Abigail McDonald

October 17, 2022

FLUX Voter Analysis for Targeting

Introduction:

The FLUX voter file is helpful for modeling the predictive behavior of eligible voters. A variety of models can be created from the demographics and survey data contained in the file. These models will help the Republican candidate campaign team determine a volunteer base, voters to target based on persuadability and likely votes, probable partisan voters to be targeted during Get Out The Vote, and which messaging will be most helpful in targeting those voters.

Summary:

In order to effectively target voters, five models were created from the FLUX voter file. The first model is Republican_partisanship_prediction which predicts whether a voter is a Republican or not. Next is the Candidate_Support_Prediction model, which predicts a voter's support for the Republican candidate. Then, the Voter_Persuadability model predicts whether a voter is persuadable to the Republican candidate. In other words, whether the voter moves to be more supportive of the Republican candidate between the first and second identification waves. The fourth model, the Turnout_prediction model, predicts whether a voter will vote in the upcoming election. The final model, Msg_A_B_moved, predicts whether a voter will change their vote based on message A or message B or is not likely to change their vote based on those messages. With trained models, new voters can be scored and added to one of the following lists: volunteer, persuadable voter, responsive to message A or B, or get out the vote (GOTV).

Recommendations:

Together the new predicted model scores allowed for the creation of campaign recommendations. A volunteer list was created from the partisanship model and the candidate support model. The list shows voters who are predicted to both support the candidate and identify as Republicans. In addition, the list shows all voters who have a greater than 60% score in both metrics. However, a cutoff can be imposed at any percentage level (higher is better) in reaching out to potential volunteers.

Voters who show greater than 60% probability for both party and candidate support

VOTER_ID	Republican Party	Support Candidate	Potential Voluntee
569	61%	100%	Y
806	66%	68%	Y
1203	62%	74%	Y
1612	61%	100%	Y
1819	62%	100%	Y
2014	66%	100%	Y
2087	63%	74%	Y
2276	62%	66%	Y
2278	64%	66%	Y
2389	60%	88%	N
3159	61%	66%	Y
3422	60%	71%	N
3764	65%	99%	Y
3854	60%	88%	N
4492	64%	90%	Y
4721	61%	88%	Y
4746	66%	100%	Y
4914	66%	71%	Y

Potential Volunteers

	A	B	C	D
1	Voters by Persuadability score (target range >=.50)			
2	VOTER_ID	Persuadable	Voted in Similar Election	Persuadable Likely Voter
3	476978	0.94	0.9	Y
4	378572	0.91	0.9	Y
5	502498	0.91	0.89	Y
6	67724	0.85	0.95	Y
7	120445	0.88	0.89	Y
8	234557	0.85	0.92	Y
9	570289	0.92	0.85	Y
10	576455	0.83	0.94	Y
11	92528	0.87	0.89	Y
12	345437	0.87	0.89	Y
13	343688	0.86	0.9	Y
14	380869	0.86	0.9	Y
15	246688	0.86	0.89	Y
16	29024	0.9	0.85	Y
17	272637	0.85	0.9	Y
18	594306	0.79	0.96	Y
19	156480	0.86	0.88	Y

Persuadable Voters

Next, a persuadable voter list was created by multiplying the voter's persuadability score by their turnout score, providing an overall persuadable likely voter score (recommended target range >50%). Persuadable voters should be contacted by all available means.

	A	B	C	D
1	Which message will help to persuade voters?			
2	VOTER_ID	message A	message B	Which Message?
3	2	-0.02	0	-
4	84	0	-0.02	-
5	95	0.01	-0.02	A
6	146	-0.01	-0.01	-
7	179	-0.01	0	-
8	234	0.03	-0.03	A
9	368	0.01	-0.01	A
10	371	0	0	-
11	401	0	-0.01	-
12	431	0	-0.01	-
13	448	0	-0.01	-
14	500	-0.01	0.01	B
15	511	0	0.02	B
16	549	0.01	-0.01	A
17	550	0.05	0.02	A
18	586	0	0	-
19	694	-0.01	-0.01	-
20	705	0	0	-
21	802	0	0.01	B

Messaging uplift

Additionally, the Msg A/B moved list shows potential voters predicted to respond positively due in part to either message A or B. Thus, these voters should be targeted with messaging of the appropriate type.

	A	B	C	D
1	Voters who likely support the candidate (>50%) but are moderately likely to vote (between 30% and 60%)			
2	VOTER_ID	Support Candidate	Voted in Similar Election	GOTV Candidate
3	2	92%	32%	Y
4	6	92%	29%	Y
5	7	49%	81%	N
6	8	44%	44%	N
7	9	11%	31%	N
8	10	16%	31%	N
9	12	11%	25%	N
10	13	71%	29%	Y
11	14	100%	88%	Y
12	16	100%	58%	Y
13	17	64%	30%	Y
14	18	98%	84%	Y
15	20	38%	38%	N
16	21	22%	89%	N
17	22	100%	94%	Y

GOTV

Finally, the GOTV list shows voters who are likely for the candidate but only moderately likely to vote. This list was created by looking for voters with high candidate support >0.5 and medium turnout score (0.3 to 0.6). These voters should be targeted during the campaign's Get Out The Vote phase.

Technical Details:

Recommendation files could be formed because all target variables (except for messaging) were derived from the same base variables and thus had similar null fields. SQL queries with table joins would have been necessary otherwise. For all models, all dependent variables except for the target variable were dropped before modeling. Additionally, all null record values in the target variable field were dropped. Yes and No variables were turned into numerical 0/1s to allow regression modeling. Then, a train/test split was used for the data frame based on the set numbers available in the voter file (sets 1 and 2 were the training set, and set 3 was the testing set). The train/test split was not arbitrary for each model further aiding in the creation of recommendation files.

A decision tree model was used for candidate support predictions, with the target variable CAND1_LR2. After the model was built, it was evaluated with a classification report. The model has an f1-score of 73%. Additionally, HH_ND at 47.8% and HH_NR at 11.87% were the features that most contributed to the model. As this was a model to predict candidate support, this makes sense. The number of registered Democrats and Republicans in a household would indicate support preferences. The decision tree model was scored and output in a CSV file.

The messaging model focuses on people who received a message and moved to be more supportive of the Republican candidate. The predictor variable used was MOVED_AR. In addition, a Random Forest model was built to allow for tweaking the uplift for each message with MSG_A and MSG_B as indicator variables. Uplift was calculated separately for each indicator and then exported to CSV files.

A combined model was used to predict Republican partisanship (party affiliation). The two models in the ensemble were a decision tree and a logistic regression. Initially, in addition to dropping all dependent variables but the target, R2 and party affiliation variables were dropped from the predictors. Then, the models were built independently and evaluated (DT f1 score of 64% and LR f1 score of 65%). Then, using the max voting method, the models were combined into an ensemble model. This ensemble was evaluated for lift by quintile, scored, and output to a CSV file.

The fourth model, predicting voter turnout, used the dependent variable VG_14_DV. This turnout is for a non-presidential election year, so all variables derived from the current year, presidential elections, and future elections were dropped. A decision tree model was fit to the data and evaluated (f1 score 75%). The model was then scored and exported to a CSV file.

The final model focuses on people who moved toward the Republican candidate between waves. The predictor variable used was MOVED_AR. Movement between waves indicates persuadability. A Random Forest model was fit to the data and had an f1score of 91%. The model was scored and exported to a CSV file.

After opening all pertinent CSV files, specific predictors were joined with others to indicate needed voter lists (see Recommendation section). Next, Excel formulas (mostly

IF/AND) combined data to create specific voter target lists. All new files were then saved as Excel workbooks preserving the formulas for future use.