Democrat or Republican?

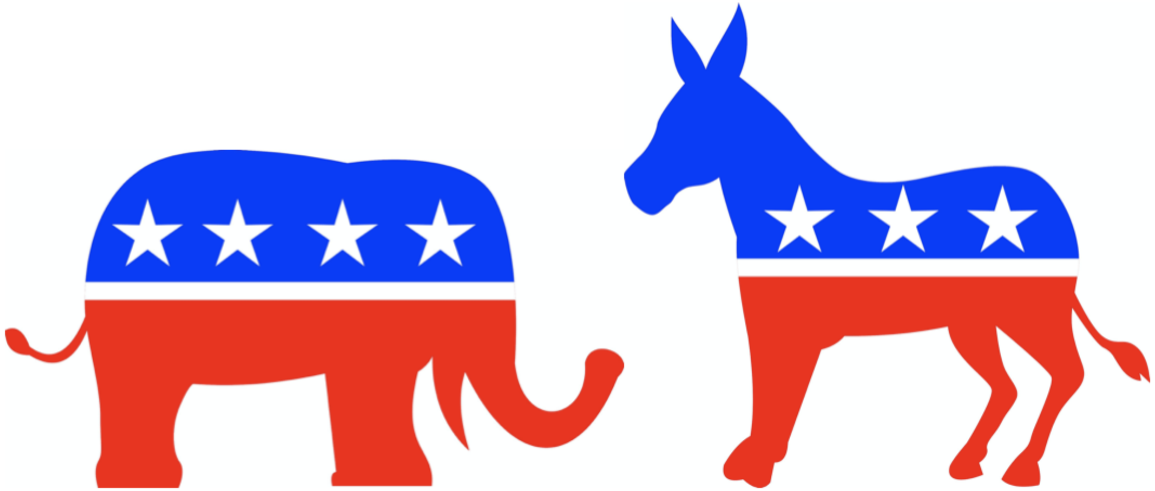
Varun Reddy

Data Scientist

Background and Problem Statement

There has been great polarization in our American political system with often two completely different sets of news and facts being reported to people depending on their political affiliation.

We were asked to investigate whether we can determine a person's political affiliation by the title posts on Reddit.



Our Process

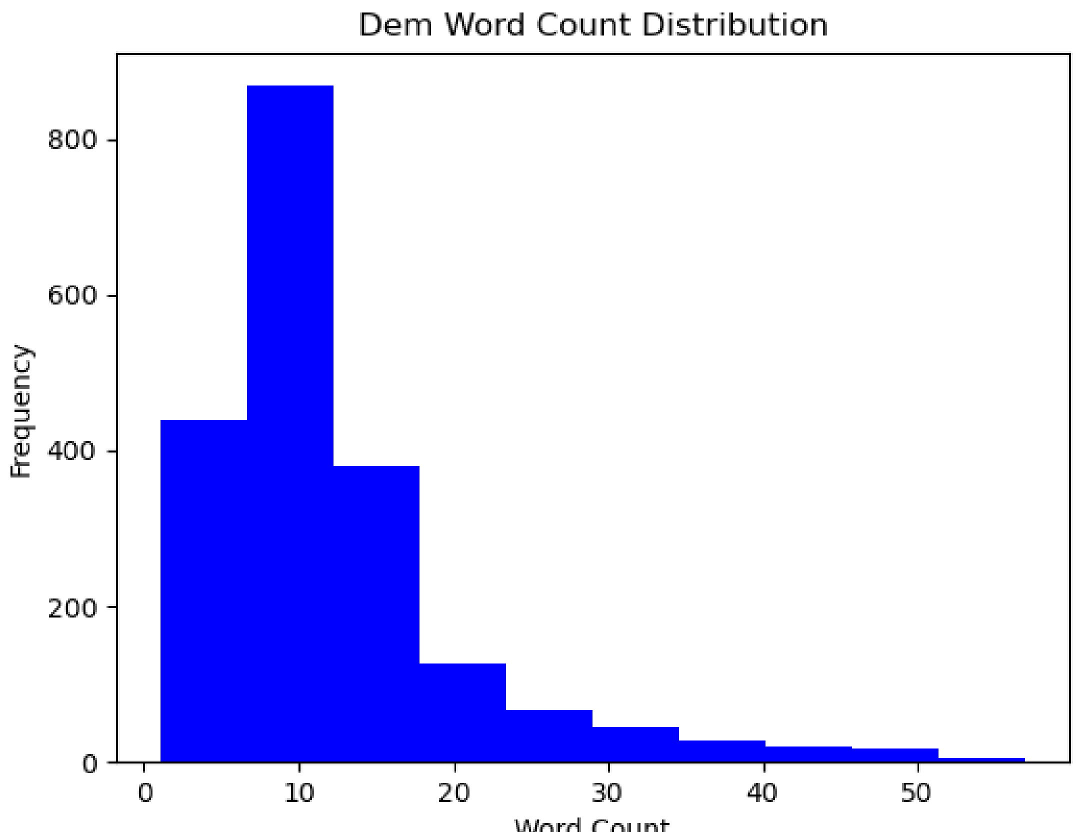
Took data from Republican and Democrat subreddits.

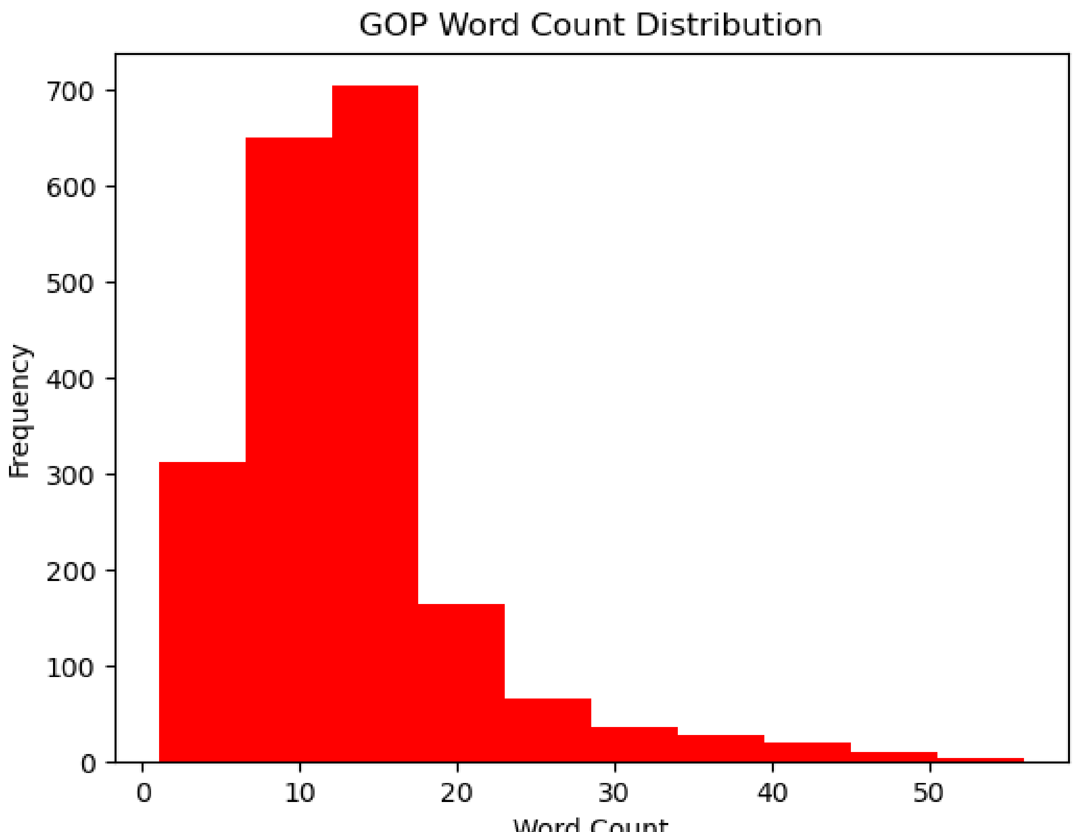
Spliced all subreddit posting titles by words.

Conducted word analysis and deployed models for prediction.

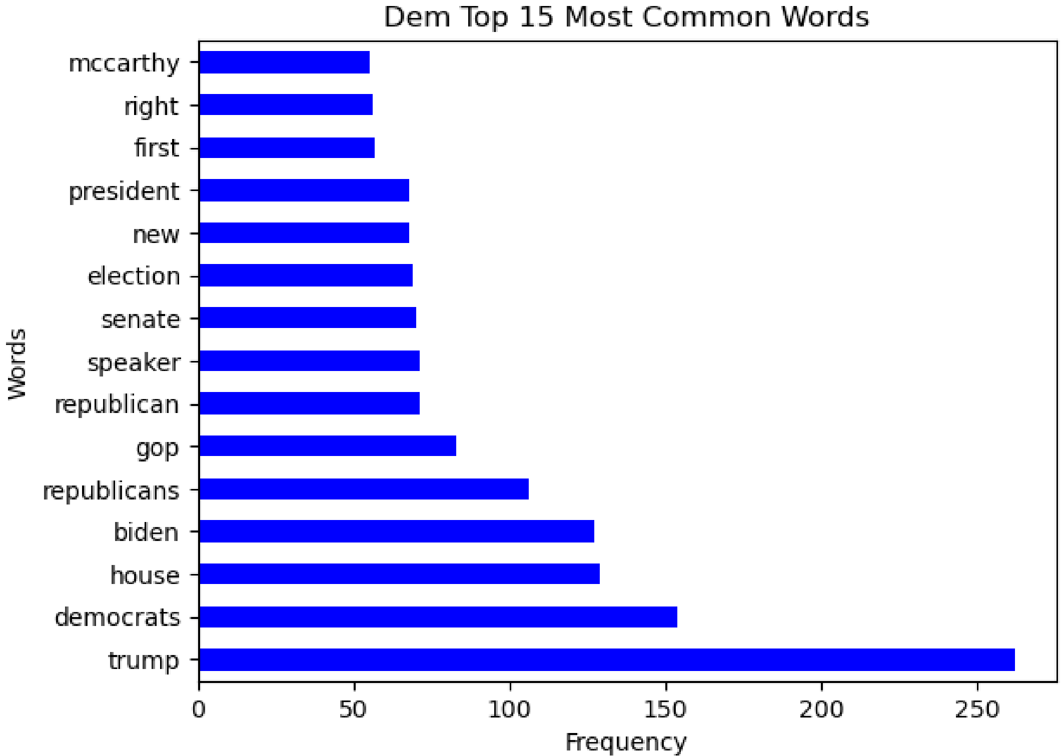
Subreddit Information

subreddit	# Posts	Date/Time	Avg. Words/Post
r/democrats	2,000	11/15/22 - 1/9/23	12.2
r/republican	2,000	11/13/22 - 1/9/23	12.9

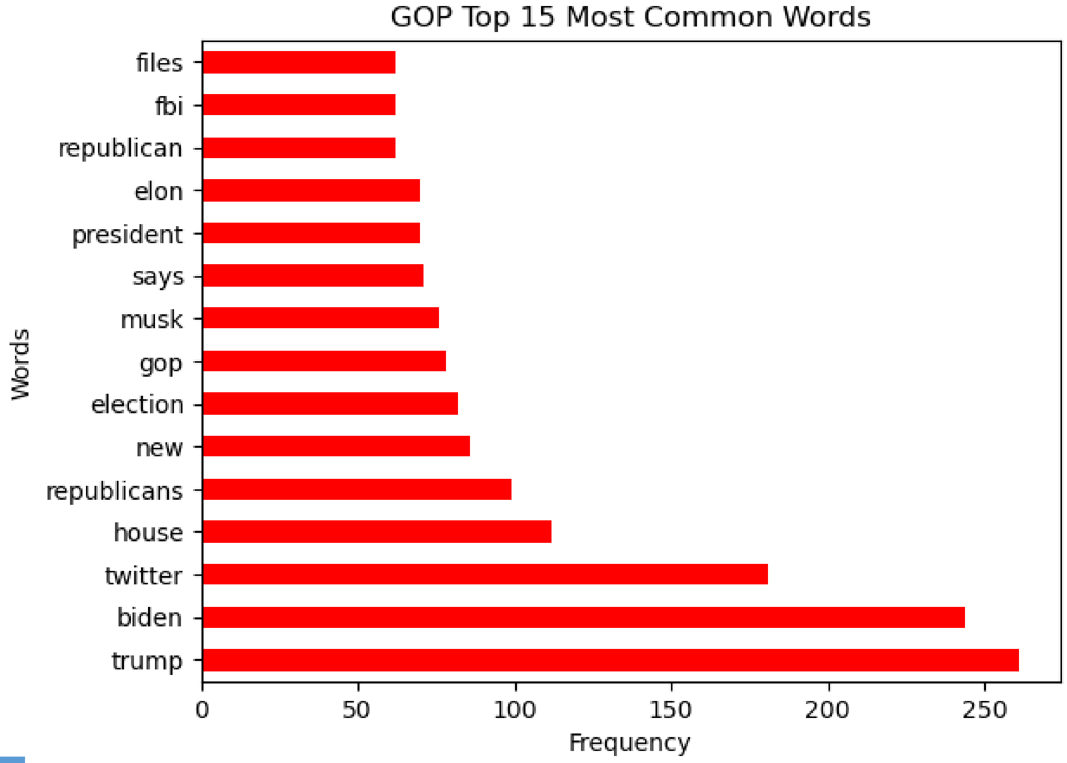




Most Common Words by Party



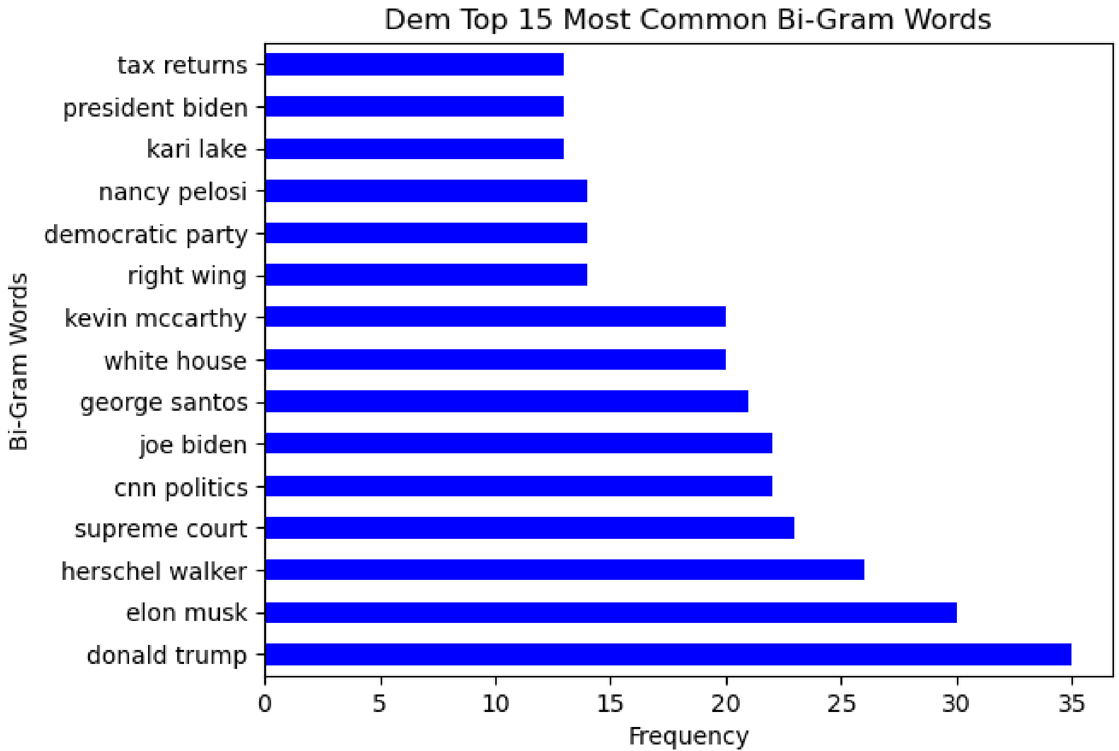
Dem Top 15 Most Common Words

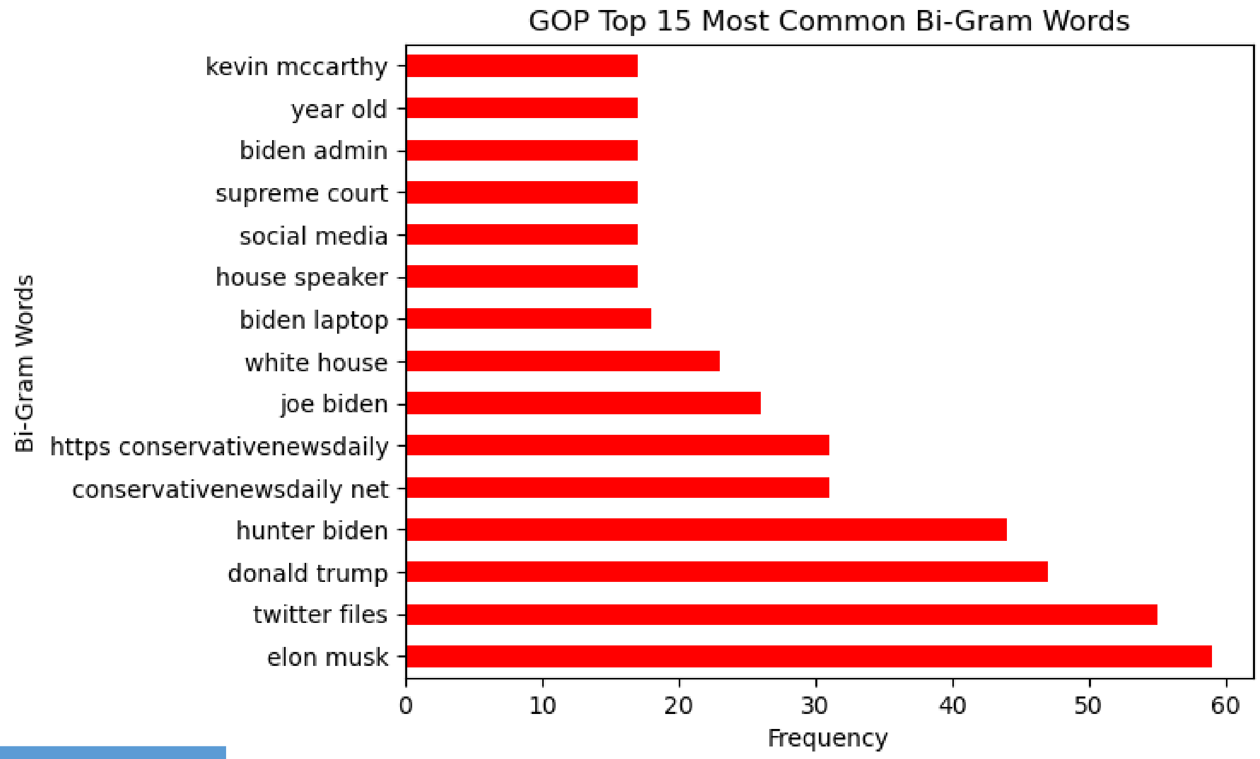


GOP Top 15 Most Common Words

Overlap Words
trump
biden
house
gop

Most Common Bi-Gram Words by Party


Dem Top 15 Most Common Bi-Gram Words


GOP Top 15 Most Common Bi-Gram Words

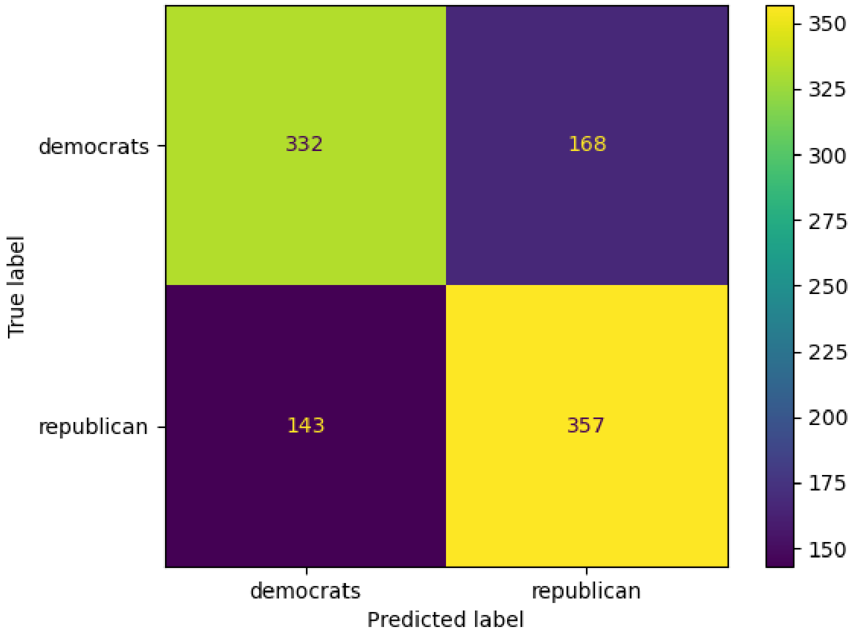
Overlap Bi-Grams
donald trump
elon musk
joe biden
white house

Model Performance

Model	Test Score
Naïve Bayes	.69
Stacked (NB, RF, LR)	.66
KNeighborsClassifier	.62
ADA Boost – Random Forest	.61
Random Forest	.61
Logistic Regression	.61

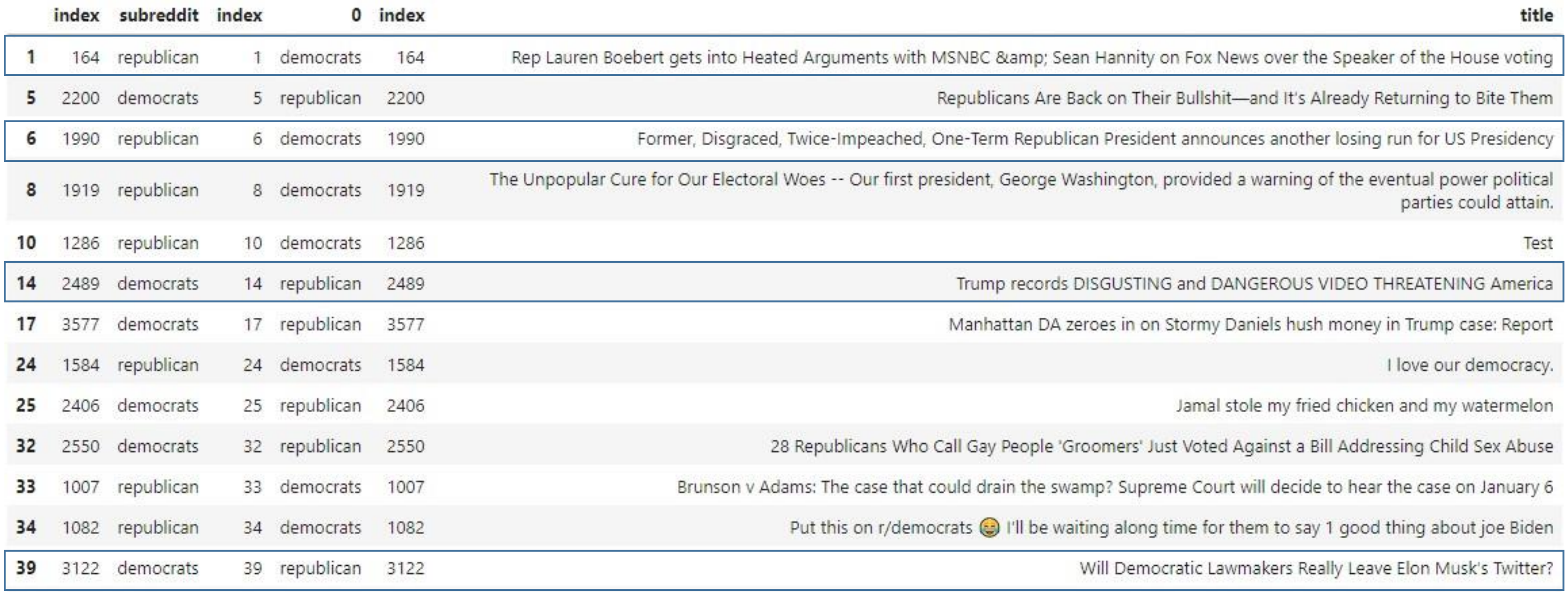
Naïve Bayes Parameters	
vectorizer	TF-IDF
nb__alpha	2
tvec__stop_words	'english'
tvec__max_features	None
tvec__min_df	0
tvec__max_df	0.3
tvec__ngram_range	(1,2)

Confusion Matrix



Metrics	
Accuracy	69%
Recall	71%
Precision	68%

Misclassification Analysis



	index	subreddit	index	0	index	title
1	164	republican	1	democrats	164	Rep Lauren Boebert gets into Heated Arguments with MSNBC &amp; Sean Hannity on Fox News over the Speaker of the House voting
5	2200	democrats	5	republican	2200	Republicans Are Back on Their Bullshit—and It's Already Returning to Bite Them
6	1990	republican	6	democrats	1990	Former, Disgraced, Twice-Impeached, One-Term Republican President announces another losing run for US Presidency
8	1919	republican	8	democrats	1919	The Unpopular Cure for Our Electoral Woes -- Our first president, George Washington, provided a warning of the eventual power political parties could attain.
10	1286	republican	10	democrats	1286	Test
14	2489	democrats	14	republican	2489	Trump records DISGUSTING and DANGEROUS VIDEO THREATENING America
17	3577	democrats	17	republican	3577	Manhattan DA zeroes in on Stormy Daniels hush money in Trump case: Report
24	1584	republican	24	democrats	1584	I love our democracy.
25	2406	democrats	25	republican	2406	Jamal stole my fried chicken and my watermelon
32	2550	democrats	32	republican	2550	28 Republicans Who Call Gay People 'Groomers' Just Voted Against a Bill Addressing Child Sex Abuse
33	1007	republican	33	democrats	1007	Brunson v Adams: The case that could drain the swamp? Supreme Court will decide to hear the case on January 6
34	1082	republican	34	democrats	1082	Put this on r/democrats 😂 I'll be waiting along time for them to say 1 good thing about joe Biden
39	3122	democrats	39	republican	3122	Will Democratic Lawmakers Really Leave Elon Musk's Twitter?

Model has trouble predicting with the similar words and is unable to interpret tone or sentiment.

Conclusion and Next Steps

- Conclusion
    - Naïve Bayes Model score of 69% outperforms the baseline of 50% and is effective at predicting political affiliation.
    - Both Democrats and Republican commonly used words are based on negativity.
- Applications
    - Monitoring the magnitude of news/disinformation along party lines.
- Next Steps
    - Naïve Bayes model was overfit with a 96% training score. Further exploration into reducing variation is needed.
    - Incorporate sentiment into the models.
    - Determine feature importance.