Smart Data Analytics: Prediction of Bitcoin and Ethereum Returns using Subreddits

Luca Riboni, Daniil Bulat

December 2020

Introduction

In our project we want to test the predictive power of Reddit submissions on the returns of cryptocurrencies. First, submission data from the Reddit websites Bitcoin and Ethereum Subreddits was collected from 2015 until December 2020. Then a sentiment analysis on the submission titles and the submission texts was run using the XXYY dictionary. As there are numerous submissions per day, the daily / weekly / monthly sentiment score is the mean sentiment of all submissions per day / week / month, weighted by the number of "upvotes" each submission received. In a final step OLS regression with three independent variables is used to test the predictive power of Reddit submissions on daily Bitcoin and Ethereum returns. The returns being the dependent variable and daily, weekly and monthly sentiment scores being the independent variables. Todays returns are regressed on yesterdays, previous weeks and previous months sentiment score.

Reddit

Reddit is a social news aggregation and discussion website. It is a major discussion forum for crypto related content. Users have only limited restrictions to contributing content on the website allowing a very free sharing of information. Due to the lack of cryptocurrencie coverage by traditional media, many investors seek advise and information regarding cryptos on Reddit and similar websites. Therefore, Reddit submissions were chosen as predictors for this project.

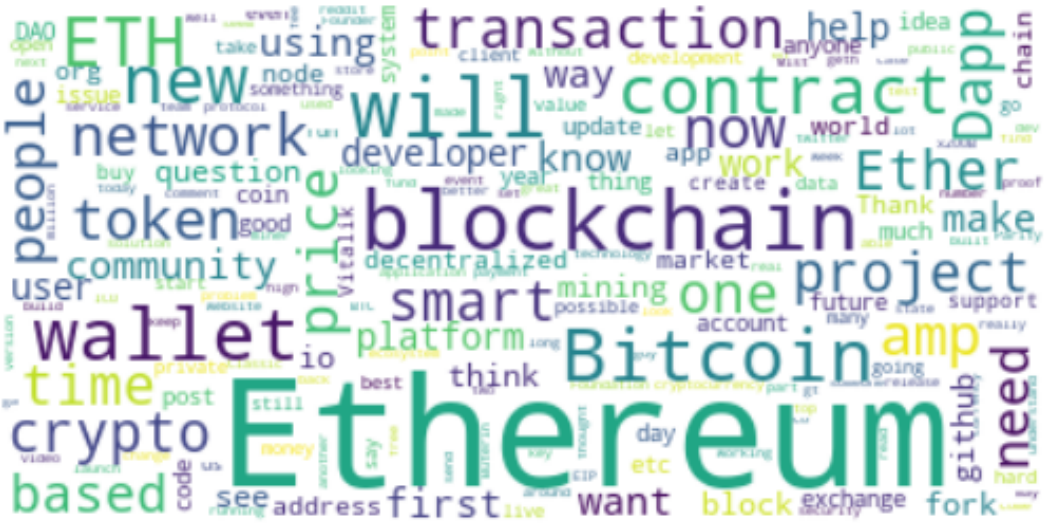
Sentiment Analysis

Word Clouds

The wordclouds on the following slides show accumilations of the most used words in the Bitcoin and Ethereum Subreddits from 2015 to 2020, with the exclusion of selected stopwords.

Bitcoin Wordcloud

Ethereum Wordcloud



OLS Regression

Bitcoin OLS Regression

OLS Regression Results

Dep. Variable:	Close	R-squared:	0.021
Model:	OLS	Adj. R-squared:	0.019
Method:	Least Squares	F-statistic:	15.08
Date:	Wed, 23 Dec 2020	Prob (F-statistic):	1.03e-09
Time:	22:12:49	Log-Likelihood:	3972.4
No. Observations:	2132	AIC:	-7937.
Df Residuals:	2128	BIC:	-7914.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0026	0.001	2.759	0.006	0.001	0.004
daily score	0.0721	0.011	6.552	0.000	0.051	0.094
weekly score	0.0044	0.026	0.173	0.863	-0.046	0.055
monthly score	-0.0674	0.037	-1.841	0.066	-0.139	0.004

Omnibus:	350.377	Durbin-Watson:	2.110
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6312.095
Skew:	-0.112	Prob(JB):	0.00
Kurtosis:	11.426	Cond. No.	50.3

Figure 1: OLS Regression Bitcoin

Ethereum OLS Regression

OLS Regression Results

Dep. Variable:	Close	R-squared:	0.008
Model:	OLS	Adj. R-squared:	0.006
Method:	Least Squares	F-statistic:	5.106
Date:	Wed, 23 Dec 2020	Prob (F-statistic):	0.00161
Time:	22:14:03	Log-Likelihood:	2558.8
No. Observations:	1943	AIC:	-5110.
Df Residuals:	1939	BIC:	-5087.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0002	0.003	-0.056	0.955	-0.006	0.006
daily score	0.0265	0.009	3.002	0.003	0.009	0.044
weekly score	-0.0513	0.025	-2.068	0.039	-0.100	-0.003
monthly score	0.0973	0.045	2.168	0.030	0.009	0.185

Omnibus:	446.344	Durbin-Watson:	1.865
Prob(Omnibus):	0.000	Jarque-Bera (JB):	19101.106
Skew:	0.179	Prob(JB):	0.00
Kurtosis:	18.356	Cond. No.	32.4

Figure 2: OLS Regression Ethereum

Results