Biased Language in News Media

Tom Wang

Contents

About

- Data Scientist
- Interested in how bias can influence language and the news

- Question: Can NLP techniques detect differences in language used when comparing different news media sources?

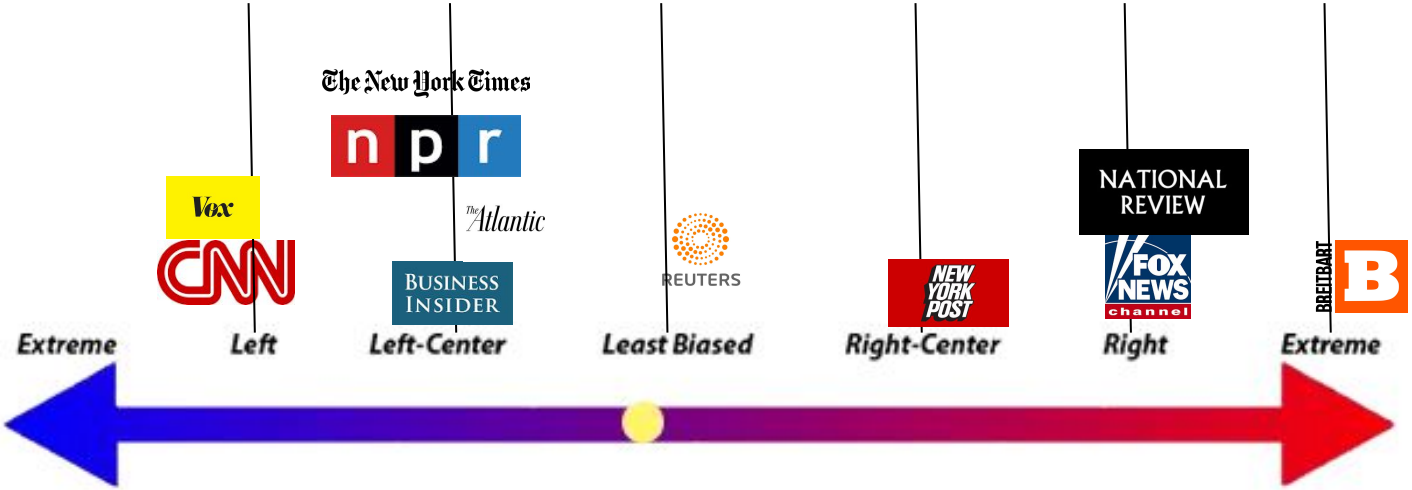
The Datasets

- News Articles Dataset
  - ~140k News articles scraped from 15 different news media websites
  - 2015 - 2017

- Bias Lexicon
  - List of 654 bias-inducing lemmas
  - Stanford researchers derived from Wikipedia edits corpus

MediaBiasFactCheck.com

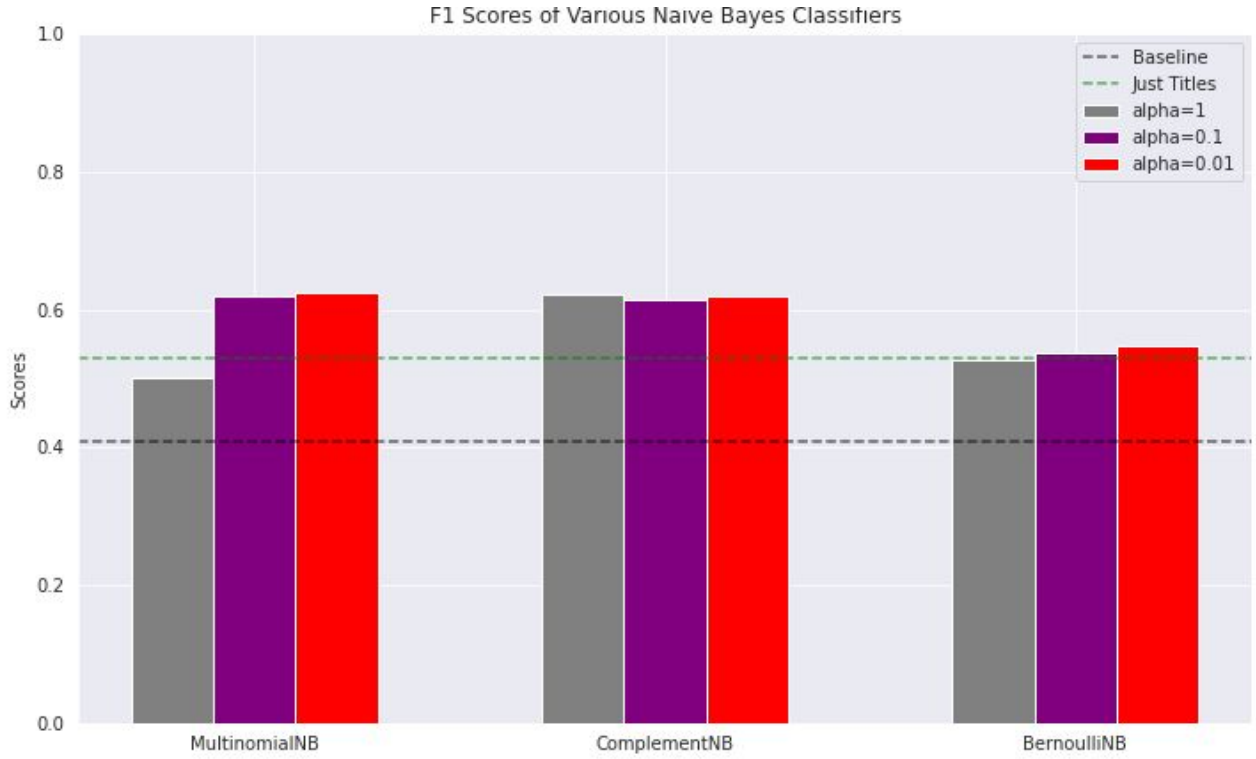
"Media Bias/Fact Check (MBFC), founded in 2015, is an independent online media outlet. MBFC is dedicated to educating the public on media bias and deceptive news practices." [1]



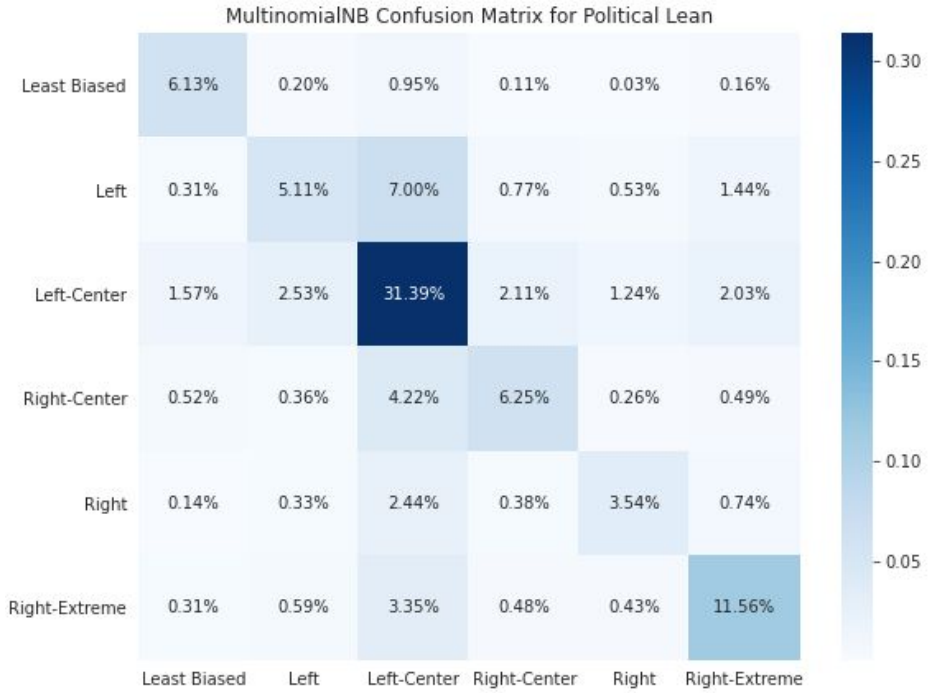
[1] - https://mediabiasfactcheck.com/about/

Naive Bayes Classifiers - F1 Score

- TF-IDF Vectorized article text
- Multinomial/Bernoulli/Complement Naive Bayes



Naive Bayes Confusion Matrix



Topic Modeling on News Set

Unsupervised Feature Generation - Topics / Subject Matter

Methodologies

1. Gensim - Latent Dirichlet Allocation (LDA)
2. NMF - Non-negative Matrix Factorization

Model	Training time (seconds)
LDA	334.6
NMF	3.6

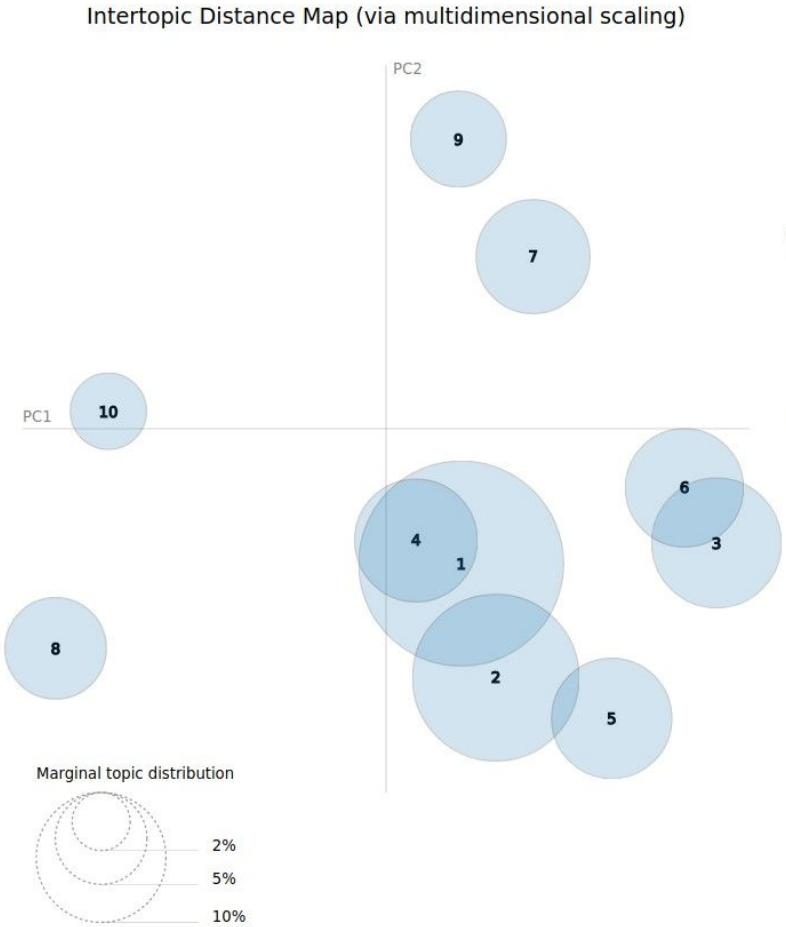
Topic Modeling on News Set

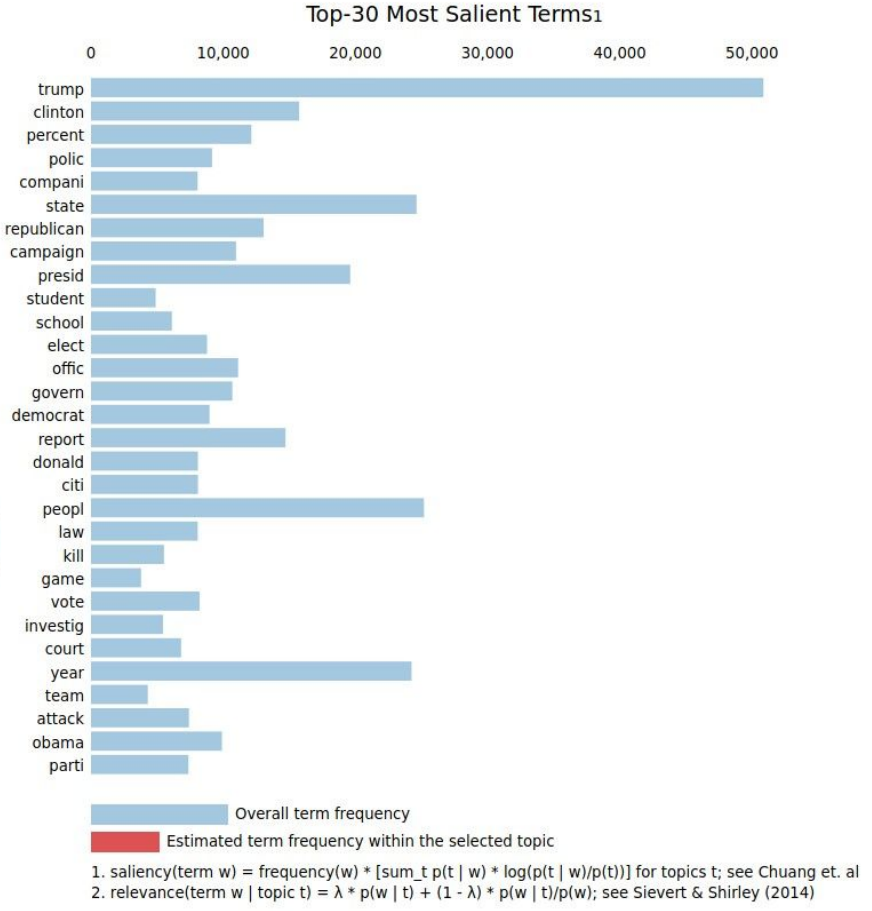
Unsupervised Feature Generation - Topics / Subject Matter

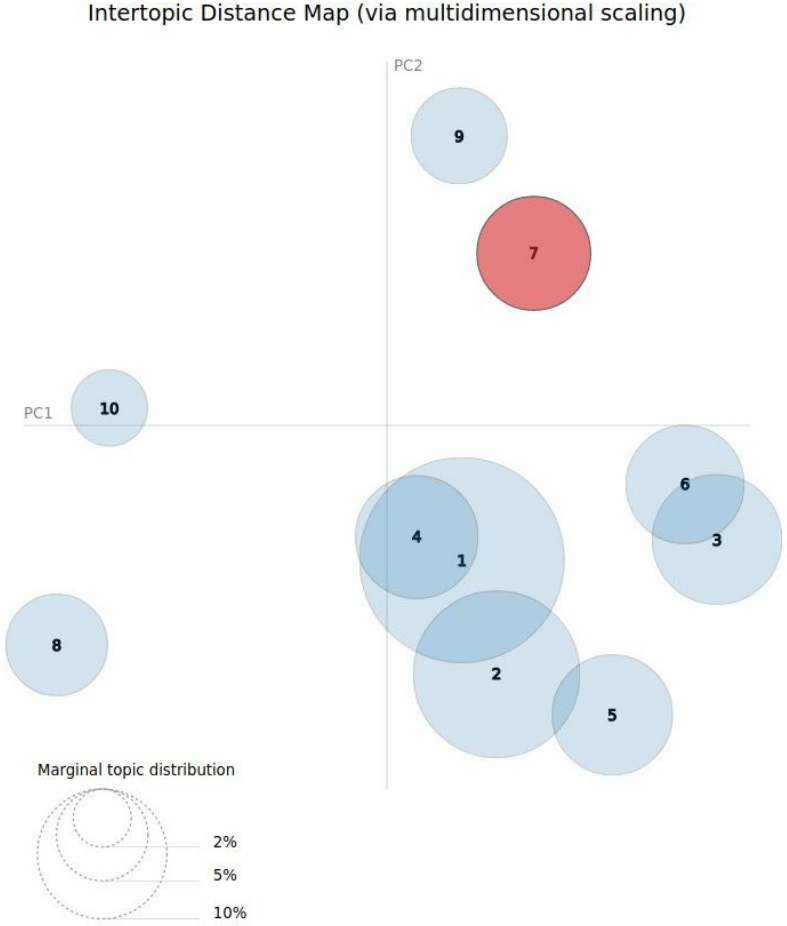
Methodologies

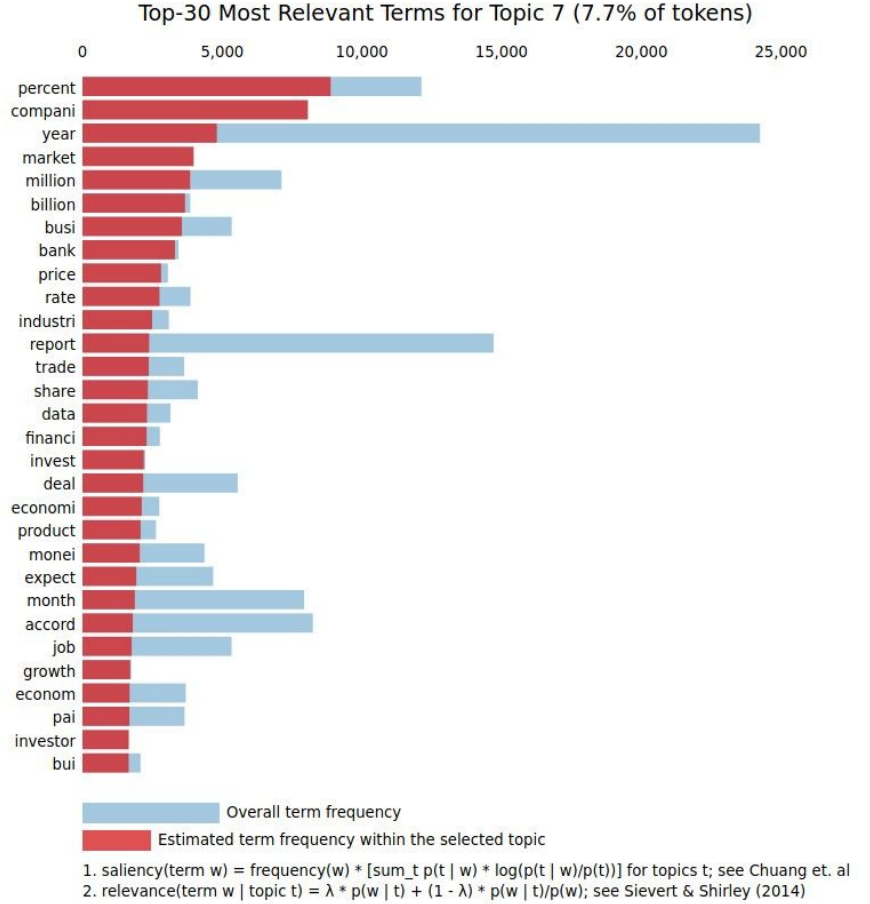
1. **Gensim - Latent Dirichlet Allocation (LDA)**
2. NMF - Non-negative Matrix Factorization

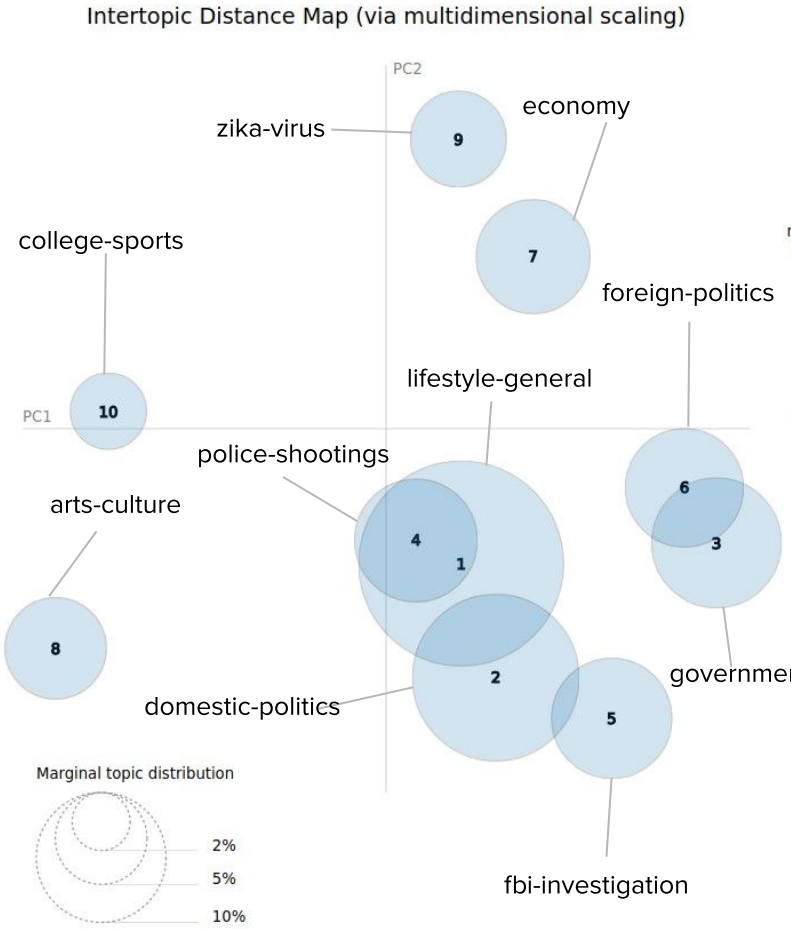
Model	Training time (seconds)
LDA	334.6
NMF	3.6

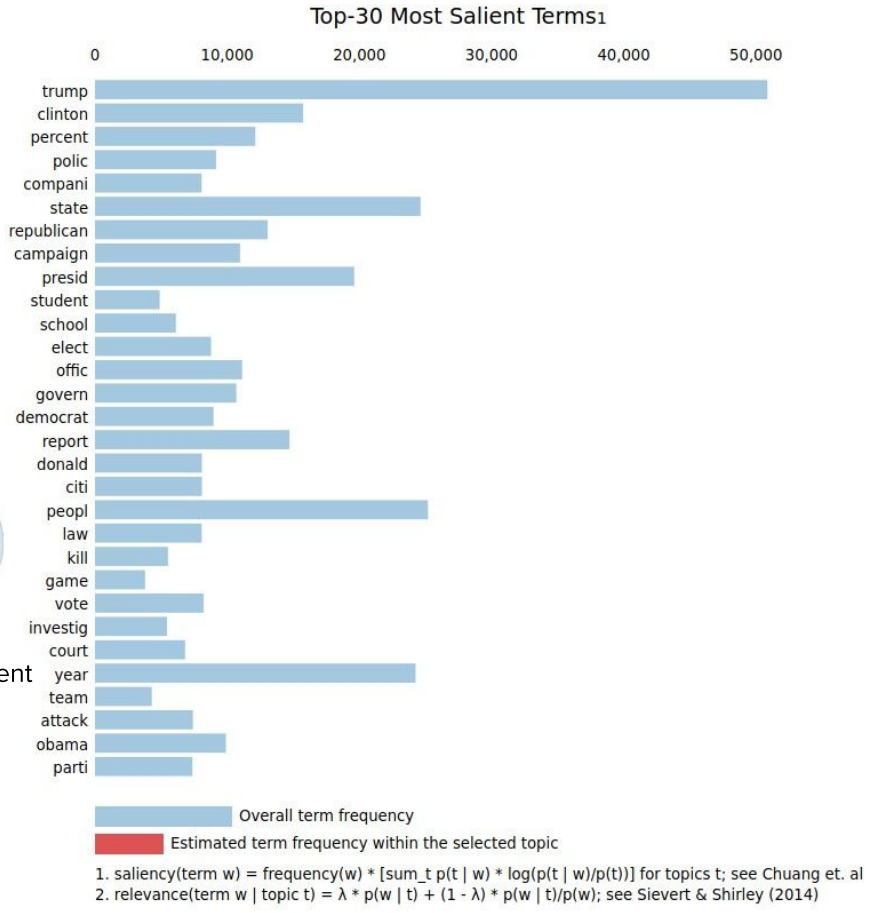












Top 3 News Topics by Political Lean

Least Biased	
economy	39.7%
foreign-politics	19.7%
domestic-politics	9.4%

Left	
domestic-politics	25.8%
lifestyle-general	21.5%
police-shootings	10.8

Left-Center	
lifestyle-general	33.9%
domestic-politics	16.65%
government	8.7%

Right-Center	
lifestyle-general	32.8%
domestic-politics	16.65%
government	8.7%

Right	
domestic-politics	31.0%
lifestyle-general	26.1%
government	11.5%

Right-Extreme	
domestic-politics	37.1%
police-shootings	14.8%
lifestyle-general	14.0%

What is biased language?

Examples [2]

"Shwekey's albums are arranged by many talented arrangers."

"The first research revealed that the Meditation technique produces a unique state fact."

"Marriage is a holy union of individuals."

[2] - Pryzant et al., Automatically Neutralizing Subjective Bias in Text. AAAI 2020: 480-489

What is biased language?

Examples

"Shwekey's albums are arranged by many **talented** arrangers."

"The first research **revealed** that the Meditation technique produces a unique state fact."

"Marriage is a **holy union** of individuals."

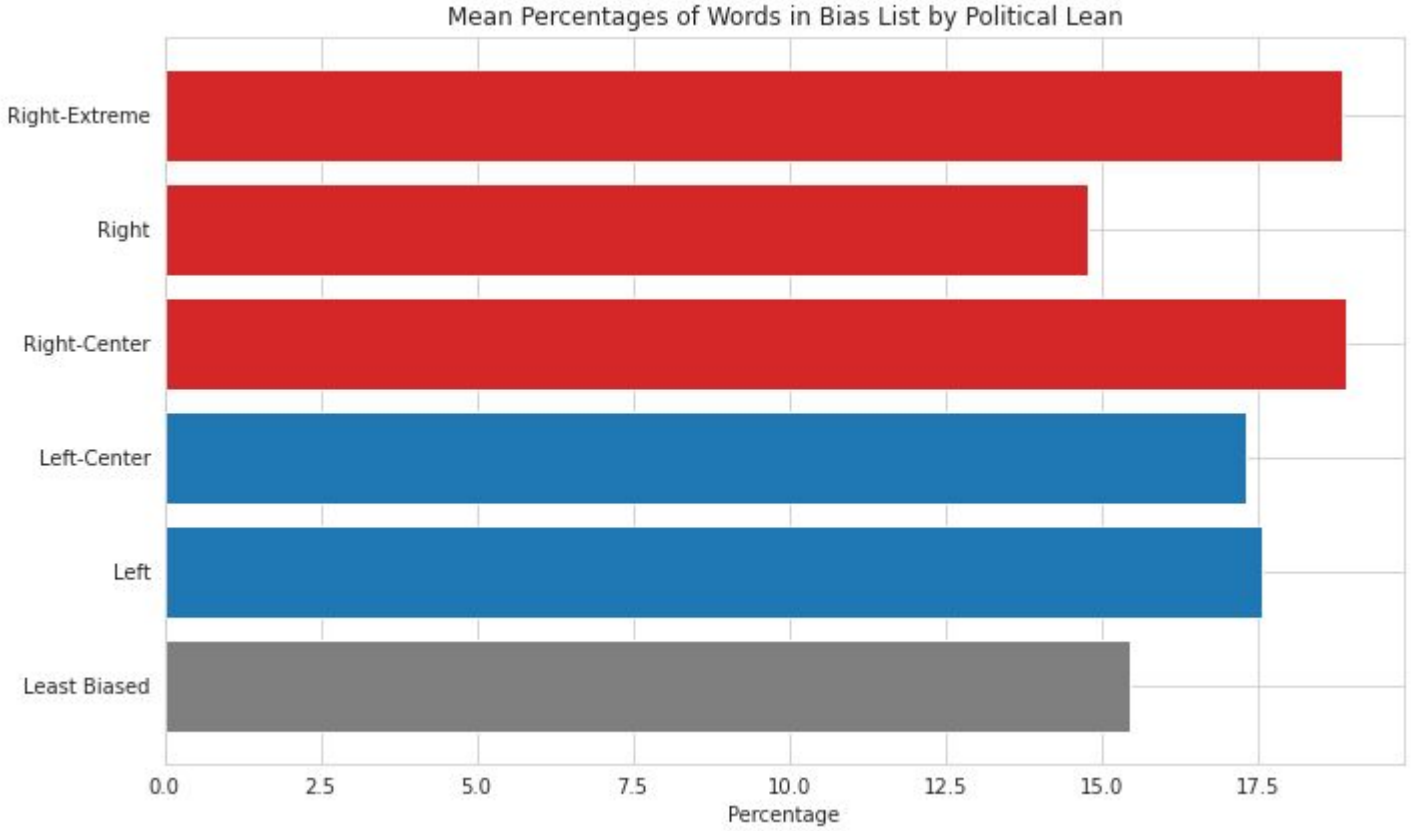
What is biased language?

Examples

"Shwekey's albums are arranged by many different arrangers."

"The first research indicated that the Meditation technique produces a unique state fact."

"Marriage is a personal union of individuals."



Conclusions

- Both topic and type of language used tend to influence the degree of bias found in news media texts
- An NLP model that could accurately detect biased language would be useful for real-world editors of encyclopedias, neutral news sources, etc.

Future Work

- Conduct the same analysis on the corpus of Wikipedia articles that have been marked “Controversial”

Questions/Comments/Feedback