

Bitcoin vs Ethereum:

Operation Much Chaos



Sam Lindy

Problem Statement

I'm an operative for Dogecoin, trying to create a rift between Bitcoin and Ethereum users by sending them targeted ads to inflame tensions. Amidst this chaos, Dogecoin will rise to the top of the crypto world and reign supreme, fren.

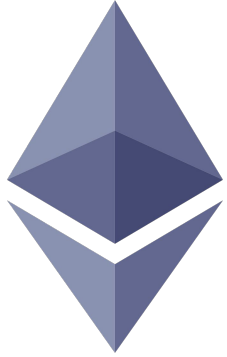
Let's create a model that will help us serve ads to the right users to maximize our limited budget.



Quick Context



\$54,290



\$2,778

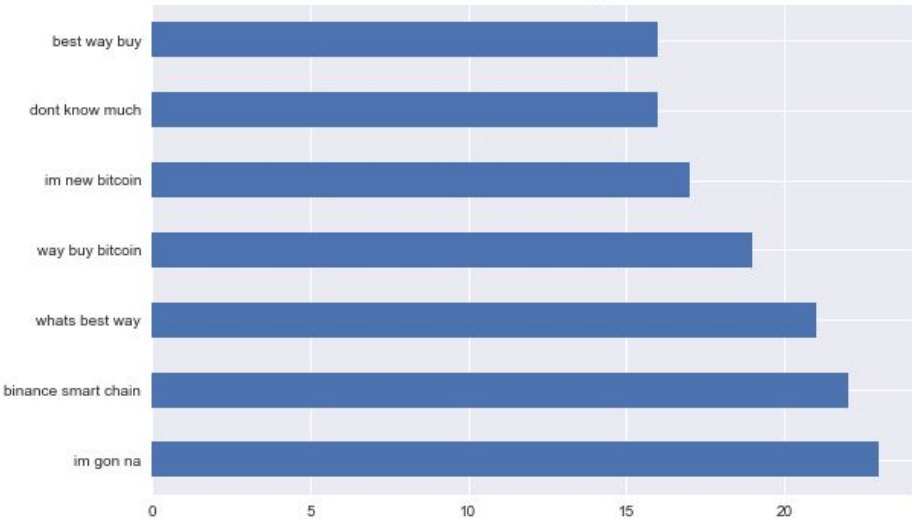


\$0.27

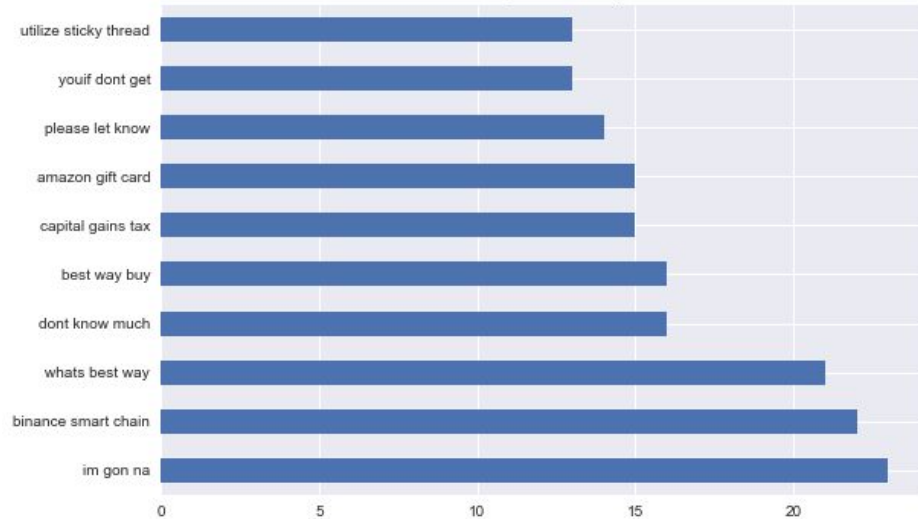
much potential

Exploratory Data Analysis - BTC Trigrams

BTC Most Common Trigrams

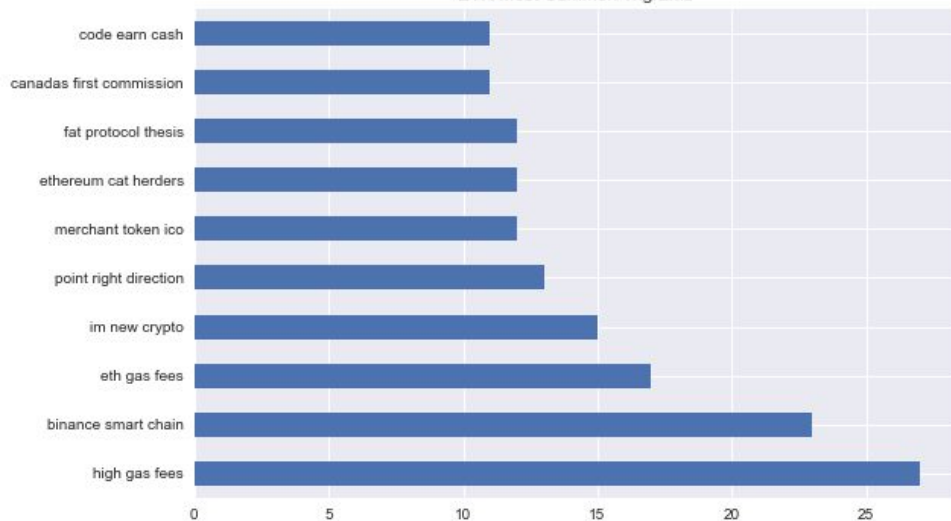


BTC Most Common Trigrams with Top Words Removed

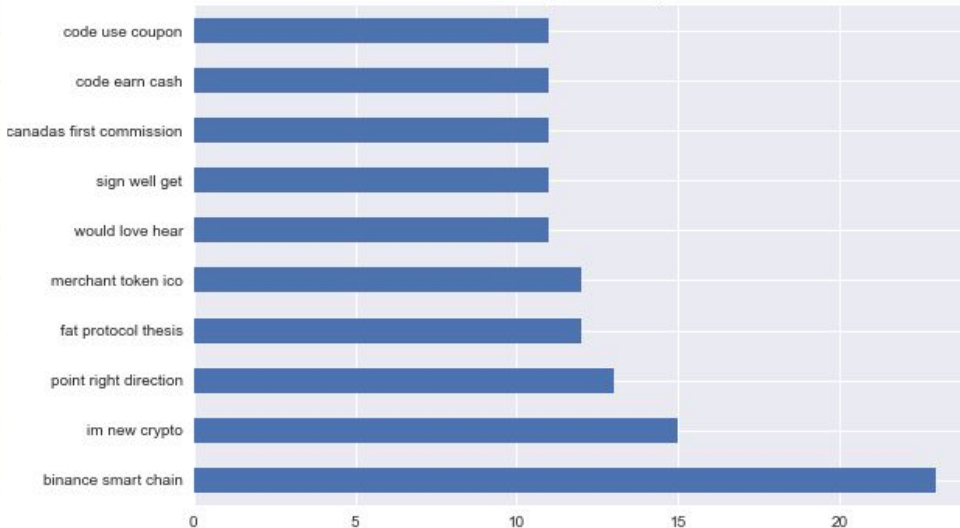


Exploratory Data Analysis - ETH Trigrams

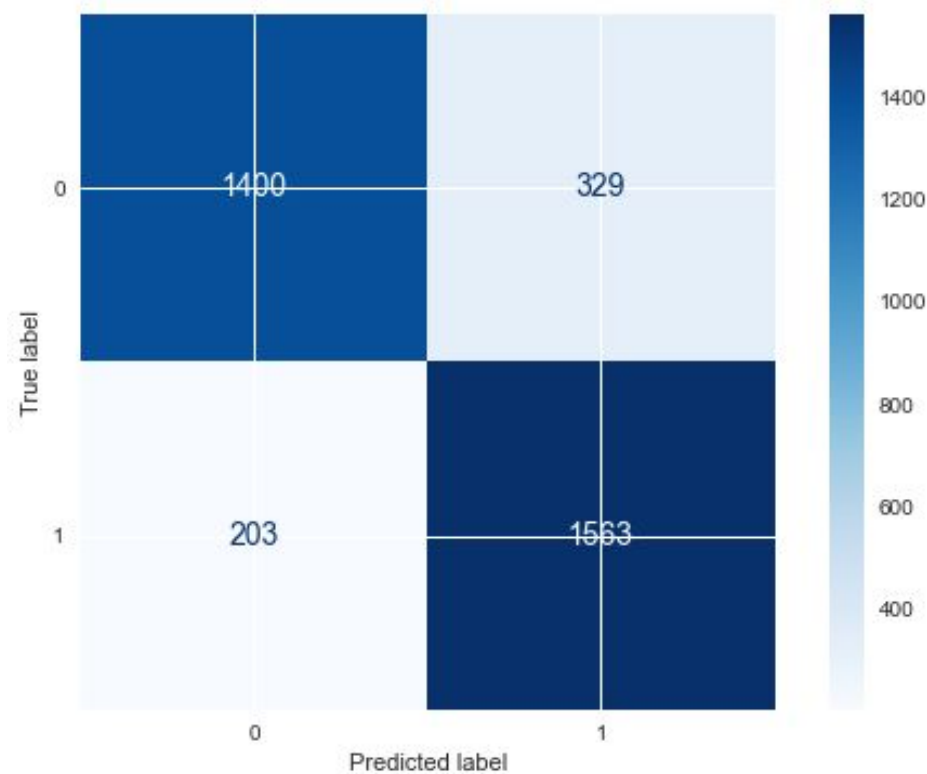
ETH Most Common Trigrams



ETH Most Common Trigrams with Top Words Removed

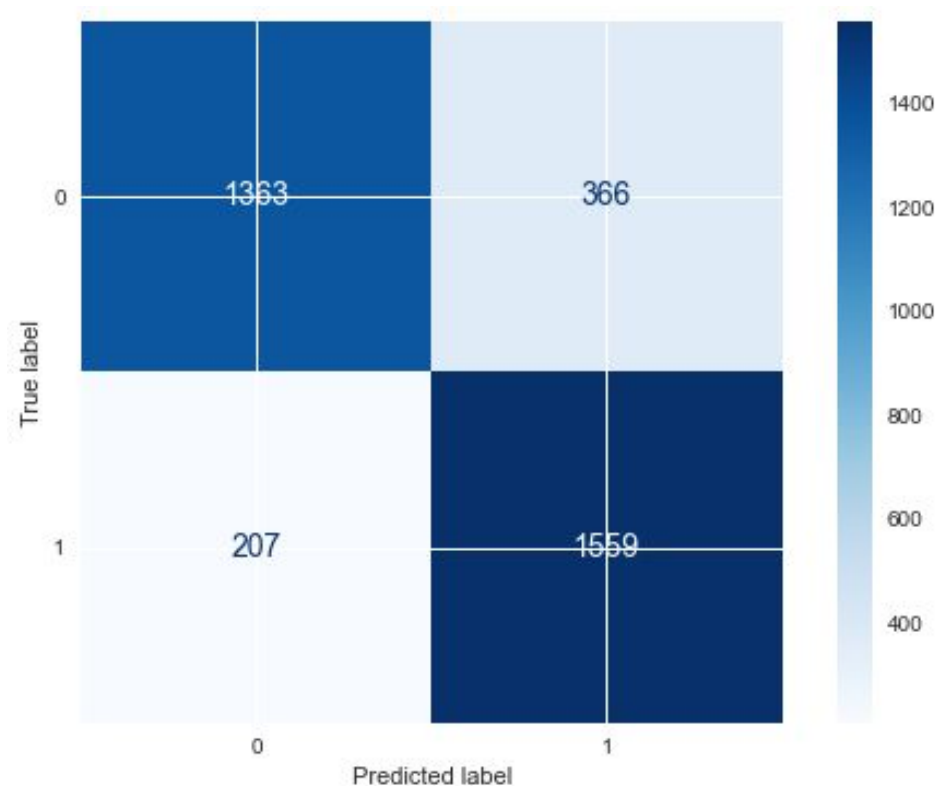


Model 1: Count Vectorizer and Naive Bayes



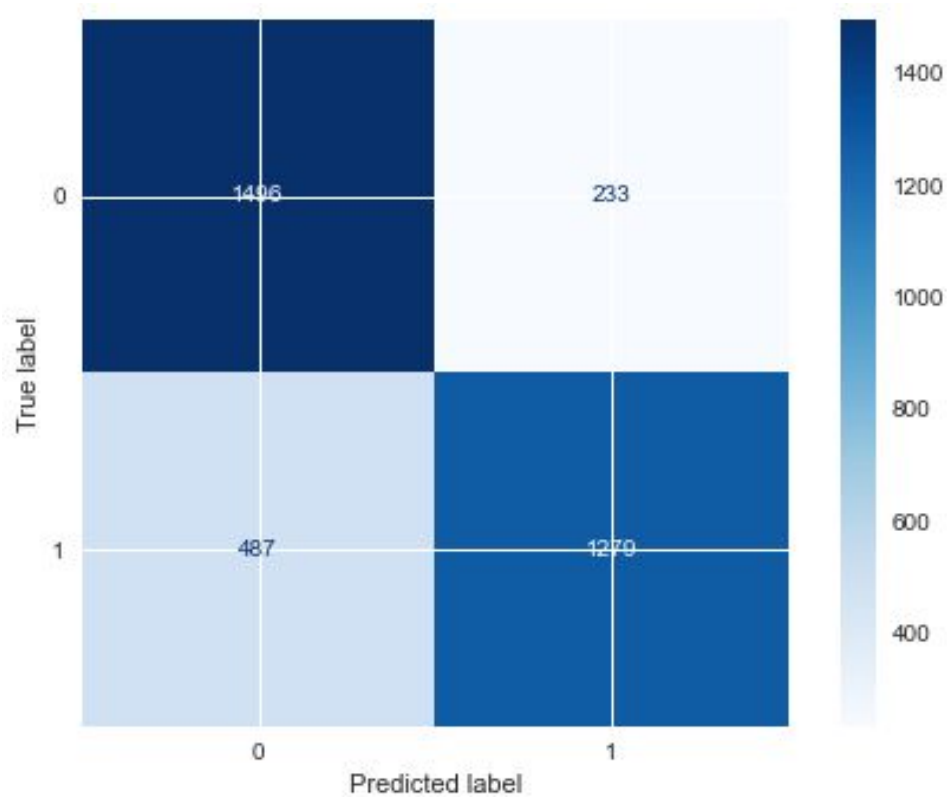
83%
Precision Score

Model 2: tf-idf and Naive Bayes



81%
Precision Score

Model 3: Random Forest



85%
Precision Score

Model Comparison: AUC-ROC Scores

93.1%

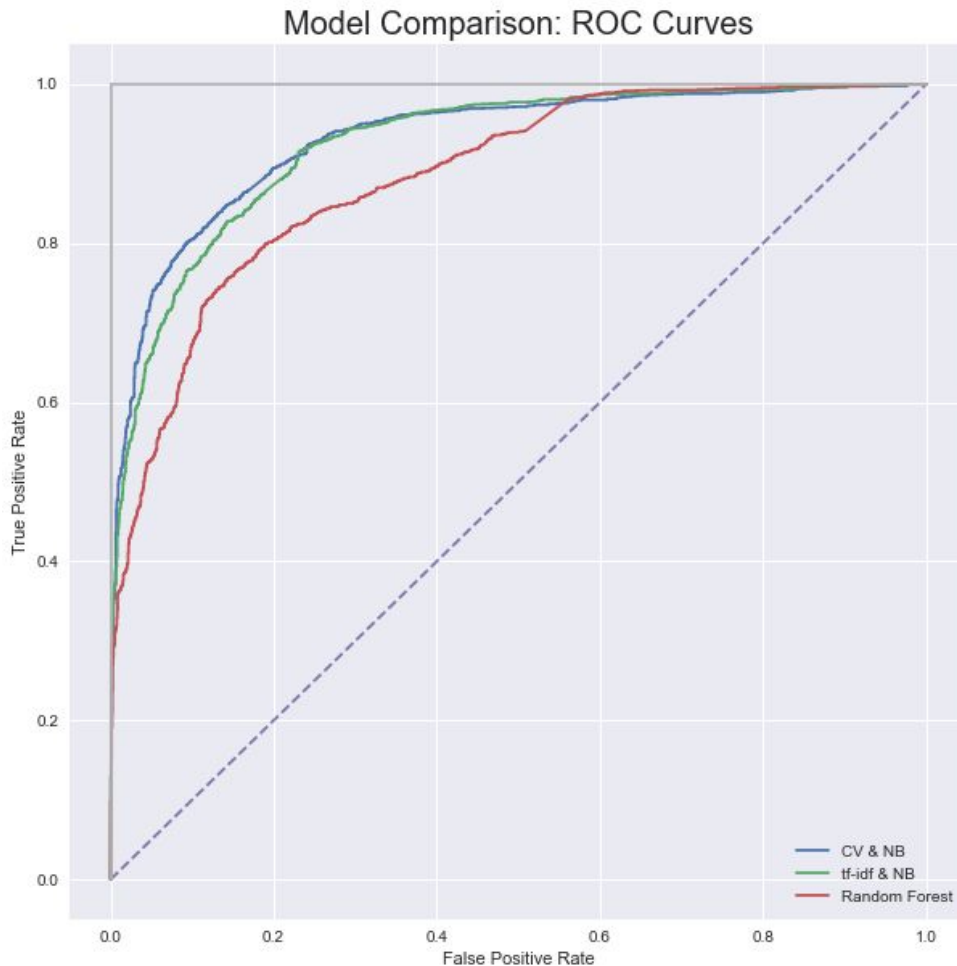
Model 1

92.5%

Model 2

88.2%

Model 3



Random Forest: Feature Importance



Incorrectly Categorized Posts

Kadudu888:

What should we tell new panicking newbie who's selling coz of the 25% dip?

Extremely-Bad-Idea:

Thodex and Vebitcoin have been shut down by Turkish authorities, arrests have been made, and the boss at Thodex is on the run with \$2 billion of bitcoins

DeadPrezFolder

Boomer Bullshit:

Someone missed out

\n\n<https://cryptoslate.com/bitcoin-is-an-open-p-onzi-scheme-says-black-swan-author/>

mambell

Privilege:

Does anyone else sometimes remember how lucky we are to own an asset like bitcoin? A community with some of the smartest and forward-thinking minds, defending and fighting for technology so destructive it can change the course of human civilization forever. Removing yourself from the USD price and recognizing the technology for what it is always reminds me to be grateful. I hope in this dip we can reflect on why we are here. Peace

Twelver313:

The first 10 comments get free sample packs of marijuana and THC gummies by Quad.WTF

Conclusions and Takeaways

- Random Forest had best precision score but worst AUC ROC score.
- CV & Naive Bayes less likely to deliver ads to incorrect users.
- EDA shows top phrases are users asking for advice or talking about specifics pertaining to the currency.
- Consider using sentiment analysis to pull out more differences between subreddits. These findings could be used as features to improve model performance.
- Model still guessed wrong when obvious words like Bitcoin were in the post.



