Final Project

Rajveer Singh
CSCI 381 - Applied Data Science

Project Goals!

Understanding the Dataset and motives

- The election dataset contains county level demographics and past election data on federal elections from 2012-2016
- The dataset was put together for analyzing election results in 2018, so features correspond to 2018 data

How can the turnover rates from past elections help determine upcoming election results?

- Turnover rate is the percentage of people that voted out of the population that is eligible to vote
- Every 4 years there is a presidential election to vote for a new president in the United States
- Determining turnover rates for each state is crucial in the world of politics as this would give us an indicator of which states are more likely to vote and which are not
- Helps us understand voter engagement

Hypothesis

- We hypothesis that election turnover rates from previous years will strongly indicate how turnover rates will trend in upcoming elections

Caveats

- We must understand that predicting turnover rates is not a simple science, there are many factors that goes into it that are out of statistical means
- Current Voting Age Population (CVAP) is based off data from 2018 while metrics given are from elections in 2012 and 2016
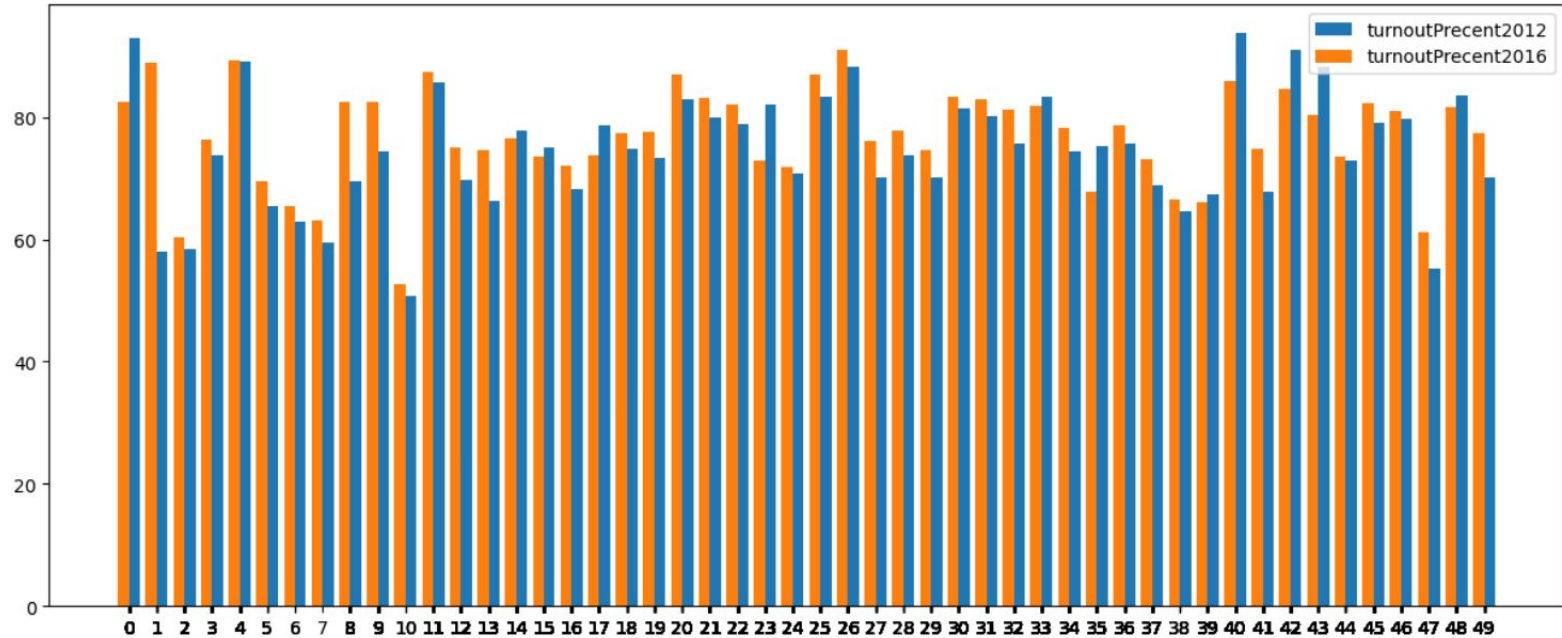
Aims

Aim 1: Analyze the relationships between turnover rates in 2012 election and 2016 election and see how when predicting with new data compares to previous elections.
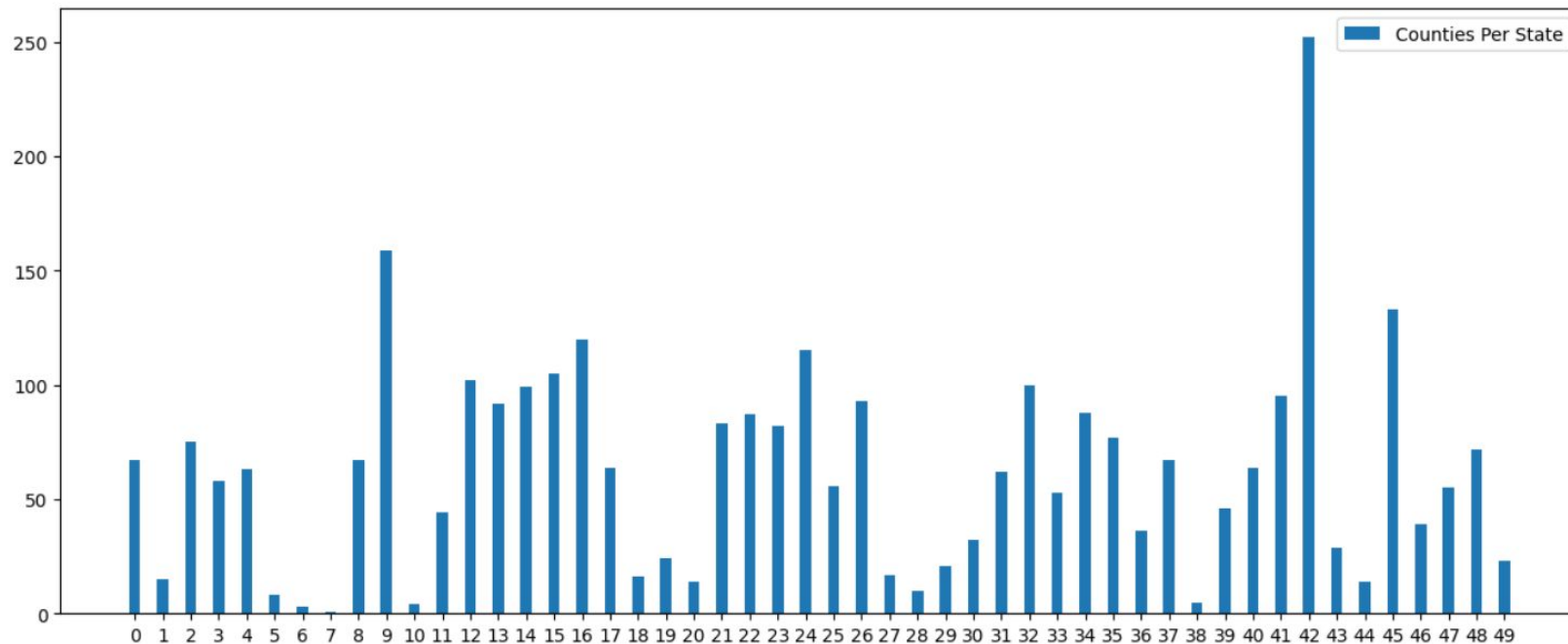
Aim 2: Analyze the feature importance in our models to see which feature impacts our predictions the most.

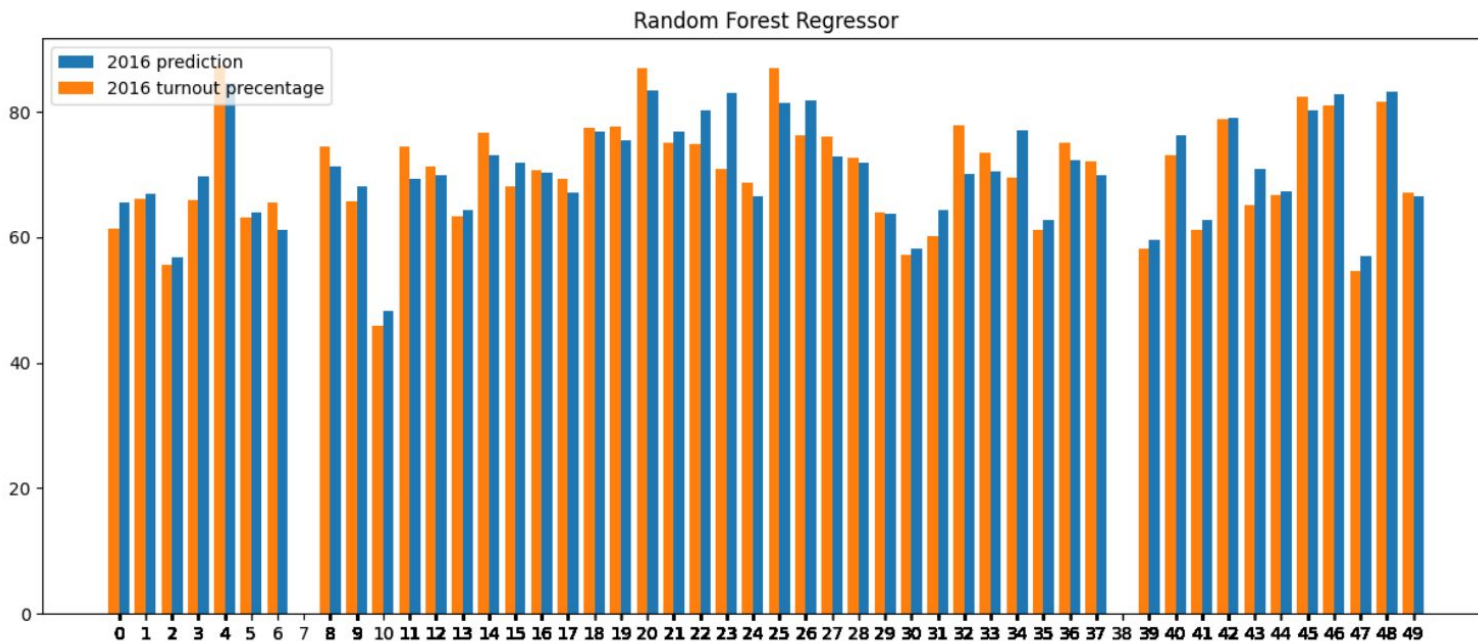# Pre – Model Analysis!

Establishing Relationship

Counties Per State

Model Results!

Linear Regression !!

{'mse': -6.625020116573754,
'mae': -1.7623719735710308,
'r2': -0.9218344514845682}

Linear Regression

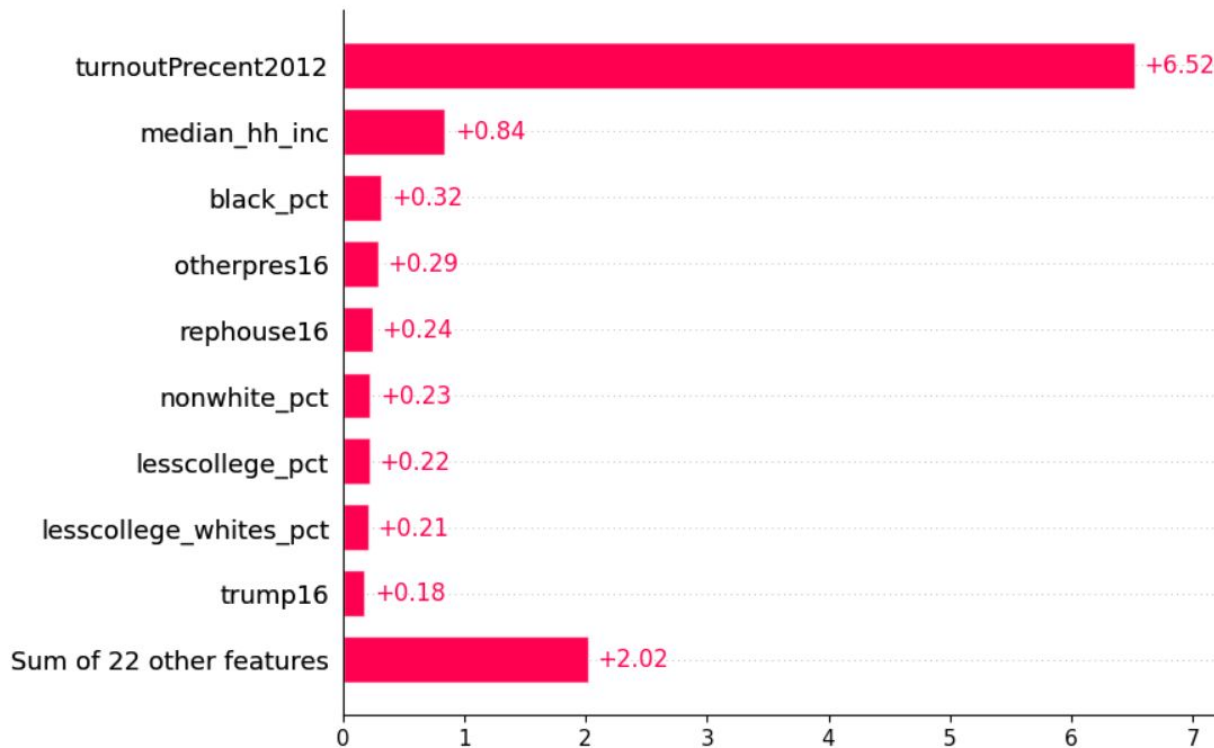# Random Forest Regressor!!

Random Forest Regressor

{'mse': 12.59627083421516, 'mae': 2.6616094241956625, 'r2': 0.8555071557155209}
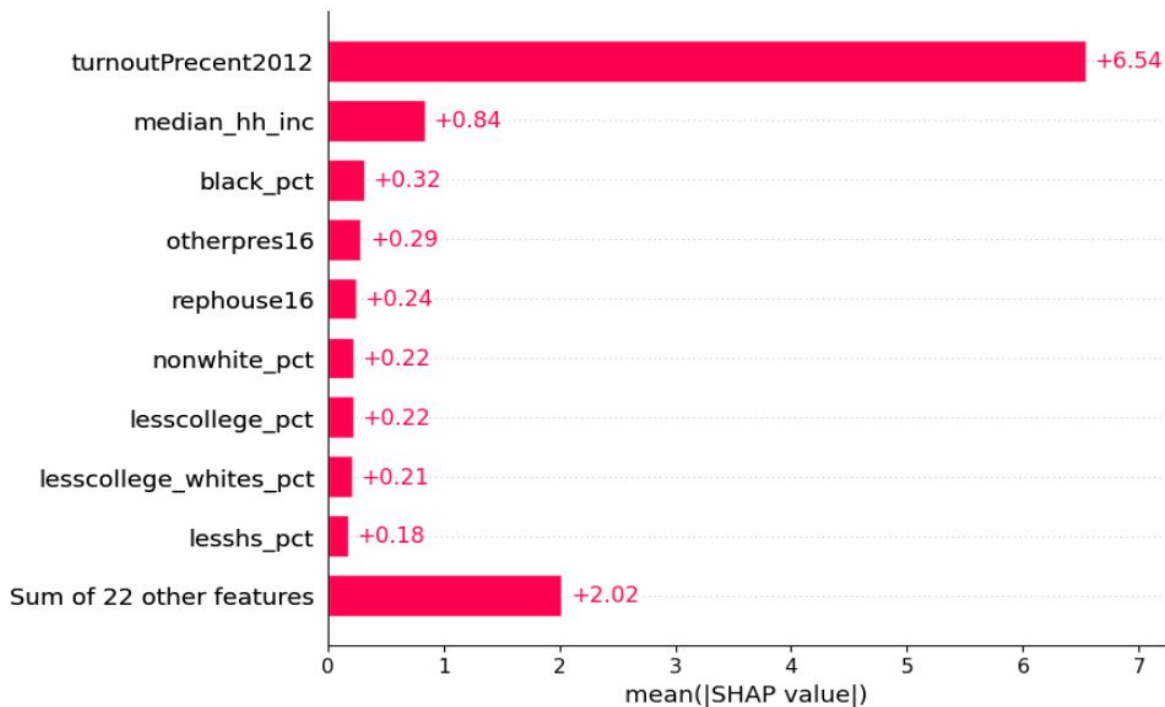
Post - Model Analysis!

Regression SHAP analysis

Random Forest SHAP analysis

Final Thoughts

While the models did predict somewhat well it's clear that it has generalized too much. This is primarily due to only having one past election 2012 in the features.

With the results from 2016 and 2012 they did appear to trend similar to each other which is why the after running the models we see that our results for turnover rates are similar to the previous years.

However it is inconclusive to say that only going off this is a good indicator on how turnover rate percentages can be predicted with only two elections to base it off.