

Determining Media Bias in Articles Using NLP

Andrew Berry

*“Information overload and awareness of fake news have been dominating the news headlines the past few years. It’s a safe assumption that each news outlet may have a certain agenda in reporting certain stories. They may under-emphasize or overemphasize certain facts to push the reader into a certain direction of thought. This project aims to provide readers on potential media bias in what their reading.....with a little help from **Natural Language Processing Methodologies** and **Latent Dirichlet Allocation (LDA) Topic Modeling**”*

Key Technologies Used:

- Spacy (NLP library for Python)
- NLTK (NLP library for Python)
- Gensim (NLP library for Python, used this mainly for LDA topic Modeling)
- TextBlob (NLP library for Python)
- plyDavis (Python Visualization library)
- Tableau (Visualization Software)

Workflow Pipeline:

1. Deployed a web-scraper to gather articles from various news outlets.
2. Text Pre-Processing
 1. Tokenization (Spacy)
 1. Data Cleaning achieved during tokenization
 2. Removing Stopwords (NLTK)
 3. Lemmatization (NLTK - WordNetLemmatizer)
 4. Bigram Identifier (Gensim)
3. LDA Topic Modeling (Gensim)
4. Sentiment Analysis (TextBlob)
5. Tableau for Visualizations

Curious about my code? Check out my Github repository for this work.

<https://bit.ly/2KbunTR>