Project Proposal

Team members:

- Heta Desai -> Captain
  - NetId : hetahd2
- Harpreet Siddhu
  - NetId: hsiddhu2

Topic:

Option 1 : Reproducing a paper - Casual Topic Mining
Hyun Duk Kim, Malu Castellanos, Meichun Hsu, ChengXiang Zhai, Thomas Rietz, and Daniel Diermeier. 2013. Mining causal topics in text data: Iterative topic modeling with time series feedback. In Proceedings of the 22nd ACM international conference on information & knowledge management (CIKM 2013). ACM, New York, NY, USA, 885-890. DOI=10.1145/2505515.2505612

Language :

We are planning to use python for the implementation of this paper.

Dataset:
We will need the following dataset for the implementation that is also used in paper
- New York Times articles from May through October of 2000 : For this dataset we registered at https://catalog.ldc.upenn.edu/account?show_gdpr=true with organization as UIUC. But we are waiting for the admin permission to be able to access the dataset.
  - But we can access the dataset for the news article from the following site:
    - https://spiderbites.nytimes.com/2000/
We can scrape the articles from this site and work on it.
- Iowa Electronic Markets (IEM)3 2000 Presidential Winner-Takes- All Market : We are able to access this website and see the data. We will either scrap this website to obtain data or simply copy the prices and paste in a file. Following is the site for the data :
  - https://iemweb.biz.uiowa.edu/pricehistory/pricehistory_SelectContract.cfm?market_ID=29