**BIA-660D Final Project Proposal - Team 5**

**Members:** Akshay Gupta, Chengzhi Ni, Abhay Sharma, Kratika Yadav, Li Zhang

**Project initiative**

Since 2017, Cryptocurrencies like Bitcoin, Ethereum have again become the vibe and draw enormous attention from up to Financial industry down to the general public. Thus, a wide range of discussion about this new phenomenon has raised from all perspective. Using the knowledge and tools we are going to learn from this course, we wish to gain a deeper understanding of this phenomenon.

**Goal**

The goal of our project is to understand the correlation between social media activity around cryptocurrencies and their price trend. Specifically, the number of posts, follow up comments, as well as their corresponding sentiments underlying. On top of that, we are interested in building a machine learning, or time series model, that incorporating such knowledge, to analyze the behavior of crypto markets according to the twitter activities.

**Data source of cryptocurrency currency**

Coindesk<https://www.coindesk.com/>(Through crawling)

Coinmarketcap<https://coinmarketcap.com/currencies/streamr-datacoin/>(Through crawling)

Possibly Bloomberg terminal as cryptocurrencies, as a financial asset, is highly likely to correlate with other financial instruments and overall market performance.

**Data source of social media**

Twitter<https://twitter.com/?lang=en> (Through crawling and API)

Bitcoin Forum <https://forum.bitcoin.com/> (Through crawling)

Market Watch <https://www.marketwatch.com/investing/cryptocurrency/btcusd> (Through crawling)

**Analytics to be involved**

The analytic process can be decomposed into three parts. In order to solve distinct problems that rise from the project initiative.

The first being to conduct exploratory data analysis on coin price data and social media data, to gain summary statistics such as volume, mean, median, volatility, correlation from a various perspective and even dive into weekly trend, monthly trend, etc.

The second step is to perform sentiment analysis on text data we collected from social media website, to use topic modeling tools such as Latent Dirichlet Allocation to summarize the sentiment of text data.

The last step is to combine the two data sources together, then to use machine learning or time series algorithms to produce a predictive model of the future price of cryptocurrency.

The last step is to combine the two data sources together, then to use machine learning or time series algorithms to produce a predictive model of the future price of cryptocurrency.