Group 42 Progress Report:
Left and Right Leaning Bias Detection in Media Headlines

Ahren Chen, Fei Xie, Grace Xiao
{chena125, xief17, xiaot13}@mcmaster.ca

1 Task Introduction

Our project builds a Natural Language Processing model that classifies news/media headlines by political leaning (left vs. right). This task is significant because headlines heavily shape first impressions of an event, and an automatic "leaning" label can support media literacy tools, dataset exploration for researchers, and other media systems that want balanced coverage. It's challenging because political framing in headlines is often subtle and context-dependent: choices like which facts are revealed, what groups are named or hidden, and the use of terms can signal ideology without using obviously emotional language. Unlike sentiment analysis, our focus is ideological framing and viewpoint, not whether the headline sounds positive or negative. To keep our labels consistent, we'll write a short annotation guide with specific cues to look for, then have everyone in the team label the same starter set of headlines to line up our interpretations. Our goal is to reach agreement that's clearly better than chance, while recognizing that some headlines will still be genuinely hard to classify.

2 Task Definition

Task definition. Our dataset consists of media headlines, so each input example is a short piece of text. The task is a supervised classification problem: given a headline, the model predicts the headline's political leaning based on its framing. We use two classes (left leaning and right leaning) making this a binary classification setup. Each headline receives one label only (single-label), meaning it is annotated as either left or right (not both).

3 Dataset

3.1 Plans for Collection

We will use an existing annotated dataset *allsides_balanced_news_headlines-texts.txt* from https://github.com/irgroup/Qbias/tree/main . Each instance contains at minimum an English news **headline** and a political-leaning label (left/right), and may also include metadata such as the news outlet/source, event/topic title, topic tags, publication date, and a short article text/snippet (depending on the fields available in the release).

We will use only publicly available data for this course project and follow the dataset's license/usage terms. We do not plan to perform additional web scraping. Prior to modeling, we will apply basic cleaning (removing empty entries, de-duplication, and text normalization). We will also remove the center class and formulate the task as a binary classification problem (left vs. right).

3.2 Expected Dataset Size

After removing the `center` class, we expect to use $N = 17{,}534$ labeled examples in total, with $10{,}296$ left-leaning and $7{,}238$ right-leaning headlines. We will create train/dev/test splits with an approximate 80/10/10 ratio (with the exact proportions determined by the outlet-based grouping constraint).

As required, we provide three example data points with labels assigned by our group:

1. **Event/Topic:** Gun Violence Over Fourth of July Weekend
   Headline: "Chicago Gun Violence Spikes and Increasingly Finds the Youngest Victims"
   Outlet: New York Times (News)
   Group label: `left`

2. **Event/Topic:** Gun Violence Over Fourth of July Weekend
   Headline: "Dozens of shootings across US mark bloody July 4th weekend"
   Outlet: New York Post (News)
   Group label: `right`

3. **Event/Topic:** Yellen Warns Congress of "Eco-

nomic Recession" if Debt Ceiling Isn't Raised
Headline: "Federal Government Will Run Out of Cash on Oct. 18 If Debt Ceiling Isn't Raised: Treasury Secretary"
Outlet: The Epoch Times
Group label: `right`

4 Team Contract

The team contract outlines the expectations of the team for the duration of the project. Submission of this project proposal implies that all members listed in the author section has read and agreed to the following contract.

4.1 Team Purpose

Our purpose is to apply the content and knowledge we learned during the SFWRENG 4NL3: Natural Language Processing class towards our project, classifying if a given news headline can indicate if the article is more right or left leaning.

4.2 Team Goals

- Complete all deliverables on time and to a high degree of quality.
- Maintain equal participation and contribution from all members
- Develop skills in collaboration, problem-solving, and accountability.
- Create an interesting project based on the course content we learn in class.

4.3 Communication Norms

Primary platform: Discord
Response expectation: Reply to discord messages within 24 hours.
Decision-Making: Strive for consensus when possible; else, majority vote will be used.

4.4 Ground Rules and Expectations

- Be respectful of all ideas and opinions.
- Come prepared for meetings.
- Complete/submit assigned work on time.
- It is up to the individual assigned to the task to inform the rest of the team if the task cannot be completed on time.
- All team members are expected to attend tutorial time [Wednesday, 4:30-5:20 PM], and to inform the team beforehand if they cannot make it.

4.5 Conflict Resolution

We recognize that conflicts may arise, particularly around workload distribution. To handle these situations:

1. Direct Conversation: The concerned member(s) will address the issue privately and respectfully with the person involved.
2. Group Discussion: If the problem continues, the whole team will discuss it openly and constructively, focusing on solutions (e.g., redistributing tasks, adjusting timelines).
3. Clear Expectations: If a team member repeatedly does not contribute, responsibilities will be formally reassigned, and the lack of contribution will be documented.
4. Escalation: If the issue remains unresolved, the team will notify the course instructor/TA with a written summary of the situation.

Signed by: Ahren Chen, Fei Xie, Grace Xiao