a. High level description of what you plan to do, and why you think it is interesting.

Our topic is cryptocurrency. Cryptocurrency is a form of payment that can be exchanged online for goods and services. It exists in the network but not in the real world. Cryptocurrencies are becoming more and more popular nowadays. As time goes by, cryptocurrency will become more important in our life because of technique development. We will get more exploration in virtual reality and may need a currency to exchange. Cryptocurrencies like bitcoin are already very famous today, and may grow more in the future.

For our topic, first we will connect some tweets through twitter. We will collect their comments and analyze what are the top 10 cryptocurrencies they have discussed. From these 10 types of cryptocurrencies, we will try to find the topics that most connect to them (like NFT may connect to metaverse). After that, we will predict on classifying new tweets or news which are connected to cryptocurrencies.

b. A detailed description of the methods you will use to achieve this:

What modeling approach do you intend to use?

For find the top 10 cryptocurrencies, we may use Naive Bayes. This model can find the most frequent word from tweets.

For define the cryptocurrencies' topics, we will use BOW features and LSA model to find the top 10 words connected to these cryptocurrencies'.

Finally, we will use Ngram to predict the new tweets that belong to what kind of topics.

What data do you intend to use?

We will use Twitter API and Kaggle to gather data. We want to get the data that shows how people think with cryptocurrency. We will gather some new tweets to predict what topic they are discussing.

How will your system be evaluated and what are the evaluation criteria?

We will evaluate the top 10 cryptocurrencies based on the word frequencies from tweets. After we use LSA models training our data, we will visualize the top n words that connect to the cryptocurrencies we have chosen. We need to choose the words that are important with these cryptocurrencies by ourselves.

Are there any special computational/hardware considerations?

There are no special hardware considerations by using the model we have discussed before. It may take some time to get the results if we have thousands of tweets through.

What are the biggest unknowns that might dictate the success or failure of this project?

When we try to find top 10 popular cryptocurrencies, we need to know the term is a kind of cryptocurrency. After we get many word and word frequencies, we need to distinguish them with cryptocurrency's name and other information we do not need right now.

c. How will the results of your work be presented? Will this be a live demo, a written report, a slide deck + oral presentation? Any of these are acceptable! Demos can be given along with reports/presentations

We will use a slide deck and oral presentation to show our work.