





# Importance of Word Choice in Political Discourse

Eric Laverdiere

7/23/20



# Problem Statement

Politics in America are very divisive. At least that is how it is often portrayed. Can we use data to try and prove if this division is as great as it seems or more overblown, and can this help the media to try and maintain neutrality.

My team has been assigned by Major News Network to try to find the level of division in discourse on the internet.

In this way we can try to look at the political difference based on what people are saying not just polls or voting numbers. We will also see how language can make a difference and what might cause problems when trying to maintain neutral reporting.

# Data

The data is taken from 2 subreddits

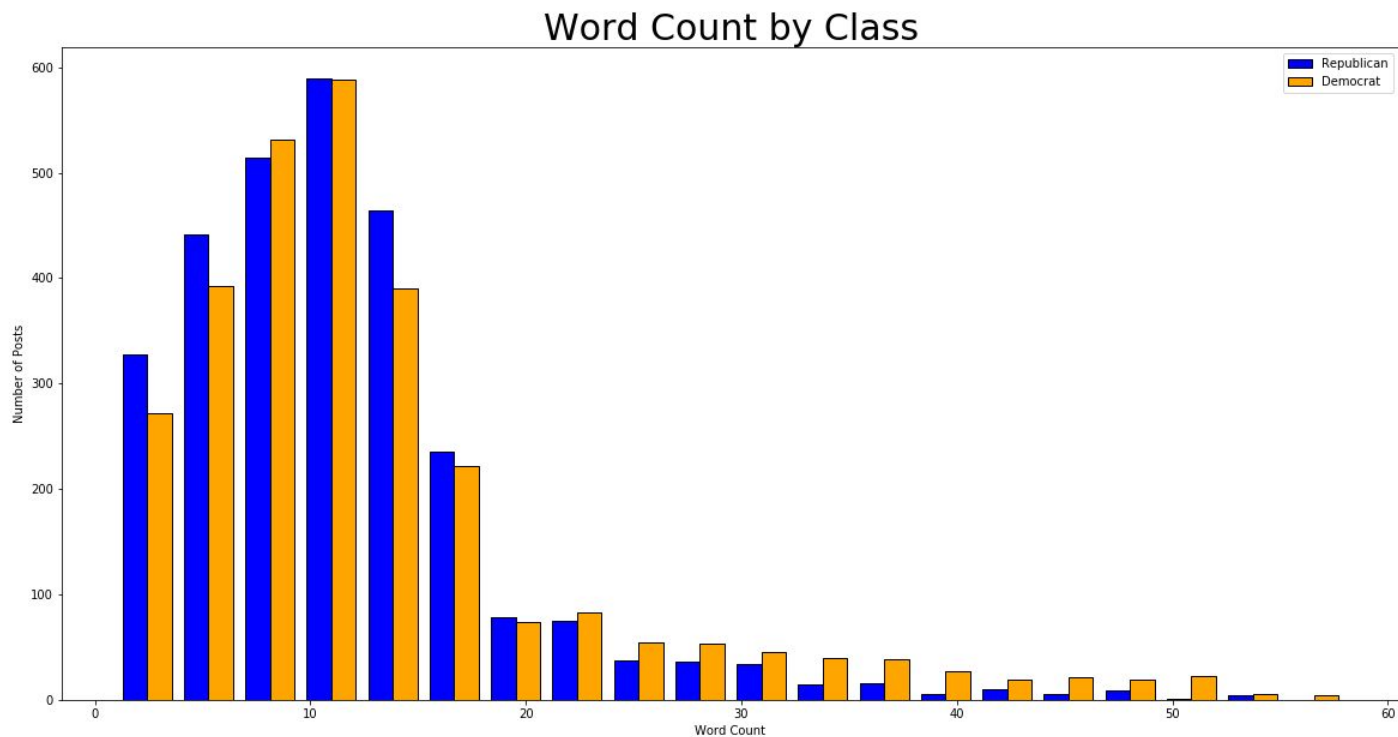
r/democrats

r/Republican

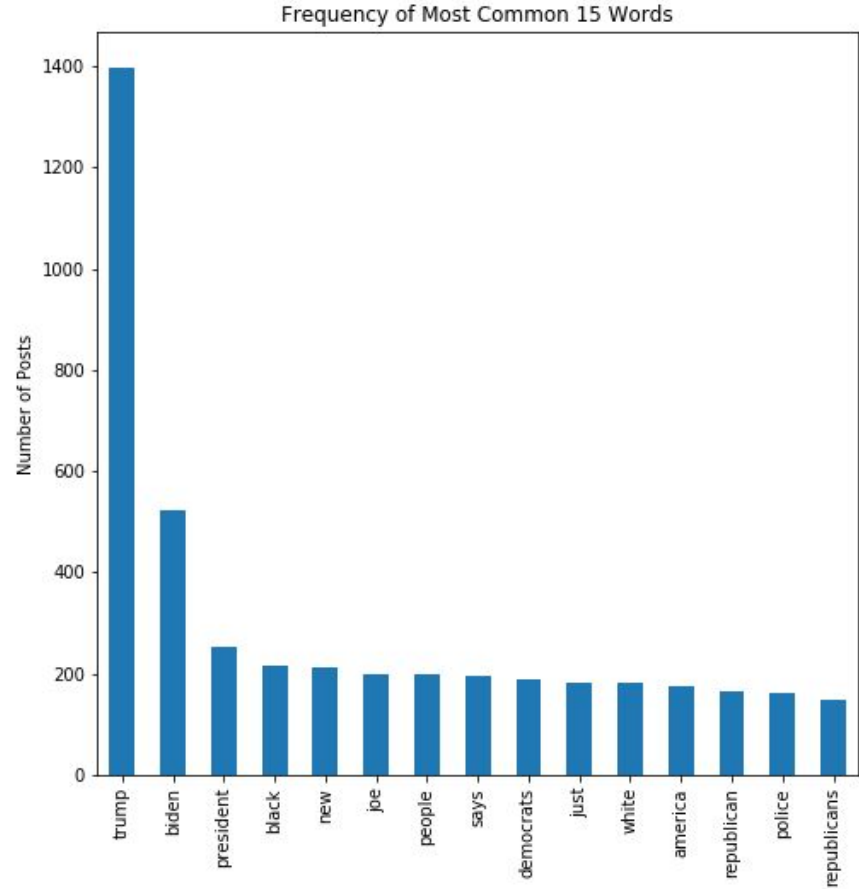
3000 posts from each subreddit both the title and text

Most posts were videos or articles so I kept just the titles

# Points of Interest



Most commonly used words are related to current news.



# Common top Words

Checking the top 20 common words from each subreddit it seems that they have 12 in common.

This could be a problem for modeling with so many in common.

'trump',  
'biden',  
'president',  
'joe',  
'democrats',  
'new',  
'says',  
'america',  
'people',  
'just',  
'republican',  
'black'

# Initial Model

Random Forest

Train Accuracy = 98.7%

Test Accuracy = 65.6%

K Nearest Neighbors

Train Accuracy = 68.4%

Test Accuracy = 56.3%

# Voting Model

Used grid search over 4 different models to tune the hyperparameters.

Random Forest, Decision Tree, Ada Boost, Gradient Boost

Reduced overfit but did not improve accuracy.

Train Accuracy = 70%

Test Accuracy = 63.8%



# What Does This Mean?

While partisanship exists both sides use many common words when discussing current events

Differences are related to what is important for that side

Word choice of articles becomes very important

# Example

Trump Plans to Expand Federal Invasion of American Cities.

<https://www.motherjones.com/anti-racism-police-protest/2020/07/trump-border-patrol-cities-portland-chicago/>

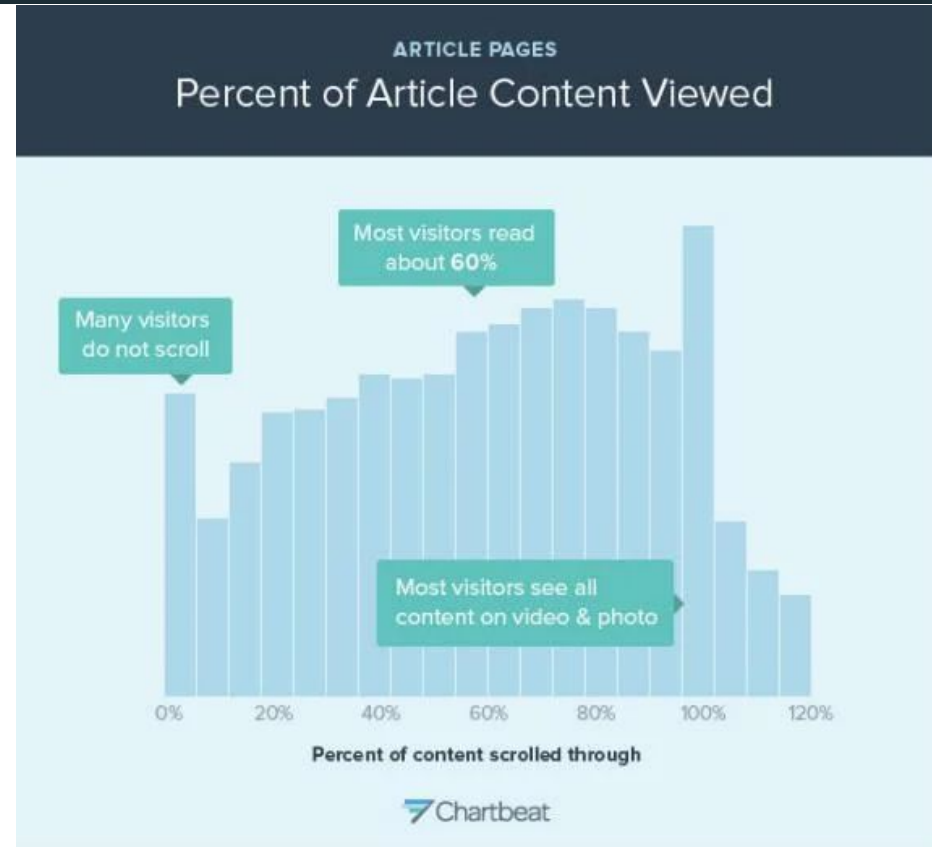
Trump vows to send feds into cities facing violence, as DHS reportedly drafts Chicago deployment plan

<https://www.foxnews.com/politics/trump-vows-to-send-federal-agents-dhs-chicago>

# Article Titles Are Important

Most people do not read a full article

Many do not get past half or even start scrolling



<https://slate.com/technology/2013/06/how-people-read-online-why-you-wont-finish-this-article.html>

# Conclusion

Models used in this work have trouble finding the difference between subreddits because of similar keywords.

Make sure to be careful of showing bias by word choice.

Titles can be just as important as content because of the way people consume media.

With improvement the model could help predict greater bias and help to move to a more neutral conversation.

# Reference

<https://www.reddit.com/r/democrats/>

<https://www.reddit.com/r/Republican/>

Manjoo, Farhad (June 06, 2013)

<https://slate.com/technology/2013/06/how-people-read-online-why-you-wont-finish-this-article.html>