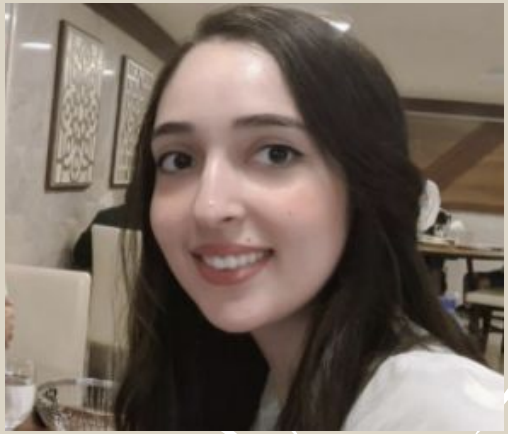
U.S.A.

Debating Debates

By: Lara Fares, Chris Rattigan, JJ Reyes

Who are we?

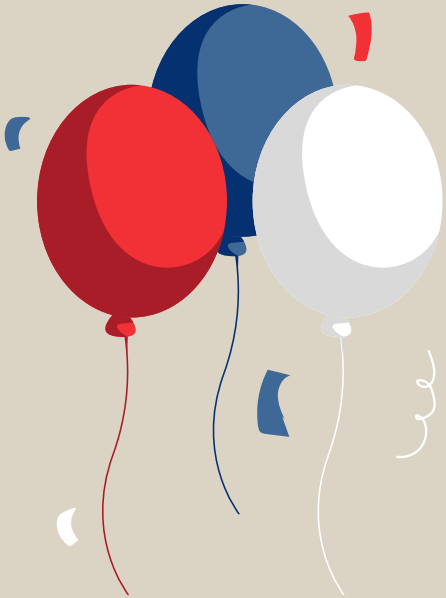
Lara Fares



Chris Ratigan

JJ Reyes

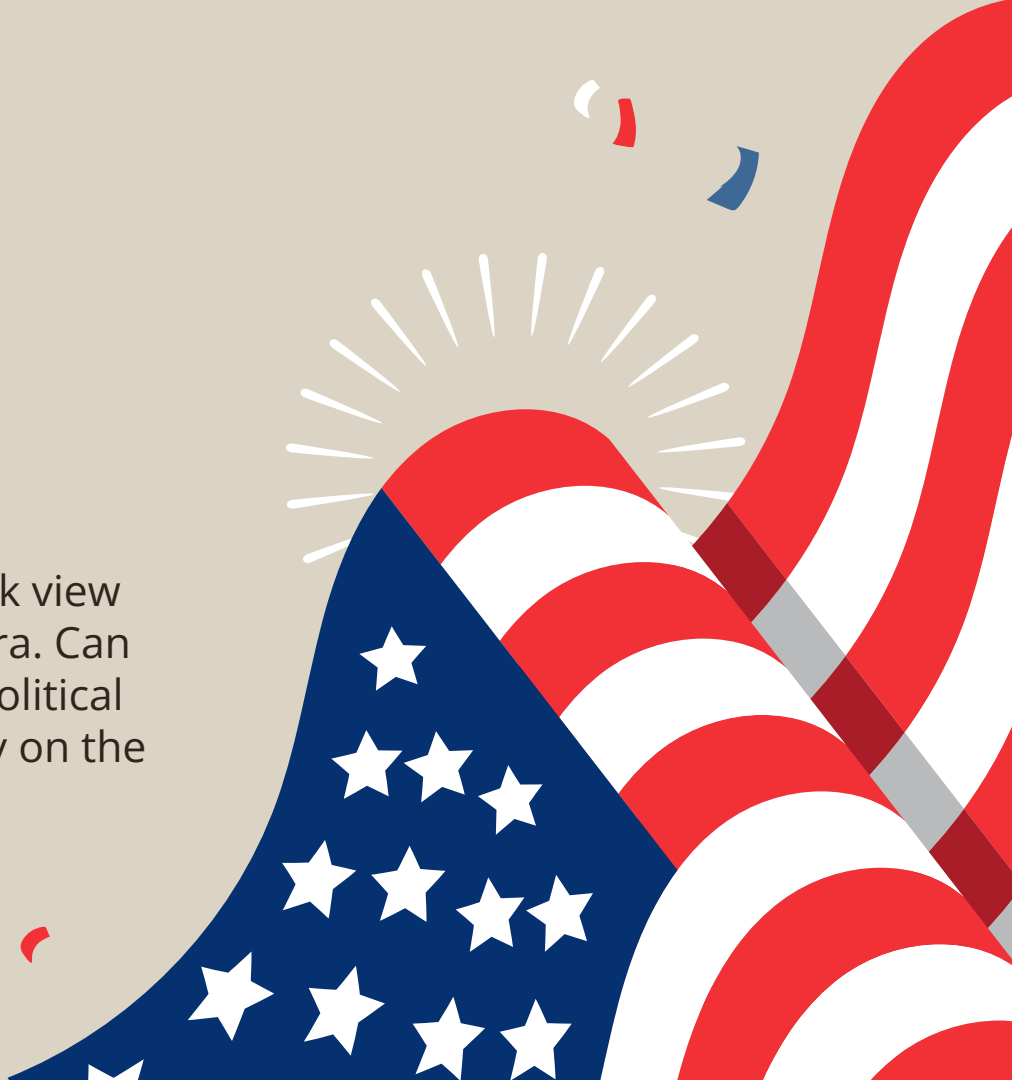
Presentation Overview



1. Raw Data
2. Data Cleaning
3. EDA
4. Models and Results
5. Conclusion

Problem Statement

Presidential debates provide a quick view into the zeitgeist of any particular era. Can we classify the candidates and/or political parties in these debates based solely on the transcripts?



Data Info

https://www.debates.org/

Unique Texts: Over 9000

1960 to 2020

45 Urls Scraped

PARTICIPANTS:
Former Vice President Joe Biden (D) and
President Donald Trump (R)

MODERATOR:
Chris Wallace (Fox News)

WALLACE: Good evening from the Health Education Campus of Case Western Reserve University and the Cleveland Clinic. I'm Chris Wallace of Fox News and I welcome you to the first of the 2020 presidential debates between

Data Diagram

Raw Web Scraped Data

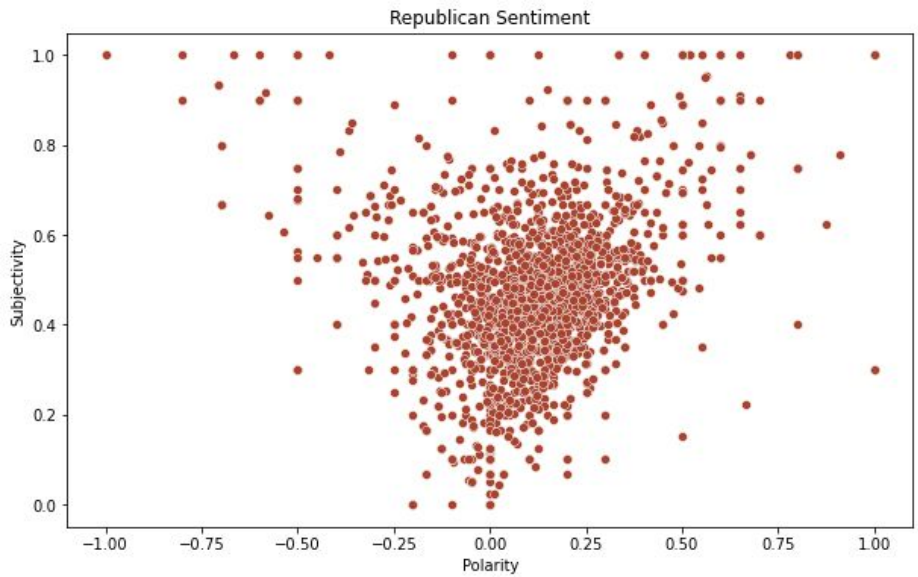
- **Speaker**
  - Moderator (removed)
  - Republican (55%)
  - Democrat (45%)
- **Debate Text**
- **Sentiment**
  - Subjectivity
  - Polarity

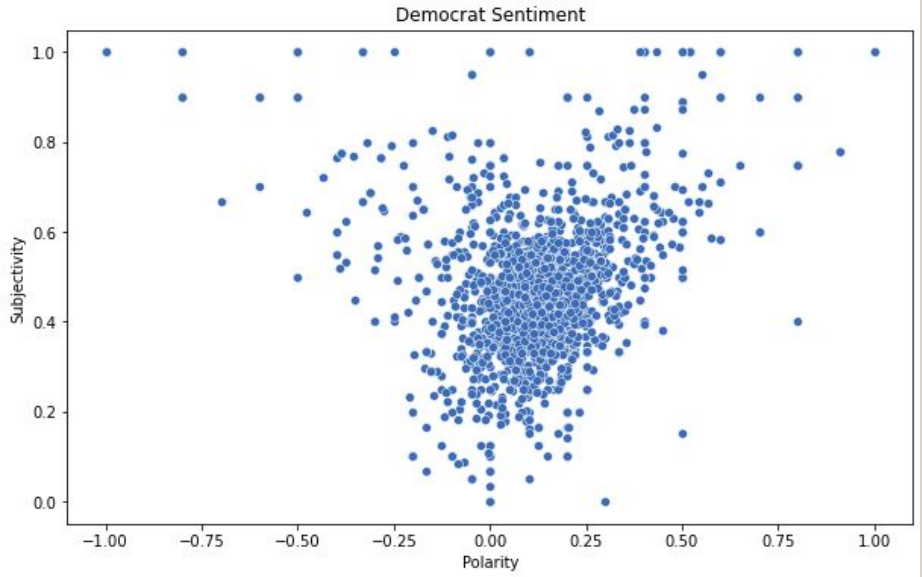
Data Cleaning: Levenshtein Distance!

		k	i	t	t	e	n
	0	1	2	3	4	5	6
s	1	1	2	3	4	5	6
i	2	2	1	2	3	4	5
t	3	3	2	1	2	3	4
t	4	4	3	2	1	2	3
i	5	5	4	3	2	2	3
n	6	6	5	4	3	3	2
g	7	7	6	5	4	4	3

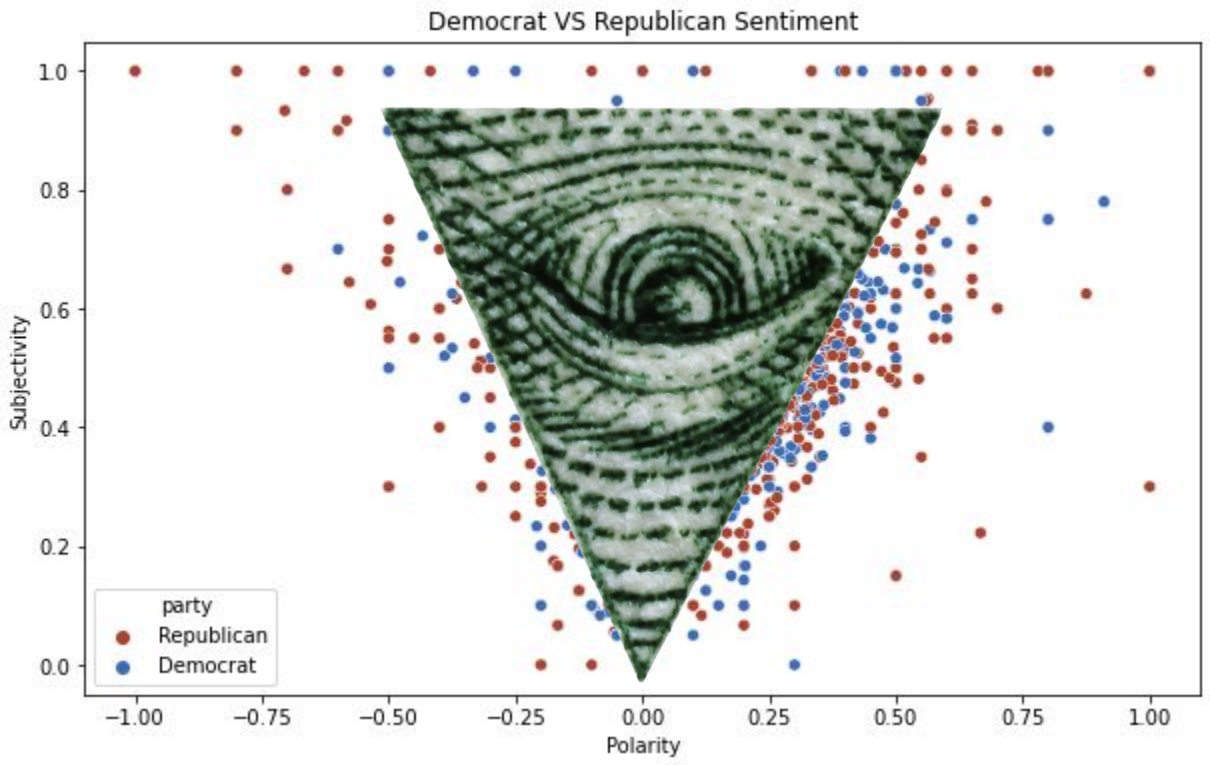
```
('[*]CROWLEY', 'CROWLEY') 3
('[*]SCHIEFFER', 'SCHIEFFER') 3
('MR.FORD', 'FORD') 3
('KONDRACKE', 'MONDALE') 4
('OTIS', 'BUSH') 4
('JOHNSON', 'CLINTON') 4
('OREGONIAN', 'REAGAN') 5
('OBAM', 'OBAMA') 1
('ROMNEHY', 'ROMNEY') 1
('SM1TH', 'SMITH') 1
('REAGAV', 'REAGAN') 1
```

Exploratory Data Analysis: Sentiment

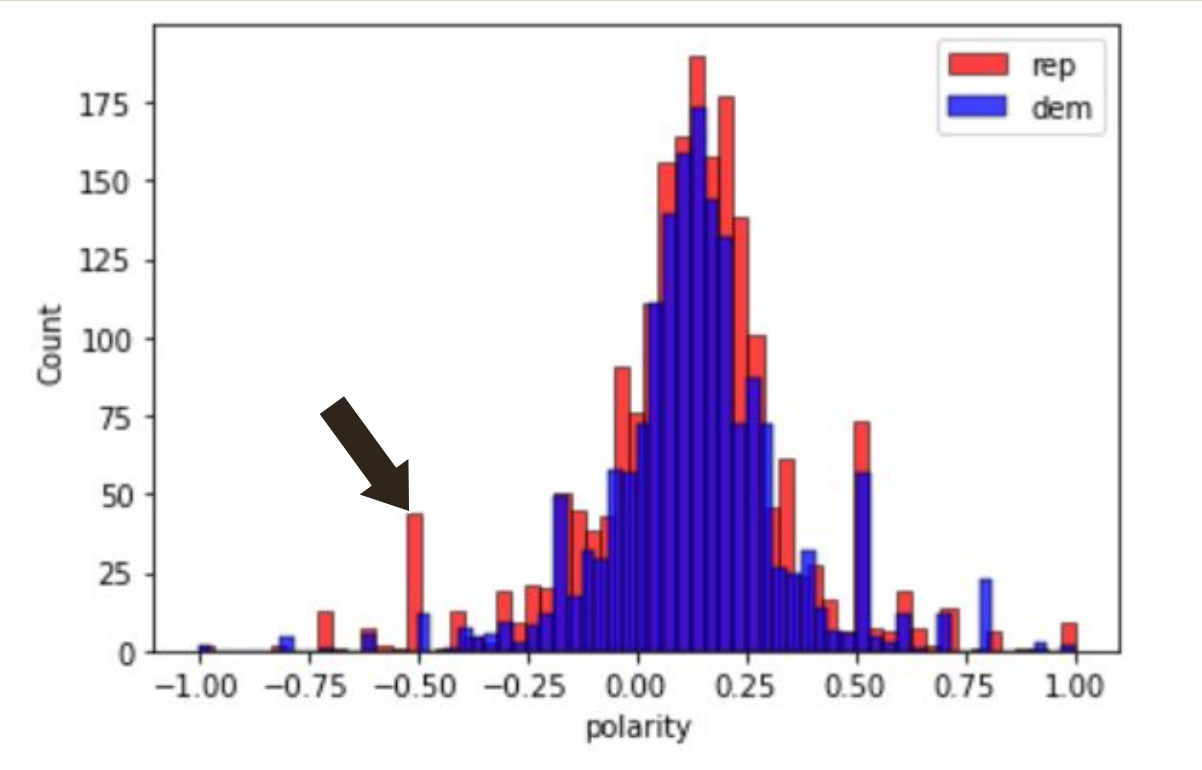




EDA continued



EDA: Polarity by Party



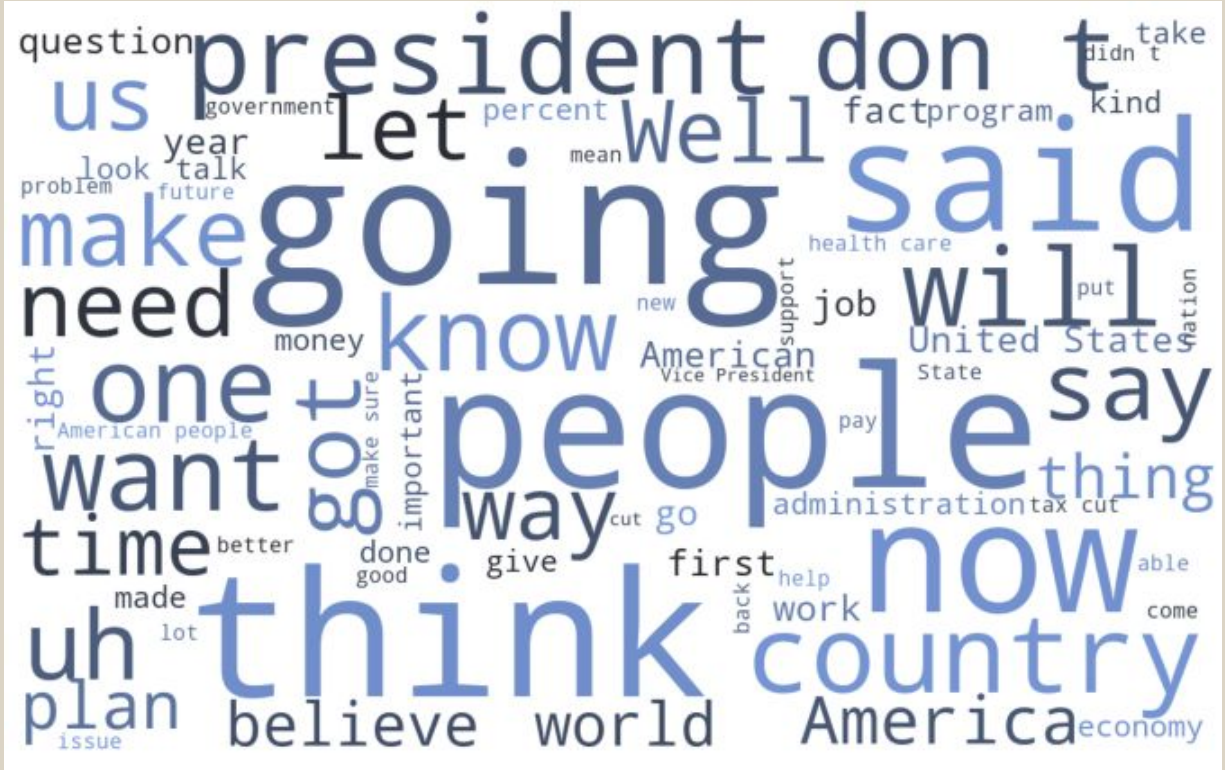
Outlier Example 1

"I get treated worse than the Tea Party got treated. A lot of people in there, deep down in the IRS, they treat me horribly. We made a deal, it was all settled, until I decided to run for president. I get treated very badly by the IRS, very unfairly, but we had a deal all done. As soon as we're completed with the deal, I want to release it. But I have paid millions and millions of dollars. And it's worse than paying. I paid in advance. It's called prepaying your taxes." - Donald J. Trump

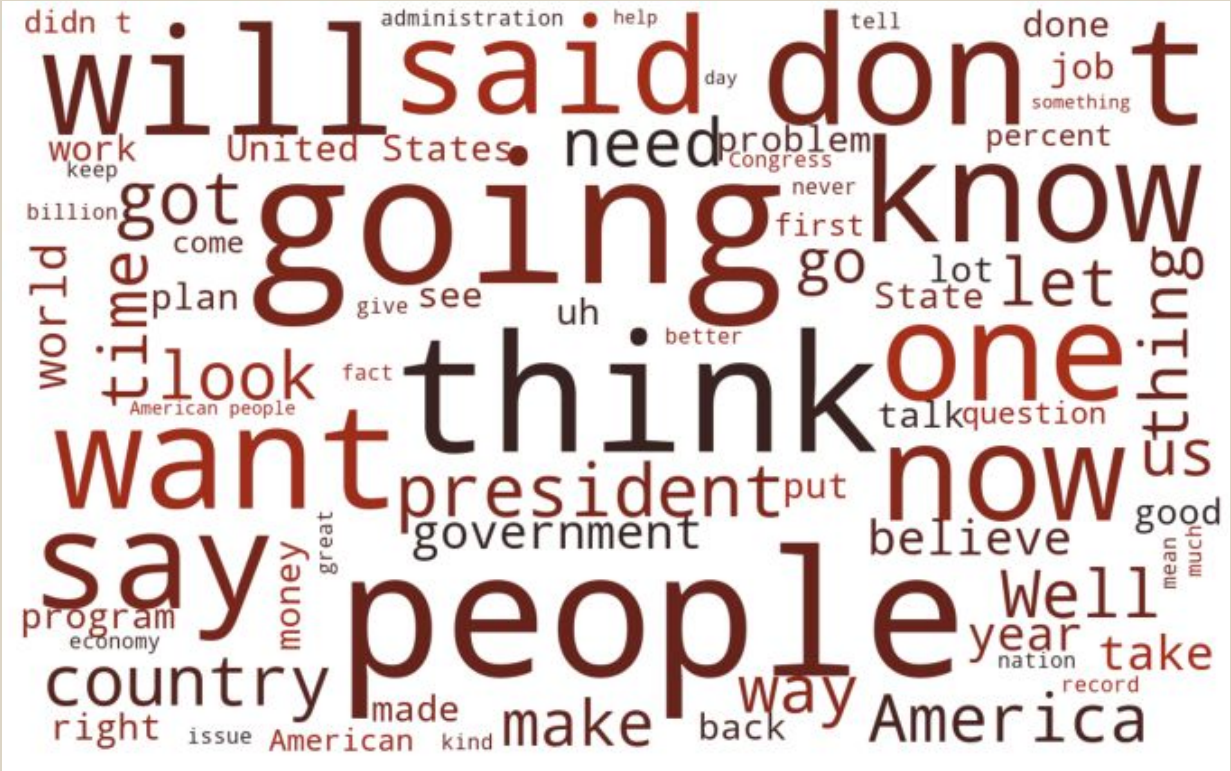
Outlier Example 2

"Wrong" - Donald J. Trump (~10% of the -0.5 polarity cohort)

EDA: Democrat Word Cloud



EDA: Republican Word Cloud



Model Testing Primaries

	Original		Lemmas		Stems	
	Cvect	Tfidf	Cvect	Tfidf	Cvect	Tfidf
MultinomialNB	**0.71**	0.64	0.70	0.64	**0.71**	0.65
LogisticRegression	0.69	0.69	0.69	0.70	0.70	**0.71**
BaggingClassifier	0.64	0.64	0.64	0.64	0.64	0.64
RandomForestClassifier	0.68	0.69	0.69	0.70	0.70	0.68
ExtraTreesClassifier	0.70	0.69	0.70	0.70	0.70	**0.71**
KNeighborsClassifier	0.55	0.68	0.56	0.68	0.56	0.68
SGDClassifier	0.67	0.70	0.70	0.70	0.70	0.70

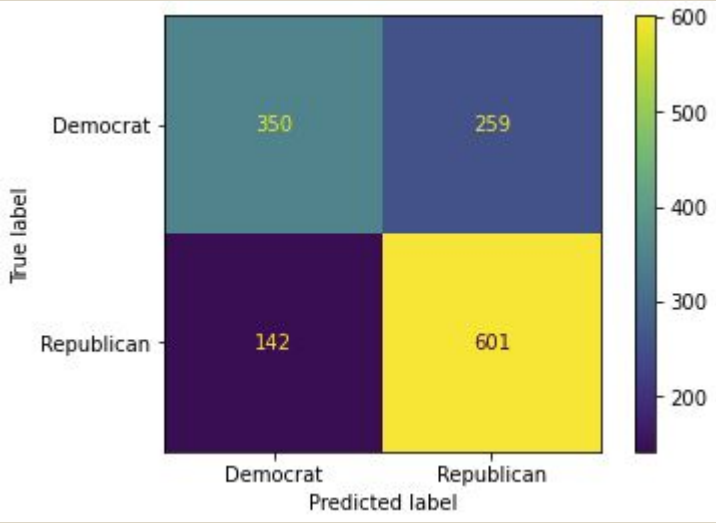
The Winner!

- **LogisticRegression** and **TfidVectorizer** via GridSearch
- LogReg C Value: **2**
- LogReg Fit Intercept: **True**
- Tfid binary: **True**
- Tfid ngram range: **1 - 4**

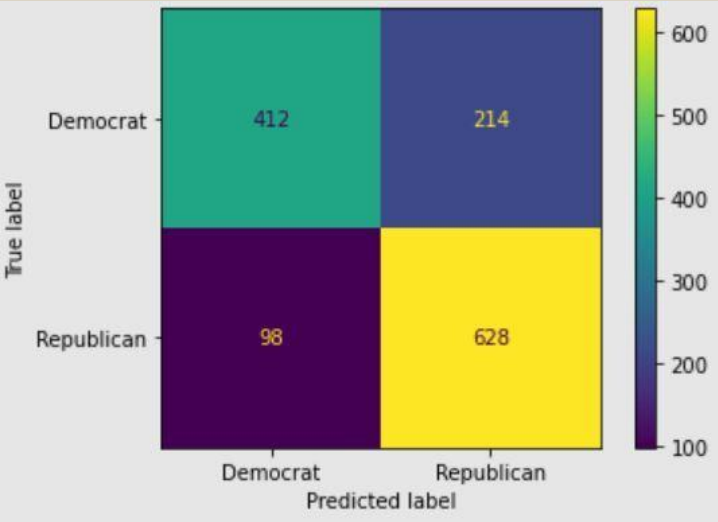
Accuracy Score: **0.76**

Modeling Results: Confusion Matrix

Before GridSearch



After GridSearch



Conclusion

- Democrats and Republicans are actually more similar than different when it comes to debates!
- Republicans have a marginally more negative polarity (especially in recent years).
- Polarity is more split in recent year data vs older debate data which is more down the middle.