

# ETHalysis: Basic Analysis of the relation between Ethereum cryptocurrency value and Twitter tweets on Etheruem

Ghaith Arar

Stevens Insitue of Technology  
CPE-551 Spring 2021 - Final Project  
garar@stevens.edu

**Abstract**—We analyze the statistical relations between Etherurm cryptocurrency Ether value and the tweets posted on Twitter mentioning Ethereum. We use TWINT - Twitter Intelligence Tool to scrap tweets since a given date to current date at the time of execution. Then we use Natural Language Toolkit (NLTK) library to perform Sentiment Analysis on tweets. We perform some statistical analysis to study the relationship between the score of the tweets and the change of the cryptocurrency value. This project could be the core for larger and more advance application.

## I. INTRODUCTION

Ethereum is a decentralized, open-source blockchain with smart contract functionality. Ether (ETH) is the native cryptocurrency of the platform. It is the second-largest cryptocurrency by market capitalization, after Bitcoin. Ethereum is the most actively used blockchain. Studying the statistical relation between social media and Ether value could be useful given the high volatility of cryptocurrincies.1

## II. OUR PROJECT

This project aims to answer the following question: Is there an obvious relationship between the value change of Ethereum and the the interaction of Twitter users who tweet about it.

### A. Description of Dataset

We used TWINT - Twitter Intelligence Tool to scrap and collect all tweets that contained the terms 'Ethereum' and 'ETH' between a date the user provides and the current date at code execution. Twint is an advanced Twitter scraping tool written in Python that allows for scraping Tweets from Twitter profiles without using Twitter's API. It utilizes Twitter's search operators to allow users to scrape Tweets from specific users, scrape Tweets relating to certain topics, hashtags trends, or sort out sensitive information from Tweets like e-mail and phone numbers.

### B. Sentiment Analysis

We use Natural Language Toolkit (NLTK) which is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning,

wrappers for industrial-strength NLP libraries, and an active discussion forum. We used the pre-trained Sentiment Analyzer provided by the NLTK library. The analyzer evaluate each tweet and gives it three scores, negative, neutral, and positive. All scores are between 0 and 1.

### C. Statistical Analysis

We calculate the average score of tweets on each day. Then we download the Ethereum historical data using Historic Crypto library. This data provided the open, close, high, and low value of the cryptocurreny of each day of the specified period.

We used panda library to load the datasets and perform the statistical analysis. The data set can be analyzed in multiple ways. One approach is to handle the data set as a time series and perform time series analysis. Another approach is treat the data set as a table and find the correlation between the score and the value. It's also possible to feed the data to a machine learning algorithm to extract non-obvious relations between the features. However, this approach is beyond the scope of this project.

## III. CONCLUSION

In our basic approach, we could find obvious correlation between the features. This result is predictable since the value of cryptocurrcies are very complex and difficult to predict. However, this project could be seen as a first step towards a more sophisticated approach e.g. deep learning model to predict the value based on features scrapped from the internet in real time.