Project Proposal

Mrunalini Devineni and Mohammad Maaz

What problem did you select and why did you select it?
We will be investigating how emotion plays a role in the historical price of cryptocurrency by analyzing the emotion of comments on the popular subreddit r/cryptocurrency.

What database/dataset will you use?
We will scrape 12 million comments over the past year from r/cryptocurrency. In order to assign emotion, we will use a pre-trained transformer, fine tune it on the emotion dataset on huggingface, and repurpose it to predict emotion on the reddit comments.

What NLP methods will you pick from the concept list? Will it be a classical model or will you have to customize it?
We will use transformers, which we will fine tune.

What packages are you planning to use? Why?
- Composing data | pmaw, PushShift
- Cleaning and analyzing | pandas, matplotlib
- Fine tuning classifier and running inferences | pytorch, transformers

What NLP tasks will you work on?
The NLP tasks involved in this project are text classification and zero shot classification.

How will you judge the performance of the model? What metrics will you use?
We will use accuracy and f1 score on the test set of the emotion dataset to gauge the performance of the model before running inferences.

Provide a rough schedule for completing the project.
The project should be completed by 6/2/2022