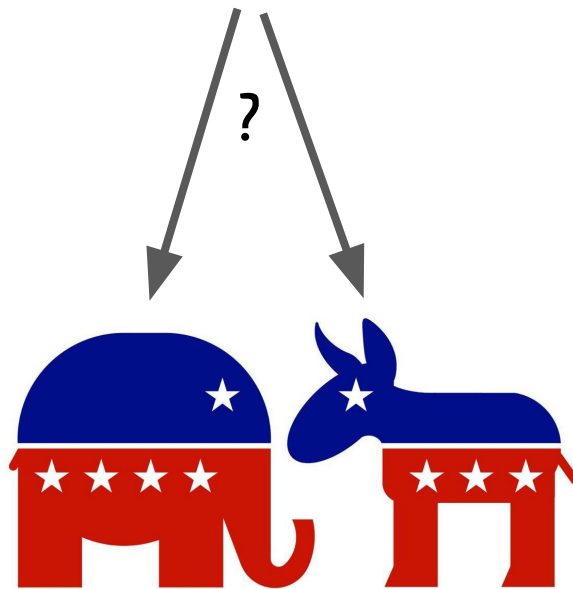# Classifying Voters

Samir Thanedar

The Problem

Can we predict which party someone will vote for based on demographic information and their answers to some personal questions?
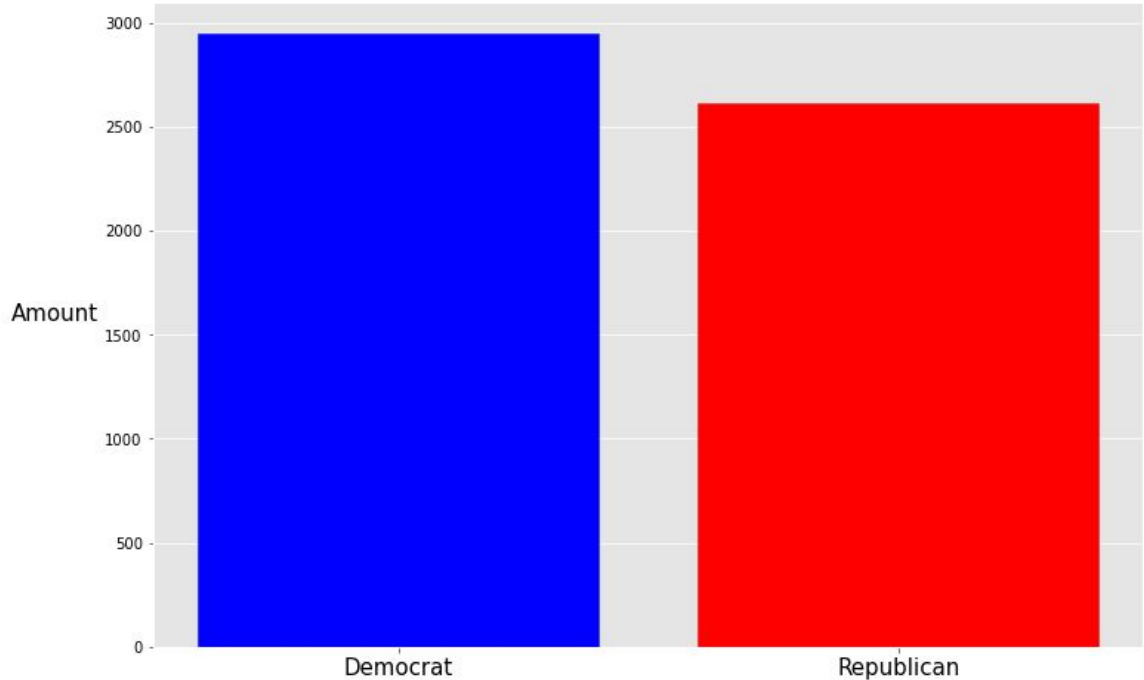
Main Use Case: Political Campaigns

Questions:

1. Where should the candidate spend their time campaigning?
2. Who should we advertise to?

Exploring the Dataset

53% Dem -- 47% Rep

61% Male -- 39% Female

Median Year of Birth: **1983**

Methodology & Workflow

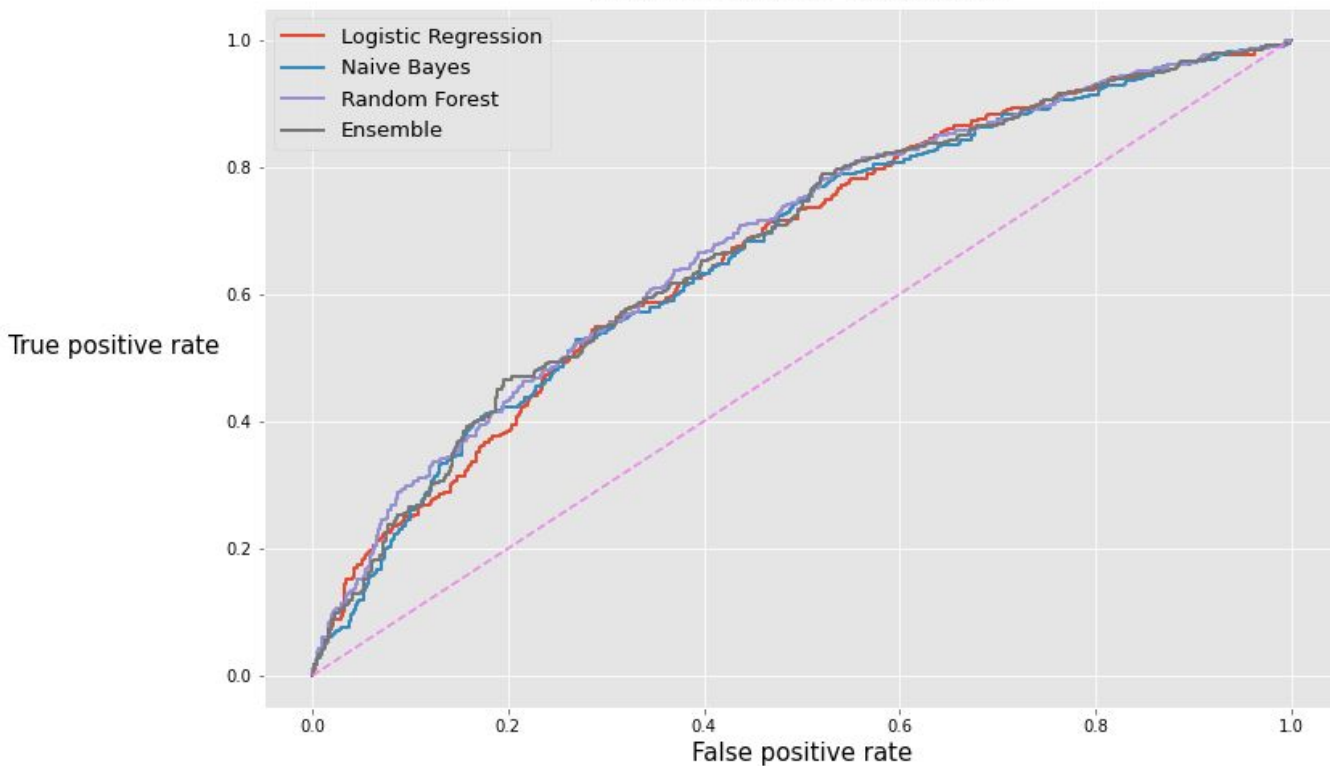Data Source: **Kaggle dataset from Show of Hands polling app**

Amount: **~6000 respondents** were asked three types of questions:

1. Basic Demographic Questions (Age, Gender, Education, Marital Status, & Income)
2. ~100 Personal questions (e.g. Are you an idealist or a pragmatist?)
3. Will you vote democrat or republican in the next election

Classification models used: **Naive Bayes Bernoulli, Logistic Regression, Random Forest**
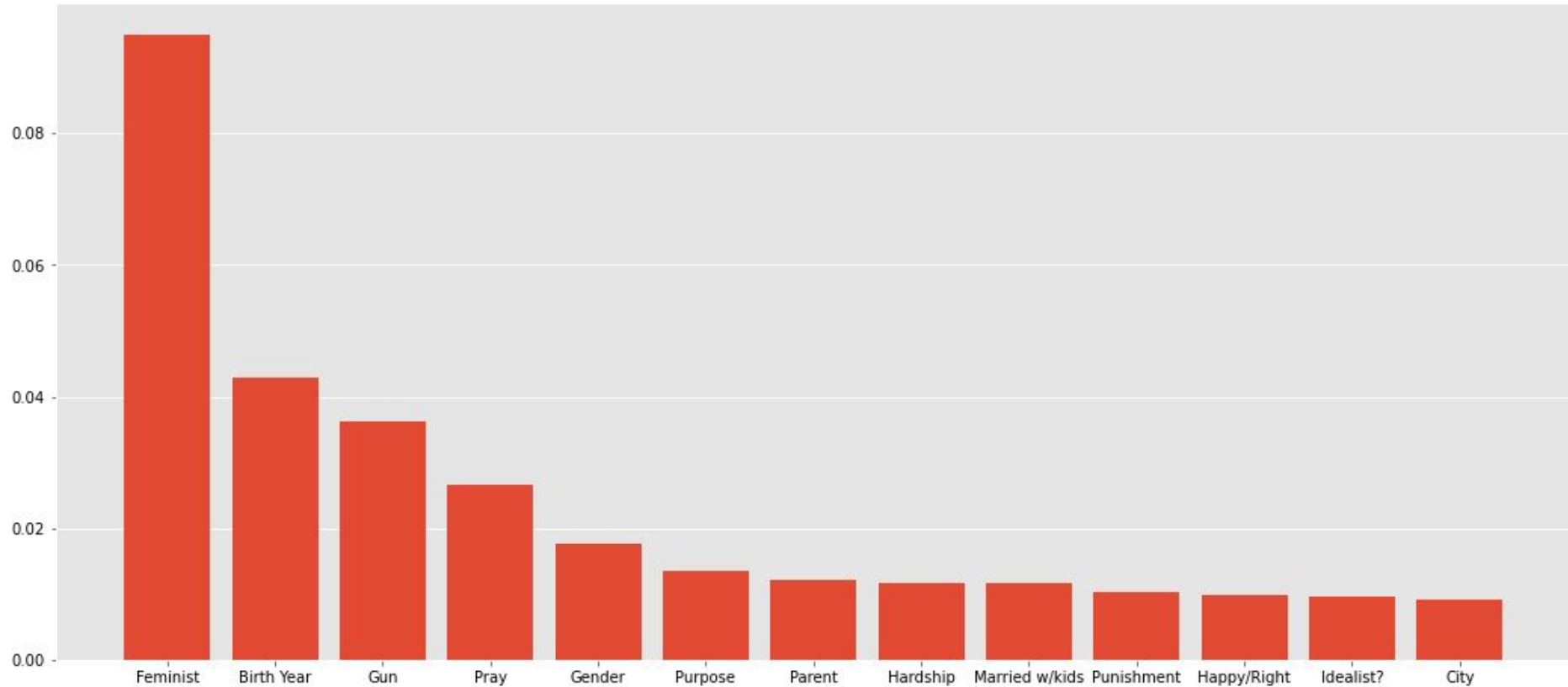
Initial Results

ROC curves for all models

Model	AUC	F1
LR	66.88%	65.47%
RF	**68.10%**	**67.88%**
NB	66.76%	65.40%
ES	67.56%	66.32%

Feature Importance:

Conclusions

- Random Forest is the best model but it overfits slightly so Ensemble model is next best option
- Should we limit false positives or false negatives?
  - It depends. For now we want to do both equally so we choose models with high F1 scores.
- Promising results! With more representative sample & better formulated problem we could generate actionable insights.

Future Work

- Get location data to this dataset. American presidential elections are all about states!
- Develop a better understanding of what needs to be optimized

Thanks!

Questions?

Also - check out my web app and let me know if I predicted the party you intend to vote for this fall. No need to share what that party is!

Appendix