Project Proposal   
DS 5220

**Analyzing Impact of social media on Cryptocurrency**   
**Members**: Parth Bhatt, Poojan Thakkar

**Summary:**

The rise of Cryptocurrency is a highly polarizing topic. This highly volatile digital asset is designed to work as a medium of exchange. It utilizes strong cryptography primarily in the form of block-chain or public ledger to secure financial transactions, control the creation of additional units, and verify the transfer of assets. Due to this fact, the fluctuations in cryptocurrencies in a matter of months, and in some cases even days, would be enough to entail the fluctuations in stock over years.

Twitter is an online news and social networking platform. It evidently serves as one of the primary channels of communication regarding any changes in cryptocurrency prices. At times it has been observed that a thread of tweets alone was sufficient to bring major changes in the market.

Through this project we would be trying to assert the influence of twitter on the prices of cryptocurrencies like Bitcoin and Doge.

**Dataset:**

We have decided to utilize the [Twitter API](https://developer.twitter.com/en/docs/twitter-api) to collect the most recent tweets and then retain only the ones that are related to cryptocurrencies. In order to do the same, we would be tokenizing the data and based on the keyword inclusion, extract only the relevant tweets. (It is limited to 450 requests of maximum 100 tweets per 15 minutes with the App login. It can only retrieve tweets 7 days old at most). While most of the publicly available datasets are from 2009-2019, using the same would not be much fruitful, however in case if any data vendor publishes the dataset for tweets based on cryptocurrency online, we would be inclined on using the same. For collecting the historical data for cryptocurrencies, we would be using [Bitfinex exchange’s API](https://docs.bitfinex.com/docs), this gives us the detailed data in windows of 60 secs starting from 2009 with at max a day’s lag.

**Algorithms:**

Since our data is going to be extracted through scraping APIs, we would be preprocessing the data and remove unwanted outliers (for example missing data between the opening and closing timeframes) as well as irrelevant tweets (images, irrelevant tags, hyperlinks, etc.). Moreover, for visualization purposes we would be summarizing the data to showcase three different versions: daily trends for over a month, weekly trends over a course of 3 months, and monthly trend. We then would be devising a metric based on the likes of the tweets and number of followers of the users, to judge the impact of the tweet. Applying correlation on the series is not enough, we would be needing cross-correlation technique to determine the influence. For that we have planned to use the best from following approaches: 1) Polarity classification, 2) Lexicon based approach, 3) Vader Algorithm for the analysis of sentiment from the tweets.

**Libraries and Tools:**

1. Python - Building blocks for the source code
2. Twython - A Python library used for ease of access to twitter data
3. Plotly - Python library used to create interactive visualizations
4. Vader Sentiment - Tool used for lexicon-based sentiment analysis
5. statsmodel - Python tool to be utilized for statistical data exploration
6. Pandas - To manipulate data in our Python code
7. SciPy - A Python library facilitating usage of scientific methods.
8. Cuda (usage likely) - Application used for parallel computing elements
9. Rapids (usage likely) - Used on top of Cuda to make operations faster
10. Amazon SageMaker - Incase additional computation power is required

**Results:**

Through this undertaking, we wish to quantify if a correlation exists between public activity on twitter regarding cryptocurrency, and the fluctuations observed in cryptocurrency prices. In doing so, we may be able to understand better the volatile nature of cryptocurrencies. These results may also be helpful for people who invest in these technologies, by serving as an additional tool to devise better investing strategies.

**References:**

* <https://developer.twitter.com/en/docs/twitter-api>
* <https://docs.bitfinex.com/docs>
* <https://github.com/cjhutto/vaderSentiment>