**TOPIC MODELING (MEHTAB)**

The goal of this part of the project was to identify and illustrate the recurring themes in a sizable collection of articles, with an emphasis on how the media shapes and ranks subjects over time. I aimed to illustrate patterns of media attention and thematic saturation, that is, what subjects are regularly emphasized and how that focus changes in response to events, by examining topic modeling results and creating visual representations of topic distribution and temporal trends.

A topic-labeled CSV file from a corpus of news articles served as the dataset. To ensure data clarity, I filtered out all articles labeled with topic -1, which usually marks irrelevant or uncategorized documents. I also removed topics that were strongly skewed by pronouns.

The cleaned data was utilized to determine the top 10 most common topics, ordered by article frequency. These were represented in a horizontal bar chart to display which themes prevailed over the media corpus. To study narrative development, I also concentrated on the top 5 topics and graphed their article frequency by time via a 3-month rolling average. It smoothed out, enabling me to highlight long-term tendencies above short-term fluctuations and making changes in media attention easier to depict.

The central argument my findings make is that media coverage is not thematically neutral; it has recurring patterns of attention, saturation, and neglect. Some themes keep on occupying attention, framing public debate through iteration and prioritization. By monitoring these trends over time, my work proposes that media narratives are patterned and structured. This visualization-based inquiry facilitates critical consideration of how subjects emerge, settle, or dissolve in the public eye.