Final Project: Proposal

#### Tushar Chand Kapoor, Mehak Parashar, Syed Ikram

### **1. List 3 questions that you intend to answer (1 point)**

1. How does cryptocurrency market behave?
2. Can we predict the future price of different crypto currencies using historical data and sentiment analysis?
3. How to visualize live data, visualize effects of different parameters that affect cryptocurrency prices, and also to visualize cryptocurrency exchanges worldwide by volume.

### **2. List all the datasets you intend to use (1 point)**

We intend to use data from various sources and combine in a single database such as Cassandra. The following are various data sources we are going to use:

* CryptoCompare.com
* Google Trends
* Google Geocoding API
* CoinApi
* News API
* Twiter API

Glimpse about how the above data sources will be used:

* We will use CryptoCompare.com and CoinApi.com APIs to collect data about the prices, exchanges, trades, market cap etc., to make predictions and form visualizations.
* Google Trends API is a good source to check the popularity of a subject we will use the Google Search data to make predictions and perform analysis.
* Google Geocoding API will be used to fetch the Geo coordinates of various exchanges across the globe.
* We will live stream current prices, volumes etc. from CryptoCompare.com streaming API for analysis and live visualizations.
* We will use News API data to do sentiment analysis on news articles related to the crypto currency.

### **3. Give us a rough idea on how you plan to use the datasets to answer these questions. (2 points)**

* Data Collection: Where/how to get data
  + Data extraction using different web API and combing them in a single database
  + Live Streaming Twitter Data
  + Live Streaming CryptoCompare API data
* Data Exploration: Do you need to conduct EDA in order to understand the data?
  + In order to analyze and do visualizations using D3 and other JavaScript libraries, we need to gain a deep understanding of data. Therefore, we plan to make it more interactive by doing it with live data by doing multivariate EDA.
* Data Cleaning: Do you need to clean data? How to clean them?
  + We plan to clean the data using data cleaning techniques to deal with missing or the duplicate data and also checking the data for the consistency.
  + Removing the outliers to get the more well-defined data set
* Data Integration: Do you need to integrate data from multiple sources?
  + Since we plan to collect the data from various sources, we will have a lot of data which needs to be integrated together to form relations between various sources which we plan to integrate into a single dataset, also integrating into a single dataset leads to redundant data which will be cleaned.
* Data Analysis: What analysis do you intend to do? (e.g., SQL, Statistics, Deep Learning) How to evaluate your analysis results? (e.g., evaluation metrics, confidence intervals, benchmark)
  + Sentiment analysis using state-of-the-art Natural Language Processing techniques for textual data of news articles etc.
  + Deep Learning models such as LSTM, RNN etc.
  + Machine Learning models such Random Forest, GBT etc.
  + Visualizations using D3 and JavaScript
  + Evaluation Metrics – Precision, Recall, F1 score and MSE
* Data Product:
  + Interactive Web app

### **4. Think about that once your project is complete, what impacts it can make. Pick up the greatest one and write it down. (1 point)**

Once the project is complete it will be able to predict the future prices of several crypto currencies in the market and show the trends of cryptocurrency with the help of easy to understand visualization. This will help the buyers make sound decision if they want to buy any cryptocurrency because it will not only forecast the future price but also help them understand the trends in the market. This platform will be designed in such a manner that it will help the curious people looking out to gain knowledge about the crypto currency market. Thus, it will have an impact on various strata of society.