Forecasting the 2022 U.S. Senate Midterm Elections

Madeleine Roberts

Explanatory Variables

01 Polling

02 Historical Voting Patterns

Polling

- Raw data from FiveThirtyEight
- Cleaned data to achieve a training set of nearly 4,000 observations
- Implemented a weighted polling average
  - Recency
  - Sample size
  - Pollster Ranking

Recency Weight:

$$W_r = e^{-1 * 0.01*(Days\ between\ the\ poll's\ conclusion\ and\ election\ day)}$$

Sample Size Weight: $W_s = \left(\frac{Sample\ Size}{600}\right)^{0.5}$

Pollster Ranking Weight:

$$W_p = if(ranking = A+, 13/91,$$

$$if\ (ranking = A, 12/91,$$

...

$$if\ (ranking = C/D, 1/91)$$

Historical Voting Patterns

- Raw data data from MIT Election Lab.
- Democratic two-party vote share in the prior elections for the respective Senate seat, plus special elections.
- Goal: Capture the electoral tendencies of a Senate seat that may not be sufficiently explained by polling or fundraising.

Results

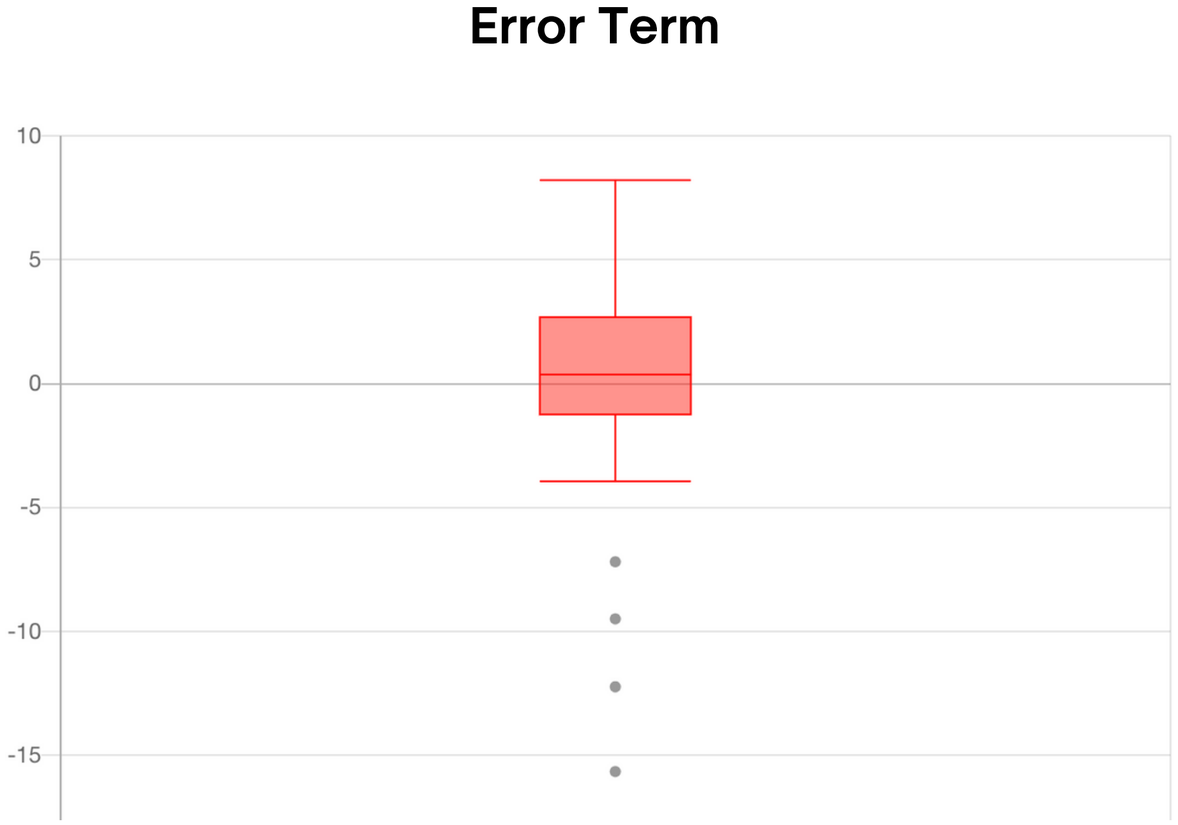
State	Candidate	Predicted Democratic Vote Share	Observed Democratic Vote Share	Obs - Predicted	.	State	Candidate	Predicted Democratic Vote Share	Observed Democratic Vote Share	Obs - Predicted
ALABAMA	WILL BOYD	32.064009	31.62743091	0.43657809		MISSOURI	TRUDY BUSCH VALENTINE	40.601207	43.23770492	-2.63649792
ALASKA	PATRICIA CHESBRO	18.535181	10.32818533	8.20699567		NEVADA	CATHERINE CORTEZ MASTO	49.408816	50.25799794	-0.84918194
ARIZONA	MARK KELLY	51.37999	52.50255363	-1.12256363		NEW HAMPSHIRE	MAGGIE HASSAN	53.18844	54.64759959	-1.45915959
ARKANSAS	NATALIE JAMES	35.295966	32.02479339	3.27117261		NEW YORK	CHARLES E. SCHUMER	62.758338	56.63983903	6.11849897
CALIFORNIA	ALEX PADILLA	61.922278	59.1	2.822278		NORTH CAROLINA	CHERI BEASLEY	47.999326	48.36400818	-0.36468218
COLORADO	MICHAEL BENNET	53.574238	57.510288	-3.93605		NORTH DAKOTA	KATRINA CHRISTIANSEN	18.65358	30.88235294	-12.22877294
CONNECTICUT	RICHARD BLUMENTHAL	59.528393	57.5	2.028393		OHIO	TIM RYAN	46.837988	46.94694695	-0.10895895
FLORIDA	VAL DEMINGS	48.129089	41.71717171	6.41191729		OKLAHOMA	MADISON HORN	35.564696	33.29875519	2.26594081
GEORGIA	RAPHAEL WARNOCK	50.640958	50.45965271	0.18130529		OKLAHOMA	KENDRA HORN	38.450586	36.28865979	2.16192621
HAWAII	BRIAN E. SCHATZ	79.840156	76.74180328	3.09835272		OREGON	RON WYDEN	58.55348	57.04008222	1.51339778
IDAHO	DAVID ROTH	30.637241	46.28820961	-15.65096861		PENNSYLVANIA	JOHN FETTERMAN	52.0428698	51.74180328	0.30106652
ILLINOIS	TAMMY DUCKWORTH	55.121935	57.78229908	-2.66036408		SOUTH CAROLINA	KRYSTLE MATTHEWS	38.731538	38	0.731538
INDIANA	THOMAS MCDERMOTT JR.	45.681715	39.2746114	6.4071036		SOUTH DAKOTA	BRIAN BENGS	33.412876	26.14583333	7.267042667
IOWA	MICHAEL FRANKEN	44.602542	43.6	1.002542		UTAH	NA_UT	5.853716	NA	#VALUE!
KANSAS	MARK HOLLAND	36.85949	38.1443299	-1.2848399		VERMONT	PETER WELCH	61.388658	70.88082902	-9.49217102
KENTUCKY	CHARLES BOOKER	42.898871	38.2	4.698871		WASHINGTON	PATTY MURRAY	56.161343	57.01402	-0.852677
LOUISIANA	MULTIPLE DEMS	33.457296	33.91933816	-0.46204216		WISCONSIN	MANDELA BARNES	51.446856	49.498998	1.947858
MARYLAND	CHRIS VAN HOLLEN	58.614327	65.8	-7.185673						

Skills Used

- Data Collection and Cleaning
  - Managing Data Quality/Missing Observations
- Ordinary Least Squares Regression and Analysis
- Two-fold Cross Validation

Final Results

- Average Predicted Error = 1.7%
- Second Best Model in Class of Second-Year Students
- Interviewed for a Harris Profile on the Assignment



Population size: 34
Median: 0.368822305
Minimum: -15.65096861
Maximum: 8.20699567
First quartile: -1.3284198225
Third quartile: 2.89129668
Interquartile Range: 4.2197165025
Outliers: -15.65096861 -12.22877294 -9.49217102