

Across the Sea

Peter Penev ppenev1@jhu.edu (601.466)

Abe Hou [bhoul4@jhu.edu](mailto:bhou4@jhu.edu) (601.466)

The focus of this project is on classifying news articles that are based on abortion on their political bias. After classifying the article as either liberal, conservative, or neutral, the project will then crawl the web to find more news articles based on abortion. Finally, after articles are found that are of different political lean compared to the original, information on said articles will be printed and the algorithm will be complete.

(2) Please provide a brief guide describing in sufficient detail how a user could run your code and/or use your provided web interface, especially if using a non-standard platform or interface.

For a user to run the code, first the user must find a news article that is on the topic of abortion. Then they must run the `crawl.py` file with the input of `python crawl.py url`. This will then crawl the web starting from the websites “<https://www.catholicnewsagency.com/tags/35/abortion>” and “<https://www.nbcnews.com/politics/abortion-news>” and then the crawler will find the targeted news articles and print the relevant information. Some important files to keep track of is the `extracted.txt` file and the `visited.txt` file. The `extracted.txt` file shows the extracted files and the score of the document, its url, and the content of the news article. The `visited.txt` file shows a list of all of the url that the crawler visited.

(3) Give an itemized list of your project’s particular achievements, strengths and selling points, especially things that may not be obvious upon inspection of your code or output, as well your personal assessment of their complexity, scope, and success.

One of the strengths of the project is the high precision and recall for the algorithm to actually find an article that has the expected bias and avoid mislabeling a conservative article as liberal or vice versa. Another strength of the project is the efficiency of the project, especially with the crawler and classifier, as the code was written in a way to avoid getting the same information multiple times while also being easy to follow and being time and space efficient. The algorithm is also notably more accurate compared to the POLITICS model while also being able to build on the topic and find key information in an efficient manner. The web crawler is also a strength due to the weighting schema making it much easier to find more relevant articles in as few cycles as possible.

(4) Give a brief itemization of your project's limitations and list suggested possibilities for improvement or worthwhile extension with additional time.

The main limitation of the current project is that we are currently restricted to only the topic of abortion. Some possibilities for improvement with extra time would be to increase the scope of the topics to either all news or most major news issues. Another possibility is also having the news articles be ranked on a scale from very liberal to very conservative and having the algorithm attempt to find an article of the opposing viewpoint with an almost equal amount of bias instead of labeling some articles as either conservative or liberal. These extensions could be made and would increase the scope and use of the project.

(5 and 6) To both document your projects achievements, as well as the methods of use and expected output(s), please also include a range of samples/screenshots of your project output on a range of project functionalities and input conditions or search queries. In the spirit of all assignments in the class, empirical evaluation appropriate for the nature of the project should also be included.

Liberal base file:

```
(base) cli@nokia:/mnt/c/Users/peter/across_the_sea$ python crawl.py https://www.nbcnews.com/politics/politics-news/abortion-civil-rights-groups-aim-put-abortion-floridas-2024-ballot-rcna83081
The stance of the article on abortion you are reading is: liberal
with extremeness: 0.7181333303451538
alternative article with conservative ideology found with an extreme score of 0.6235834360122681
alternative article with neutral ideology found with an extreme score of 0.40291929244995117
alternative article with conservative ideology found with an extreme score of 0.6235834360122681
text: ByPeter Pinedo Since a federal judge's Friday ruling called the legality of the abortion drug mifepristone into question, the governors of New
url: https://www.catholicnewsagency.com/news/254070/democratic-governors-stockpile-abortion-pills-in-response-to-texas-fda-ruling
alternative article with neutral ideology found with an extreme score of 0.40291929244995117
text: ByKatie Yoder The Food and Drug Administration (FDA) lifted restrictions on mifepristone, a drug approved for use in medical abortions, on Thu
url: https://www.catholicnewsagency.com/news/249915/breaking-fda-allows-women-to-get-abortion-pills-by-mail
https://www.catholicnewsagency.com/tags/35/abortion
https://www.catholicnewsagency.com/news/254070/democratic-governors-stockpile-abortion-pills-in-response-to-texas-fda-ruling
https://www.catholicnewsagency.com/news/254065/a-texas-judge-suspended-abortion-pill-approval-so-does-that-mean-they-are-illegal
https://www.catholicnewsagency.com/news/253431/biden-vows-to-increase-abortion-pill-access-on-the-50th-anniversary-of-roe-v-wade
https://www.catholicnewsagency.com/news/252709/fda-warns-against-provision-of-abortion-pills-to-women-before-pregnancy
https://www.catholicnewsagency.com/news/249915/breaking-fda-allows-women-to-get-abortion-pills-by-mail
https://www.nbcnews.com/politics/abortion-news
```

Neutral base file:

```
(base) cli@Nokia:/mnt/c/Users/peter/across_the_sea$ python crawl.py https://www.foxnews.com/politics/florida-pro-life-pregnancy-center-targeted-with-decapitated-chicken-mutilated-lamb-ritualistic-attack
The stance of the article on abortion you are reading is: neutral
with extremeness: 0.6771071553230286
alternative article with conservative ideology found with an extreme score of 0.6235834360122681
alternative article with liberal ideology found with an extreme score of 0.5900959968566895
!
alternative article with conservative ideology found with an extreme score of 0.6235834360122681
text: ByPeter Pinedo Since a federal judge's Friday ruling called the legality of the abortion drug mifepristone into question, the governors of New
url: https://www.catholicnewsagency.com/news/254070/democratic-governors-stockpile-abortion-pills-in-response-to-texas-fda-ruling
alternative article with liberal ideology found with an extreme score of 0.5900959968566895
text: ByPeter Pinedo Texas judge Matthew Kacsmaryk suspended the FDA's approval of the abortion drug mifepristone Friday on the grounds that approva
url: https://www.catholicnewsagency.com/news/254065/a-texas-judge-suspended-abortion-pill-approval-so-does-that-mean-they-are-illegal
|
https://www.catholicnewsagency.com/tags/35/abortion
https://www.catholicnewsagency.com/news/254070/democratic-governors-stockpile-abortion-pills-in-response-to-texas-fda-ruling
https://www.catholicnewsagency.com/news/254065/a-texas-judge-suspended-abortion-pill-approval-so-does-that-mean-they-are-illegal
https://www.nbcnews.com/politics/abortion-news
```

Conservative base file:

```
(base) cli@Nokia:/mnt/c/Users/peter/across_the_sea$ python crawl.py https://edubirdie.com/examples/the-issue-of-abortion-in-christianity/
The stance of the article on abortion you are reading is: conservative
with extremeness: 0.8555071353912354
alternative article with liberal ideology found with an extreme score of 0.5900959968566895
alternative article with neutral ideology found with an extreme score of 0.40291929244995117
!
alternative article with liberal ideology found with an extreme score of 0.5900959968566895
text: ByPeter Pinedo Texas judge Matthew Kacsmaryk suspended the FDA's approval of the abortion drug mifepristone Friday on the grounds that approva
url: https://www.catholicnewsagency.com/news/254065/a-texas-judge-suspended-abortion-pill-approval-so-does-that-mean-they-are-illegal
alternative article with neutral ideology found with an extreme score of 0.40291929244995117
text: ByKatie Yoder The Food and Drug Administration (FDA) lifted restrictions on mifepristone, a drug approved for use in medical abortions, on Thu
url: https://www.catholicnewsagency.com/news/249915/breaking-fda-allows-women-to-get-abortion-pills-by-mail
|
https://www.catholicnewsagency.com/tags/35/abortion
https://www.catholicnewsagency.com/news/254070/democratic-governors-stockpile-abortion-pills-in-response-to-texas-fda-ruling
https://www.catholicnewsagency.com/news/254065/a-texas-judge-suspended-abortion-pill-approval-so-does-that-mean-they-are-illegal
https://www.catholicnewsagency.com/news/253431/biden-vows-to-increase-abortion-pill-access-on-the-50th-anniversary-of-roe-v-wade
https://www.catholicnewsagency.com/news/252709/fda-warns-against-provision-of-abortion-pills-to-women-before-pregnancy
https://www.catholicnewsagency.com/news/249915/breaking-fda-allows-women-to-get-abortion-pills-by-mail
https://www.nbcnews.com/politics/abortion-news
```

(7) If your code submission includes components that were (a) written by people not on your team or (b) written by yourself for another course, for prior research and/or employment, please very clearly explain/document which components were done elsewhere. Such outside borrowing is permitted if clearly documented, and you are permitted to build upon prior accomplishment, just like you may utilize software libraries from elsewhere, but your project will be evaluated on the original work done for this course.

We utilized a repository of abortion information from <https://arxiv.org/pdf/2302.01439.pdf> and we also used a pretrained POLITICS model that was significantly modified to help achieve the project goal. However, no code was duplicated or copied from any source excluding the template code for hw4 for this class for the crawler.