

# Political Party Affiliation

---

Shreya Patel

April 26, 2018

Data Science for Economists

# Introduction

My final project seeks to predict the political affiliation of a person given specific data. The political views of a consumer is highly sought-after information, ranging from public and private organizations across industries and geographical locations.

Social media platforms allow third party vendors to access such information. Recently, many Facebook users learned that the social media giant was categorizing them as

- "Very Conservative"
- "Conservative"
- "Moderate"
- "Liberal"
- "Very liberal"

Facebook allows you to check what category its algorithm puts you in. This algorithm mislabeled my views, which is what got me thinking about this topic. The political parties being predicted are Democrat or Republican currently. However, this may be changed to Liberal or Conservative.

# Data Collection

The data initially collected for this project was from the University of California - Irvine (yay UC!) Machine Learning Respository. There is a set of data on house votes with 435 instances (rows) of data. However, I am currently in the process of searching for data that is more specific to Oklahoma.

At the simplest level, this project is in need of binary classification. This project initially started with a decision tree to predict whether a given person is a Democrat or a Republican based off of voting data. However, single decision trees have a known tendency to over-fit data. In an attempt to correct such behavior while sticking to a singular model, a random forest will be implemented.

As this project is in actively being changed, I will refrain from stating findings for previous versions. However, I do hope that implementing a random forest will increase model accuracy. This will be an interesting comparison.

# Conclusions

Information about political views is highly sought after. This information is picked up and used by social media platforms like Facebook and Twitter to frame what content you see, what advertisements you get, and more. In a previous internship, I used Twitter semantics to learn about social views. Looking at semantics would be interesting to add to this project.



These are some links I found interesting during my research:

- <http://electionanalytics.cs.illinois.edu/>
- <http://time.com/money/5212501/how-facebook-tracks-me/>
- <https://www.smithsonianmag.com/science-nature/study-predicts-political-beliefs-with-83-percent-accuracy-17536124/>
- <http://labs.time.com/story/can-time-predict-your-politics/>
- <https://www.sciencedaily.com/releases/2017/02/170202141851.htm>