2018 House of Representatives Election

By: Noah Hunsicker and Julie Leung

Data Sources

- 2018 MIT Election Data and Science Lab
- Dave Leip's Atlas of U.S. Elections, specifically the 2018 House of Representatives election

Business Proposition

How well can we predict the outcome of the 2018 House of Representatives election?

Criterias:

- The ability to predict the raw outcome (who wins, who loses)
- The ability to predict the winner and the margin of victory for a county

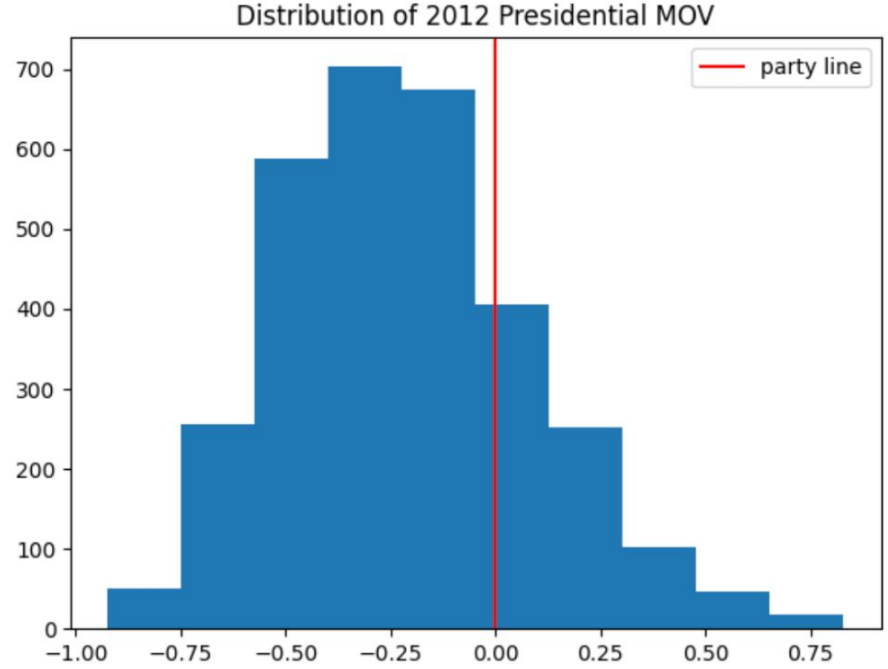
Evaluation

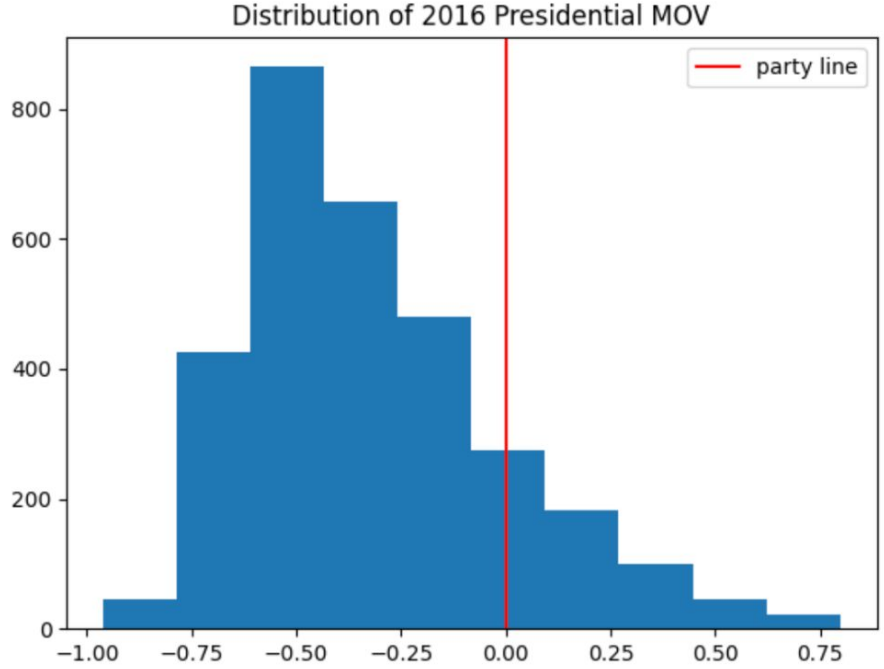
- Created a Margin of Victory scale to determine how likely each political party will win a county
  - MOV = the vote difference between the two major political parties then divided by the total number of votes

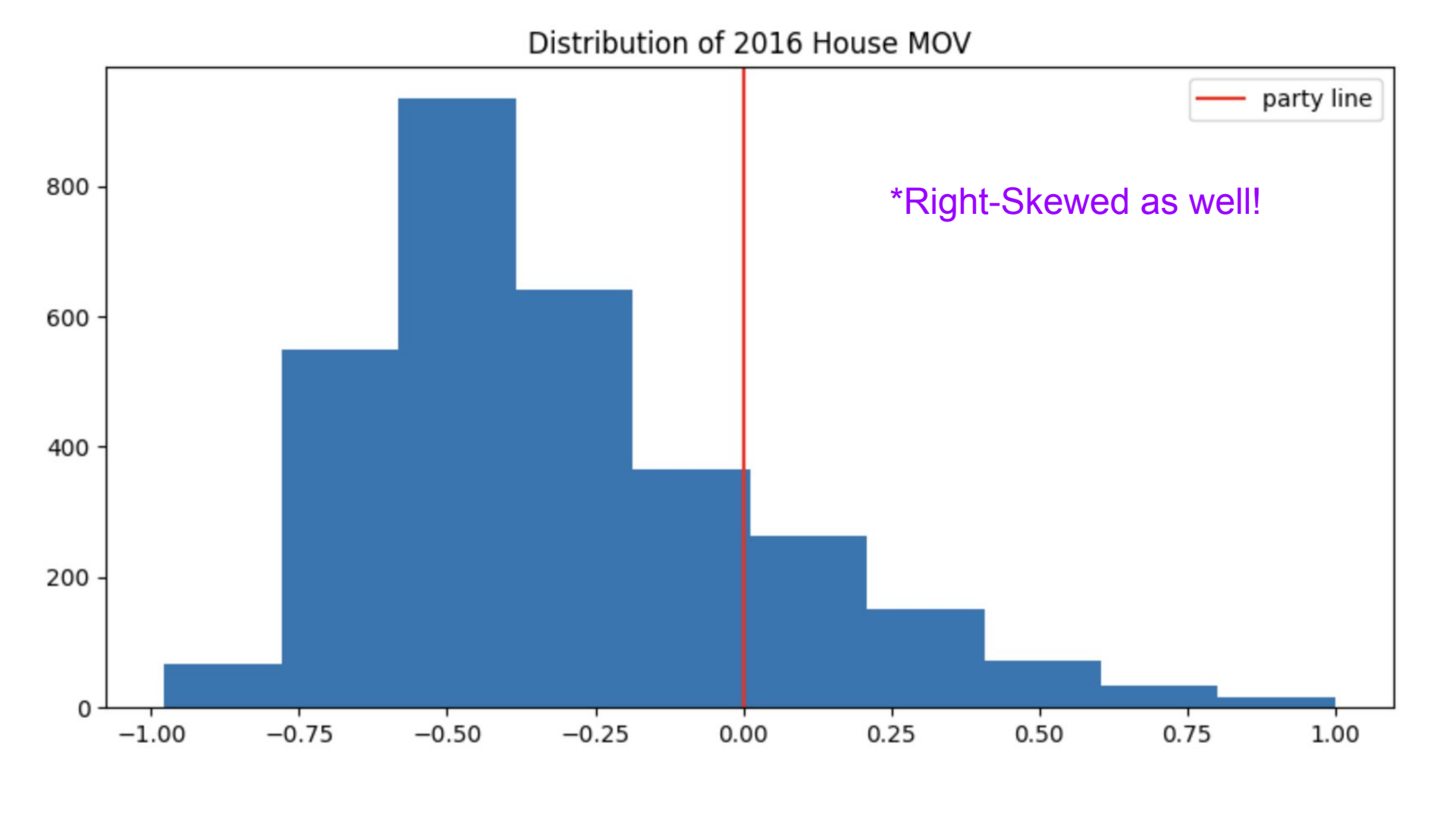
1 [-0.20, -1]	2 [-0.10, -0.20)	3 (-0.10, 0]	4 (0, 0.10]	5 (0.10, 0.20]	6 (0.20, 1]
High MOV for Republican party	Middle MOV for Republican party	Low MOV for Republican party	Low MOV for Democratic party	Middle MOV for Democratic party	High MOV for Democratic party

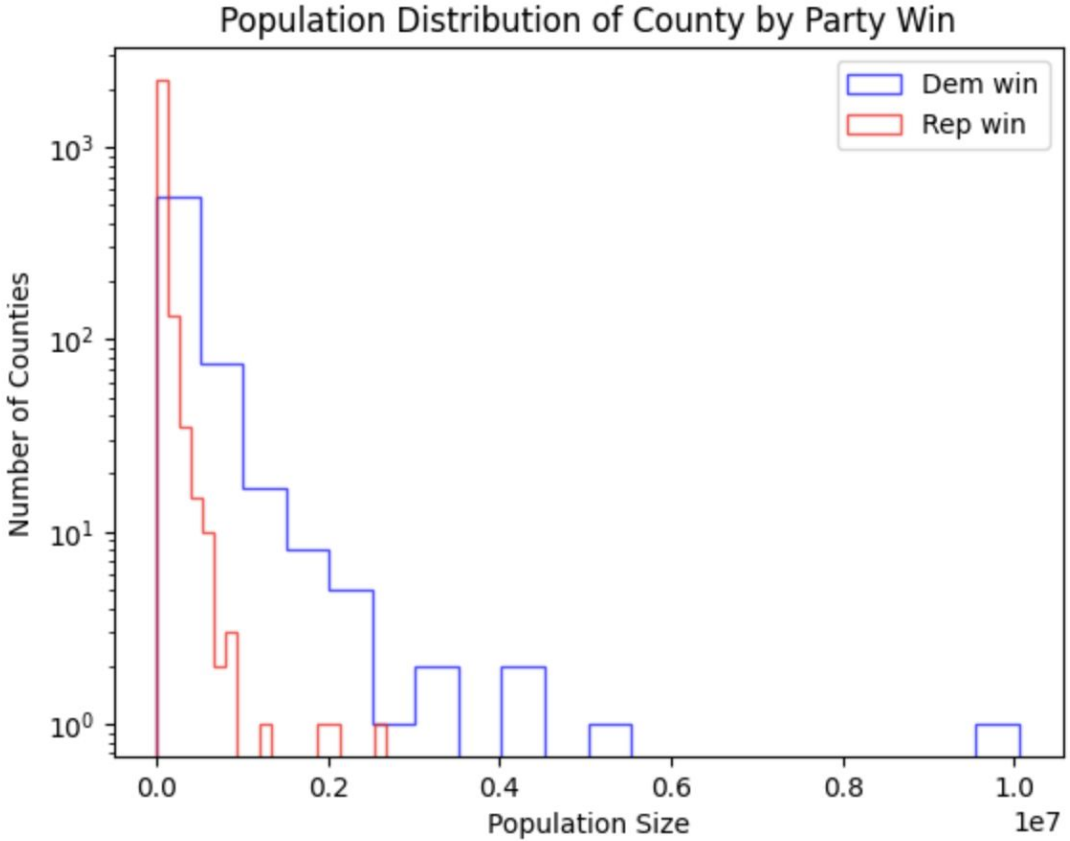
Distributions of MOV in Previous Presidential Elections

*Notice: they are
both right-skewed






Distribution of 2016 House MOV
party line
*Right-Skewed as well!
800
600
400
200
0
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00



- Based on the results, you would assume Republicans are performing much better than Democrats.

- However, it's important to note that the winner of a congressional election is determined by district-wide votes, not the individual county votes. Many small counties might be outvoted by a single large county.

Modeling

We performed the following:

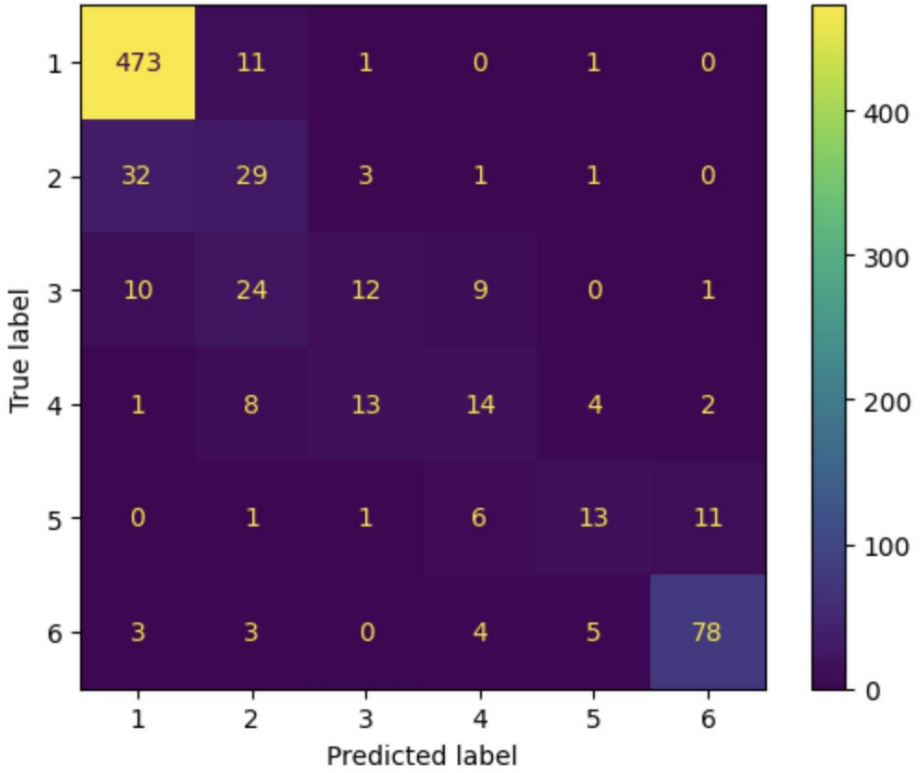
- ❏ Logistic Regression
- ❏ Decision Trees
- ❏ K-Nearest Neighbors

Best Features

- Percentage of female population per county
- Percentage of people with no college degree per county
- Margin of Victory for 2016 Presidential Election
- Margin of Victory for 2012 Presidential Election
- Margin of Victory for 2016 House Election
- Percent of Voting Age Population

Best Model

- Best Model: KNN with an accuracy score of ~80% using the reduced feature sets
- Decision Trees had an accuracy score of ~74.5%
- Logistics Regression with ~74%



KNN Confusion Matrix

Next Steps

- Developing an unsupervised model that is able to accurately predict an election that has not happened yet.
- Make a similar model for Presidential election outcome by county and for Senate election outcome by county.
- Build a model that predicts outcome for the district as a whole, not just predicting the outcome by county.

VO
TE

THANK YOU!