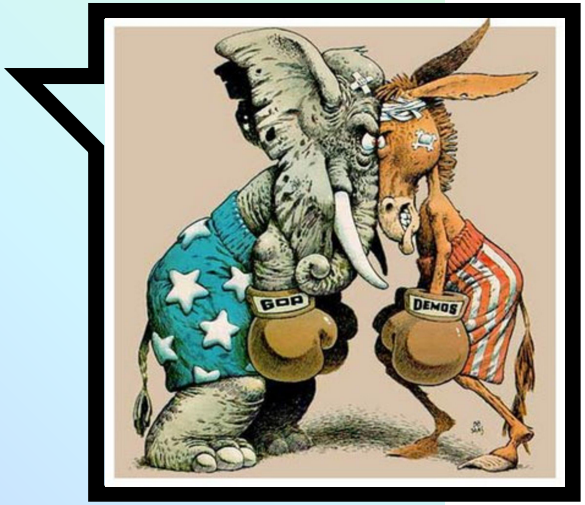
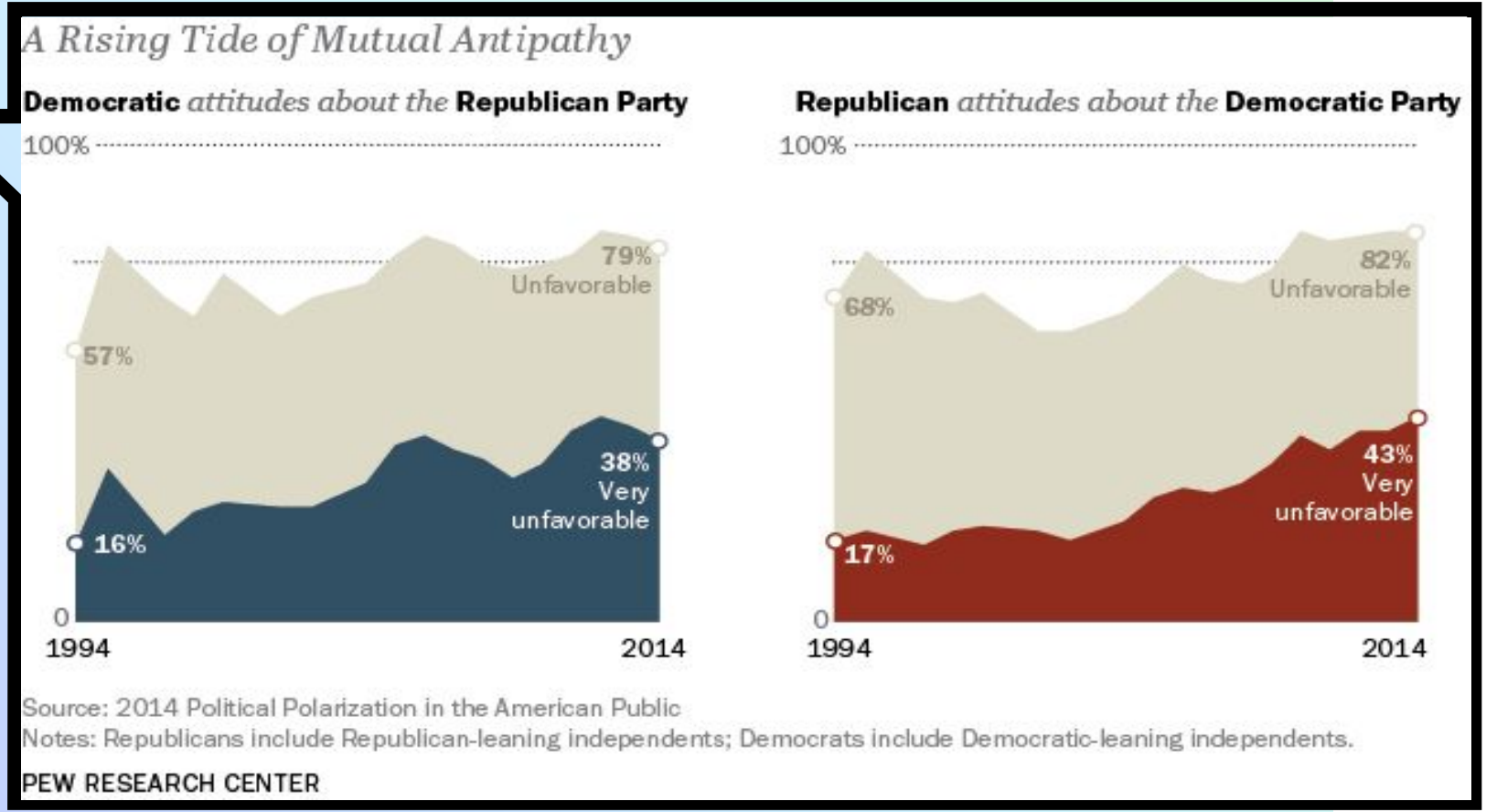
Reddit: Democrats & Republicans

Michael Ruggiero



Problem Statement

Given the current partisan climate, understanding the linguistic differences between the Democrats and Republicans may facilitate future communication among dissonant groups. This project aims to both predict the origin of a partisan subreddit and also understand the different uses of language within these groups. Various predictive models will be considered and data will be harvested with the Reddit API Praw.


A Rising Tide of Mutual Antipathy
Democratic attitudes about the Republican Party
100%
79%
Unfavorable
57%
38%
Very
unfavorable
16%
0
1994
2014
Republican attitudes about the Democratic Party
100%
82%
Unfavorable
68%
43%
Very
unfavorable
17%
0
1994
2014
Source: 2014 Political Polarization in the American Public
Notes: Republicans include Republican-leaning independents; Democrats include Democratic-leaning independents.
PEW RESEARCH CENTER

Partisans more likely to participate

Voting Rates Highest Among the Most Polarized

Cell entries show the percent of partisans in each quadrant who say they always vote

Among Democrats:

Ideology	View of the Republican Party: Very unfav	Mostly unfav	Total
Consistently liberal	70	51	60
Mostly liberal	55	46	47
Mixed	49	41	41
Total	58	46	49

Among Republicans:

	View of the Democratic Party: Very unfav	Mostly unfav	Total
Consistently conservative	83	70	79
Mostly conservative	69	56	62
Mixed	48	41	42
Total	68	50	56

Source: 2014 Political Polarization in the American Public
Note: Republicans include Republican-leaning independents; Democrats include Democratic-leaning independents (see Appendix B). Groups too small to analyze (e.g. Republicans who have a favorable view of the Democratic party and conservative Democrats) are included in totals, but not shown in individual cells.

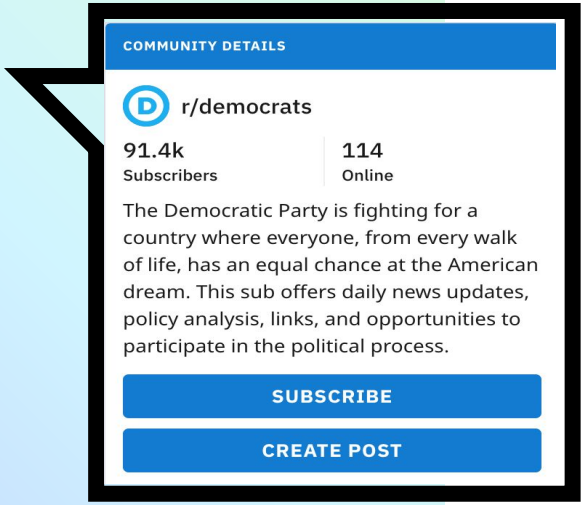
PEW RESEARCH CENTER

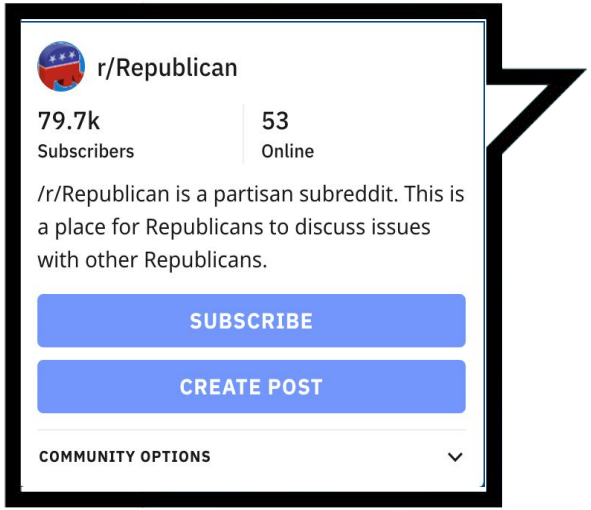
1.

Data Collection

This modeling uses Reddit API: PRAW

1) Easy
2) Ethical
3) Efficient





Key numbers: Among the 31 states (plus D.C.) with party registration, there are nearly **12 million** more registered Democrats than Republicans. 40% of all voters in party registration states are Democrats, 29% are Republicans, and 28% are independents. Jul 10, 2018

First look: Democrats crush GOP in party registration - Axios
https://www.axios.com/democrats-crush-republican-party-registration-2018-midterms-8...

Data Collection

Gathered top 1000 titles

Collected all comments

Collected Vital Stats

2.

EDA
Data Cleaning

Republicans have an inconsistent robot!

> /r/Republican is a partisan subreddit. This is a place for Republicans to discuss issues with other Republicans.
>
> Out of respect for this sub's main purpose, we ask that unless you identify as Republican that you refrain from commenting and [leave the vote button alone]...*I am a bot, and this action was performed automatically. Please [contact the moderators of this subreddit]

3.

Model Selection

```
models = {
    "lr" : LogisticRegression(n_jobs=3),
    "svc": SVC(),
    "ada": AdaBoostClassifier(),
    "bnb": BernoulliNB(),
    "gbc": GradientBoostingClassifier()}

params = {
    "lr" :{
        "solver"            : ["newton-cg", "lbfgs"],
        "multi_class"       : ["ovr", "multinomial"]},
    "svc":{
        "gamma"             : ["scale"]},
    "ada":{
         "n_estimators"     : [50, 500],
         "learning_rate"    : [1,.5]},
    "bnb":{},
    "gbc":{}}
```

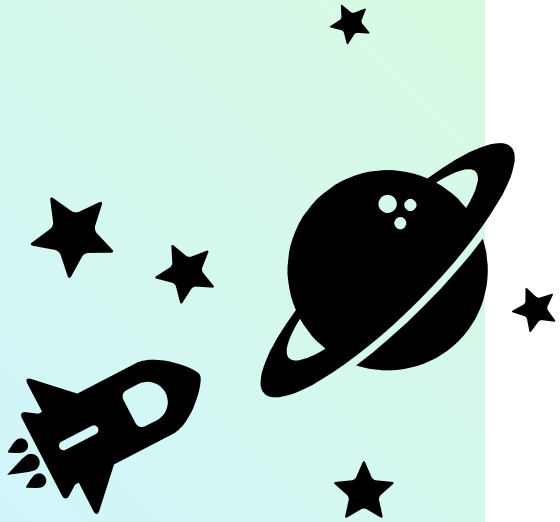
```
for model in models:
    pipe = Pipeline([
        ("cvec", CountVectorizer()),
        ("grid", GridSearchCV(models[model],
                              params[model],
                              cv = 3,
                              n_jobs = 3,
                              verbose = 2))
    ])

    cvec_params = {
        "cvec__stop_words":[None],
        "cvec__max_features":[7_500,10_000, 25_000],
        "cvec__stop_words": [(1,1),(2,2)],
        "cvec__max_df": [.8],
        "cvec__min_df": [.01]}

    gs = GridSearchCV(pipe,
                      param_grid=cvec_params,
                      cv = 3,
                      n_jobs = 3,
                      verbose = 2)

    gs.fit(X_flat_train, y_train)
    print(model, gs.best_score_)
    print(gs.best_params_)
```

AdaBoost | Logistic

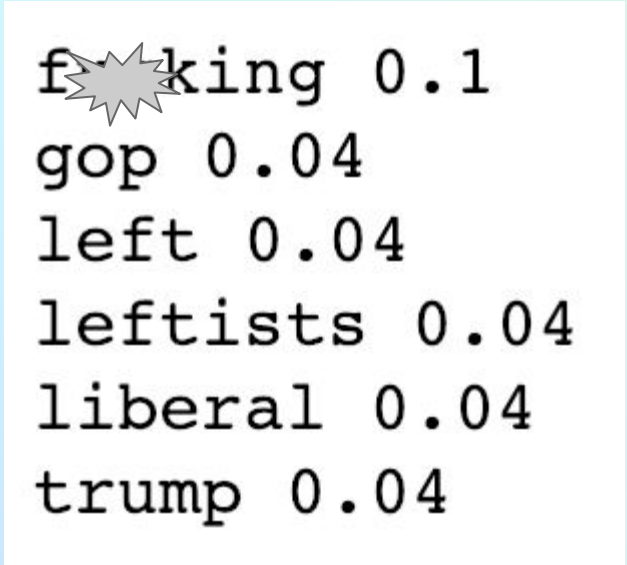


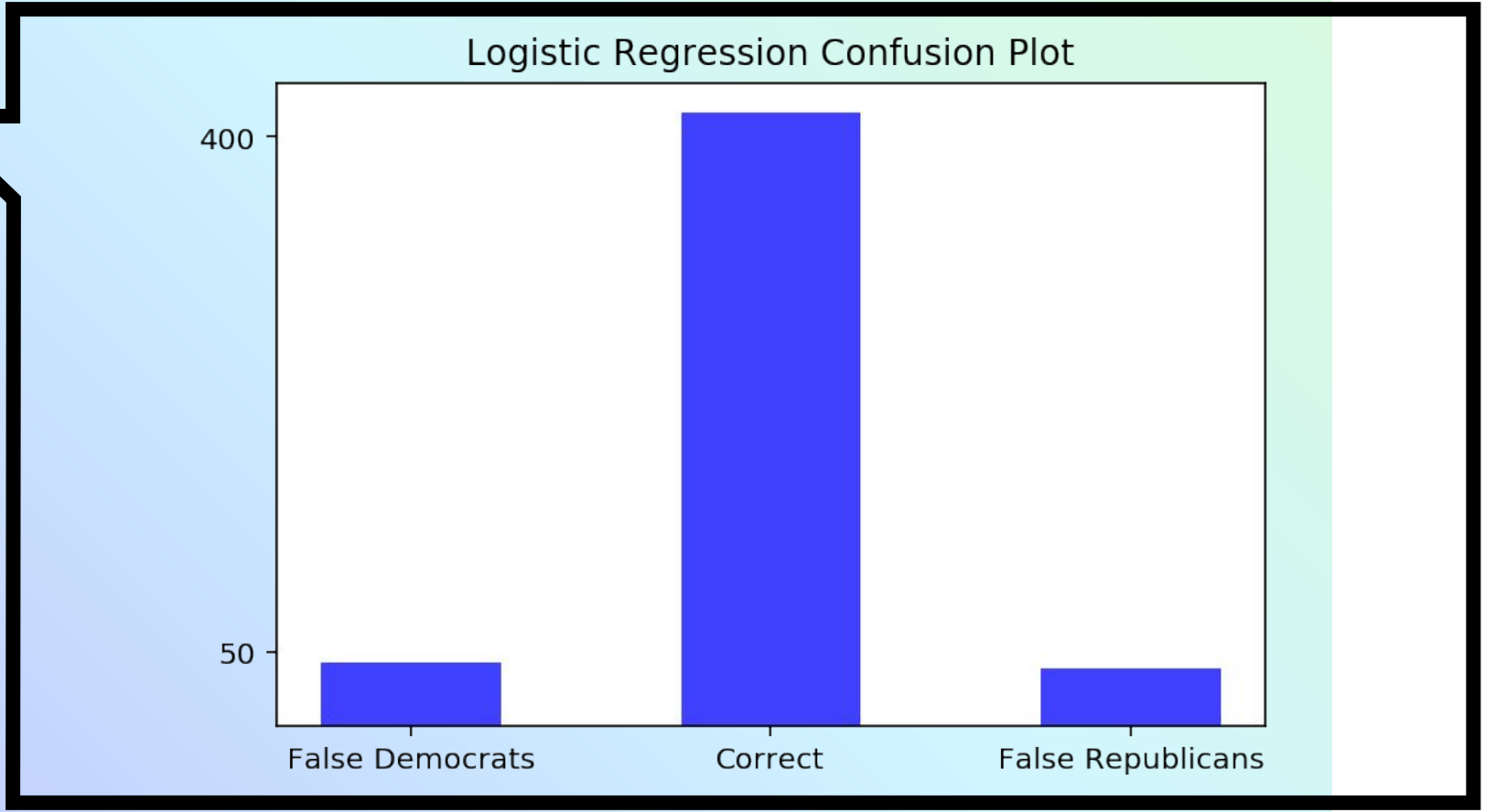
While AdaBoost had the best predictive scores ~85%, we will talk about Logistic Regression for ease of reading betas.

Political Jeopardy!

	F-bomb	“politics”	
Dems			
Reps			

Ada Classifier most important words




Logistic Regression Confusion Plot
400
50
False Democrats
Correct
False Republicans

Partisans likely to “totemize” each other

What Makes Online Content Viral?

17 Pages • Posted: 27 Dec 2009 • Last revised: 30 Aug 2012

Jonah A. Berger
University of Pennsylvania - Marketing Department

Katherine L. Milkman
University of Pennsylvania - The Wharton School

Date Written: December 25, 2009

Abstract

Why are certain pieces of online content more viral than others? This article takes a psychological approach to understanding diffusion. Using a unique dataset of all the New York Times articles published over a three month period, the authors examine the link between integral affect (i.e., the emotion evoked) and whether content is highly shared. Results suggest a strong relationship between emotion and virality, but indicate that this link is more complex than mere valence alone. Positive content is more viral (than negative content), as is content that inspires awe. But while sad content is less viral, anger or anxiety inducing articles are both more likely to make the paper’s most emailed list. These results hold controlling for how surprising, interesting, or practically useful content is (all of which are positively linked to virality), as well as external drivers of attention (e.g., how prominently articles were featured). The findings shed light on why people share online content, provide insight into how to design effective viral marketing campaigns, and underscore the importance of individual-level psychological processes in shaping collective outcomes.

Keywords: Social Transmission, Word-of-Mouth, Social Epidemics