ADS-509 Text Mining

**Team 9**

Problem Statements

Assumptions:

1. All PSs involve one or more news sites, accessible via API or web scraping. The machine learning methods must include classification per the Project Assignment instructions.
2. We want to limit data based on keywords associated with:
   1. Politics and/or opinion in non-editorial journalism.

Things to consider:

1. ML type (i.e., un-, semi-, or supervised).
2. Using meta data as part of analysis, including date/year.
3. Sources:
   1. NewsAPI.org (<https://newsapi.org/>)
   2. TBD x1 or 2

**Problem Statement #1 [Dave to research academic sources]**

Find article political leaning based on reference of keywords from academic source(s). Methods could include bag-of-words, sentiment analysis, or both.

**Problem Statement #2 [Aaron to research]**

Identify specific news sources that have political leaning indices based on external rating, e.g., academic paper or bipartisan organization that scores news sources based on the political bias of articles. Use their data to determine bias [e.g., in general or on particular issue(s)].

**Problem Statement #3 [Susy to research]**

Opinion journalism vs. bias free reporting:

Looking at non-editorial news, pulling/finding obvious sentiment or opinion-related words. What constitutes opinion words or sentiment?