

# Political Alignment Classification

**Using Reddit Comments**



# PROBLEM STATEMENT

- Can machine learning predict the political alignment of Reddit posts and users?
- This model aims to help political organizations deliver targeted ads by analyzing online discourse.



# Data Collection



- r/AskALiberal
- r/AskConservatives
- r/AskPolitics
- r/AskTrumpSupporters



# User Flairs

Select your community flair

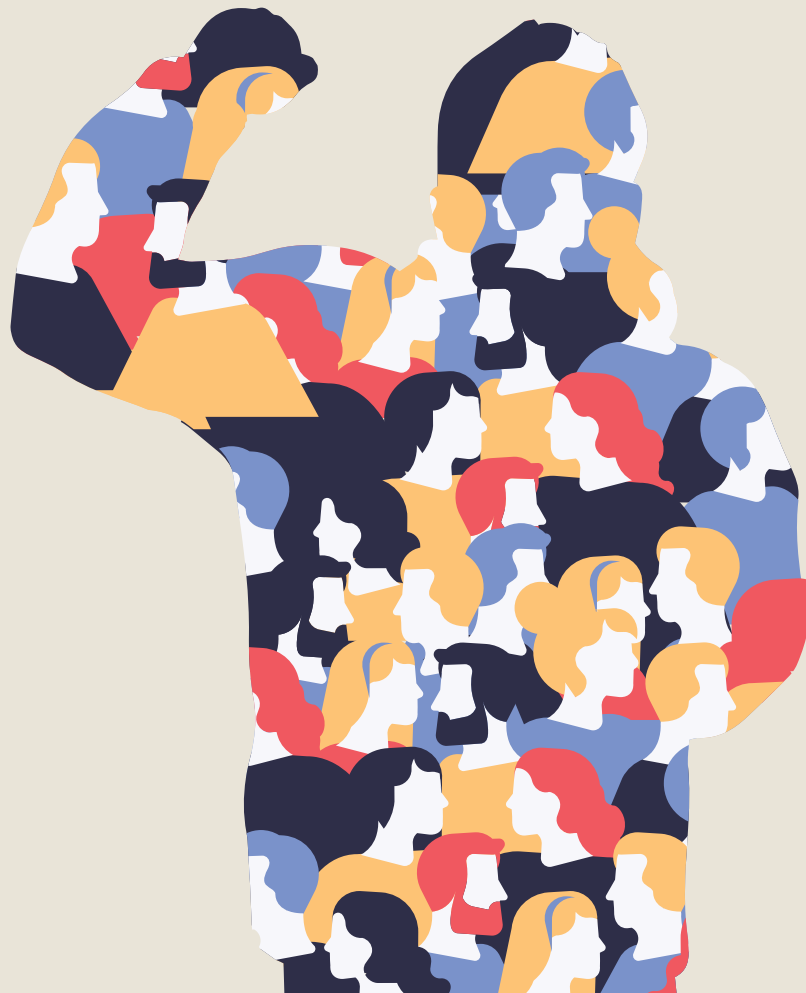
- ☐ Left Libertarian
- ☐ Democrat
- ☐ Liberal
- ☐ Leftwing
- ☐ Progressive
- ☐ Social Democracy
- ☐ Socialist
- ☐ Democratic Socialist
- ☐ Neoliberal
- ☐ Leftist

Select your community flair

- ☐ Conservative
  - ☐ European Conservative
  - ☐ Constitutionalist
  - ☐ Nationalist
  - ☐ Free Market
  - ☐ Social Conservative
  - ☐ Republican
  - ☐ National Minarchism
  - ☐ Rightwing
  - ☐ Monarchist
- 

## Democrat-Aligned ( 0 )    Republican-Aligned ( 1 )

Democrat	Republican
Democratic Socialist	Conservative
Far Left	Right-leaning
Liberal	Neoconservative
Progressive	Paleoconservative
Social Democrat	Nationalist
Social Liberal	Social Conservative
Neoliberal	Constitutionalist
Social Democracy	Rightwing
Leftist	Conservatarian
Pragmatic Progressive	Tea Party
Leftwing	Trump Supporter
Left-leaning	Far Right



# What defines a useful post/comment?

Rightwing

In these particular instances, they both are.

Center-left

I can agree.

Republican

None.

Liberal

Sneeches get steeches

Short posts often lack sufficient context and meaningful information...



### Conservative

That's not how it works. The constitution is a higher authority, and it does not give congress any authority to regulate education. Because of the tenth amendment, that authority is delegated to the state government.

To give an example from the other side of the aisle, congress can't just pass a law to force state and local police agencies to deport illegal immigrants. That's because states have the right to enforce the law how they want, and congress has no authority to compel them to do so. Exactly the same with education.

### Progressive

I'm glad to see someone who did probably disagree with on a lot of issues have a reasonable take on DEI lol. Nail on the head there. Personally, I think our body of representatives and judiciary should reflect our population (I dream of the day it's 50/50 men and women lol) because it is better representation.

I wish all the dumb money that's spent bribing government people could be funneled in to programs getting younger people interested and educated on becoming a part of our government at a higher level. Idk what that program would look like or if it would even make sense. It's just a fun hypothetical. And also might mean less geriatrics making decisions for us all

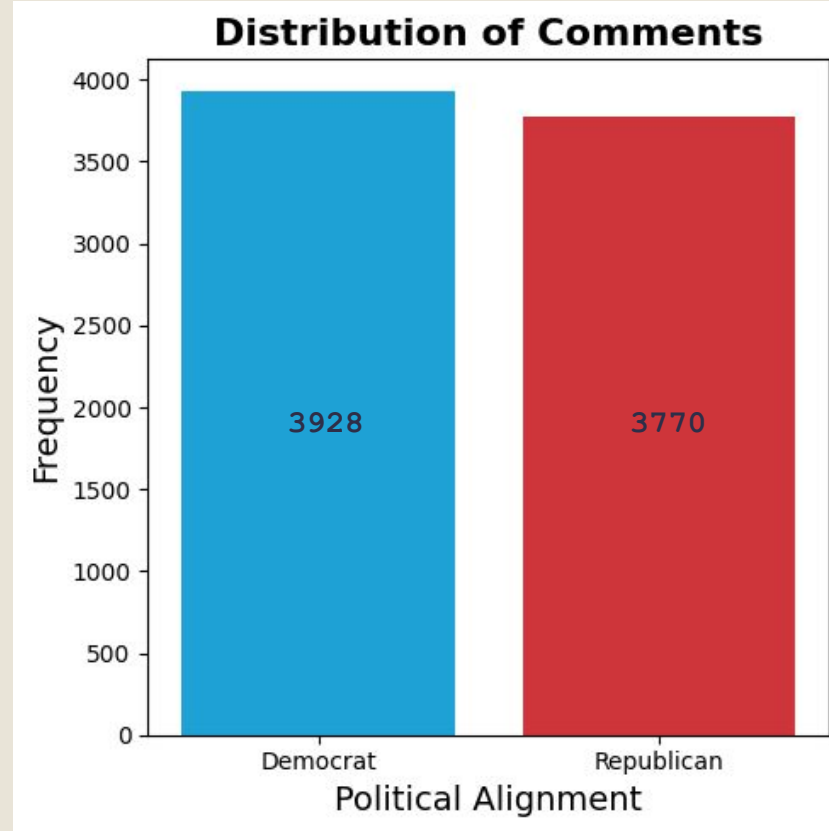
**Longer posts/comments are typically more detailed and better reflect a user's opinion.**

## Collected Entries

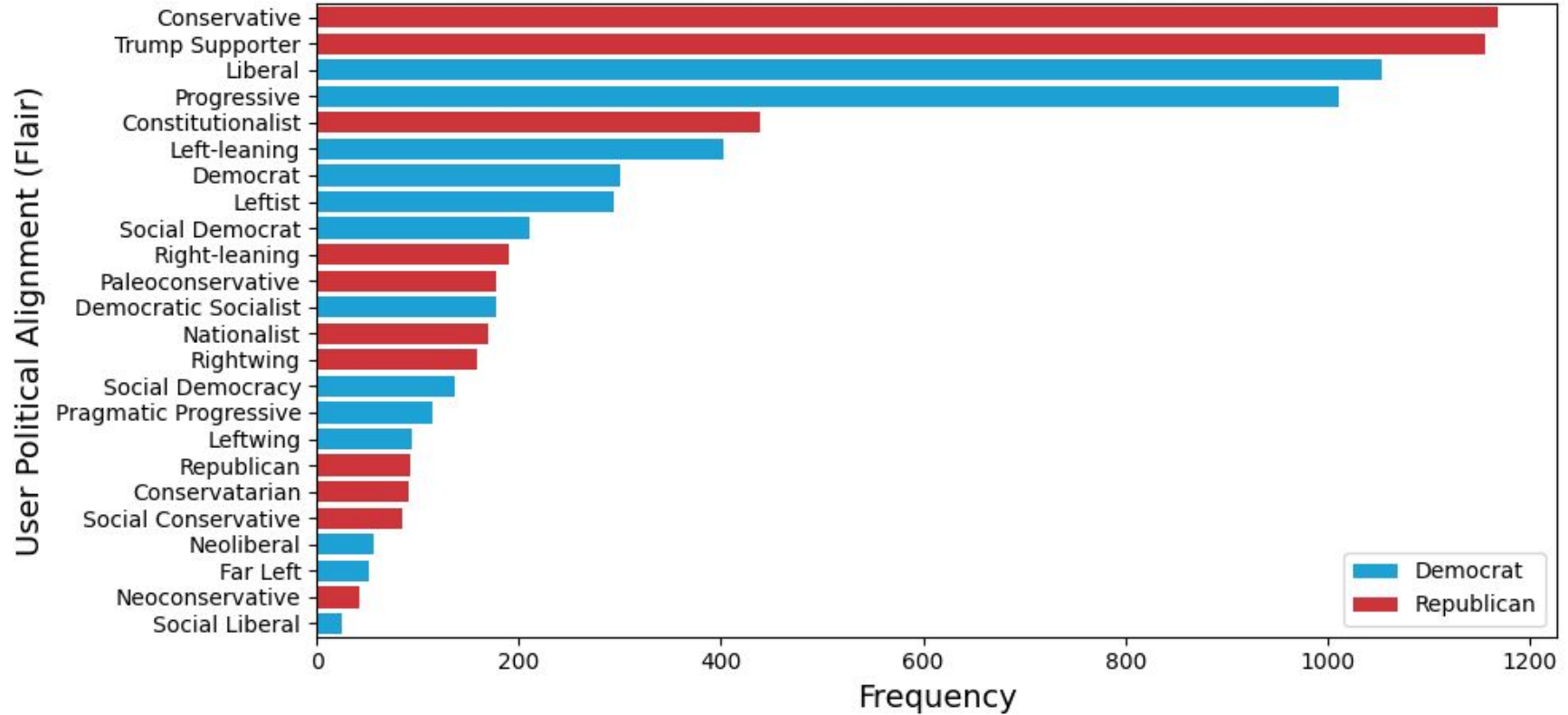
- Collected highly upvoted entries.
- Filtered entries with a minimum of 400 characters.
- Republican-aligned posts were harder to find, so the length requirement was lowered to 250 characters.



Democrat 3928  
Republican 3770



## Distribution of User Political Alignment by Flair

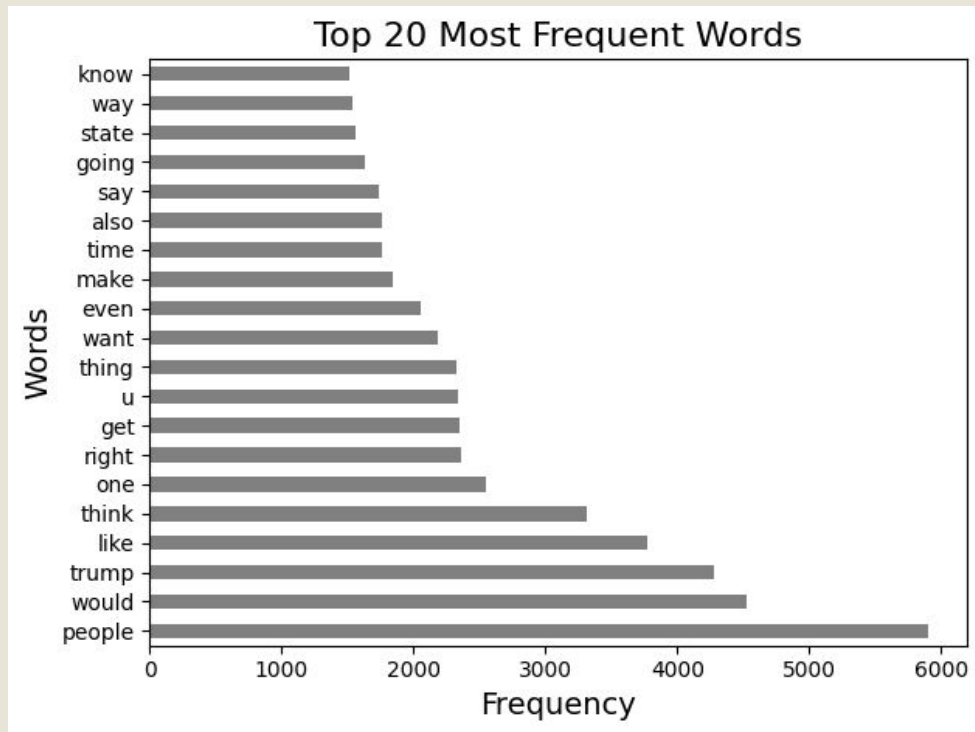


# Natural Language Processing

- **Text Normalization:** Converted text to lowercase, removed URLs, mentions, quote blocks, hashtags, numbers, contractions, and non-alphabetical characters.
- **Stop Word Removal:** Removed common words that don't carry meaningful information (e.g., "the", "is", "and").
- **Tokenization:** Split text into individual words or tokens.
- **Lemmatization:** Reduced words to their base or dictionary form.
- **N-grams:** Created bigrams and trigrams to capture word combinations and context.

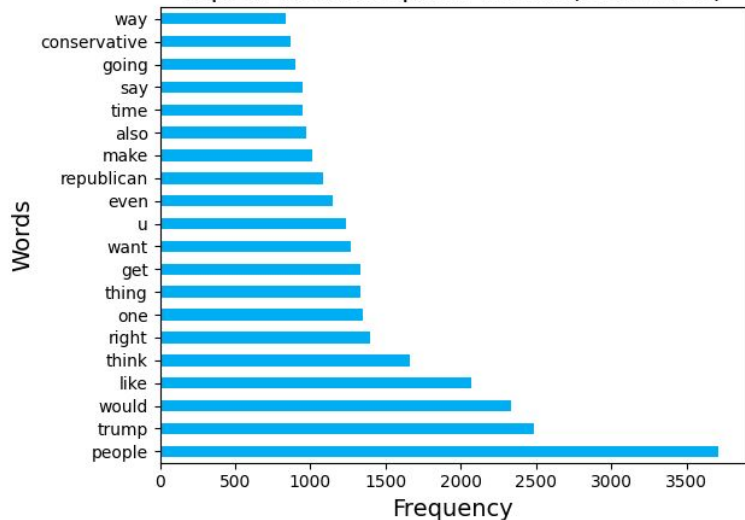


# Most Common Words

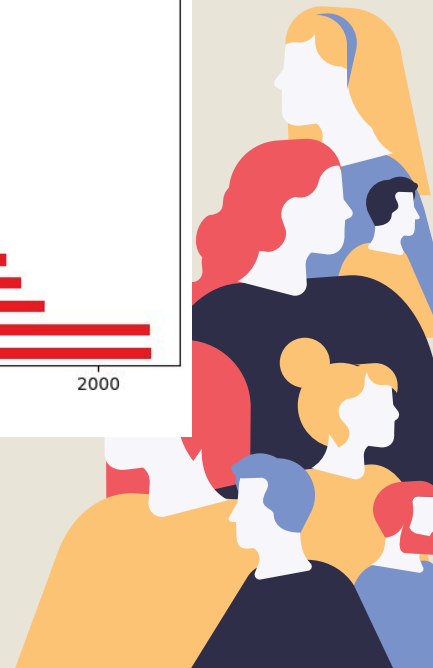
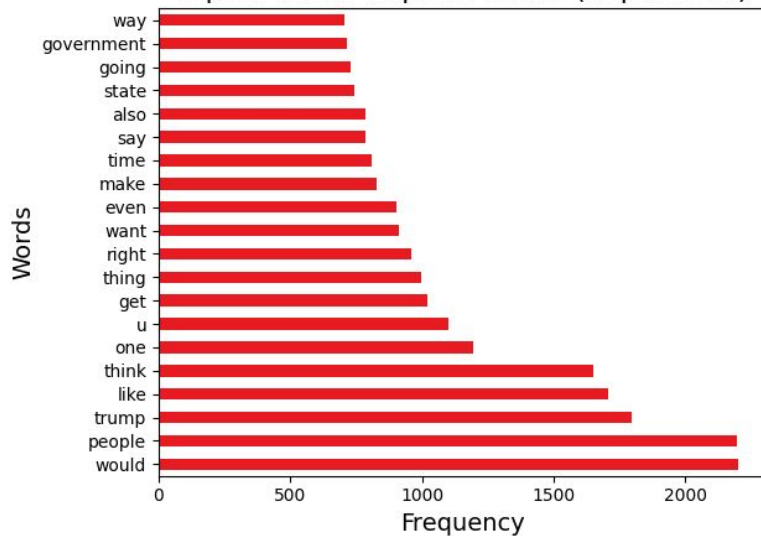


# Most Common Words

Top 20 Most Frequent Words (Democrat)

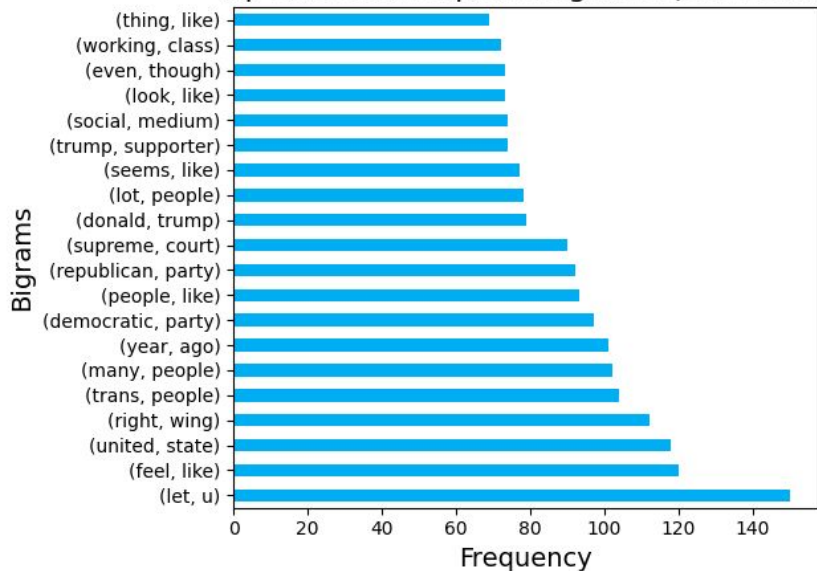


Top 20 Most Frequent Words (Republican)

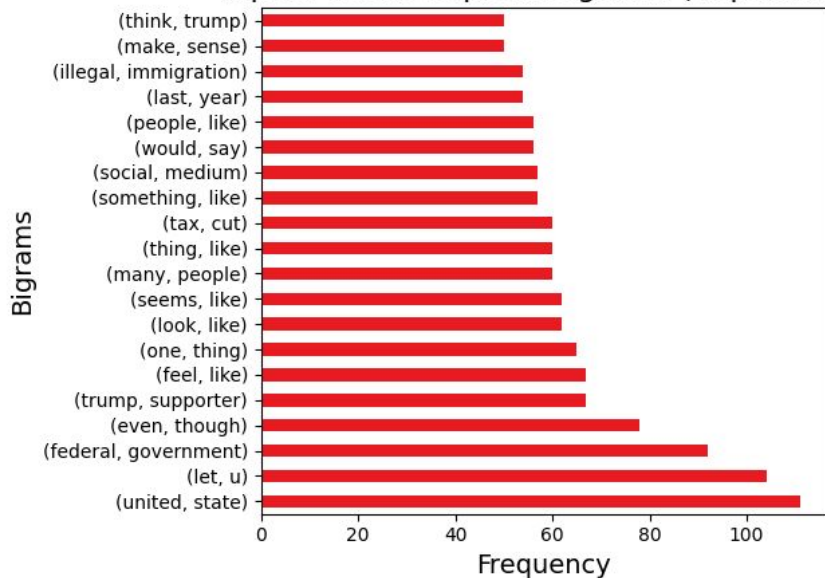


# Most Common Words

Top 20 Most Frequent Bigrams (Democrat)

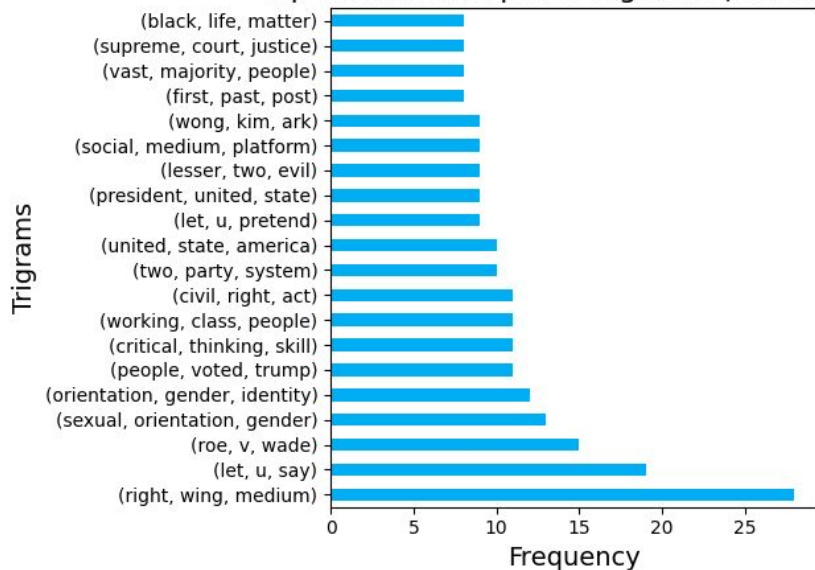


Top 20 Most Frequent Bigrams (Republican)

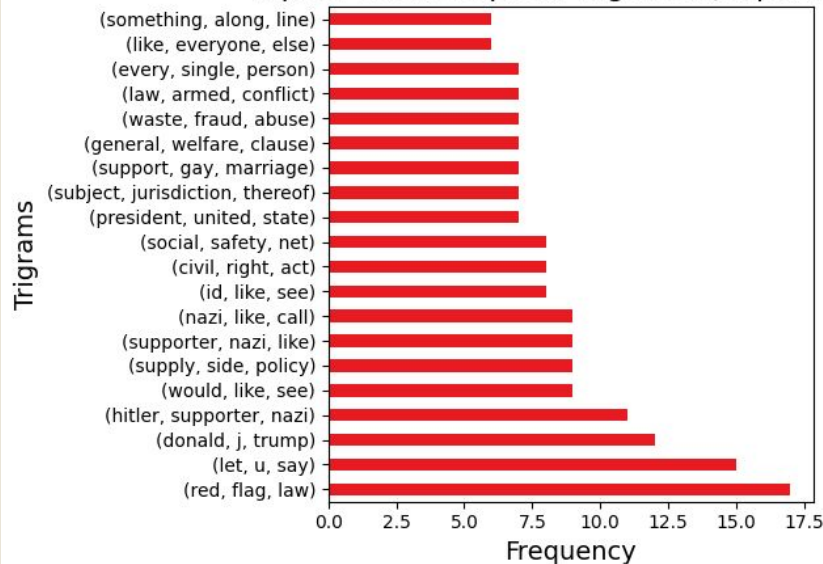


# Most Common Words

## Top 20 Most Frequent Trigrams (Democrat)



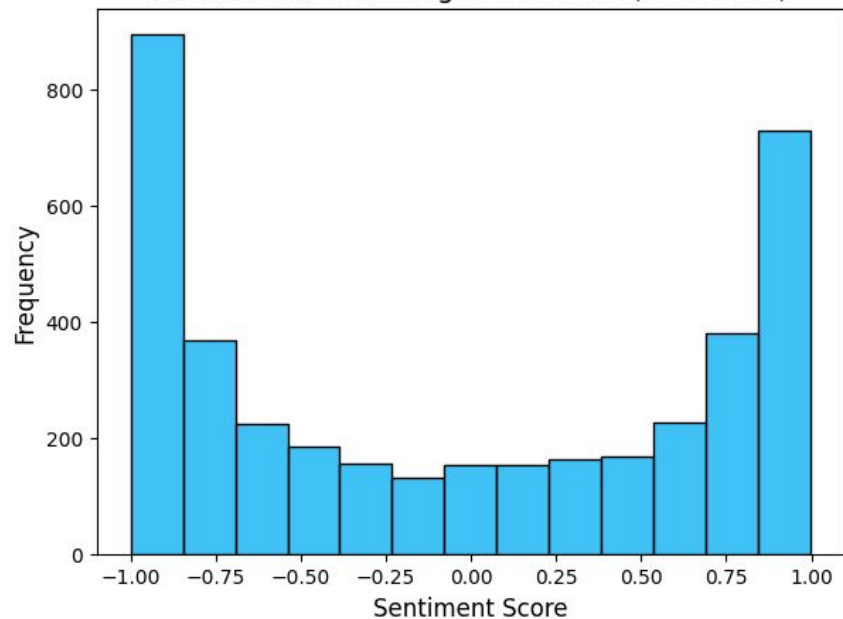
## Top 20 Most Frequent Trigrams (Republican)



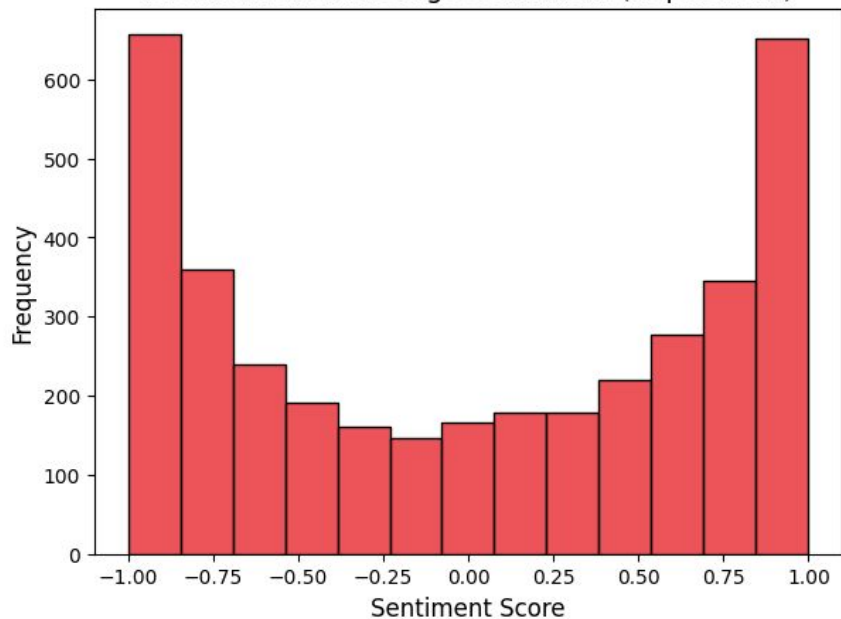
# Sentiment Analysis

- VADER Sentiment Analysis

Distribution of Average Sentiment (Democrat)

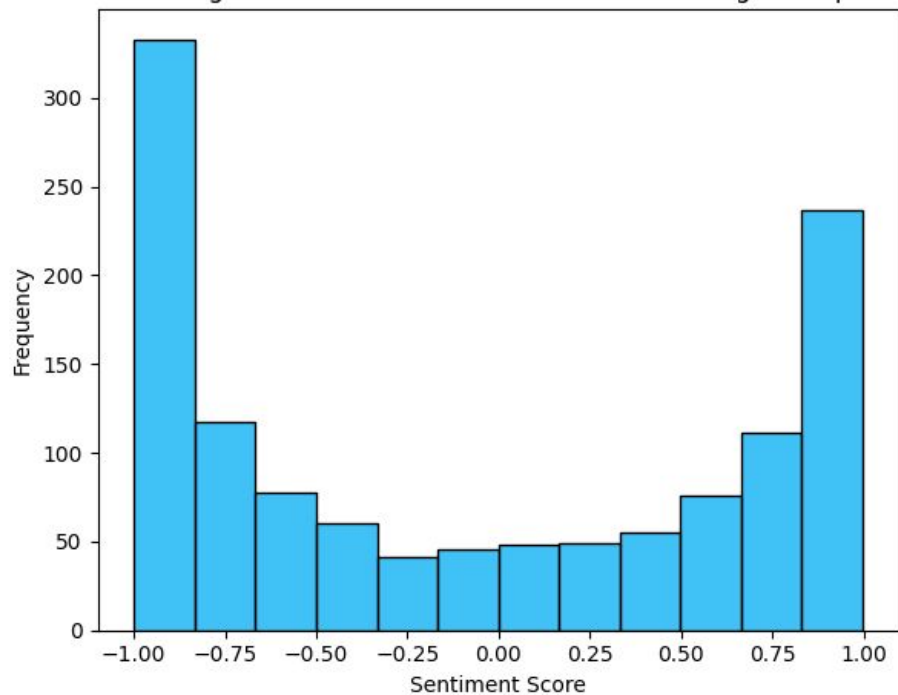


Distribution of Average Sentiment (Republican)

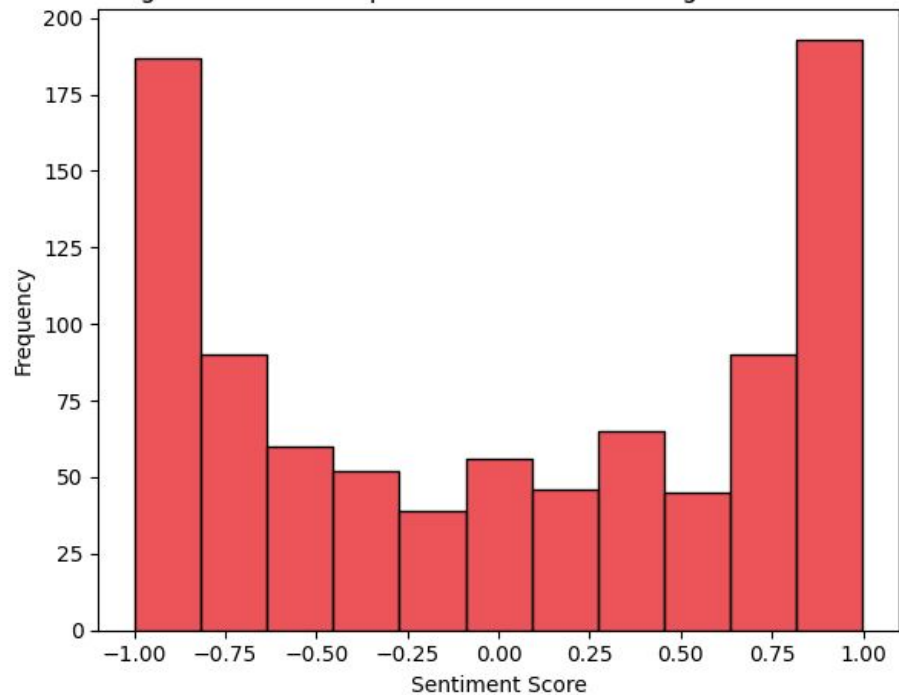




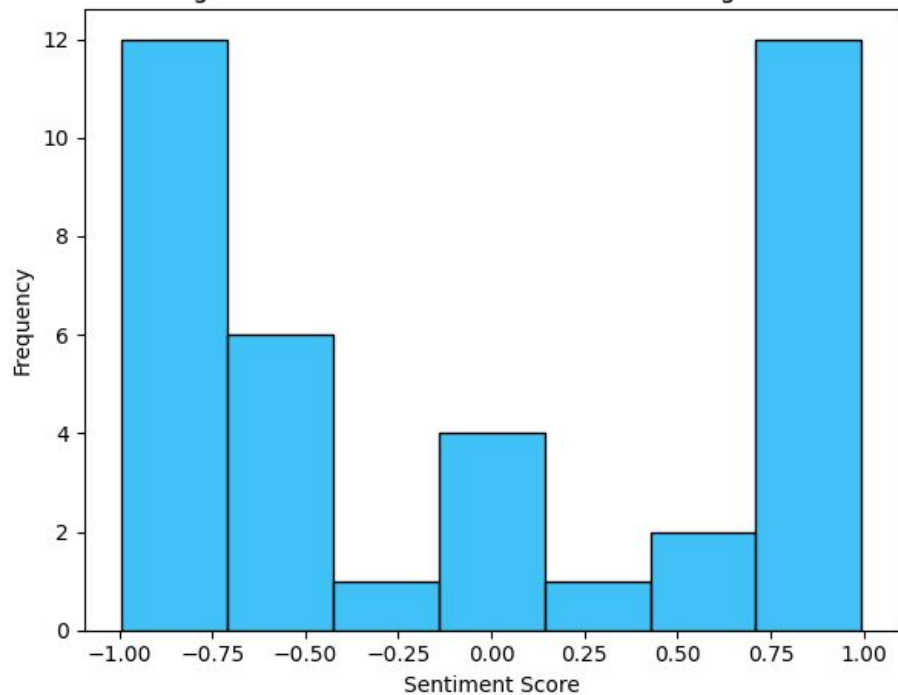
Average Sentiment - Democrat Posts Containing "Trump"



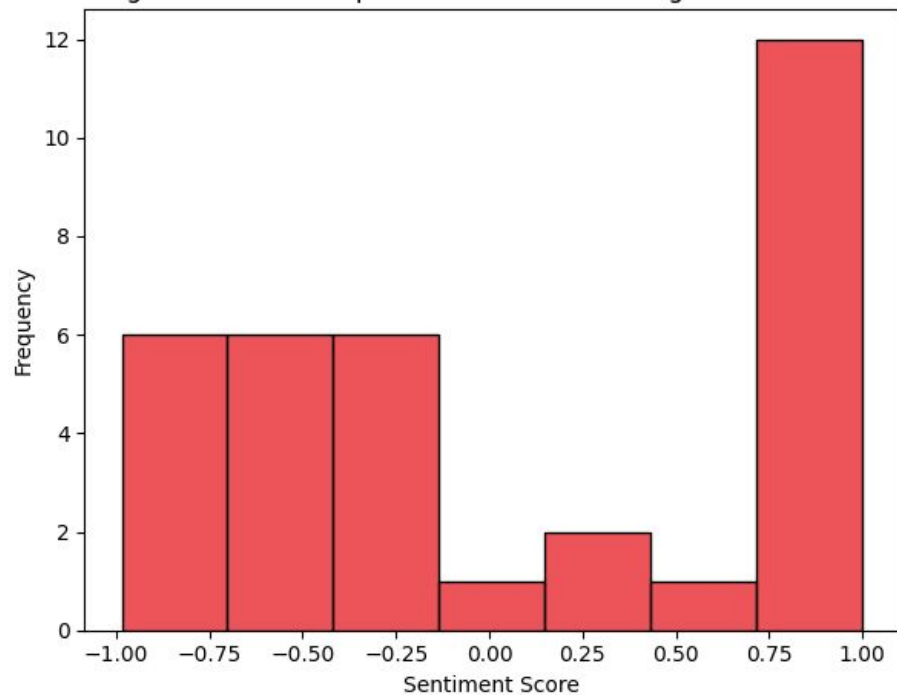
Average Sentiment - Republican Posts Containing the word "Trump"



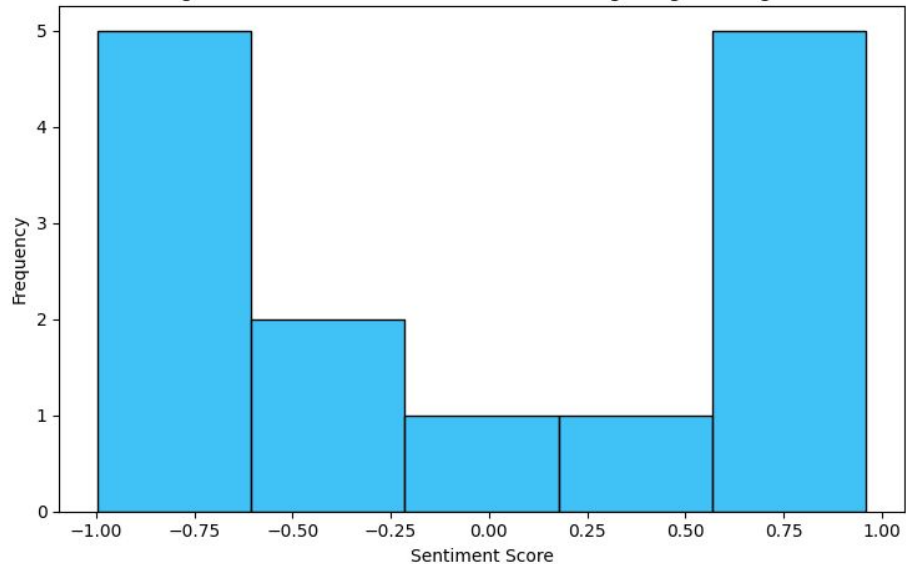
Average Sentiment - Democrat Posts Containing "Tax Cut"



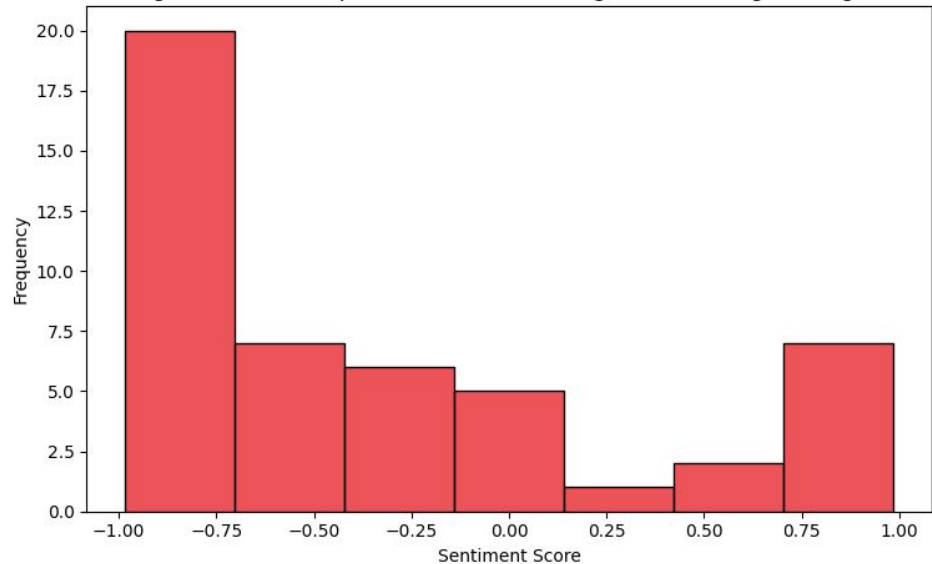
Average Sentiment - Republican Posts Containing the word "Tax Cut"



Average Sentiment - Democrat Posts Containing "Illegal Immigration"



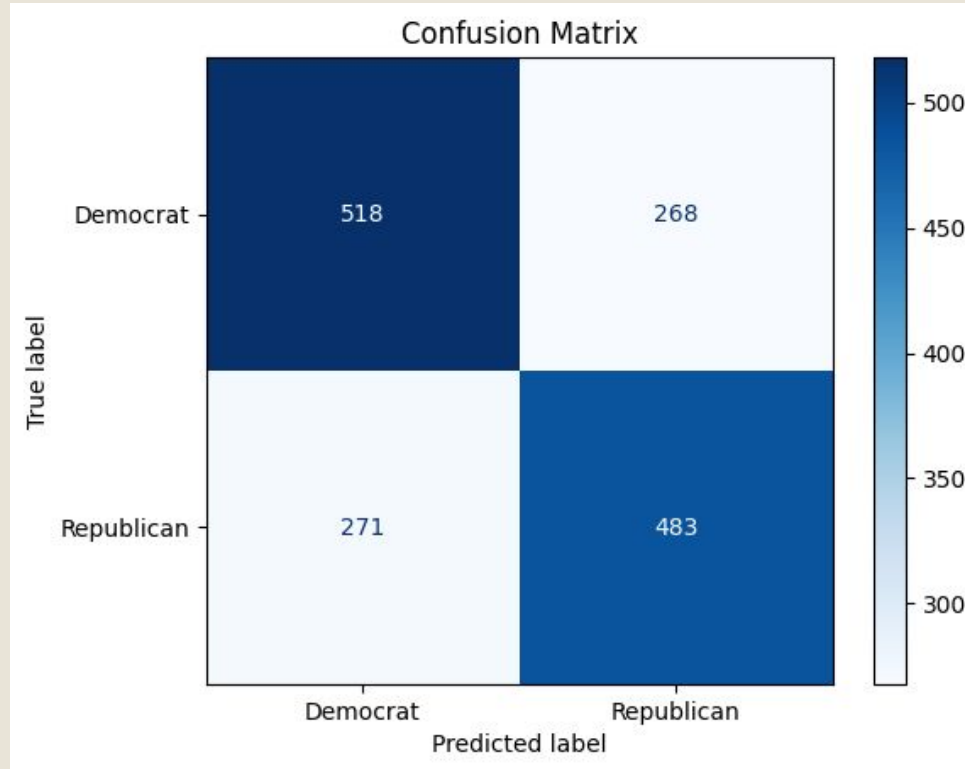
Average Sentiment - Republican Posts Containing the word "Illegal Immigration"



# Modelling

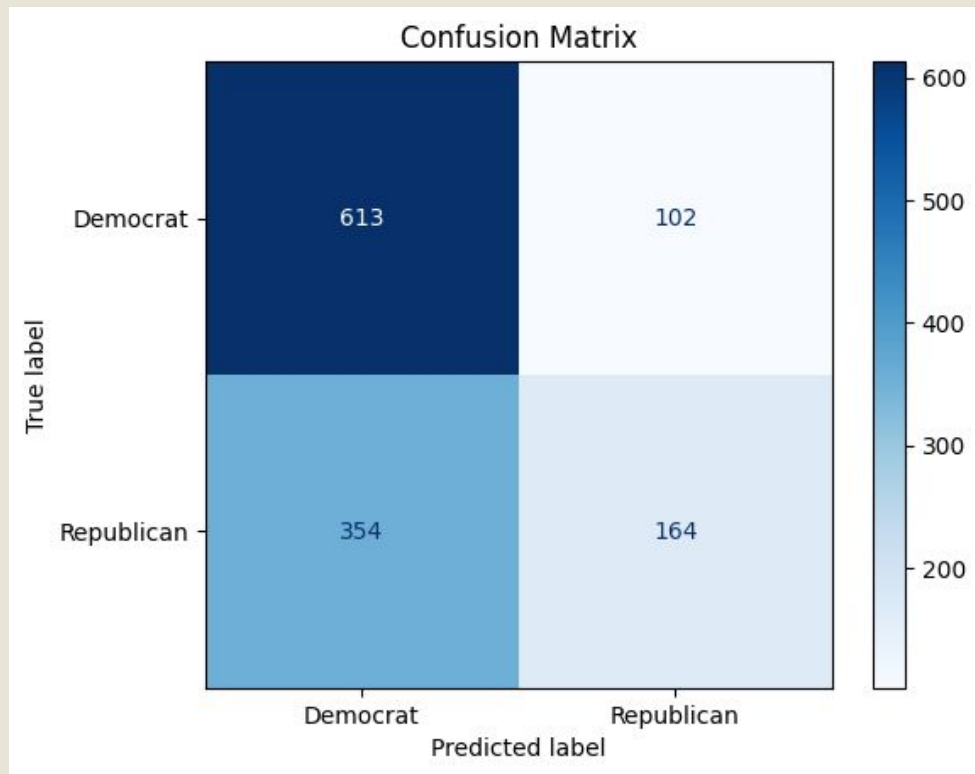


# Logistic Regression



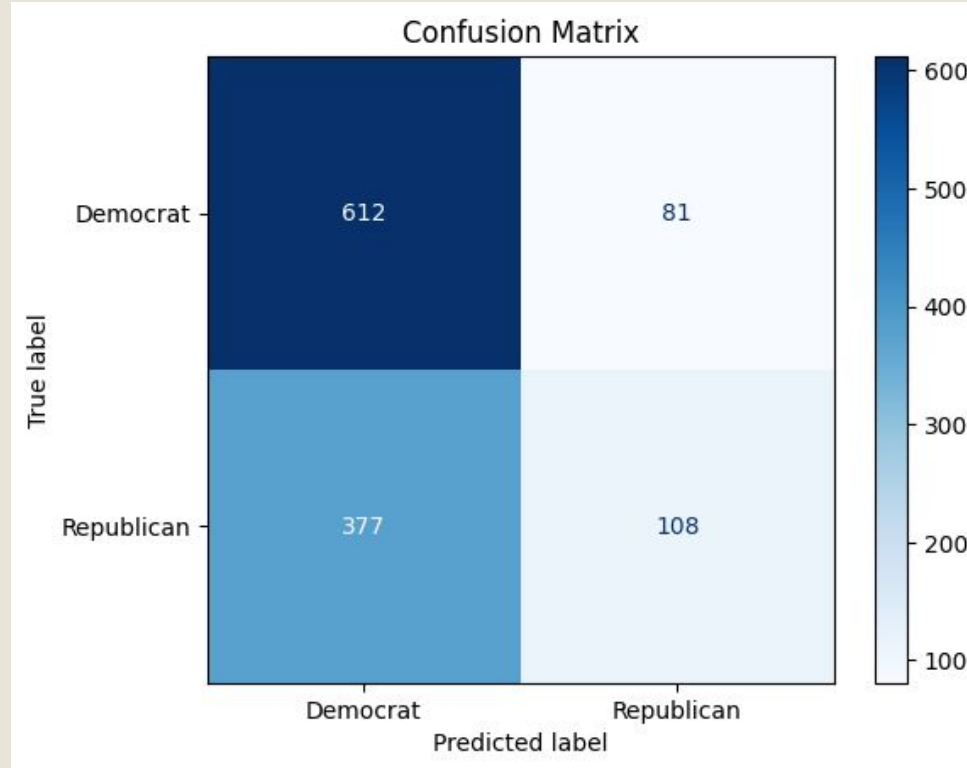
Accuracy: 0.66

# Linear Support Vector Classification



Accuracy: 0.63

# Gradient Boosting Classifier



Accuracy: 0.61

## Recommendations:

- **User flairs were not a fully reliable indicator of political alignment.**
- **Political opinions are nuanced** and may vary across different topics.
- **Some users may express views that don't align with their flair** , making labels inconsistent.
- **Sentiment analysis revealed overlaps** in tone and sentiment across political groups.
- **Short posts likely hurt model performance** due to lack of context or clear opinion.
- **A cleaner, manually labeled dataset** would improve classification accuracy and model reliability.

