

Annotator Guidelines for Labeling Political Tweets and Articles

Objective

This annotation task aims to label political content from various media sources to analyze sentiment disparities between Democrats and Republicans. Some tweets and articles may overlap across annotators to assess inter-annotator agreement (IAA).

Dataset Information

You will work with a dataset of tweets and their linked article content. Each entry contains information such as PostId, PostUrl, ChannelName, and extracted_article_content. The `extracted_article_content` field is where you'll find the primary text to annotate.

Preprocessing Steps (Already Completed)

1. Filtering for Valid Content:

- a. Only records with non-empty article content have been selected for annotation. Rows with placeholder text or non-informative content, like "Heading:\n\nArticle content:\n" have been removed.
- b. Posts that contain links to videos (like YouTube or Vimeo) without relevant textual content have also been excluded.

2. Random Selection and Distribution:

- a. The dataset was shuffled to randomize the entries and ensure no bias in the distribution.
- b. Each annotator has received 200 posts, with two sets:
 - i. Unique Posts (150): Unique entries that only you will label.
 - ii. Overlapping Posts (50): These are shared with one other annotator for IAA calculation. They allow us to measure consistency in labeling across annotators.

Annotation Task

1. **Annotation Sections:** You will see two sections in each row of your Google sheet:
 - **For Tweet Text Only:** `post_body_text` (green section).
 - **For Extracted Articles Only:** `extracted_article_content` (yellow section).
2. **Annotate Only the Allowed Cells:** Only annotate in the specified columns for each section. Do not edit other columns or cells to avoid accidentally modifying the original content.

3. **Annotation Labels and Definitions:** Use the following definitions to assign labels (only 'yes' or 'no') based on the overall sentiment and political focus:
- **pro-republican:** If the overall conclusion of the content is positively inclined towards Republicans.
 - **pro-democrats:** If the overall conclusion of the content is positively inclined towards Democrats.
 - **con-republican:** If the overall conclusion of the content is negatively inclined towards Republicans (clearly critical or against Republicans).
 - **con-democrats:** If the overall conclusion of the content is negatively inclined towards Democrats (clearly critical or against Democrats).
 - **neither republican nor democrats:** If the content is political but not specifically related to the Republican or Democratic parties.
 - **not-relevant:** If, after all filtering, the post or article content is not related to politics.

Important Notes

1. **Read Context Carefully:** For each entry, read the entire content of `post_body_text` and `extracted_article_content` to assess the overall tone and conclusion.
2. **Do Not Assume:** Avoid assumptions or inferences beyond the provided content. Focus solely on the context provided.
3. **Save Your Work:** Save the file periodically and ensure the annotations are complete before submission. Although Google Sheets automatically saves, just ensure it is indeed saving your annotations.
4. Emory University provides free access to the following news outlets. In case you find that the `extracted_article_content` column does not contain article text, please do the following:
 - a. Annotators have no access to edit columns other than the ones they are annotating. So, please raise this issue to the sheet owner and they can manually add the article content.
 - b. Emory University ID provides us access to various news outlets. So if some content is not available, try to find it via the following Emory URL:
 - i. [New York Times](#)
 - ii. [Wall Street Journal](#)