



# 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. Today's returns are regressed on yesterday's, 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 cryptocurrency coverage by traditional media, many investors seek advice and information regarding cryptos on Reddit and similar websites. Therefore, Reddit submissions were chosen as predictors for this project.

# Sentiment Analysis

# OLS Regression

# Results