Christopher Kruger

EM 624

Exercise 9: Analyze News – Narrative

The original json data set involved ~200,000 recorded entries of articles posted by the Huffington Post between the years of 2012 and 2022. The data contained in the json file included information such as the category of news, the headline, the authors, the link to view the article, a short description/abstract of the article, and the date of publication. This 10-year slice of time is a perfect decade to go through news articles for each year as social media and the interconnectivity of everyone began to skyrocket throughout this time.

For the articles posted between 2012-2022, the below image is the top 20 words used in the short descriptions amongst all articles posted. For articles posted between 2012-2016, common words included verbiage to describe the American people and American culture as a whole, such as “people”, “time”, “life”, “love”, and other similar words. Only until 2016 did the Huffington Post begin to seemingly solely focus on political topics. It is clear that articles posted in 2016 primarily reported on the upcoming election, with words such as “Trump, “Clinton”, etc. Between 2016 and 2020, the two most used words were “Trump” and “said”, implying that a large majority of the articles would report a synopsis of whatever Trump had recently said. The 2020 also began to discuss largely about the emerging COVID-19 pandemic as well as the upcoming election between Biden and Trump. Similar reporting on Biden and Trump was made from 2020 – 2022, and for 2022 you can see reporting on the emerging Russian/Ukrainian conflict that began to be mass reported on.

A black keyboard with white text

Description automatically generated

Regarding sentiment analysis, the Huffington Post is mostly neutral in regard to verbiage in their reporting. It is important to note however that although the Huffington Post is neutral in tone, it does not imply that they are neutral in the topics they report on. Overtime, the positive sentiment has gone down with the negative and neutral sentiment going up. A possible conclusion for this is that as the Huffington post began to post more political topics, the subjectivity went away for objective reporting which often has negative and neutral verbiage.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2012 Articles | 2013  Articles | 2014 Articles | 2015 Articles | 2016 Articles | 2017 Articles | 2018 Articles | 2019 Articles | 2020 Articles | 2021 Articles |
| Positive Sentiment | 25.4% | 23.6% | 23.7% | 19.5% | 19% | 16.7% | 12.2% | 12% | 13.8% | 13.9% |
| Negative Sentiment | 10.7% | 10.1% | 11.8% | 14.7% | 13.8% | 16% | 15.8% | 16% | 13.7% | 13.8% |
| Neutral Sentiment | 63.9% | 66.3% | 64.5% | 65.9% | 67.3% | 67.3% | 72% | 72.1% | 72.6% | 72.4% |

|  |
| --- |
| 2022 Articles |
| 11.8% |
| 15.4% |
| 72.9% |

A Wordcloud was then generated to show the most used words in the short descriptions for each year’s articles.

A close up of words

Description automatically generated

2012

A word cloud with different colored text

Description automatically generated

2013

A close up of words

Description automatically generated

2014

A close up of words

Description automatically generated

2015

A close up of words

Description automatically generated

2016

A close up of words

Description automatically generated

2017

A close-up of words

Description automatically generated

2018

A close up of words

Description automatically generated

2019

A close up of words

Description automatically generated

2020

A close up of words

Description automatically generated

2021

A close up of words

Description automatically generated

2022

To discuss just three examples of the Wordclouds, in 2017 you see for the first time a political name such as Trump being mentioned. This marked the moment that Huffington Post began to report on more political news as this matches with the first year of Trump’s presidency. This is further proven in the 2017 sentiment that shows a drop off in positive verbiage due to reporting on more political news. In the 2020 Wordcloud, you also see for the first time COVID-19 and the word Pandemic being used. This marks the same year that COVID-19 was discovered within the United States. In 2020, the sentiment calculations also show a drop in negative verbiage and an increase in positive verbiage. This is most likely due to news corporations at the time trying to find the good news in the constant news of more COVID-19 cases occurring. It is also clear in the 2020 Wordcloud that the words “police” and “black” emerge, and this was due to the Black Liver Matter movement that began to gain worldwide coverage. In the 2021 Wordcloud, one word that stands out to me is “Capitol” which alone doesn’t mean anything, but it was mentioned due to the extensive news articles that occurred during the January 6th United States Capitol attack.