Feb 25, 2021

1. Intro about the project
   1. News articles from 10 mainstream media
   2. Covering how the White House delivers messages about COVID-19.
   3. Key words search (“white house/trump/pence” & “covid/coranavirus”)
   4. Drawback of the key words search method: we cannot get rid of some articles that are not about White House briefings.
      1. E.g. “WHO held a conference criticizing White House’s responses to COVID”
   5. “related”
      1. 0-- the news articles are NOT related to White House briefings about COVID-19. For example, WHO held conferences in which they mentioned the White House and COVID. US governors held conferences in which they mentioned Trump.
      2. 1-- the news articles are related to White House briefings about COVID-19. As long as the news articles mentioned White House briefings/conferences about COVID-19, they should be counted as 1. For example, Dr. Fauci will attend the White House briefing.
2. Dataset
   1. Training/test dataset (n=1345)
   2. Raw dataset (N=8045)
3. For the Exploratory Data Analysis assignment
   1. text-preprocessing. That is, tokenize the titles into words/features so that machine learning models could use for further analysis. (Yue: by Thursday/Friday early morning)
   2. Everything before plots (Bobby: by Friday night/Saturday early morning)
   3. Plots (Haihang: by Saturday night/Sunday morning)
   4. PCA (Archit: by Sunday morning/Sunday afternoon)
   5. Report (Lauren: by Sunday night)