

CS410 Course Project Proposal

US News Political Bias Detector

Team Name: Bias Detectives

Names: Anh Nguyen, Muhammad Rafay, Nicholas Bachman

NetID's: anh4@illinois.edu, mrafay2@illinois.edu, bachman5@illinois.edu

Team Captain: Nicholas Bachman

Free Topic: Sentiment Analysis

Description: Our free topic is to design a system that can examine mainstream news headlines and determine if they contain any political bias. The bias would be determined using sentiment analysis and NLP techniques.

Important or Interesting: The political tension between the Democrat and Republican parties in the US has never seemed higher. Mainstream news media reports on events with a palpable bias that slants heavily toward one party or the other. Many Americans want unbiased, factual news reports to avoid being manipulated. An example visualization of this political bias from AllSides.com is shown in Figure 1. These visualizations are created from community feedback data which can also be bias / not normalized. <https://www.allsides.com/media-bias/media-bias-ratings>



Figure 1 Media Bias Visual

Planned Approach: Our approach will be to text mine news headlines from 2020 to create a data set or find an existing dataset that has already been labeled. We will then build or modify an existing Sentiment Analysis Model and tune it for political bias. Other models that we might try in case we don't get good results are clustering based on latent Dirichlet allocation (LDA), logistic regression and deep learning. We then plan to visualize the results to see how various media outlets rank across a bias spectrum. Left and Right will be the sentiment classes. We might add Lean Left and Lean Right classes if deemed appropriate.

Tools, systems, datasets: NLTK is a suite of python libraries that can be used for classification, tokenization, stemming, tagging, parsing and NLP. Google's Named Entities Sentiment Beta API for entity sentiment. Might use TensorFlow if we tried Deep Learning model. Gensim for LDA. We will high likely use Hybrid sentimental analysis algorithms which include stemming, tokenization, lexicon along with some automatic approaches available for public use like BERT. The final report might be displayed in form of BI visualization with Tableau or D3.js.

Expected outcome:

- Input
 - A New Article Headline
 - A fixed set of classes $C = \{c_1, c_2, \dots, c_n\}$
- Output
 - A predicted class $c \in C$

Project Timeline:

Milestone	Due Date
Submit Proposal	Oct 25th
Working Prototype	Nov 22th
Project Progress Report	Nov 29th
Project Completion and Submission	Dec 13th

Programing Language: Python and NLTK (<http://www.nltk.org/>)

Workload Justification:

N = 3 team members

3 * 20 = 60 hours

Task	Estimated Hours
Build Text Mining / Web scrapping for News Headlines	5
Collect and Label Test / Training Datasets (Corpus)	5
Build / Tune Sentiment Analysis Model	20
Train Sentiment Analysis Model (Iterative)	15
Test Sentiment Analysis Model	15

Visualization to display results (Tableau or Website)	8
Develop Software Documentation	2
Create Video Demonstration	2
Total	72