Using Reddit and Markov Chains to Time Bitcoin

Motivation

- Camou (2022) shows that a sentiment analysis based on Reddit posts about cryptocurrencies has **predictive power** over crypto **volatility**. However, the results are **mixed** for predicting **returns**.
- Poyser (2018) tests the hypothesis that cryptocurrency prices are driven by **herding**. For that, he studies herding behavior under different conditions, including the **Markov-Regime-Switching** approach.
- In this work, we use **sentiment analysis** and the **Hidden Markov Model** (Baum & Petri (1966)) to get more (less) exposed to Bitcoin when there is a higher (lower) chance of the market being more bullish (bearish) in the next period.

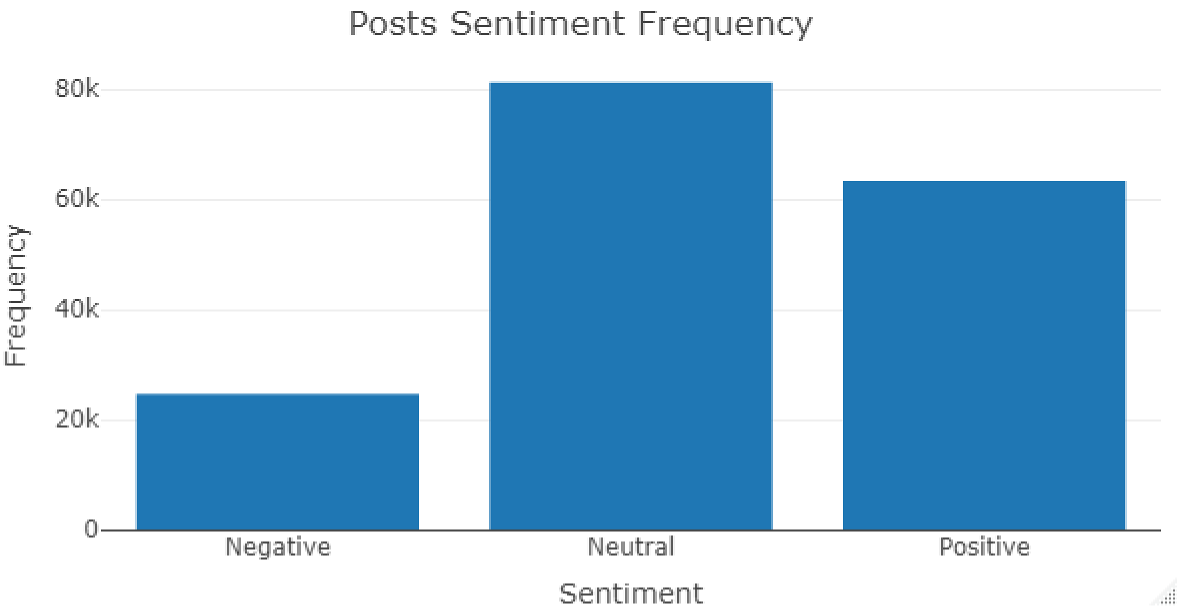
Data & Metodology

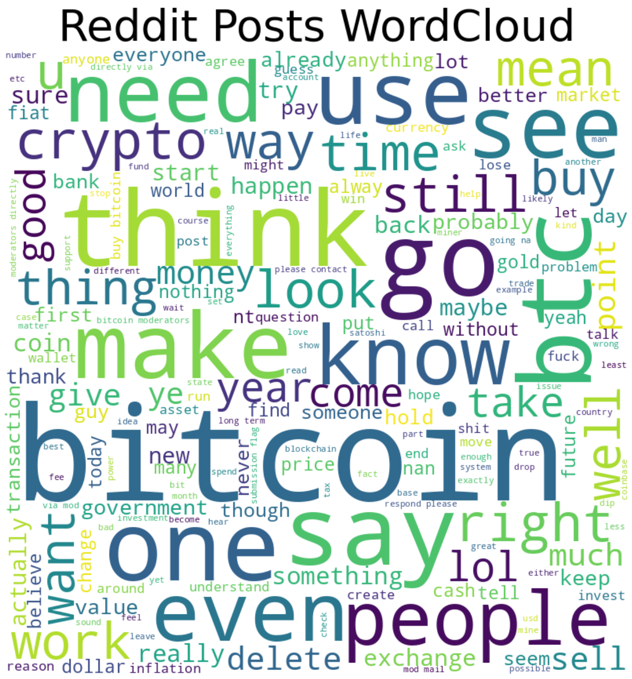
- We collect data regarding **Reddit posts** and comments from **2021-01-01** to **2022-10-01**. In total, we have **169155 text documents**, as well as information about comments and upvotes.
- First, we **preprocess** our text data, removing stopwords, punctuations, URLs, numbers, emojis, etc.
- After that, as in Camou (2022), we apply **VADER analysis** (Hutto & Gilbert (2014)) to assign an **intensity score** to every observation, indicating if it is positive or negative.

Data & Metodology

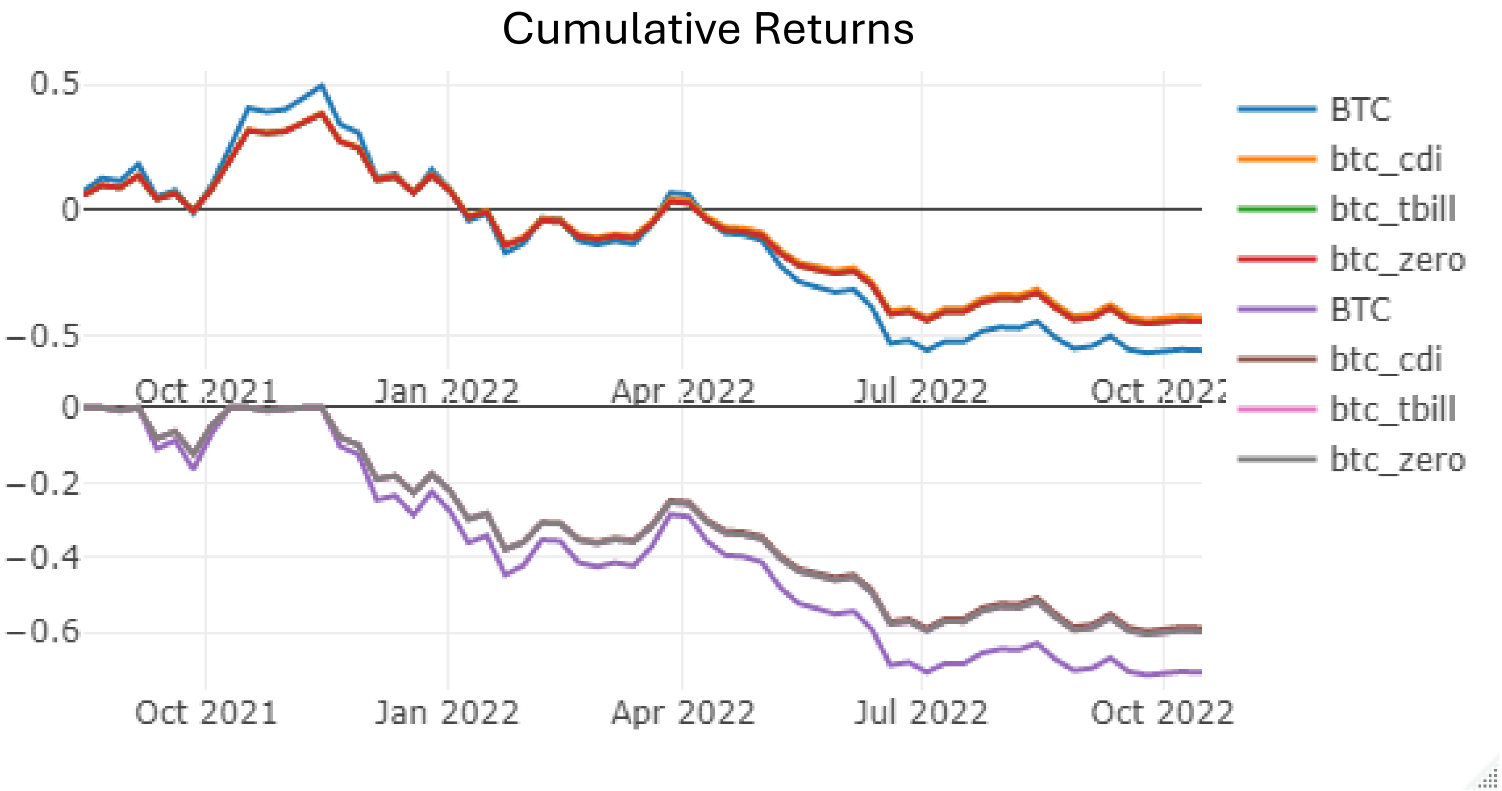
- Following, for **every week** in our sample, we take the **weighted average** of the **VADER score** based on the **post score** (difference between upvotes and downvotes).
- Thenceforth, we use this weekly score to model the Bitcoin returns' **hidden states**. We specify that there are **two hidden states**: a **bullish** and a **bearish** one.
- Finally, we go **long X%** in **Bitcoin** and **1 – X% long** in **cash** or the **risk-free** rate. In this case, X is the **probability of being in the bullish state** in the next week.

NLP Analysis


Posts Sentiment Frequency


Reddit Posts WordCloud

Results


Cumulative Returns
0.5
0
−0.5
Oct 2021
Jan 2022
Apr 2022
Jul 2022
Oct 2022
0
−0.2
−0.4
−0.6
Oct 2021
Jan 2022
Apr 2022
Jul 2022
Oct 2022
BTC
btc_cdi
btc_tbill
btc_zero
BTC
btc_cdi
btc_tbill
btc_zero

Results

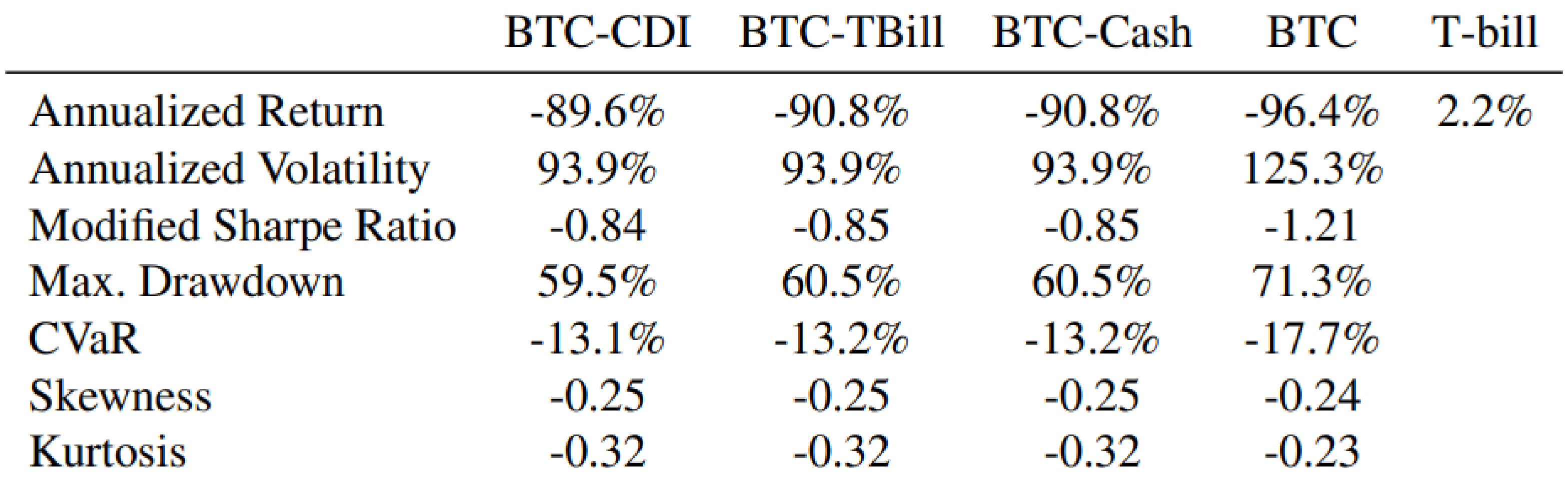


Table 1 – Portfolio Return Statistics

	BTC-CDI	BTC-TBill	BTC-Cash	BTC	T-bill
Annualized Return	-89.6%	-90.8%	-90.8%	-96.4%	2.2%
Annualized Volatility	93.9%	93.9%	93.9%	125.3%	
Modified Sharpe Ratio	-0.84	-0.85	-0.85	-1.21	
Max. Drawdown	59.5%	60.5%	60.5%	71.3%	
CVaR	-13.1%	-13.2%	-13.2%	-17.7%	
Skewness	-0.25	-0.25	-0.25	-0.24	
Kurtosis	-0.32	-0.32	-0.32	-0.23	

Conclusion and Future Developments

- Since the **scrapping** process (API limits) and **NLP** analysis (preprocessing) are **slow**, we had to limit our sample to just under **two years**. On account of this fact, the **inferences** are **limited**. Despite this warning, using sentiment analysis and the HMM, we were able to form **better portfolios** in comparison to the plain Bitcoin one.
- Interested researchers can improve the model by adding **more data** (longer time horizon) and by exploring **different structures**. The work could also be expanded by adding **other cryptocurrencies** and getting data from **other social platforms**, like Twitter.

Open Science

References

- CAMOU, L. A. L. Reddit as a prediction tool for crypto-assets. Brazilian Review of Finance, v. 20, n. 1, 2022.
- POYSER, O. Herding behavior in cryptocurrency markets. arXiv preprint arXiv:1806.11348, 2018.
- BAUM, L. E.; PETRIE, T. Statistical inference for probabilistic functions of finite state Markov chains. The annals of mathematical statistics, JSTOR, v. 37, n. 6, p. 1554–1563, 1966.
- Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model forsentiment analysis of social media text,Proceedings of the InternationalAAAI Conference on Web and Social Media, Vol. 8.
- ISRAELSEN, C. L. et al. A refinement to the sharpe ratio and information ratio. Journal of Asset Management, v. 5, n. 6, p. 423–427, 2005.