Reema Domadia

Professor Abdi Awl

DATS 6501

January 14, 2020

***Are 21st Century News Sources Propaganda Machines?***

**Introduction**

Often being called “fake news” or “enemy of the American people”, the 4th pillar of democracy: The Media, has constantly been at the heart of Donald Trump’s attacks. The polarity in reporting by each source has nearly diminished the moderate class and turned liberals and conservatives against each other, destroying the unity that served at the foundation of America’s democracy. The media, whose purpose is to distribute information, is a key component in creating this polarity. Constantly in competition with social media platforms such as Facebook and Twitter, news outlets struggle to retain their viewer base resulting in their contribution to the information bubble. This means that news outlets are broadcasting information that their userbase wants to hear rather than raw facts. Over time, this bubble becomes so strong that the users do not understand the other side. While it is hard for an analyst or a journalist to prevent personal beliefs and opinions cloud reporting, it is crucial that the news broadcast only the facts and allow the people to reach their own decisions. By including an opinion segment on news shows or letting opinion creep into reporting, its audience is swayed to believe the journalist or the news anchors beliefs, converting a news outlet into a propaganda machine.

The primary objective of this project is to identify the leaning of each news source to not only be able to inform the American public of the integrity facts they will receive. In addition, the identification offers news sources feedback to implement necessary changes to offer more factual and bi-partisan news.

**Methodology**

The entirety of this project contains two phases: data gathering and analysis. The data gathering phase requires the construction of an application that will scrape the web using an API for variables such as news articles, source names, time/date, etc. Due to the time constraint, the project will utilize a pre-made dataset and focus on the second part: analysis. The dataset for this project contains 2.7 million observations ranging from 27 news sources. However, processing this sum of data request extensive computational power, which is not available for the construction of this project. As a result, the dataset contains data from 15 news sources dating back to 2013, encapsulating both a democratic and a republican administration.

There are three components to this project:data pre-processing, natural language processing, visualization. Data pre-processing and natural language processing will be done using the python programming language in the PyCharm IDE. Once the data is processed and natural language processed to identify biases, visualizations will be built using tableau. Effective visualizations are crucial in conveying a story to the members of the audience that are not technologically inclined.