis is rejected</u>**, it suggests the time series **<u>is stationary.</u>**
- After we run the ADF on bitcoin price time series, the **<u>p-value is 0.5354</u>** ($p > 0.05$) meaning null hypothesis **<u>is failed to be rejected</u>**, indicating that the time series **<u>is not stationary</u>**.
- Once we run the ADF on comment count time series, the **<u>p-value is 0.2481</u>** ($p > 0.05$), concluding with the same result the time series **<u>is not stationary.</u>**

ENSURING STATIONARITY

- To **ensure stationarity** we can use a technique called **differencing the data**. Computing the difference between consecutive observations in the series will help us to **remove seasonality and trend components** in the time series.
- After applying the differencing technique to both datasets and conducting the **ADF test again**, the results show that the p-value for the bitcoin price series is **4.98e-11**, indicating that the series is stationary (**Null hypothesis is rejected, $p < 0.05$**). Similarly, the p-value for the comment count series is **5.41e-25**, also indicating that the series **is stationary**.
- Now, we can perform the **Granger Causality Test** on the differenced data.

GRANGER CAUSALITY

```
Granger Causality
number of lags (no zero) 1
ssr based F test:         F=0.0154  , p=0.9013  , df_denom=1456, df_num=1
ssr based chi2 test:   chi2=0.0154  , p=0.9012  , df=1
likelihood ratio test: chi2=0.0154  , p=0.9012  , df=1
parameter F test:         F=0.0154  , p=0.9013  , df_denom=1456, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=4.4836  , p=0.0114  , df_denom=1453, df_num=2
ssr based chi2 test:   chi2=8.9980  , p=0.0111  , df=2
likelihood ratio test: chi2=8.9703  , p=0.0113  , df=2
parameter F test:         F=4.4836  , p=0.0114  , df_denom=1453, df_num=2

Granger Causality
number of lags (no zero) 3
ssr based F test:         F=1.3152  , p=0.2678  , df_denom=1450, df_num=3
ssr based chi2 test:   chi2=3.9647  , p=0.2653  , df=3
likelihood ratio test: chi2=3.9593  , p=0.2659  , df=3
parameter F test:         F=1.3152  , p=0.2678  , df_denom=1450, df_num=3

Granger Causality
number of lags (no zero) 4
ssr based F test:         F=2.4628  , p=0.0435  , df_denom=1447, df_num=4
ssr based chi2 test:   chi2=9.9126  , p=0.0419  , df=4
likelihood ratio test: chi2=9.8790  , p=0.0425  , df=4
parameter F test:         F=2.4628  , p=0.0435  , df_denom=1447, df_num=4
```

- The Granger Causality Test results are displayed on the left, covering a range of lag orders from one to four.
- **The first and third** results indicate **weak evidence of Granger Causality** between the dependent and independent variables, as the **p-values are greater than 0.05**.
- In contrast, **the second and fourth** results demonstrate a **strong Granger Causality** between the two variables as **$p < 0.05$**. This implies that we can build a predictive model by incorporating **either 2 or 4 lagged values of the comment count variable** to forecast the price of Bitcoin.

RESEARCH QUESTION 1 ADDRESSED

- Up to this point, we have addressed the following question:
  <u>Does a correlation exist between the number of comments made in YouTube and the price movements of Bitcoin ?</u>
- Based on the results of the Granger Causality Test conducted on our dependent and independent variables, **<u>we can confidently conclude that there is a significant Granger Causality relationship between the variables</u>**. This implies that the number of comments posted on YouTube **<u>have the ability to explain or potentially predict the price movements of Bitcoin</u>**.

FUTURE WORK

- Moving forward, our primary focus will be directed towards addressing the second question: *<u>To what extent can social sentiment on YouTube be used as a predictive indicator for Bitcoin's price action?</u>*
- The future work will be including the following steps:
  1. **<u>Preprocessing</u>** of YouTube comments for Natural Language Processing purposes.
  2. **<u>Classifying</u>** the comments to exclude those that are not relevant.
  3. Conducting a **<u>sentiment analysis</u>** to obtain sentiment scores.
  4. Performing a **<u>lead-lag analysis</u>** to explore the relationship between social sentiment and bitcoin price.

THANK YOU FOR YOUR TIME

Do you have any Questions?