

Bryce Hutton, Dan Schroeder, Hunter Hubers

Dr. Jared Moore

### Project 3: Project Proposal

For our project we intend to develop a "News Bias Topic Network", that ~~analysis~~ <sup>analyzes</sup> news articles using neural networks and topic modeling in order to identify the type of bias that the article has. We will accomplish this by training our network using a variety of different political-leaning websites, ~~and by~~ <sup>by</sup> basing our topic modeling off of those sites as well. We intend on having 6 different types of articles: extreme left, moderate left, centrist, moderate right, extreme right, and neutral. For each type of article we will use a news source that has that political bias (Such as ~~something~~ like Breitbart for far right, Fox for moderate right, etc), and for neutral we intend on using articles with no political bias (such as technical or scientific articles). Once we have trained the network, we will give the program a URL through the command line. The end result will be an average between the neural network and the topic modeling.

One of the most notable and similar projects that has been done comes from Massachusetts Institute of Technology's Computer Science and Artificial Intelligence Lab (CSAIL) and the Qatar Computing Research Institute (QCRI).

*"Researchers first took data from Media Bias/Fact Check (MBFC), a website with human fact-checkers who analyze the accuracy and biases of more than 2,000 news sites, from MSNBC and Fox News to low-traffic content farms. They then fed that data to a machine learning algorithm called a Support Vector Machine (SVM) classifier, and programmed it to classify news sites the same way as MBFC"* (Massachusetts Institute of Technology)

Though this CSAIL and QCRI model more so attempts to validate the accuracy and the truthfulness of a news article, our model will solely be focused on the bias of the article itself. We won't be gauging truthfulness (maybe as a stretch goal), but we will attempt to determine if the article is right, left, or neutral. We could potentially use this information to further make inferences about the article's writer and readers. We have also found a Journal on the topic (cited below).

For this project we will be using Python and (at least) the following libraries:

- SKLearn
- Tensorflow
- Keras
- NLTK
- Beautiful Soup
- Poppler
- Gensim

We will not need any special hardware in order to complete the project. With the work that we are doing we do not expect to need anything extra besides our own machines and EOS.

*If you need access to a GPU we have an Nvidia Titan on Ghost in the DEN.*

For research, we will have to learn how the various libraries work, and create a training dataset for the ANN. We will do this by scraping various news sites for articles and supplying those articles to the ANN with the political bias attached to them. *Is there a publicly available dataset?*

After the dataset has been created we will develop both the ANN and the topic modeling. We will train the ANN (potentially utilizing SKLearn, Keras, Tensorflow), then we will preprocess the data for topic modeling. We will use NLTK for preprocessing and normalization, such as lemmatization. After that we will model the topics with gensim, using each article as a document within gensim.

Once we have developed the ANN and topic modeling, we will create an application that will allow you to check any individual news article for its political bias. The application will prompt the user for a URL through the command line and visit it. The application will visit the URL, scrape the website for the article, and will feed it to the ANN and topic modeling. After that the application will average the results from the two methods and give back a verdict on what sort of bias it has, if it can find one.

We will reflect on our project by checking on the results of the program manually. This means reading articles that the program evaluates and determining if there is a left or right bias.

### Works Cited

Massachusetts Institute of Technology. "Machine Learning System Aims to Determine If a News Outlet Is Accurate or Biased." *Tech Xplore*, Massachusetts Institute of Technology, 4 Oct. 2018, [techxplore.com/news/2018-10-machine-aims-news-outlet-accurate.html](https://techxplore.com/news/2018-10-machine-aims-news-outlet-accurate.html).

Hamborg, F., Donnay, K. & Gipp, B. "Automated Identification of Media Bias in News Articles: an Interdisciplinary Literature Review." *Int J Digit Libr* (2018). <https://doi.org/10.1007/s00799-018-0261-y>

*Very ambitious given the timeline but it appears that you have looked into it quite thoroughly. I'd strongly encourage you to look for blog posts detailing similar projects as a starting point. Also consider searching for a publicly available dataset.  
Good Luck!*

*95%*