**Capstone Project 1**

What is the problem you want to solve?

As of October in 2017 alone, more than $3.25 billion of funding has entered into blockchain companies via an alternative funding method called the Initial Coin Offering, a.k.a. ICO. Majority of these ICO funding take place with Ether as the native currency as a source of funding in exchange for tokens based in the Ethereum blockchain generated by the respective fund-receiving companies. Similar to the traditional stocks and shares, these tokens become tradable as they get listed on exchanged. However, contrary to the traditional stocks and shares, these tokens do not fundamentally come with entitlements to the respective companies’ assets and therefore the value of these tokens are much more difficult to determine and predict. Due to the volumes and volatility involved, there is an ever-growing opportunity in the cryptocurrency market for profits from trading activities.

Ether prices, just like prices of most other cryptocurrencies, are driven by market hype. While market hype could increase awareness and excitement of Ether and other cryptocurrencies and hence increasing their prices, negative rumours are likely to drive panic sell or general selling actions from existing coin holders. On the other hand, it is also in my hypothesis that there is a fundamental relationship between the price of Ether and the prices of the tokens based in the Ethereum blockchain. The purpose of this project is to create a model that allows the user to use market hype and Ether prices to predict the likely price movements of Ethereum-based tokens. The creation of the model involves analysing Ether prices and the prices of three of the largest Ethereum-based tokens, OmiseGO, Qtum, and Golem, and the extent to which they are impacted and predicted by market hype using Twitter posts as proxy. The cross-correlations between prices of these alt-coins and Ethereum prices are also explored. These relationships are also to be explored across different time-lags; initial time-lags explored are to be 1 day, 2 days, 5 days, and 7 days but are not limited to lower time scales where opportunities arise.

It must be noted that this model is for educational purposes only and should not be used as investment or trading advice. It must also be noted that ‘price’ in this project brief refers to price in USD.

Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn’t have otherwise?

The target client for this model is to be trading portfolio managers who are looking to explore trading opportunities in the cryptocurrency space. They will then be able to use market hype and Ether prices to predict likely price movements for the Ethereum-based tokens. As such, they will be able to make long/short decisions. Further to portfolio managers, as cryptocurrencies trading is also largely open to everyday retail investors, this model can also be used to provide insights to everyday investors on the likely relationships between market hype and Ether prices and also prices of the Ethereum-based tokens.

On the other hand, the target client could also be myself. With the insights generated from this investigation, I could use it as part of a core strategy in trading Ethereum-based tokens.

What data are you going to use for this? How will you acquire this data?

Price Data (Ethereum, OmiseGo, Qtum, Golem): <https://coinmarketcap.com/>, <https://www.coinigy.com/>

Data Acquisition Method: Web scrapping or through API

Twitter Data: <https://twitter.com/>

Data Acquisition Method: API

In brief, outline your approach to solving this problem (knowing this might change later).

1. Acquire the prices and Twitter data from the sources.
2. Perform exploratory data analysis and descriptive statistics on data sets.
3. Merge datasets using Numpy or Pandas packages.
4. Determine a set of KPIs, metrics, calculations and formulate strategies in obtaining these using datasets.
5. Create exploratory visualizations to confirm initial validity of anticipated trends from my hypothesis.
6. Consider a potential machine learning or linear model system to perform analysis and prediction, including NLP and other methods or classifiers.
7. Split data into training data and test data in proportions of 70/30.
8. Perform analysis on the Twitter data and explore its relationship with prices of Ethereum, OmiseGo, Qtum, and Golem.
9. Perform cross-correlation analyses between the pricing pairs: OmiseGo/Ethereum, Qtum/Ethereum, and Golem/Ethereum.
10. Compile pricing model and test model against test data.
11. Optimise parameters and re-test.

What are your deliverables?

The deliverables of this project will include code, slide deck, and a paper stored in a github repository.