



Political Statement Lie Detector

Project McNulty | August 5, 2016 | Jessica Cox



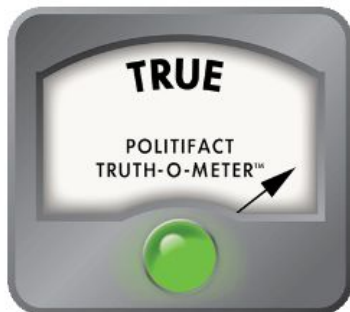
POLITIFACT

WINNER OF THE PULITZER PRIZE

- Run by the Tampa Bay Times, an independent newspaper in Florida
 - Fact checks claims by elected officials, candidates, leaders of political parties and political activists
 - Select the most newsworthy statements for fact-checking
-

POLITIFACT

WINNER OF THE PULITZER PRIZE








Central question

Can we predict if a politician's public statements are true or not, based on demographics of their electorate?



Data acquisition

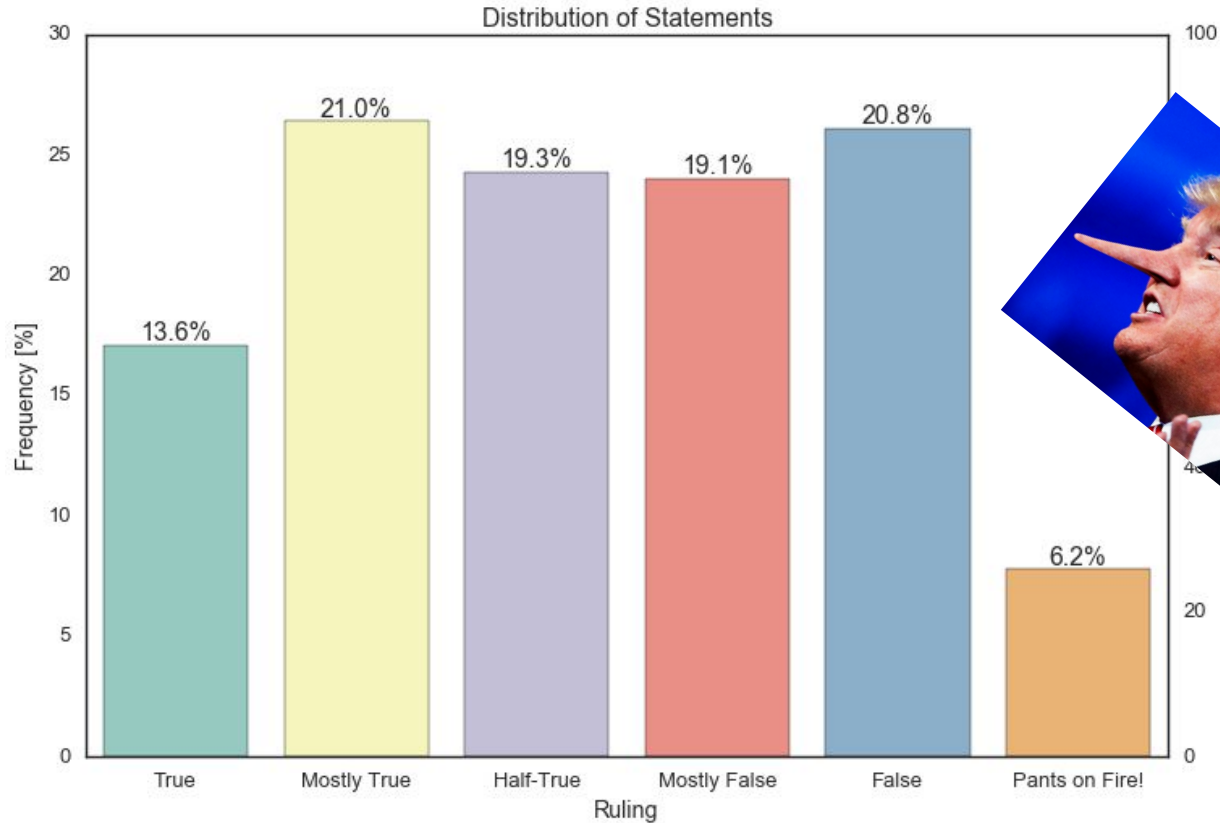
-  API returns all statements by person, statement, and date.
 -  EveryPolitician and wikipedia for database of politicians
 -  provided database of all demographic information by state and year
-



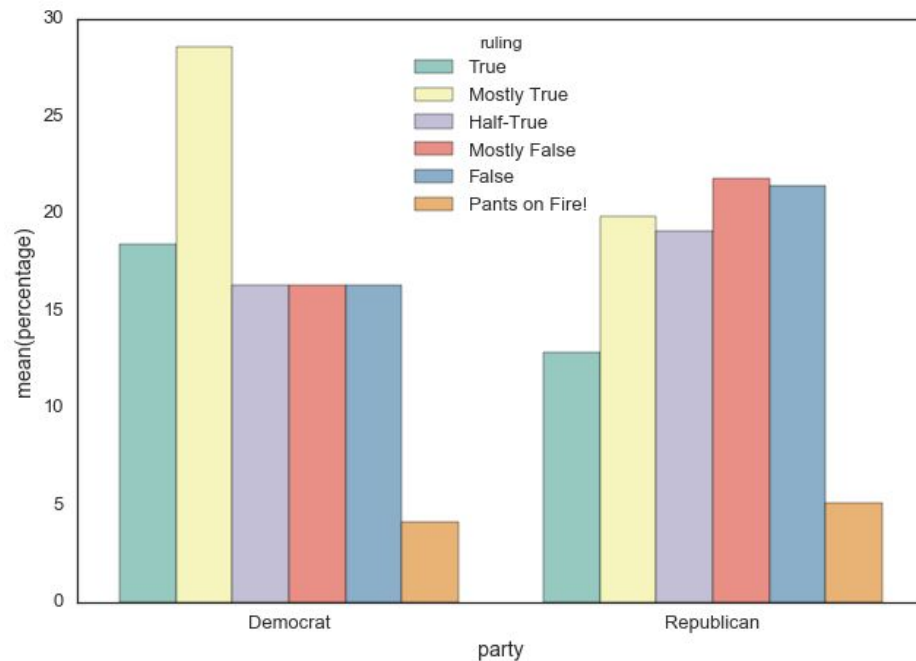
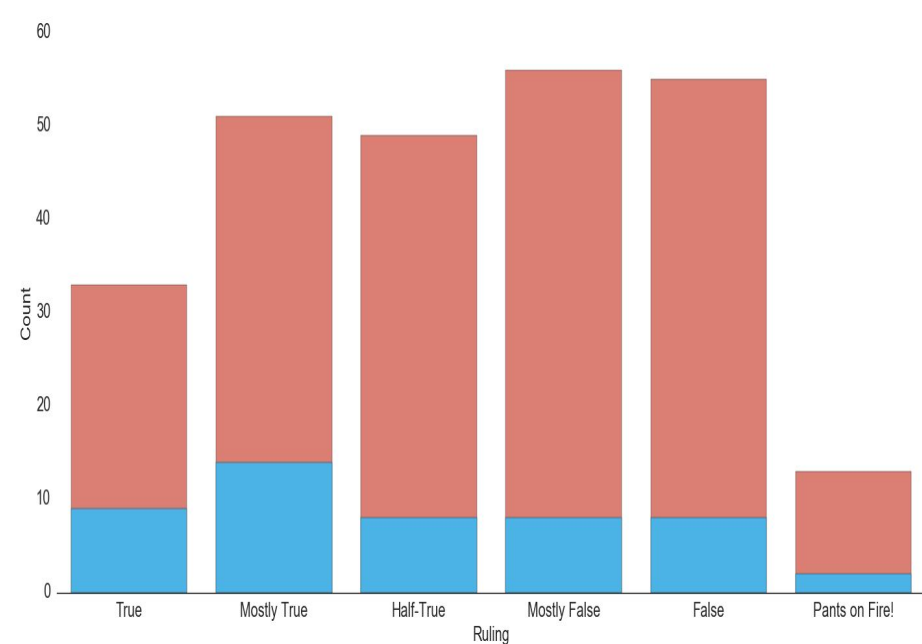
Data acquisition

- API returned ~ 3000 statements made since 1995
 - Compiled a list of senators and governors
 - Fuzzy text matching to merge databases
 - Final database resulted in **419** statements made by politicians between 2008-2014
-

Most statements are lies!

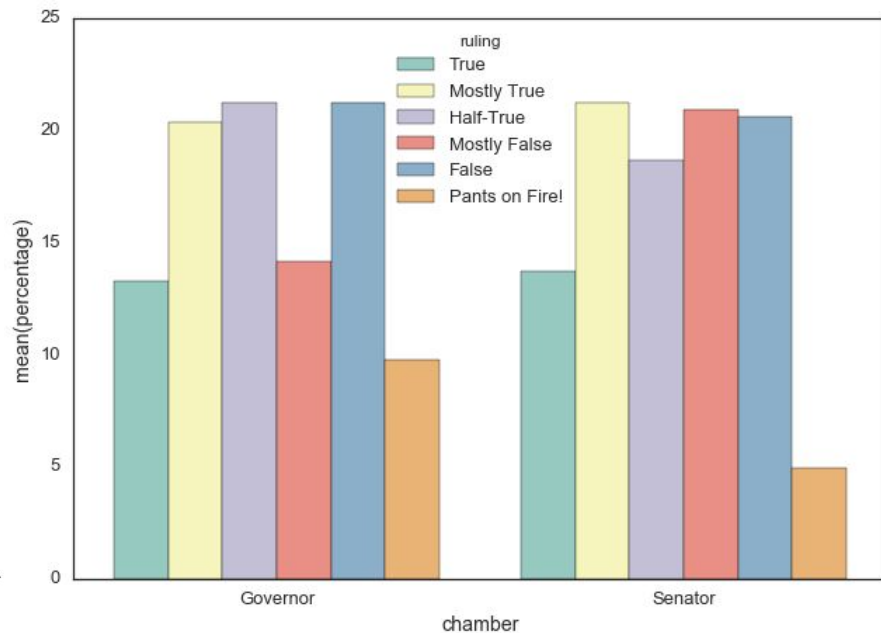
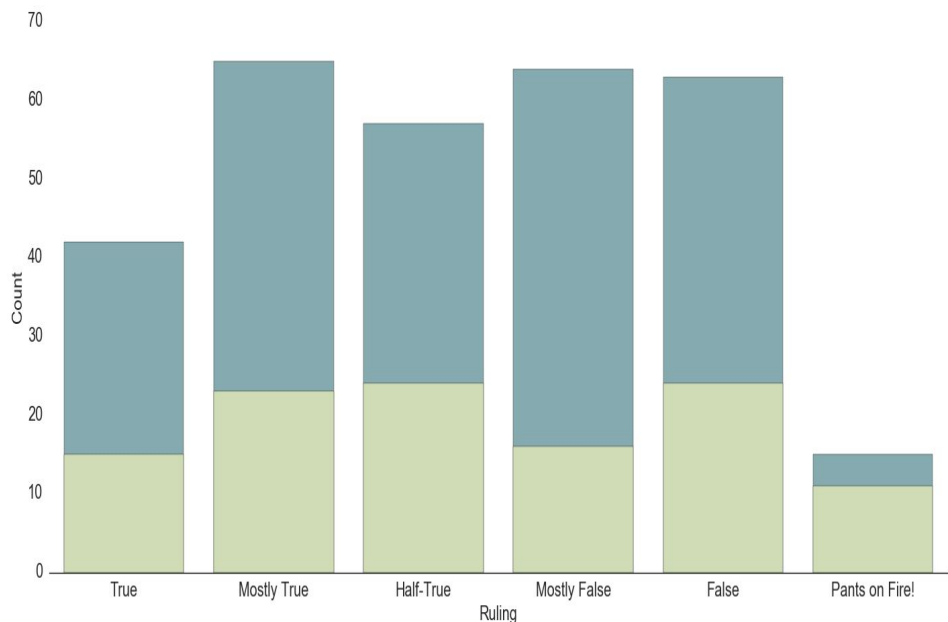


Republicans are fact-checked more...



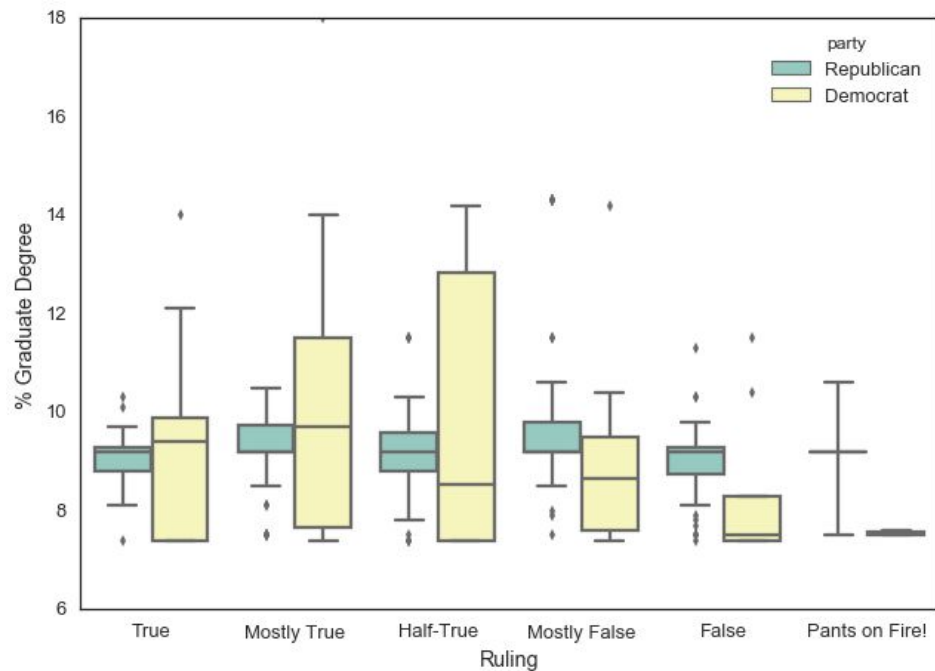
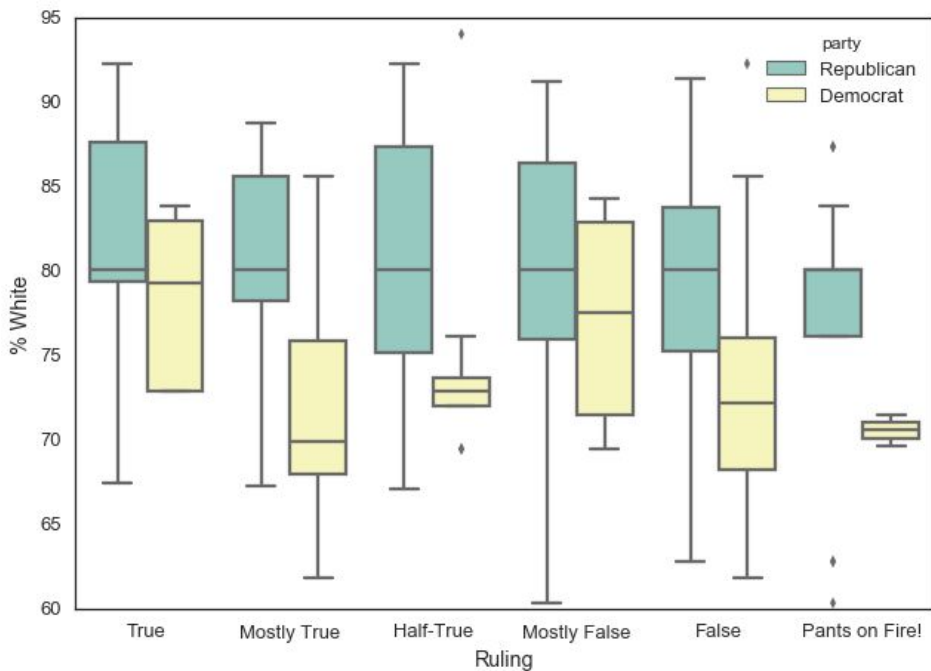
...but lie more in their checked statements.

Senators are fact-checked more...



...but Governors lie just as much!

Statements by state demographics



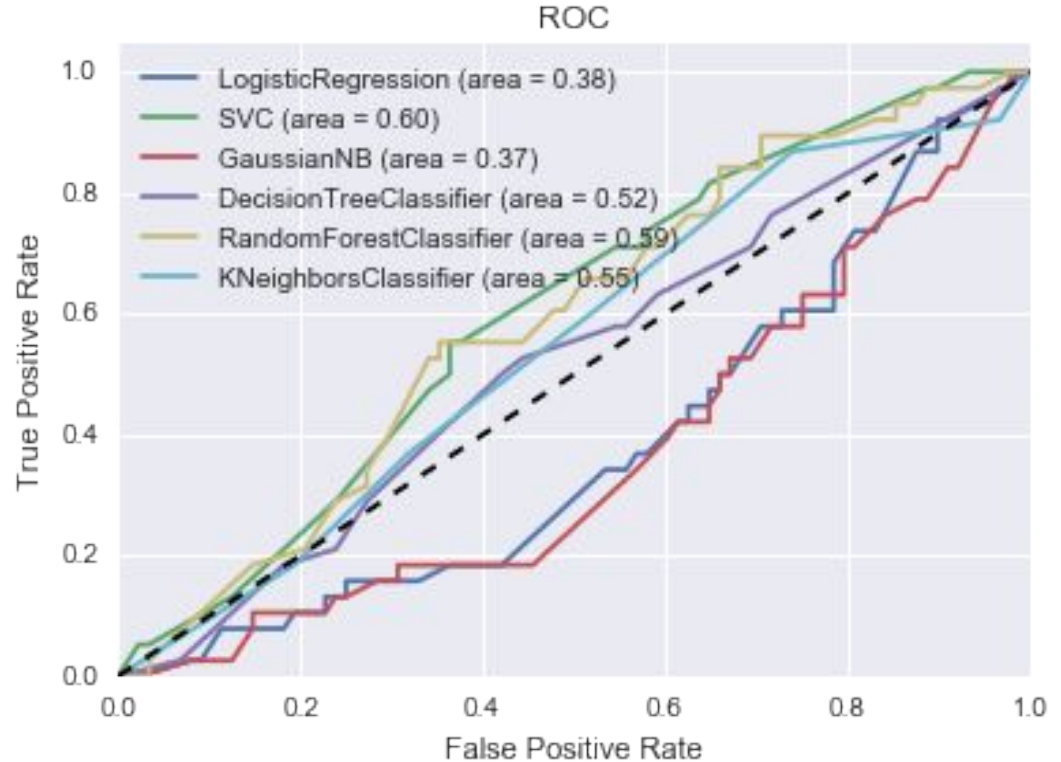
Model building



- Imbalanced classes
 - Binned 6 outcomes into:
 - True: **174**
 - False: **245**
 - Used 74 features
 - Dummy variables for states
 - State demographic information
 - Party
 - Python package: imbalanced learn
-

Model building

- Imbalance correction with cluster centroids
- Random Forest Classifier
auc = 0.59
- SVC auc = 0.60



Model building - tuning hyperparameters

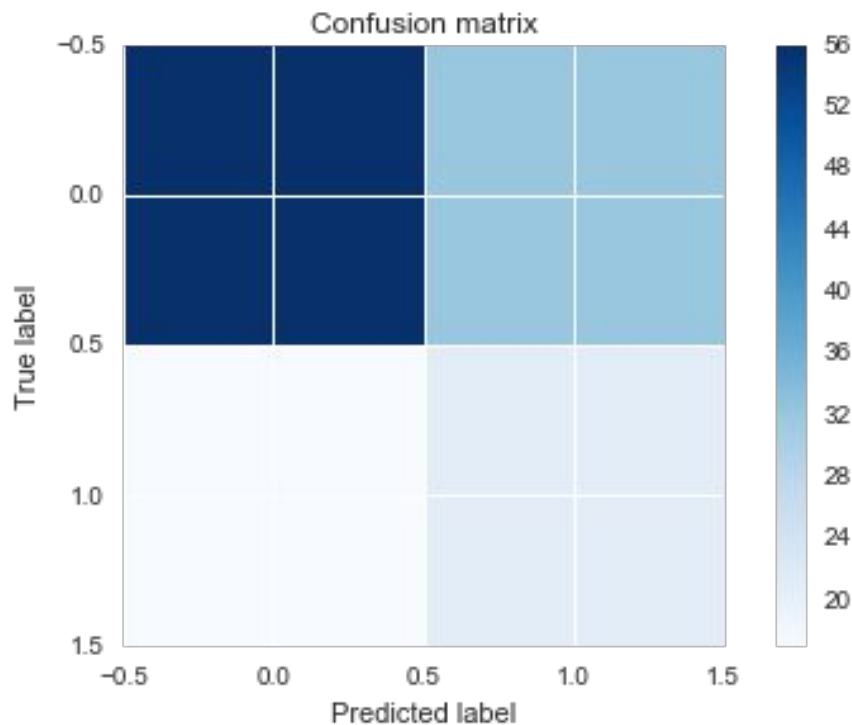
Accuracy: 61.1%

AUC: 60.4%

	precision	recall	f1-score	support
0	0.77	0.64	0.70	88
1	0.40	0.55	0.46	38
avg / total	0.66	0.61	0.63	126

```
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,  
    decision_function_shape=None, degree=3, gamma='auto', kernel='rbf',  
    max_iter=-1, probability=True, random_state=None, shrinking=True,  
    tol=0.001, verbose=False)
```

Confusion matrix



Confusion matrix

```
[[56 32]
 [17 21]]
```

D3 Visualization



Conclusions

- Difficult to predict a true statement!
 - Some limitations
 - Potential bias related to sample selection
 - Not all politicians are checked equally
 - Topic was not taken into account
 - Imbalance of some features
 - Future Directions
 - Collect more data
 - Congressional districts
 - More features reflective of politician
 - Improve D3 Visualization
-

Thank you!

© MARK ANDERSON

WWW.ANDERSTOONS.COM



"Since you're in politics, to save time we're going to use a truth detector."



POLITIFACT

WINNER OF THE PULITZER PRIZE

- Is the statement rooted in a fact that is verifiable?
 - Is the statement leaving a particular impression that may be misleading?
 - Is the statement significant?
 - Is the statement likely to be passed on and repeated by others?
 - Would a typical person hear or read that statement and wonder: is that true?
-