# Who Wrote This? Classifying the Source of News Articles

Stephen Lee

**Abstract** The 2016 United States Presidential Election, to many, raised questions as to the reliability of their news sources [1]. Since then, many social media companies and other institutions have begun public campaigns to combat the perceived threat from fake news, with arguably limited results [2]. With a goal of news article source identification, I scrape several thousand news articles from Fox, Vox, and PBS. By some estimations [3], these three represent distinct categories of news: Fox is often considered extreme “right” opinion (i.e. conservative); Vox is considered extreme “left” opinion (i.e. liberal); and PBS is considered “center” primary source news. I first calculate the top word, 2-gram, and 3-gram frequencies to better understand the dataset, and then train a Bidirectional LSTM neural network with pretrained embeddings to classify the news source. My tentative results indicate that articles from each source are nearly indistinguishable by both the top word frequency count and a simple Bidirectional LSTM classifier. While more model testing is needed, this suggests that perceived bias is likely subtle and relies heavily on context.

# References

1. Hunt Allcott & Matthew Gentzkow, 2017. "Social Media and Fake News in the 2016 Election," Journal of Economic Perspectives, vol 31(2), pages 211-236.
2. Robinson Meyer, 2018. “The Grim Conclusions of the Largest-Ever Study of Fake News,” The Atlantic
3. <https://www.marketwatch.com/story/how-biased-is-your-news-source-you-probably-wont-agree-with-this-chart-2018-02-28>