University of San Diego®
SHILEY-MARCOS SCHOOL OF ENGINEERING

ADS-509 – Team Project Status Update Form

Fill out this form and submit it by the end of Module 4 in Blackboard.

Team Number: **Team 9**

Team Leader/Representative: **Aaron**

Full names of team members:

1. **Aaron Carr**
2. **Azucena Faus**
3. **Dave Friesen**

Title of your Applied Text Mining project: **Illuminating Mainstream Media Political Bias through Text Mining**

Short description of your project and objectives:

The question of mainstream media bias is a significant area of contention in U.S. politics. At the same time, the political 'inclination' of a news outlet often aligns with the personal perspectives of its audience. This overlap provides a unique opportunity and our project's objective: To leverage sophisticated text mining methods and supporting data science techniques to classify the political bias of leading online news sources. This classification will be based on a combination of "politically polarizing" terms as identified through an impartial (academic) source, as well as sentiment analysis context around the use of these terms (Liu et al., 2022).

The objective of developing this classifier will be to advise our clients who are the executives at a news source that prides itself in being considered politically "centered." We would use this classifier to analyze their content on a continuous basis and report back the overall/average political "lean" of their articles. The details regarding the metric for overall publication lean will be based on an average of left/right leaning probabilities per article (this will be further tuned as the project moves forward). This feedback will then provide our client with actionable insights so they ensure their overall political lean metric remains centered.

University of San Diego®
SHILEY-MARCOS SCHOOL OF ENGINEERING

Name of your selected dataset: **Queried news articles from CNN, Fox News, Breitbert, and Washington Post for the month of May 2023.**

Description of your selected dataset (data source, number of variables, size of dataset, etc.):

News articles will be sourced via a REST API called NewsAPI, which logs information on "current and historic news articles published by over 80,000 worldwide sources" (NewsAPI, n.d.). Out of the many possible attributes returned for each API query, this project will use six: source name, author, title, url, publishedAt, and content. Prior to data preprocessing, an additional feature (article_text) will be used to store the scraped data from each specific URL. The current target *N* is ~5,000 articles, however the size of the final dataset will be limited by both time restrictions, as well as web scraping access to specific sites or articles. Queries for topics of political interest are used to gather articles from explicitly chosen sources. Independent studies show the political lean for each of these sources (CNN, "left"; Fox News, "right"; The Washington Post, "left"; Breitbert News, "right") and this will help with training and validation of our classifier (AllSides, 2022; Ralph & Relman, 2018).

For this project, use GitHub as a code hosting platform for version control and collaboration. Provide the link here: **https://github.com/fausa/ADS509_Final_project/tree/main**

How many times have your members met in the last two weeks? **We have formally met twice via Zoom. We are also working actively asynchronously over Slack.**

What was the agreed upon method of communication? Are you using any teamwork project management software, such as Trello or Asana? **Slack, GitHub, and Google Drive are currently our primary communication and collaboration tools (with Zoom for meetings).**

Comments/Roadblocks: **We are off to a good start on our project and, given that we're almost to Week 5, expect to accelerate progress over the next week.**

References

AllSides. (2022, March 15). *AllSides Media Bias Chart version 6: Updated ratings for NPR, Newsmax, and more*. https://www.allsides.com/blog/new-allsides-media-bias-chart-version-6-updated-ratings-npr-newsmax-and-more

Liu, R., Jia, C., Wei, J., Xu, G., & Vosoughi, S. (2022). Quantifying and alleviating political bias in language models. *Artificial Intelligence*, *304*. https://doi.org/10.1016/j.artint.2021.103654

NewsAPI. (n.d.). *Search worldwide news with code*. Retrieved June 5, 2023 from https://newsapi.org/

Ralph, P., & Relman, E. (2018, September 2). These are the most and least biased news outlets in the US, according to Americans. *Business Insider*. https://www.businessinsider.com/most-biased-news-outlets-in-america-cnn-fox-nytimes-2018-8?op=1#and-heres-how-republicans-ranked-them-from-fox-news-to-cnn-20