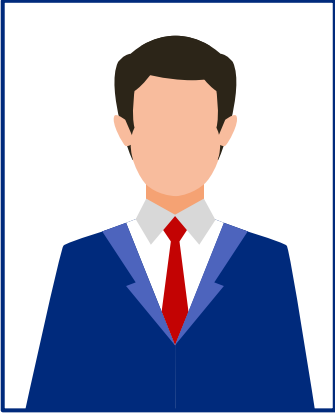# Republican v. Democrat

Subreddit NLP Analysis

Agenda

- Introduction & Problem Statement
- Data Summary
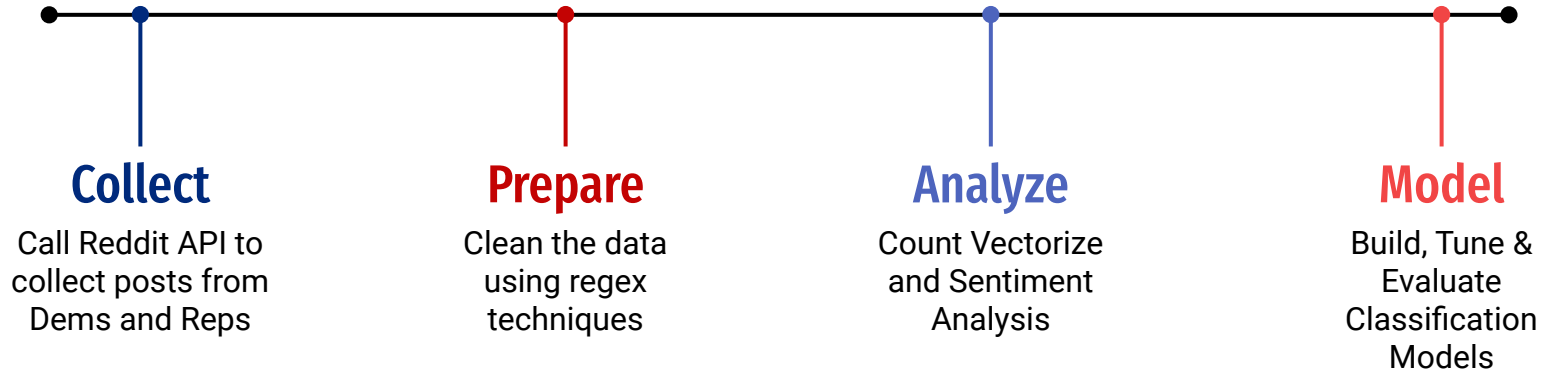- Exploratory Data Analysis
- Classification Model Results

Introduction & Problem Statement

I am a Data Scientist for the 2024 Presidential Campaign. My role is to collect and analyze data in order to identify what topics Democrats and Republicans are most concerned with so that we can formulate a winning campaign strategy.

We would also like to build a model that can identify whether someone is Democrat or Republican based on posts that they make to social media platforms such as Reddit.

Collect
Call Reddit API to collect posts from Dems and Reps

Prepare
Clean the data using regex techniques

Analyze
Count Vectorize and Sentiment Analysis

Model
Build, Tune & Evaluate Classification Models

Data Summary

Democrat

Subscribers: 429k
Submissions: 9989
Posts: 3408
Avg Word Count: 15
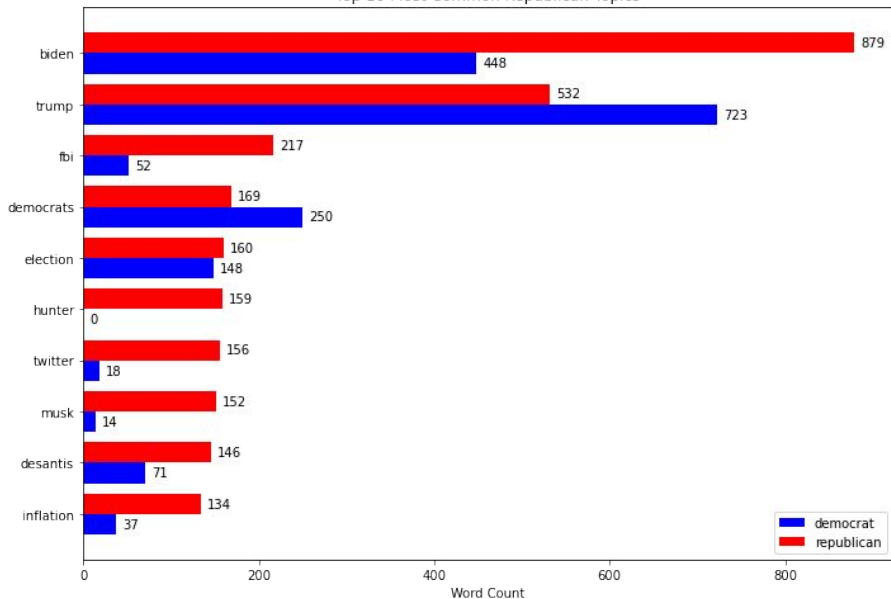Unique Authors: 591

55%

45%
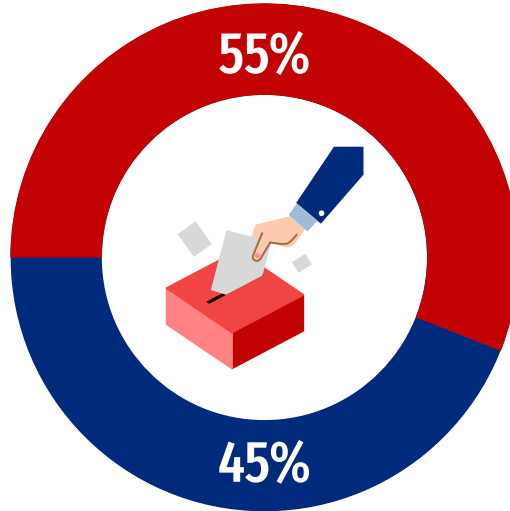
Republican

Subscribers: 182k
Submissions: 9999
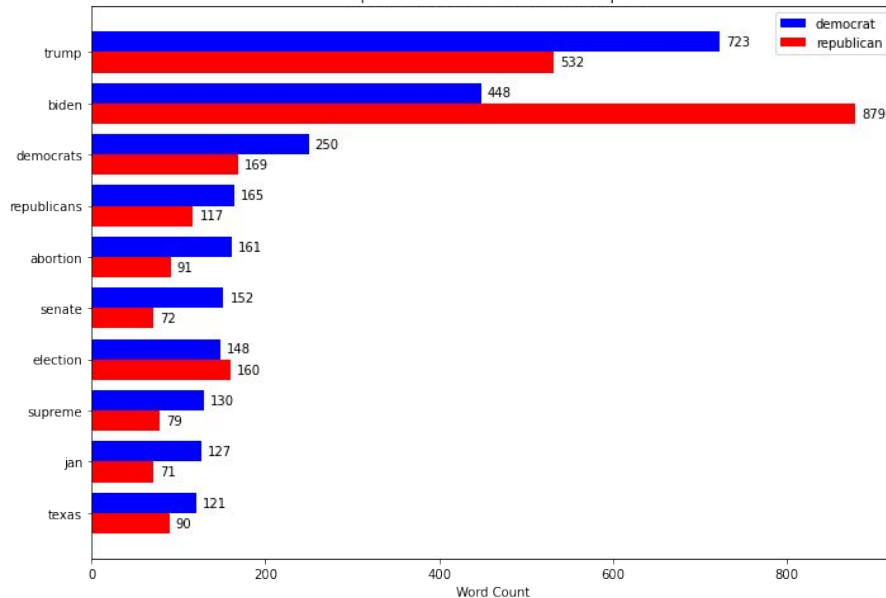Posts: 4173
Avg Word Count: 15
Unique Authors: 431
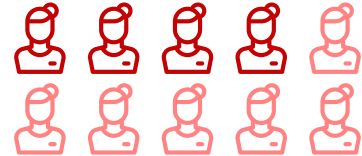
Most Popular Topics

Key Takeaway: There is a lot of overlap between what Democrats and Republicans are talking about which will make classifying between these groups quite challenging.
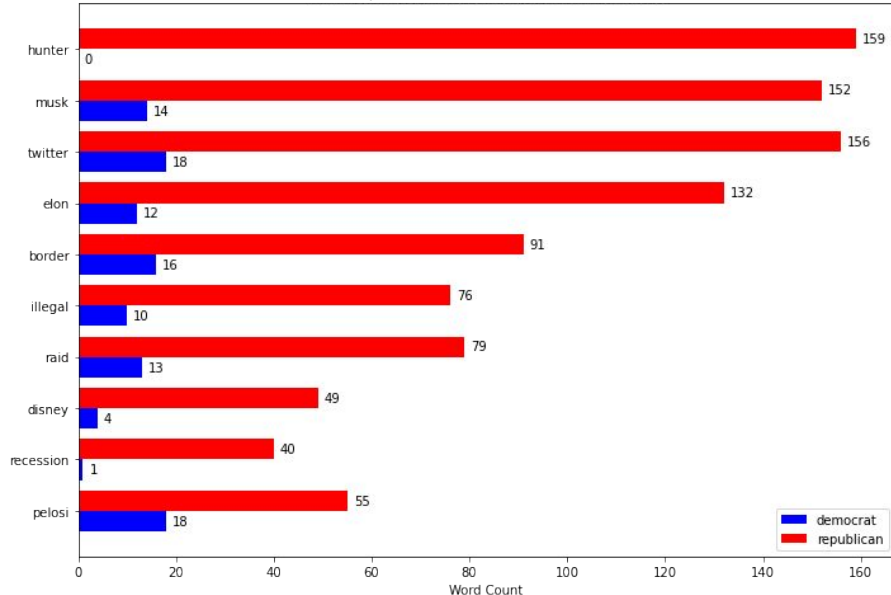
Republicans

Key Takeaway: If we win the primary it may be advantageous to identify which issues one party is concerned with that the other doesn't care much about. We can play to these topics in order to target voters from the otherside without losing our base.

What Republicans Talk about that Dems Do Not

Democrats

Model Evaluation

	train_scores	test_scores	diff	baseline
logreg	0.919789	0.716245	0.203544	0.55
rf	0.998065	0.747363	0.250702	0.55
ada	0.777485	0.685127	0.092358	0.55
gb	0.875462	0.698840	0.176622	0.55
stack	0.992612	0.733122	0.259490	0.55

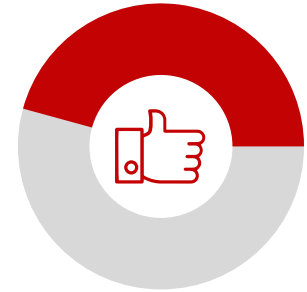|	sensitivity	specificty
logreg	0.743295	0.683099
rf	0.792146	0.692488
ada	0.768199	0.583333
gb	0.774904	0.605634
stack	0.771073	0.686620

Random Forest

Conclusions

Campaign Topics

-Top of Mind
-Overlapping
-Differing

Party Classifier

Random Forest
-75% Accuracy
-79% Sensitivity