

The Forgotten Schism

Social media engineering, censorship, and the first
economic fork

Contents

- Why?
- What are we talking about?

- EDA
- Modeling
- Model Performance
- Conclusions
- Action Items

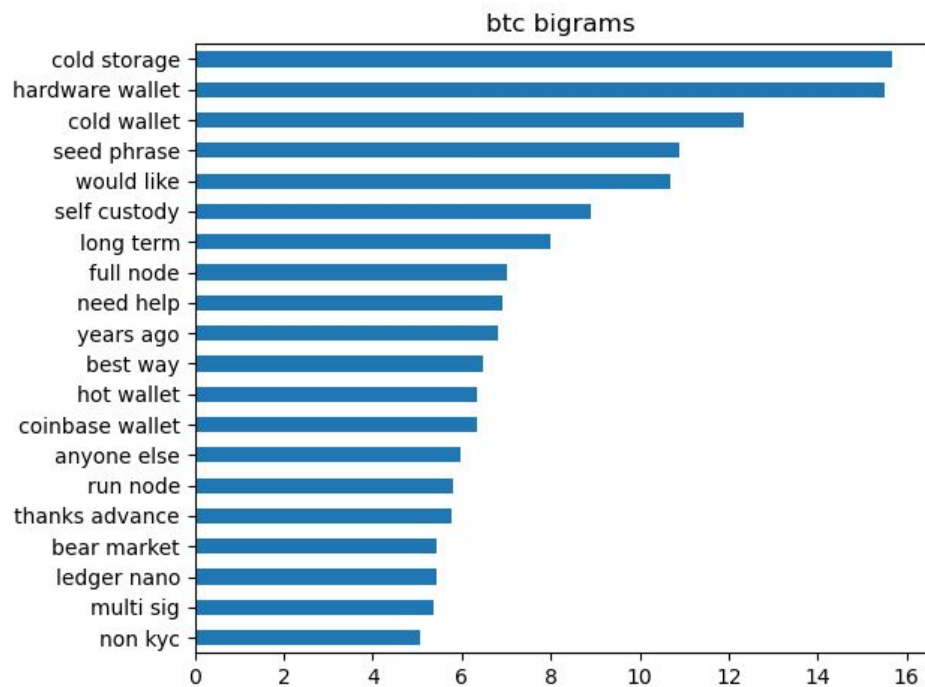
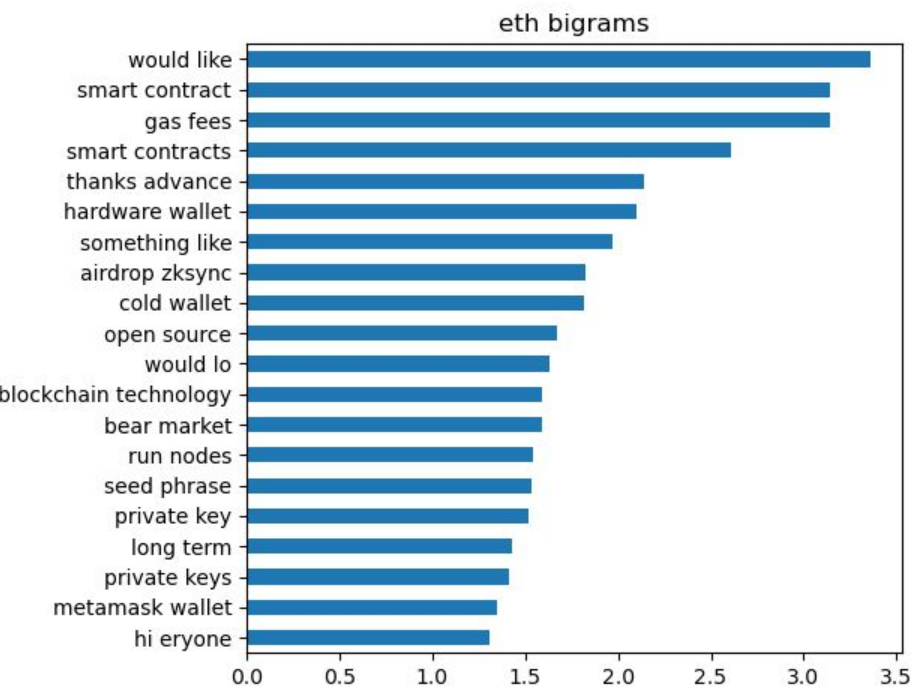
Why?

- Sentiment analysis is valuable to enrich existing data
- WeMeta is creating an analytics platform to monitor dapp repos, community forums, and publicity in addition to on-chain data.
- In pursuit of this goal, I have been tasked with performing initial EDA in an MVP context to parse content and categorize users into bitcoin and ethereum buckets.
 - “Which crypto users are more likely to adopt dapp products?”
 - “What are trends with more progressive crypto users?”

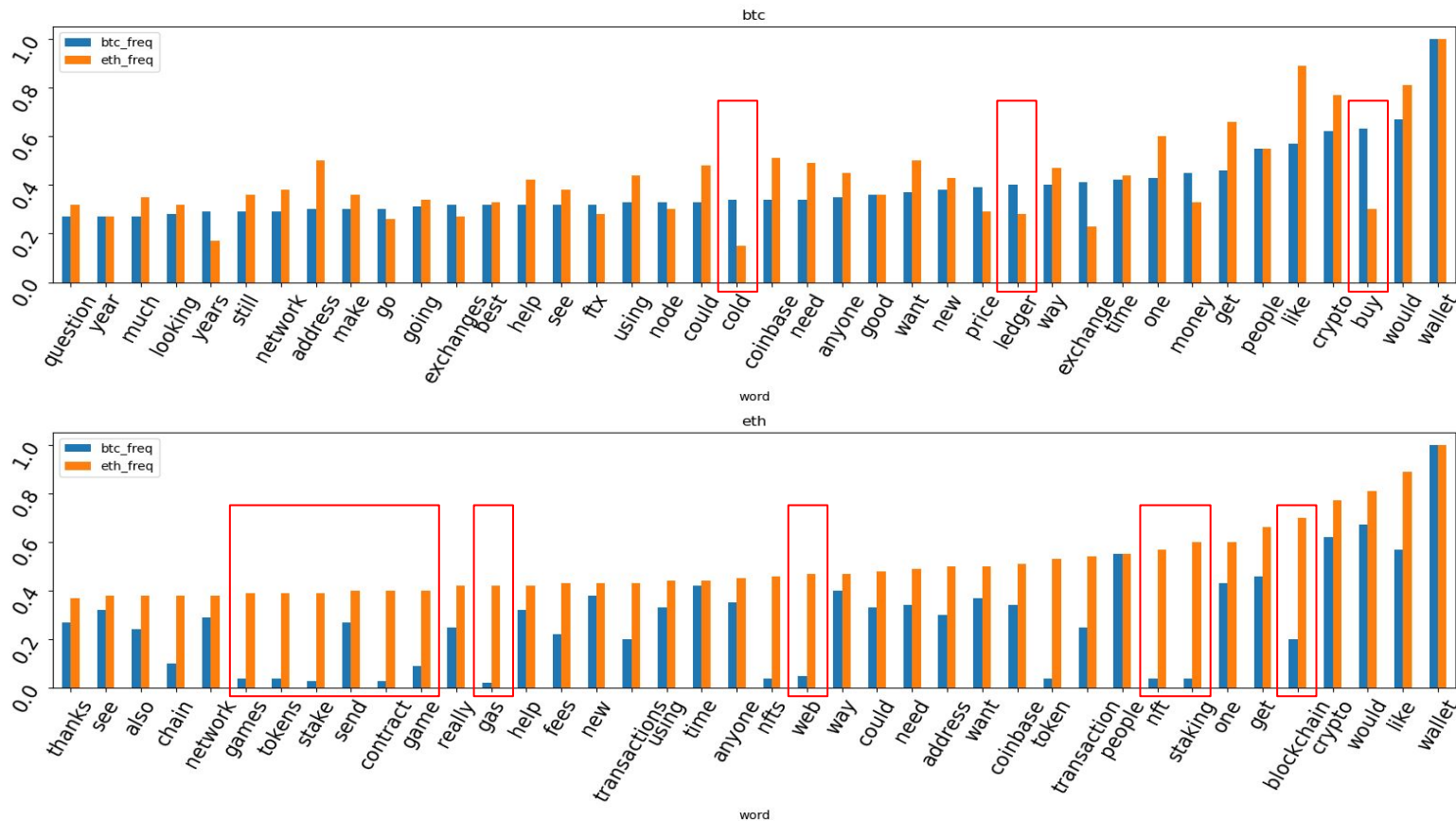
What are we talking about?

- From 2015-2017, Bitcoin was “taken over” by devs who formed a startup called Blockstream to capture BTC.
-
- Success - original devs were pushed out, Blockstream features were integrated (Segwit, RBF, lightning).
- Anon. mod Cobra-Bitcoin was key in maintaining strict censorship during the transition (and managed almost all Bitcoin-related forum traffic on earth).
- /r/bitcoin became a cult-like temple for the glorification of “right-think” about the bitcoin protocol. Culled users have moved to other platforms.
- **Have these subreddits sufficiently diverged that we can easily parse users?**

EDA: Bigrams

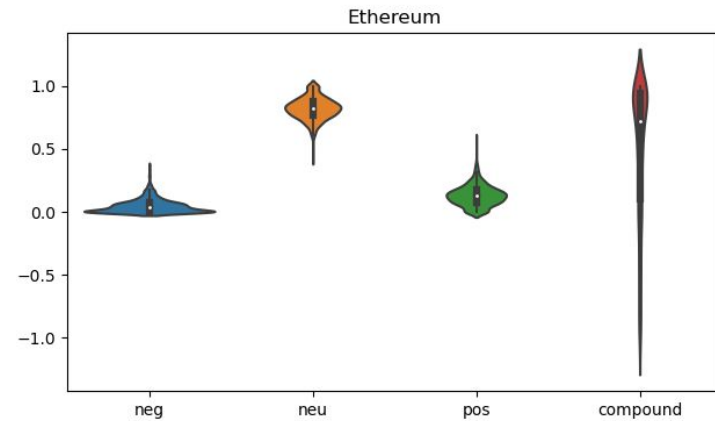
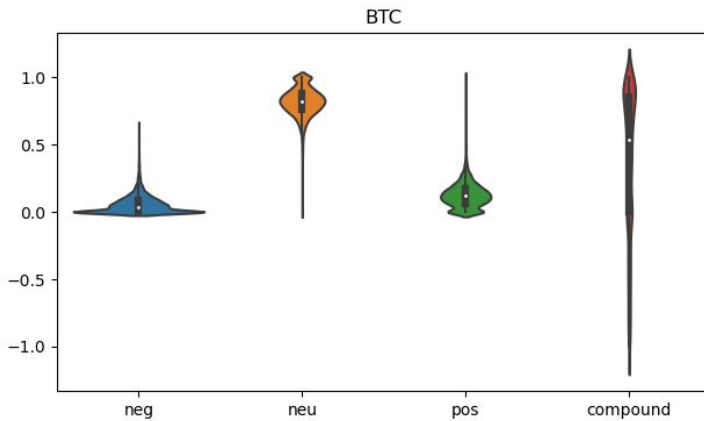


EDA: Common word Crossover



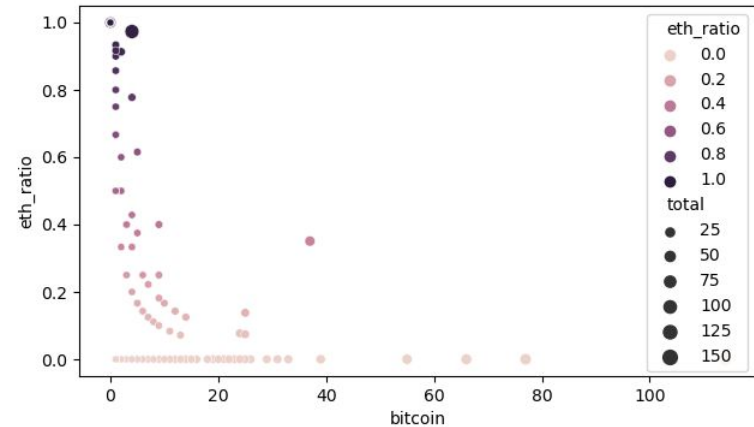
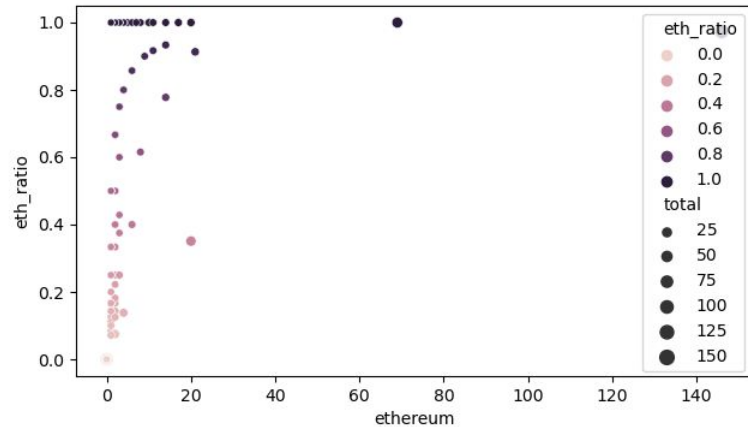
EDA: Sentiment (Vader)

- Sentiment is similar
 - Neutral, technical / infrastructure-oriented
 - ETH is slightly more positive



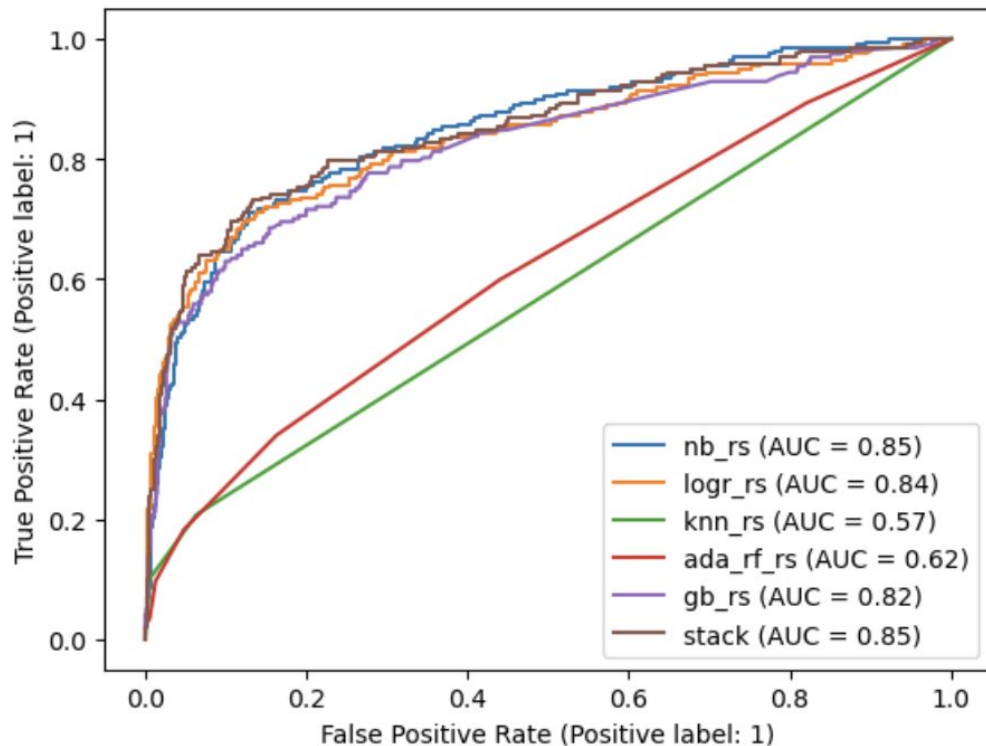
EDA: User behavior

- Crossover between subreddits is low
 - This aligns with evolution of subreddits



Modeling

- Parse, stop words, lemmatize
- Tokenize / vectorize
- Classification models
 - Naive Bayes
 - Logistic Regression
 - KNN
 - Adaboost (RF base)
 - Gradient Boost
 - Stack (nb, logr, gb, ada/rf)
- Hyperparams & best_estimator_



Model Performance

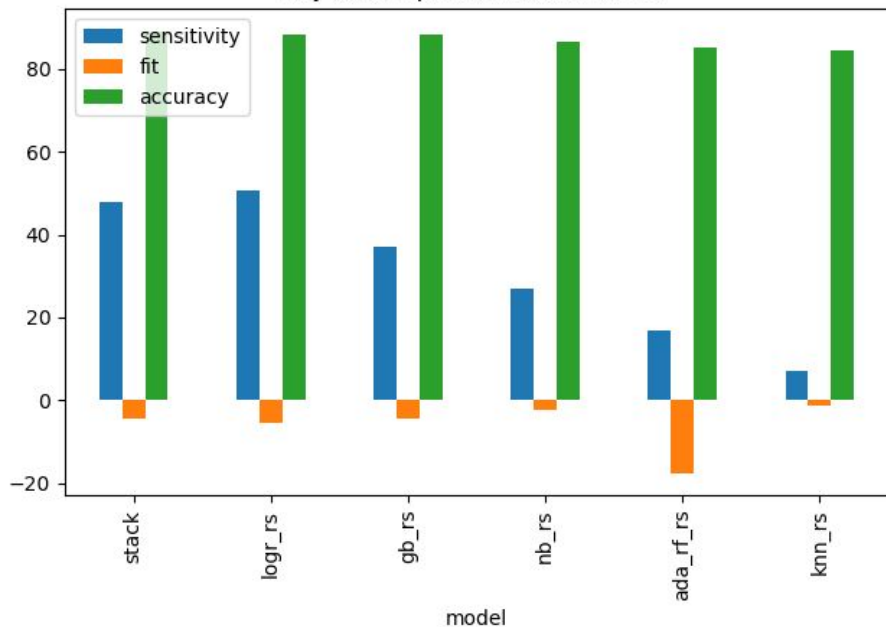
- High specificity, low sensitivity

	tn	fp	fn	tp	sensitivity	specificity	precision	neg_predictive_val	accuracy	f1	test_score	fit
model												
stack	976	28	103	94	47.72	97.21	77.05	90.45	89.09	58.93	0.89	-4.45
logr_rs	961	43	97	100	50.76	95.72	69.93	90.83	88.34	58.82	0.88	-5.62
gb_rs	986	18	124	73	37.06	98.21	80.22	88.83	88.18	50.69	0.88	-4.47
nb_rs	987	17	144	53	26.90	98.31	75.71	87.27	86.59	39.70	0.87	-2.18
ada_rf_rs	988	16	164	33	16.75	98.41	67.35	85.76	85.01	26.83	0.85	-17.56
knn_rs	1000	4	183	14	7.11	99.60	77.78	84.53	84.43	13.02	0.84	-1.25

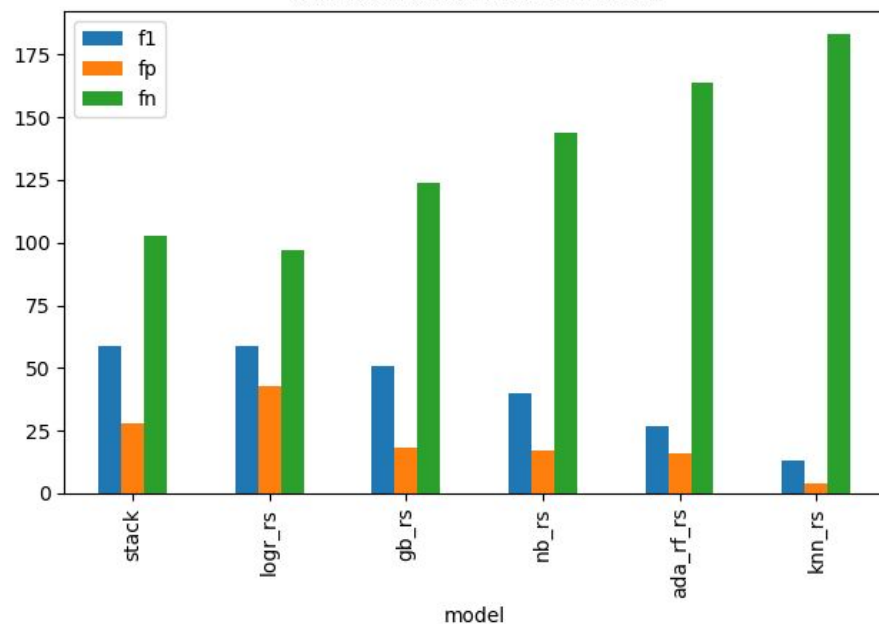
Model Performance

- FNs reduce sensitivity & F1; reduce proba threshold

key model performance metrics

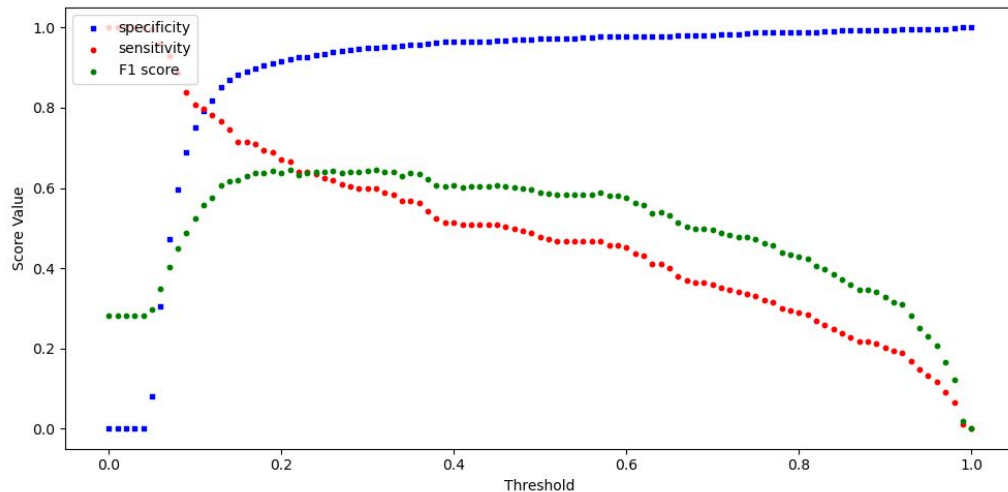
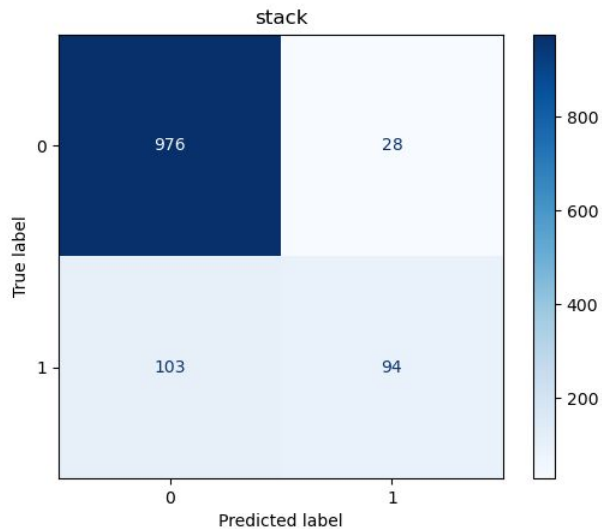


model F1 and misclassification



Model Performance: Stack

- More FNs than FPs (good true negative rate, poor true positive)



Conclusions

1. **Parsing eth users is challenging, but not impossible.** ETH is a small but growing subset of crypto, not the most publicized
2. **Topics are different, tone is similar.**
 - BTC: Hodling, cold storage, trading / markets, monetary policy, “cultier” / “us vs. them”
 - ETH: Tokens / NFTs / Web3 / games, ZKPs / infrastructure innovation, tokenomics
 - Crossover: Node operation, self-custody
3. **Classification is easy with keywords.**
 - When using basic stopwords, these topics provide a clear signal for classification

Text Samples

FNs

- i know i'll get roasted but is there anyone willing to allow me to borrow btc i gambled with the wrong people and now they are threatening me my family i had planned on getting it by today but it will be late tonight but for sure tomorrow i will pay anyone back who help me if i have to pay interest
- unpopular opinion get usb hardware wallet there i said it
- i think this need to be said with more adoption comes more stupidity and trolling on this sub we hodlers must remind ourselves of this it really doesnt matter to bitcoin so it shouldnt to u adoption is good no matter how ignorant the reason people give

FPs

- what change will bitcoin undergo in the future other than the mining reward and size of the data that will be stored on any node what can user expect to change about the protocol and actual bitcoin infrastructure mass adoption obviously come with more external infrastructure to support use and exchange but when that time come will everyone has to use lightning transaction time are much too long to expect bitcoin in it current state to replace a credit card in a shopping situation so is lightning the final solution to that
- **not your key not your coin** forbes article on binance's asset value loss in the last day are we about to see another major crash like the ftx affair i know a lot of bitcoiners are not really into mainstream medium narrative but what are your thought

Next Steps

1. **Increased feature analysis:** Consider post scores, include comments, relationship amongst text
2. **Author analysis**
3. **Include other subreddits:** Expand classification, acquire more data / training
4. **Fine tune models:** adjust proba threshold / FN rate, hyperparams

Appendix: Key characters in /r/bitcoin

“The greatest threat to bitcoin is simply googling who maintains it, and how it is governed.”

Greg Maxwell (/u/nullc)



Adam Back



Luke Dash Jr. (/u/Luke-jr)

