CSCE 439 - Project Proposal

Andrew Wheir and Sarah Gullion

Dataset

https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset/
https://www.kaggle.com/datasets/saurabhshahane/fake-news-classification

Research Question

Being able to determine if a news article is fake or real can be surprisingly difficult in our modern era. Social media has made the spread of smaller, less reliable news sites very common, making fake news a problem in our modern society. While there are companies devoted to algorithms to cross list facts by scraping the web, there may be an easier way. Our hypothesis is that just the style of writing may be enough to determine whether or not an article is real or not.

To do this, we plan to train a Random Forest Classification algorithm to learn what kind of features are used in fake and real articles, then use it to see how good it is at being able to distinguish between the two.

Citations:

Gangireddy, S. C., P, D., Long, C., & Chakraborty, T. (2020). Unsupervised fake news detection. *Proceedings of the 31st ACM Conference on Hypertext and Social Media*. https://doi.org/10.1145/3372923.3404783

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017, September 3). *Fake news detection on social media: A Data Mining Perspective*. arXiv.org. https://arxiv.org/abs/1708.01967

Zhang, J., Dong, B., & Yu, P. S. (2019, August 10). *Fakedetector: Effective fake news detection with deep diffusive neural network*. arXiv.org. https://arxiv.org/abs/1805.08751