



Predicting  
Party  
Affiliation  
Using Social  
Media

Joseph Brown,  
Mikaela  
Jordan, and  
Adam Swayze

Polling and  
Predictions

Our Primary  
Predictions

Collecting Data  
Cleaning Tweets  
Models

Results

Democrats  
Republicans

Limitations

Future Work

References

# Predicting Party Affiliation Using Social Media

Joseph Brown, Mikaela Jordan, and Adam Swayze

Tarleton State University

April 1, 2016



# Predicting Primary Results

- In Primaries, polls change fast and are inaccurate

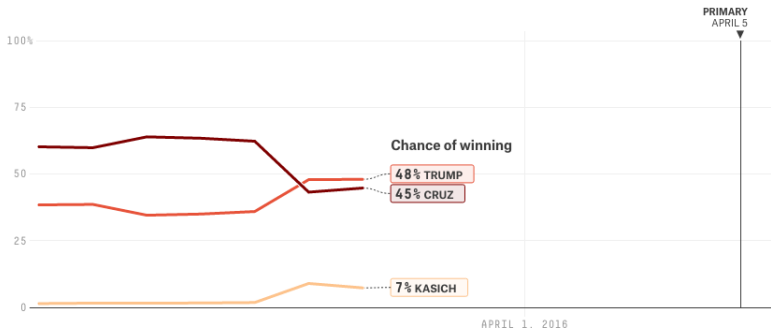


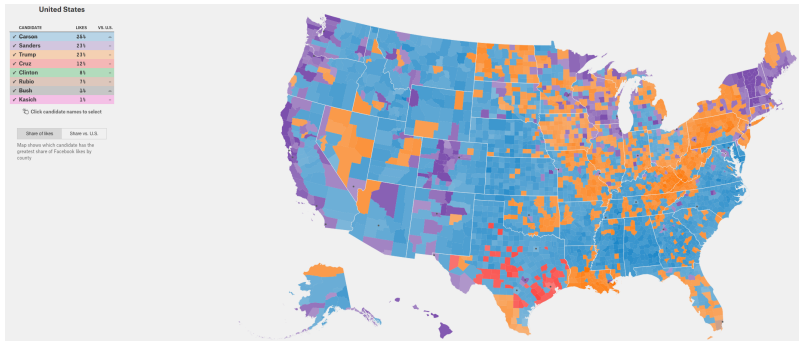
Figure: Current polls for the Wisconsin Republican Primary from 538



# Alternative Primary Predictions



- Counties colored by Candidate with most Facebook likes

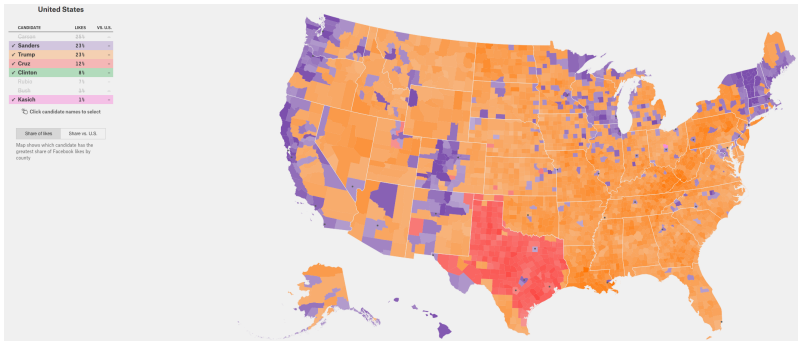




# Candidates and Social Media



- Candidates still running for President



Predicting  
Party  
Affiliation  
Using Social  
Media

Joseph Brown,  
Mikaela  
Jordan, and  
Adam Swayze

Polling and  
Predictions

Our Primary  
Predictions

Collecting Data  
Cleaning Tweets  
Models

Results

Democrats  
Republicans

Limitations

Future Work

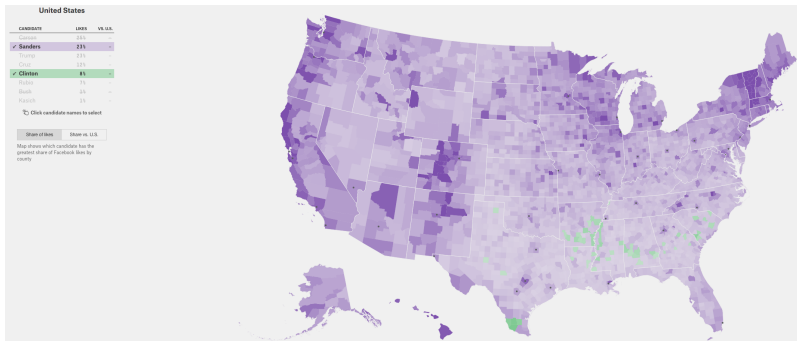
References



# Candidates and Social Media



## • Democratic Candidates



Predicting  
Party  
Affiliation  
Using Social  
Media

Joseph Brown,  
Mikaela  
Jordan, and  
Adam Swayze

Polling and  
Predictions

Our Primary  
Predictions

Collecting Data  
Cleaning Tweets  
Models

Results

Democrats  
Republicans

Limitations

Future Work

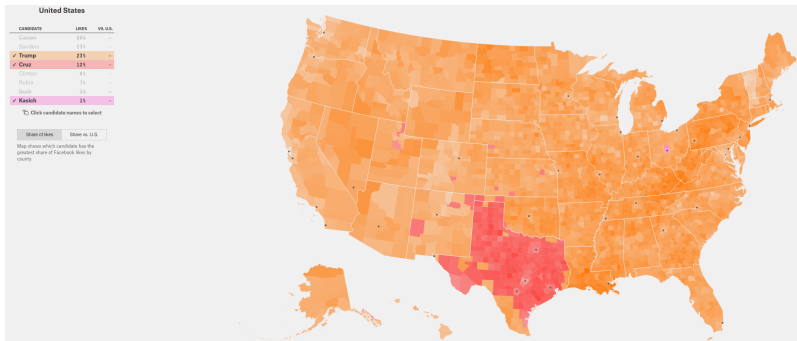
References



# Candidates and Social Media



## ● Republican Candidates





# Natural Language Processing

Predicting  
Party  
Affiliation  
Using Social  
Media

Joseph Brown,  
Mikaela  
Jordan, and  
Adam Swayze

Polling and  
Predictions

Our Primary  
Predictions

Collecting Data  
Cleaning Tweets  
Models

Results

Democrats  
Republicans

Limitations

Future Work

References

- Basic Idea
  - Computers understanding language
- Several Tasks in Natural Language Processing
  - Question Answering
  - Automatic Summarization
  - Sentiment Analysis



## Predicting Party Affiliation Using Social Media

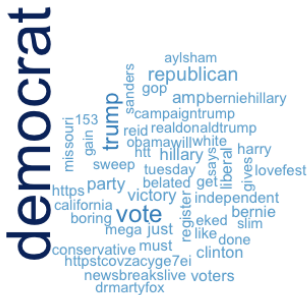
Joseph Brown,  
Mikaela  
Jordan, and  
Adam Swayze

## Our Primary Predictions

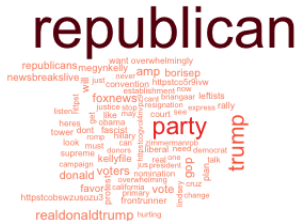
## Limitations

## Future Work

## References



(a)



(b)

**Figure:** Word Clouds for Twitter Searches with Keywords “Democrat” and “Republican”





# Design of the Primary Prediction

In a Perfect World

- Created a dataset with the names of all counties in US, their respective FIPS codes, and the recorded majority winners for each county
  - Unique identification number for each county in US
- Found a dataset with area and centroid of each county in US
- Search for tweets with specified keywords in every county of AZ, FL, IL, ID, MO, NC, OH, and UT

Candidates	Bernie Sanders	Hillary Clinton	Donald Trump	John Kasich	Ted Cruz
Keywords	"Bernie" "Sanders" "Feel The Bern" "Bernie2016"	"I'm with Her" "HillaryClinton" "Hillary2016" "Clinton"	"Trump" "Donald" "Make America Great Again" "Trump2016" "DonaldTrump"	"Kasich" "JohnKasich" "Kasich2016"	"Cruz" "TedCruz" "Trust Ted" "Cruz2016"

- Merge results data frame with tweets data frame by FIPS code

Predicting  
Party  
Affiliation  
Using Social  
Media

Joseph Brown,  
Mikaela  
Jordan, and  
Adam Swayze

Polling and  
Predictions

Our Primary  
Predictions

Collecting Data  
Cleaning Tweets  
Models

Results

Democrats  
Republicans

Limitations

Future Work

References



# Design of the Primary Prediction

In a Perfect World

Predicting  
Party  
Affiliation  
Using Social  
Media

Joseph Brown,  
Mikaela  
Jordan, and  
Adam Swayze

Polling and  
Predictions

Our Primary  
Predictions

Collecting Data  
Cleaning Tweets  
Models

Results

Democrats  
Republicans

Limitations

Future Work

References

## An Example

Here is an original tweet from our collection:

†Trusted #cruz2016 🇺🇸🇺🇸🇺🇸🇺🇸🇺🇸🇺🇸🇺🇸🇺🇸🇺🇸🇺🇸🇺🇸🇺🇸 @ Palmbay FL  
<https://t.co/SNElizEZY0L>

- Cleaning Tweets

- Remove punctuation
- Remove emojis
- Remove Stopwords
- Make everything lower case

"trusted cruz2016 palmbay fl httpstcosneliezy0l"



# Design of the Primary Prediction

In a Perfect World

Predicting  
Party  
Affiliation  
Using Social  
Media

Joseph Brown,  
Mikaela  
Jordan, and  
Adam Swayze

Polling and  
Predictions

Our Primary  
Predictions

Collecting Data  
Cleaning Tweets  
Models

Results

Democrats  
Republicans

Limitations

Future Work

References

- Create a Document-Term Matrix

Doc Name	Term 1	Term 2	...	Term $n$
Doc 1	$\text{Freq}(\text{T1 in D1})$	$\text{Freq}(\text{T2 in D1})$	...	$\text{Freq}(\text{Tn in D1})$
$\vdots$				
Doc $m$	$\text{Freq}(\text{T1 in Dm})$	$\text{Freq}(\text{T2 in Dm})$	...	$\text{Freq}(\text{Tn in Dm})$

- Attach labels to each tweet based on location of tweet
- Use machine learning algorithms to predict majority winners of each tweet's county



# Results

Predicting  
Party  
Affiliation  
Using Social  
Media

Joseph Brown,  
Mikaela  
Jordan, and  
Adam Swayze

Polling and  
Predictions

Our Primary  
Predictions

Collecting Data  
Cleaning Tweets  
Models

Results

Democrats  
Republicans

Limitations

Future Work

References

		Predicted	
		Bernie	Hillary
Actual	Bernie	431	3332
	Hillary	148	10212

Table: Support Vector Machine Confusion Matrix

		Predicted	
		Bernie	Hillary
Actual	Bernie	434	3329
	Hillary	132	10228

Table: Neural Network Size 6 Confusion Matrix

	Accuracy Rate
Support Vector Machine	75.35934%
Neural Network	75.49388%

Table: Accuracy Rates of Both Models



# Results



Predicting  
Party  
Affiliation  
Using Social  
Media

Joseph Brown,  
Mikaela  
Jordan, and  
Adam Swayze

Polling and  
Predictions

Our Primary  
Predictions

Collecting Data  
Cleaning Tweets  
Models

Results

Democrats  
Republicans

Limitations

Future Work

References

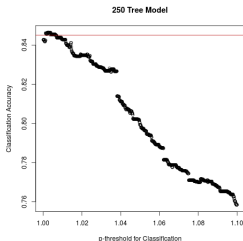


Figure: A plot of classification accuracy versus probability threshold for Random Forest Model

		Predicted	
		Trump	Not Trump
Actual	Trump	9517	42
	Not Trump	1695	58

Table: Random Forest with 250 Trees Confusion Matrix. Has an accuracy of 84.644625%



# Limitations with Social Media



## Predicting Party Affiliation Using Social Media

Joseph Brown,  
Mikaela  
Jordan, and  
Adam Swayze

## Polling and Predictions

## Our Primary Predictions

Collecting Data  
Cleaning Tweets  
Models

## Results

Democrats  
Republicans

## Limitations

## Future Work

## References

- Class imbalance - Trump is winning many more precincts than other Republican candidates
- Majority of Sanders supporters are younger and more likely to use Twitter (estimated at 88% by Pew Research Center)
- According to the Pew Research Center, only 23% of Americans use Twitter
- Twitter says that  $< 5\%$  of all tweets are georeferenced
- Can only get tweets up to a week prior of collection time
- Could not find a comprehensive list of counties and results



# Future Work

Predicting  
Party  
Affiliation  
Using Social  
Media

Joseph Brown,  
Mikaela  
Jordan, and  
Adam Swayze

Polling and  
Predictions

Our Primary  
Predictions

Collecting Data  
Cleaning Tweets  
Models

Results

Democrats  
Republicans

Limitations

Future Work

References

- Continue to update our models for each Primary
- Improve our models to account for other geographic differences
- Use debate transcripts to predict partiality of media networks using online articles
- Extend current models to predict outcome of presidential election in November



# References

We'd like to thank the Office of Student Research and Creative Activity at Tarleton State University



- Benjamin, Dan. "A Gentle Introduction to Random Forests, Ensembles, and Performance Metrics in a Commercial System - Blog & Press" *Citizen and Net*. N.p., 09 Nov. 2012. Web. 28 Mar. 2016.
- Phillips, Winfred. "Introduction to Natural Language Processing." *Consortium on Cognitive Science Instruction*. The MIND Project, 2006. Web. 16 Mar 2016.
- Raine, Lee. "Social Media and Voting." *Pew Research Center Internet Science Tech RSS*. Pew Charitable Trusts, 05 Nov. 2012. Web. 28 Mar. 2016.
- Weston, Jason. "Support Vector Machines (and Statistical Learning Theory)." *ComputerScience4701*. Columbia University, New York. *Columbia University*. Web. 28 Mar. 2016.
- "FiveThirtyEight." *FiveThirtyEight*. Nate Silver, n.d. Web. 28 Mar 2016.
- "2016 Primary Election Results: President Live Map by State, Real-Time Voting Updates" *Election Hub*. Politico LLC, 27 Mar. 2016. Web. 28 Mar. 2016.