Team Meow

Exploring Political Bias in Online Communities: Sentiment Analysis on Controversial Political Topics

Evan Carlson, Jonathan Hsia, Jared Johnson, Aaron McAughan, Tommy Nguyen, Lance Wilhelm

Summary

The US political climate has become **increasingly polarized** and often precludes broad engagement on divisive topics. Citizens usually get their information from biased sources, **infrequently expending effort** to seek contrary information. We aim to make it easier for users to **juxtapose info from a broad spectrum of political beliefs** on topics of their choice, exposing various biases and slants, and ultimately, reducing this polarization.

How?

Our data pipeline extracts keywords from comments using a name-entity recognition algorithm, before analyzing the result with our trained PyTorch BERT model to execute natural language processing and sentiment analysis on each keyword.

Why?

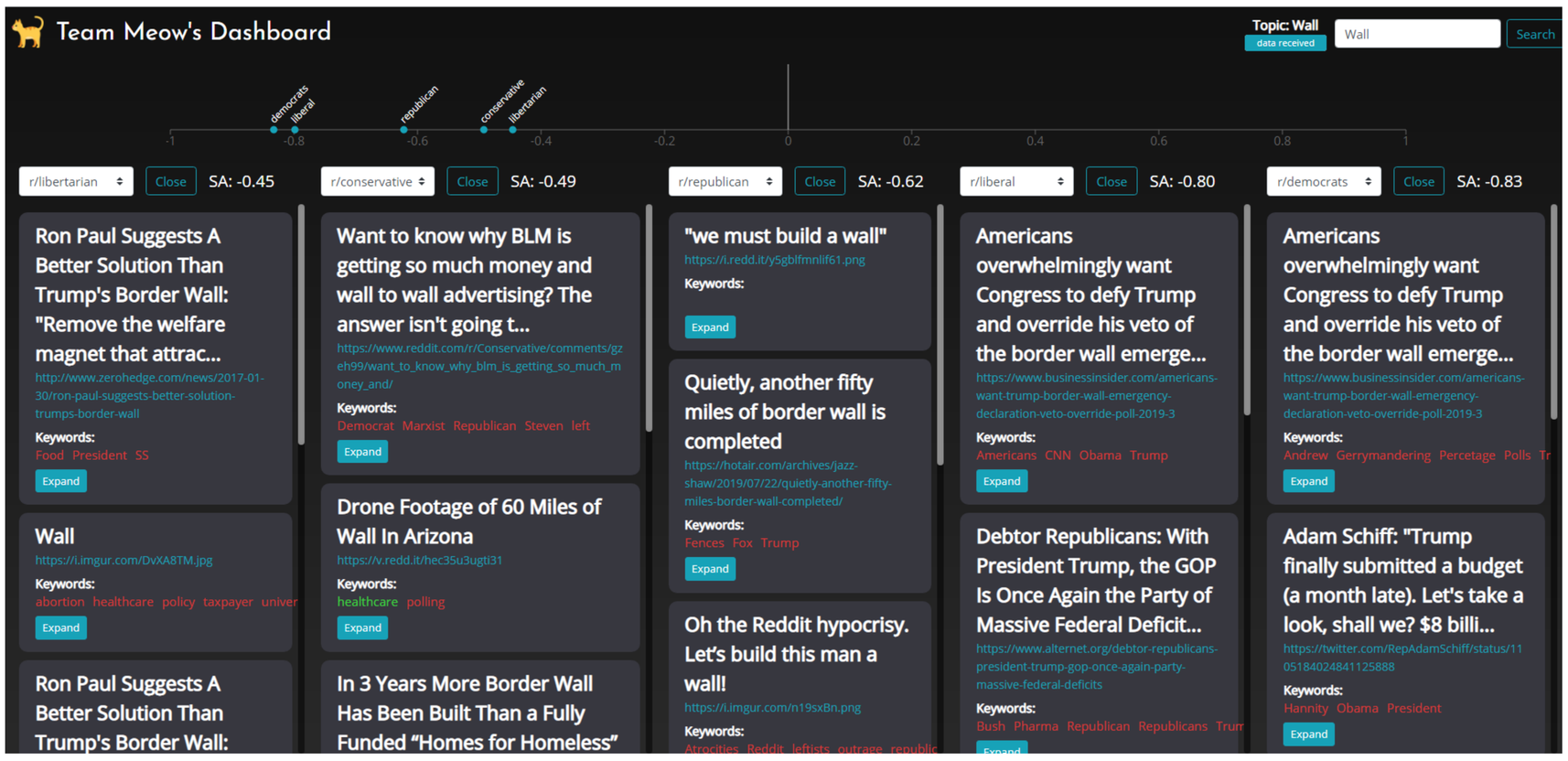
Communities have a **significant effect** on political outcomes, with misinformation easily spreading in both traditional media outlets and via social media. We want citizens to know the biases that several major communities and media outlets have so they can understand those groups' predispositions and general sentiments towards individuals or other topics of interest. We also want to make it easy for people to research and access substantially different perspectives on current topics, side by side, so they are more likely to do so. This will result in more well-informed citizens who are more aware of their, and others', inherent bias and its effect on news.

Data Sources

The project searches comments and submissions from several politically significant and active subreddits in Reddit's database, such as *r/conservative*, *r/republican*, *r/libertarian*, *r/democrats*, and *r/liberal*.

Data Stats

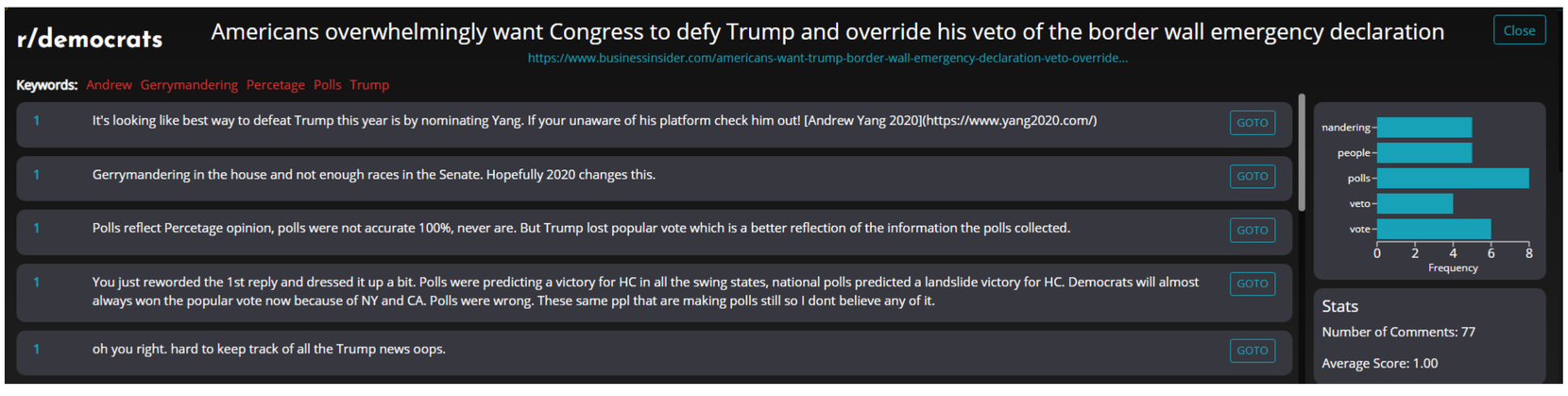
#1.47 million submissions
#20.7 million comments



The dashboard displays the top five submissions from each selected subreddit. The keywords and overall sentiment behind each keyword is displayed, with keywords with negative sentiment red and positive sentiment green. By looking into the various subreddits, we can see how a topic is viewed by a certain political community and what associated topics those groups believe is important.

Method

After a user searches a topic, our dashboard searches several subreddits for relevant submissions. The resulting data is then processed via spaCy, and the keywords are extracted using a named-entity recognition (NER) model. We pretrained a PyTorch BERT model and utilized it to calculate sentiment on each keyword, ranging from -1 (entirely negative) to +1 (entirely positive). The resulting data is sent to a UI written in React.



By selecting a submission, a user can see various stats (votes, etc.) and actual Reddit comments. This allows users to explore it in more depth within each community, or find links to media outlets posted in the submission.

Experiments

The dashboard was presented to family and acquaintances and we solicited feedback. The general overview was positive; people were open to seeing the topic from other perspectives. However, their distrust of news coming from subreddits opposing their perspectives continued. More research on how to present opposing viewpoints more effectively is needed.