Predicting the Political Leaning of Individuals in the 2016 Elections

Sachin Sridhar, Jeffrey Fossett, Saran Liukasemsarn, Elliot Smalling

April 24, 2018

Introduction

Information regarding political inclination of individuals is vital for political organizations to best understand how to identify and enhance their support base. Core concepts of generalized linear models can be applied in this regard to predict the political leaning of an individual, based on a given set of features. The objective of this project is to build a model that most accurately predicts the political inclination of an individual (Democratic or Republican), based on the individual's biographical information and some behavioral and financial attributes such as purchase preferences and social activities. Since this is a binary classification problem, techniques such as logistic regression and additive models have been applied to arrive at the best predictive model.

Data Cleaning and Exploration

The data set for our analysis comes from Blue Labs on Kaggle. The data set contains 10,439 observations, with 47 predictor variables and one outcome `support_dem`, which takes on the value 1, if the individual self-reported supporting a Democratic candidate in 2016, and 0 otherwise. Of the 10,439 observations, 3,975 (or about 38%) report supporting a Democratic candidate.

The first step of our analysis is to clean the data set to prepare it for modeling. First, we notice that several of the predictor variables are perfectly collinear. For example, each observation has a `1` for exactly one of the three predictors `density rural`, `density urban` and `density suburban`. We combine these predictors into a single categorical predictor for "density" (to which we can later apply contrast coding) so that model coefficient can be identified. Similarly, a new factor variable "Marital Status" has been developed, which takes on the values "Single", "Married" or "Unknown".

Next, we notice that many of the predictor variables have `NA` (missing) values. For quantitative predictors with missing values (`age`, `cnty_pct_religious` and `cnty_pct_evangelical`), we use the `na.convert.mean` function from previous homework assignments to impute the mean age for the missing observations, and create new predictors (`age.na`, `cnty_pct_religious.na` and `cnty_pct_evangelical.na`), which are indicators for whether the observation was missing. We notice that a substantial share of our observations (942 or around 9%) have missing `education` information. To handle this, we add a new "unknown" level to the education factor for these observations.

Exploring Quantitative Predictors

The next step is to explore the quantitative predictors in the data set. We present histograms and summary statistics for these predictors below. In the histograms, we notice that many of these variables, namely, median income, PPI, and the number of children (which is not shown), are right skewed, and thus, the log transformation has been applied on them, to give a more symmetric result.

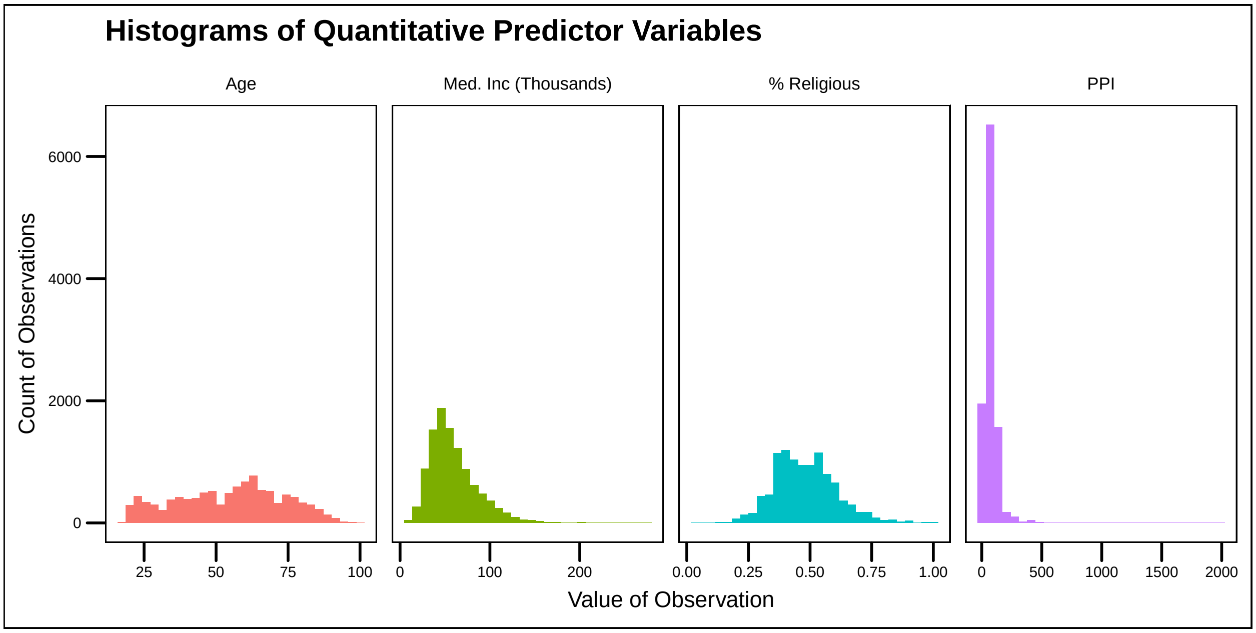
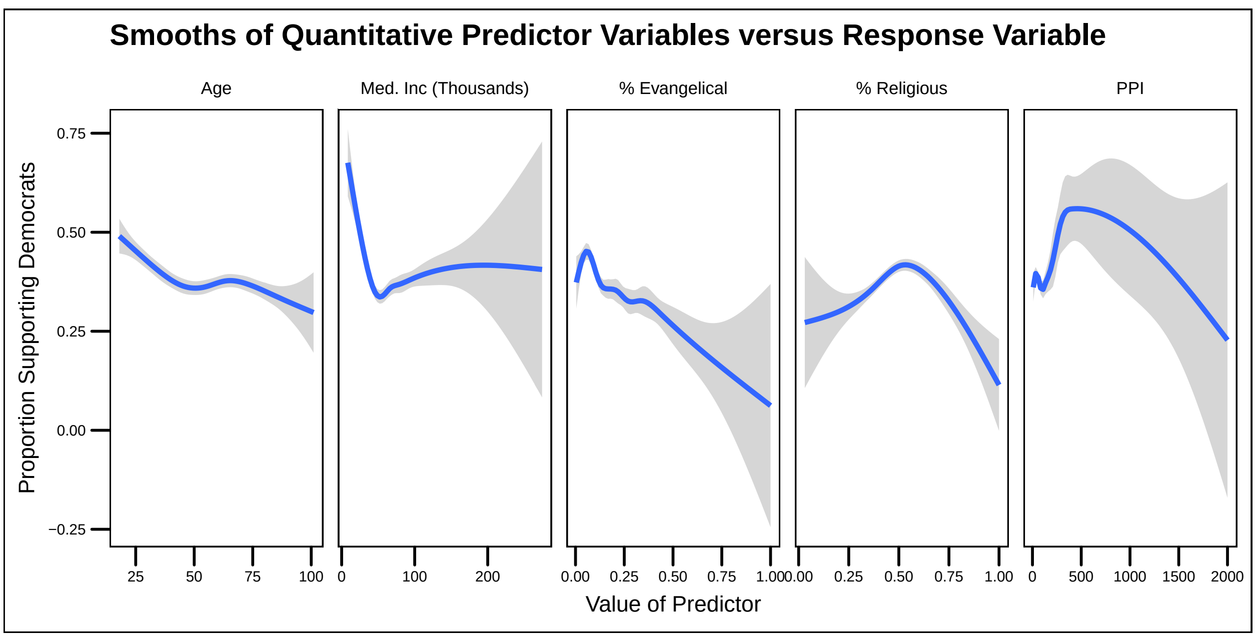


Table 1: Summary Statistics for Quantitative Predictors

Variable	Min	25th Pctile	Mean	Median	75th Pctile	Max	SD
Age	18.00	40.00	54.77	57.00	69.00	101.0	18.64
PPI	8.00	44.00	77.37	64.00	89.00	2000.0	86.60
% Evangelical	0.00	0.08	0.16	0.12	0.21	1.0	0.12
% Religious	0.03	0.39	0.48	0.47	0.56	1.0	0.13
Med Income (thousands)	8.43	40.65	59.65	53.74	72.79	274.4	27.60

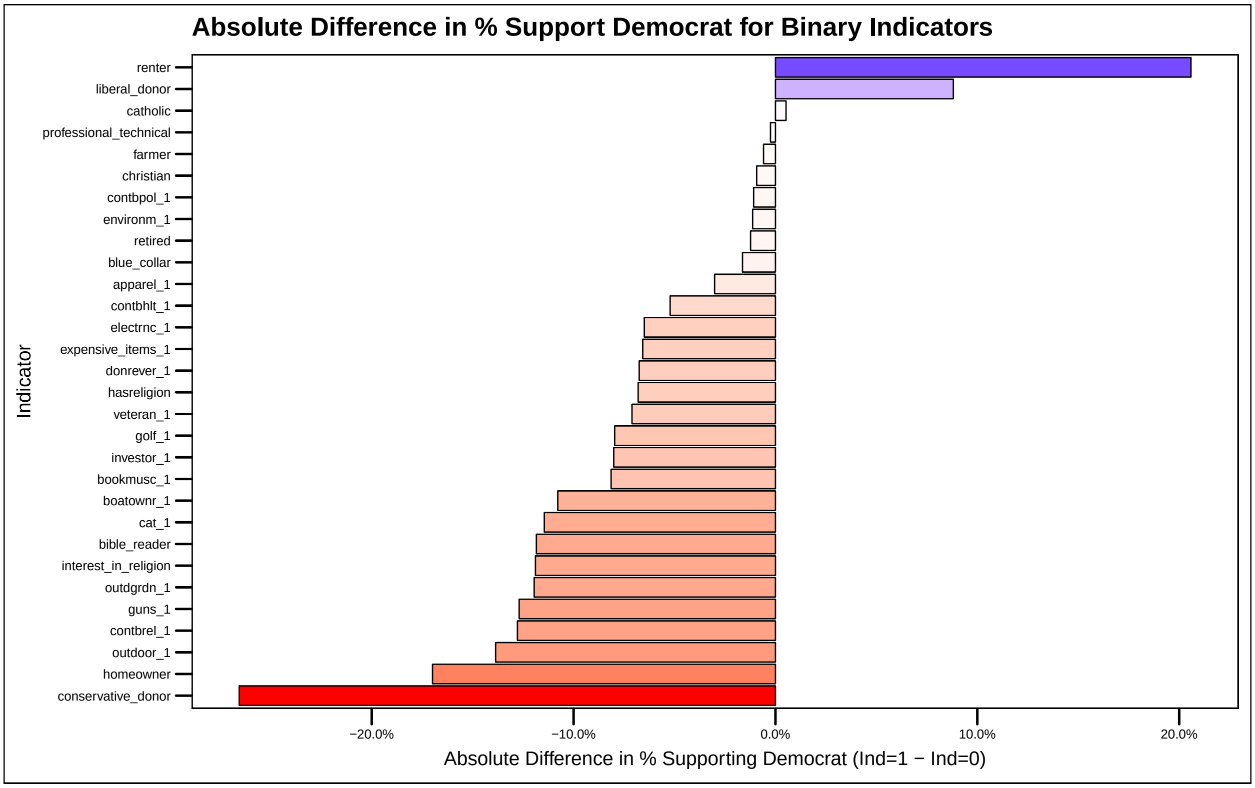
The next step is to create plots describing the association between each predictor and the outcome. Localized smoothers have been used to approximate the proportion of voters supporting democrats, at a given value of the quantitative predictor variable.



We see that for all variables there is an overall downward trend in the proportion of voters supporting Democrats as the predictor value increases. Additionally, we observe strong evidence of a non-linear relationship between the predictor variables and the response.

Exploring Binary Predictors

Next, we present summaries of binary indicator variables in the data set. First, we present a plot showing the absolute difference in the percentage of individuals who report supporting a Democratic candidate in 2016, corresponding to each of the binary indicators. The results are sorted by the strength of association between the indicator and the outcome variable. In the appendix, we also provide more detailed summary statistics on the binary outcomes.



Most results from the plot are as we would expect. For example, we see that renters (a group which presumably skews younger and more cosmopolitan) and people who are liberal donors are substantially more likely to support a Democratic candidate. By constrast, being a homeowner or a conservative donor is strongly negatively associated with supporting a Democratic candidate. Many of the other binary indicators in the dataset (e.g. relating to religion, gun ownership, consumer goods etc.) seem also to be negatively associated with supporting a Democrat.

Exploring Categorical Predictors

After combining mutually exclusive binary variables as described above, the dataset contains five categorical variables. The different levels of these categorical variables and their proportions of the dataset are presented in Table 2 below, as well as the proportion of each level that supports Democrats. Based on this analysis, ethnicity seems to be predictive of political leaning, with 69% of African Americans in the dataset supporting Democrats, but only 30% of white people. With the density variable we see a similar divide, with over half of urban residents supporting Democrats but only 24% of rural residents. The impact of Education on political leaning is less clear, with relatively similar proportions among the groups, and with `no hs degree` and `post graduate degree`, the least and most educated, being the most likely to support Democrats.

We also see the effect of missing or imbalanced categorical variables in the table below. While the proportion supporting Democrats of the `other` marital status and `Unknown` sex levels is starkly different from other levels in their respective categories, they both are underrepresented in the dataset, so it will be challenging to understand their relationship with the outcome variable.

Table 2: Summary Statistics for Categorical Variables

Variable	Level	% Total	% Support Democrat
combined_ethnicity_4way	W	0.74	0.30
combined_ethnicity_4way	H	0.13	0.53
combined_ethnicity_4way	A	0.01	0.56
combined_ethnicity_4way	B	0.11	0.69
density_clean	rural	0.25	0.24
density_clean	suburban	0.55	0.39
density_clean	urban	0.21	0.53
education	bach degree	0.21	0.36
education	high school	0.28	0.37
education	unknown	0.09	0.38
education	some college	0.25	0.38
education	post graduate	0.11	0.42
education	no hs degree	0.06	0.44
marital_status_clean	other	0.00	0.14
marital_status_clean	married	0.83	0.35
marital_status_clean	single	0.17	0.52
sex	M	0.46	0.33
sex	F	0.53	0.42
sex	U	0.01	0.53

Multivariate Analysis and Modeling

Analysis Plan

Following the data cleaning process and the exploratory data analysis, the data set is ready for modeling. To allow us to estimate out-of-sample performance, we split the Kaggle data into a training and test set in an 80-20 ratio. Our modeling strategy is to first build a baseline model, and then to improve on it by considering different link functions (comparing the results using the logit, probit and the complimentary log log links), including interaction terms, as well as using smoothers and additive models.

Building a Baseline Model

To arrive at a baseline model, we first fit a logisitic regression including all of the available predictor variables (see appendix for this model), and then use (1) the variance inflation factor (VIF) and (2) analysis of deviance to develop a more parsimonious model.

Variance Inflation Factor

We first use the variance inflation factor (VIF) to test for collinearity. Given the presence of several categorical variables, we use the adjusted VIF output to account for the predictor degrees of freedom. Conventionally, a score of 10 or above indicates that a variable is highly collinear and should be removed from consideration.

In the table below, showing the highest VIF scores in our initial model, we see that no variables reach this threshold and thus no predictors are removed based on this test.Still, while we did not remove any variables because of this, we can see, for example, that `homeowner` and `renter` may be redundant, as are the several religion-related variables also showing higher VIF scores. We would be concerned if all of these variables appeared in our final model together.

Table 3: Top 10 VIF Scores for Predictor Variables

Variable	Adjusted GVIF
homeowner	3.479027
interest_in_religion	3.362280
renter	3.101524
bible_reader	2.928526
hasreligion	2.613938
outdoor_1	2.282493
catholic	2.081634
outdgrdn_1	2.048112
donrever_1	2.026103
investor_1	1.964665

Analysis of Deviance

Our initial logistic regression model has many non-significant coefficients. Hence, we run an analysis of deviance procedure to see which variables can be excluded from the model. In particular, we use a stepwise backward selection procedure that works as follows: assume we start with a model with n predictors. Exclude each one of them, one at a time, and perform a likelihood ratio test for each of the models with $n-1$ predictors, and select the model which produces the most statistically insignificant result (highest p-value greater than 0.05). This would represent the scenario that one of the predictors has been excluded without a loss in the predictive power of the model. This process is repeated until all likelihood ratio tests performed with the exclusion of one of the predictor variables yields a statistically significant result (p-value less than 0.05), and hence, we would retain all the remaining predictors. After running this variable selection procedure, we arrive at a model with 19 predictors, which is summarized in Table 4.

Table 4: Baseline Logit Model Summary

Predictor	Est	Std Err	t-Stat	P-Value	Sig?
(Intercept)	-0.23	0.22	-1.04	0.30	
sexM	-0.40	0.05	-8.02	0.00	***
sexU	-0.27	0.26	-1.01	0.31	
combined_ethnicity_4wayB	0.92	0.21	4.30	0.00	***
combined_ethnicity_4wayH	-0.05	0.21	-0.23	0.81	
combined_ethnicity_4wayW	-0.75	0.20	-3.66	0.00	***
log_nchildren	-0.03	0.01	-2.59	0.01	**
educationhigh school	0.02	0.07	0.31	0.76	
educationno hs degree	0.09	0.11	0.78	0.44	
educationpost graduate	0.36	0.09	4.05	0.00	***
educationsome college	0.04	0.07	0.53	0.60	
educationunknown	-0.07	0.10	-0.68	0.49	
hasreligion	-0.19	0.08	-2.36	0.02	*
catholic	0.26	0.08	3.32	0.00	***
christian	0.28	0.12	2.40	0.02	*
interest_in_religion	-0.44	0.09	-5.12	0.00	***
donrever_1	0.16	0.06	2.53	0.01	*

Predictor	Est	Std Err	t-Stat	P-Value	Sig?
liberal_donor	0.89	0.09	9.55	0.00	***
conservative_donor	-1.56	0.31	-5.02	0.00	***
contbrel_1	-0.60	0.13	-4.56	0.00	***
apparel_1	0.12	0.05	2.19	0.03	*
cat_1	-0.21	0.07	-3.10	0.00	**
outdoor_1	-0.07	0.07	-0.94	0.35	
guns_1	-0.23	0.06	-3.69	0.00	***
cnty_pct_evangelical	-1.14	0.24	-4.81	0.00	***
density_cleansuburban	0.47	0.06	7.21	0.00	***
density_cleanurban	0.69	0.08	8.40	0.00	***
marital_status_cleanother	-1.41	1.11	-1.27	0.20	
marital_status_cleansingle	0.25	0.07	3.62	0.00	***

Examining the estimated coefficients in the baseline model, we see that they largely confirm our findings from our initial exploration of the data. For example, the coefficients for African American and White ethnicities are very significant and are directionally consistent with what we initially postulated: the positive coefficient for African Americans indicates an increased probability of supporting Democrats, while the negative coefficient for White indicates the opposite. The positive effect of living in an urban area on the likelihood of supporting Democrats is similarly confirmed in this model. Interestingly, of the continuous variables in the data set, only `cnty pct evangelical` was included in this model. The smooth of this variable showed an approximately linear relationship with probability of supporting Democrats, reinforcing its inclusion in the model. The other continuous variables, which were found to have a non-linear relationship to the response, are not included.

To create a benchmark to assess predictive power, below we calculate the discrepancy measure for our baseline model on both the training and the held-out test dataset.

Table 5: Discrepancy of Predictions: Baseline Model

Data	Discrepancy
Train	0.5924
Test	0.5953

We see our baseline model has a discrepancy of roughly `0.59` on the test set. We attempt to improve on this performance below.

Improving the Model (GAM and Interactions)

Recognizing that continuous predictors in our data set may have non-linear relationships to the response variable, we attempt to improve upon the logistic regression model by fitting a generalized additive model (or GAM). We again use the stepwise analysis of deviance procedure described above, including all predictors, as well as the smooths of the continuous predictors. Additionally, several interactions between the categorical variables that had the potential to be predictive have been handpicked, based on the initial data exploration.

Incorporating interactions of variables into the model proved challenging, given our treatment of missing values of the categorical predictors. For example, `sex` is a strong candidate to interact with other variables, as the way other variables influence political beliefs may affect men and women differently. However, since the observations missing `sex` had been coded as a separate category, there wasn't sufficient data to estimate the effect of this `Usex` "Unknown sex" interacted with another variable. Still, some interactions did not pose this problem and were included in the stepwise procedure. However, none were included in the final model. The results of our final GAM model are presented in Table 6.

Table 6: GAM Model Summary

Predictor	Est	P-Value	Sig?
(Intercept)	-0.25	0.27	
sexM	-0.41	0.00	***
sexU	-0.27	0.32	
combined_ethnicity_4wayB	0.90	0.00	***
combined_ethnicity_4wayH	-0.08	0.70	
combined_ethnicity_4wayW	-0.76	0.00	***
educationhigh school	-0.01	0.86	
educationno hs degree	0.03	0.77	
educationpost graduate	0.35	0.00	***
educationsome college	0.02	0.82	
educationunknown	-0.09	0.40	
hasreligion	-0.10	0.15	
catholic	0.17	0.01	*
interest_in_religion	-0.47	0.00	***
donrever_1	0.15	0.02	*
liberal_donor	0.89	0.00	***
conservative_donor	-1.58	0.00	***
contbrel_1	-0.61	0.00	***
apparel_1	0.13	0.02	*
cat_1	-0.20	0.00	**
outdgrdn_1	-0.11	0.10	
guns_1	-0.25	0.00	***
density_cleansuburban	0.47	0.00	***
density_cleanurban	0.66	0.00	***
marital_status_cleanother	-1.35	0.23	
marital_status_cleansingle	0.23	0.00	**

After arriving at a final GAM model, we again predict on the test set and present the discrepancy measure. The discrepancy measured for our GAM model (0.5950), is slightly lower than that of the baseline GLM (0.5953) when evaluated on the test dataset, suggesting that this could indeed be a better predictive model. Hence, this has been selected as our final model for submission on Kaggle.

Table 7: Discrepancy of Predictions: Final Model

Data	Discrepancy
Train	0.5888
Test	0.5950

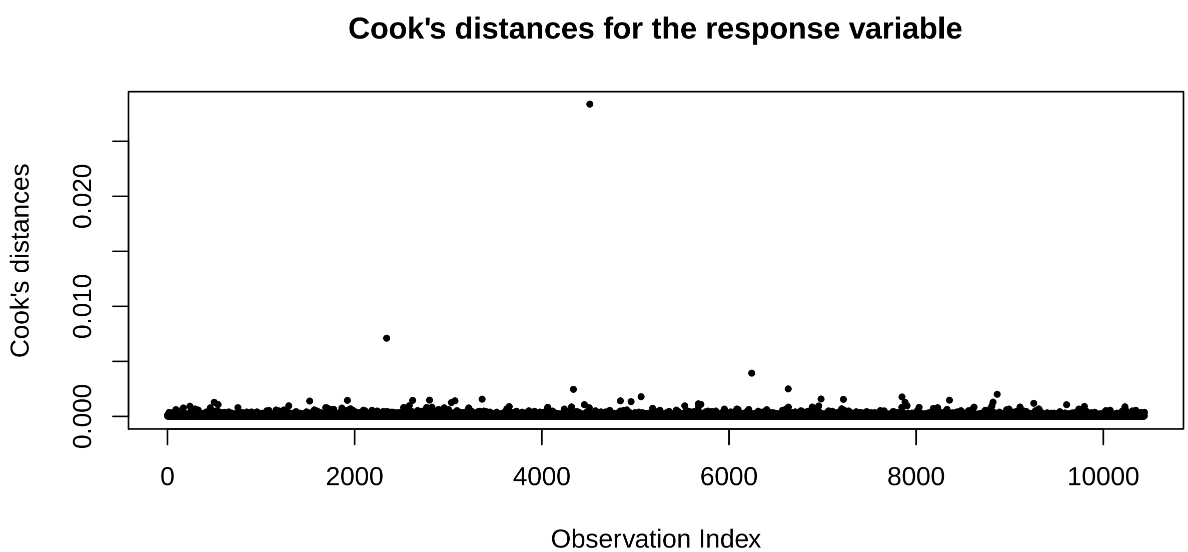
Testing Assumptions

By using the deviance-based stepwise backward selection procedure, it has been ensured that each predictor included in the model resulted in a better fit that was statistically significant. Additionally, we conduct a Hosmer-Lemeshow test on the model to confirm its goodness of fit. Below we see that for several different sized groupings of the data, the null hypothesis - that the probabilities predicted by the model are consistent with the observed data, was not rejected. This indicates that there is no systemic lack of fit.

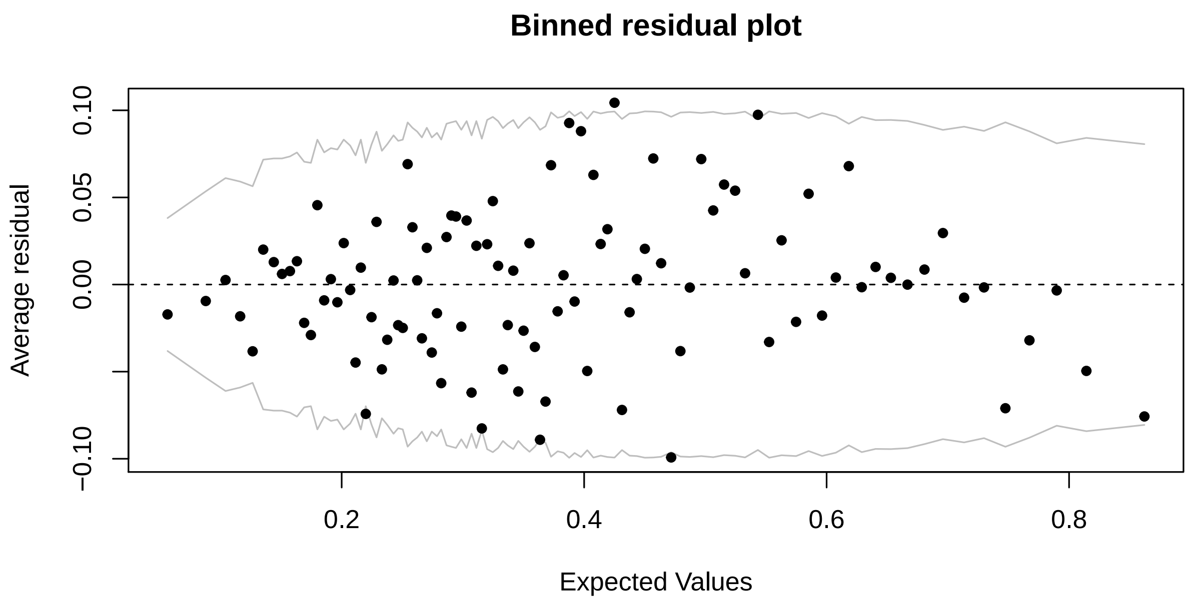
Table 8: Hosmer-Lemeshow test

groups	pvalues
5	0.2296213
7	0.2521186
10	0.4806731

The plot of Cook's distances suggests that there are no influential points to be concerned about, since the values of the Cook's distances D_i are considerably less than 1.



We get a sense of outlier points by examining a binned residual plot of the data. While there are one or two points moderately more extreme than the others, there are no glaring groups of observations with large average residuals. The points are generally evenly dispersed around the horizontal "zero-residual" line, suggesting residual assumptions are met.



Interpreting Results

Since we arrived at the final model by starting from the full model, with all predictors, and removing one variable at a time through the analysis of deviance procedure, all predictors yield statistically significant results on the likelihood ratio test. Five predictors in the final model are continuous variables: `age`, `log_ppi`, `cnty_pct_religious`, `cnty_pct_evangelical`, and `log_nchildren`, while the remaining are all either categorical or binary variables. The complete list of predictors in the final model and their types is given below:

Categorical variables

Since `sex` is a categorical variable which takes on three different values: `F` (female), `M` (male), and `U` (unidentified), it gets encoded as two binary variables, with `F` as the reference group. Holding other predictors fixed, on average, being a male corresponds to a 0.41 decrease in log odds that the person supports the Democratic party compared to if the person was a female. Similarly, holding other predictors constant, the log odds of supporting the Democratic party for a person whose sex has not been specified is 0.27 lower than that for a female on average. This follows our intuition, as the issues that Democrats support are largely in line with women's rights, hence a higher propensity for women to support Democrats.

The `combined_ethnicity_4way` coefficients can be interpretted similarly, where the reference group is Asian, and B stands for Black, H for Hispanic, and W for White. The reference group for the `education` variable is `bach degree`, and the reference group for the `density_clean` variable is `rural`. Lastly, the reference group for `marital_status_clean` is married. The direction of these coefficients are also largely in line with our expectations. The positive coefficient for African Americans versus the negative coefficient for White people reaffirms the fact that African Americans are more likely to vote Democrat, and White people to vote Republican.

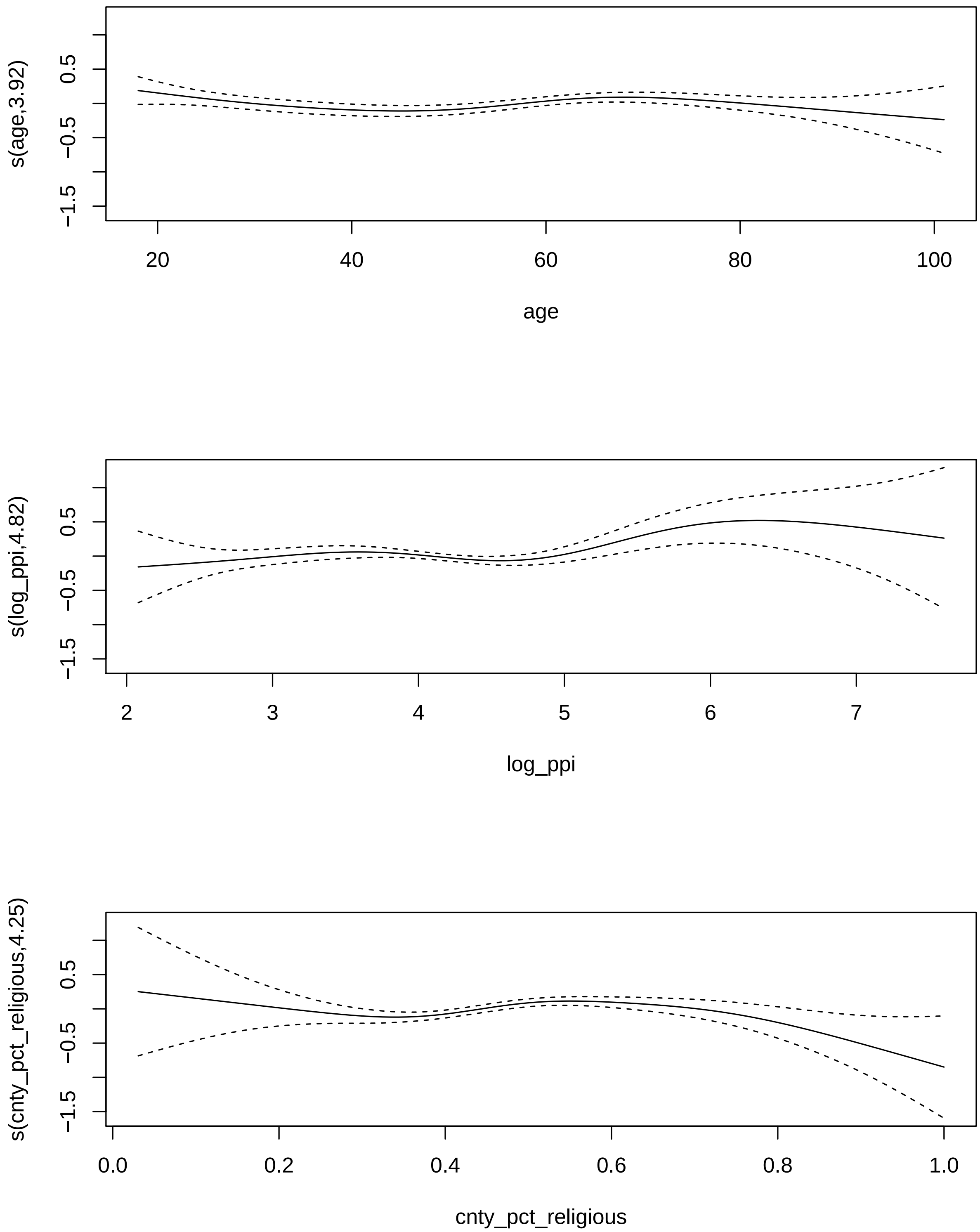
	type	variable name
1	categorical	sex
2		combined_ethnicity_4way
3		education
4		density_clean
5		marital_status_clean
6	binary	hasreligion
7		catholic
8		interest_in_religion
9		donrever_1
10		liberal_donor
11		conservative_donor
12		contbrel_1
13		apparel_1
14		cat_1
15		outdgrdn_1
16		guns_1
17	smoother	s(age)
18		s(log_ppi)
19		s(cnty_pct_religious)
20		s(cnty_pct_evangelical)
21		s(log_nchildren)

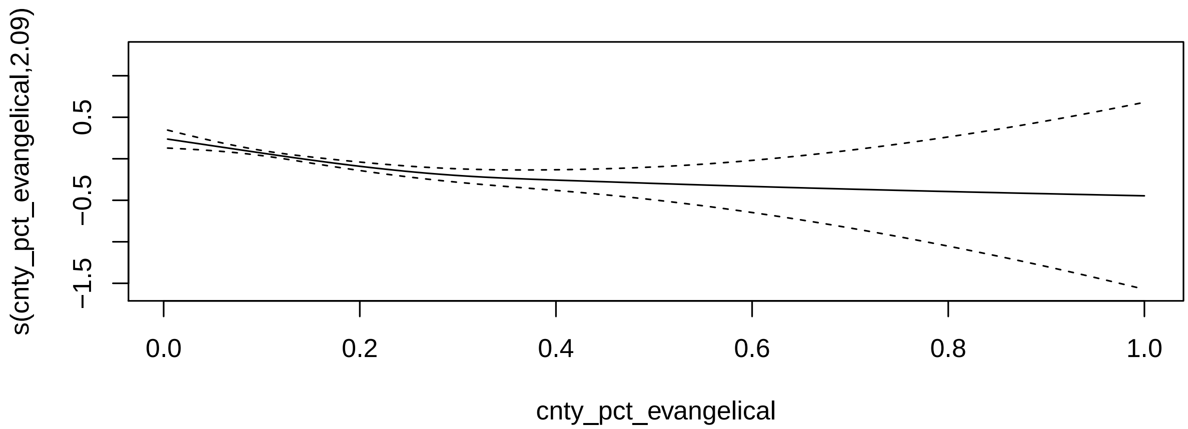
Binary variables

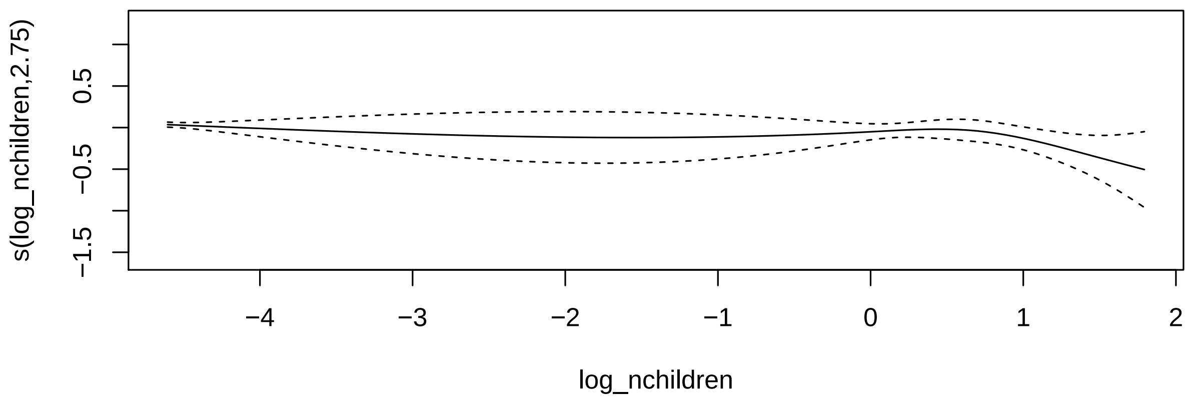
The interpretation of coefficient estimates for binary variables is similar to that for categorical variables, where the reference group is now the group that correspond to value 0 of the binary variable. To see this, consider `liberal_donor` as an example. On average, fixing other variables, the log odds of supporting the Democratic party for a person who has donated to liberal causes is 0.89 higher than that for a person who has never donated to liberal causes. Unsurprisingly, the coefficient for `liberal_donor` is positive, while `conservative_donor` and `guns` are negative. A result that we were not expecting is the relationship of owning a cat to political preference: with a negative coefficient, our model suggests that cat owners are more likely to support Republicans.

Continuous variables

While there are no estimated coefficients associated with the smooths of the continuous variables, we can understand their impact on predictions by examining plots of these smooth functions. For example, as `cnty_pct_evangelical` increases, the probability that a voter supports democrats decreases, albeit in a non-linear way. With `age` the relationship isn't monotonic; we see very young people more likely to support Democrats (as expected), but also an increase in likelihood around the age of 70.


s(age,3.92)
−1.5
−0.5
0.5
20
40
60
80
100
age
s(log_ppi,4.82)
−1.5
−0.5
0.5
2
3
4
5
6
7
log_ppi
s(cnty_pct_religious,4.25)
−1.5
−0.5
0.5
0.0
0.2
0.4
0.6
0.8
1.0
cnty_pct_religious





Conclusion

Based on the methods and considerations in this project, it is evident that generalized additive models are a powerful tool in making predictions in a binary response setting. We applied and compared various methods associated with generalized linear models and additive models to achieve the best predictive power in this binary classification problem.

Our modeling approach was to first build a robust baseline model using the stepwise backward selection procedure starting out with all predictors. Subsequently, we applied techniques such as smoothers and interactions terms to improve the predictive power of the model. In hindsight, this seemed a reasonable approach, since at each step, we observed an (albeit small) improvement on the discrepancy score. The likelihood ratio test played an important role in selecting the combination of variables that would lead to the best predictive model. We tested various aspects of fitting GLMs, such as the choice of the link function, consideration and interpretation of interaction terms, and smoothing continuous predictor variables. After identifying the model with the best predictive power, we used diagnostic tests such as the Hosmer Lemeshow Test, the binned residual plot and the Cook's distance plot to verify whether there were issues with the model fit.

Based on the low observed p-values of the predictors, it is evident that they play an important role in predicting the political leaning of an individual. We gained an appreciation for the diversity of the attributes that inform an individual's political preferences, and in a general sense, of how we can leverage data to better understand voter behavior.

Perhaps the greatest challenge we encountered in this process was the amount of missingness in the categorical predictors. Had the dataset been complete, we would have been able to test all interactions, which might have resulted in a model with better predictive power. One remedy for this issue would be to consider model-based imputation. In terms of modeling methods, more complex models such as support vector machines and neural networks could result in higher classification accuracies, but the results obtained from these techniques would be more difficult to interpret.

Overall, the analysis was a great opportunity to explore most tools related to generalized linear and additive models that have been introduced during the course.

Appendix

1. Summary of binary indicators

Table 9: Summary Statistics for Binary Indicators and Association with Outcome

Binary Indicator	% Obs. = 1	% Supp. Dem (Ind=0)	% Supp. Dem (Ind=1)	Abs. Diff in % Supp Dem
renter	0.07	0.37	0.57	0.21
liberal_donor	0.08	0.37	0.46	0.09
catholic	0.30	0.38	0.38	0.01
professional_technical	0.16	0.38	0.38	0.00
farmer	0.00	0.38	0.38	-0.01
christian	0.06	0.38	0.37	-0.01
contbpol_1	0.08	0.38	0.37	-0.01
environm_1	0.46	0.39	0.37	-0.01
retired	0.20	0.38	0.37	-0.01
blue_collar	0.22	0.38	0.37	-0.02
apparel_1	0.51	0.40	0.37	-0.03
contbhlt_1	0.16	0.39	0.34	-0.05
electrnc_1	0.66	0.42	0.36	-0.06
expensive_items_1	0.51	0.41	0.35	-0.07
donrever_1	0.56	0.42	0.35	-0.07
hasreligion	0.56	0.42	0.35	-0.07
veteran_1	0.11	0.39	0.32	-0.07
golf_1	0.26	0.40	0.32	-0.08
investor_1	0.60	0.43	0.35	-0.08
bookmusc_1	0.62	0.43	0.35	-0.08
boatownr_1	0.14	0.40	0.29	-0.11
cat_1	0.23	0.41	0.29	-0.11
bible_reader	0.11	0.39	0.28	-0.12
interest_in_religion	0.15	0.40	0.28	-0.12
outdgrdn_1	0.67	0.46	0.34	-0.12
guns_1	0.31	0.42	0.29	-0.13
contbrel_1	0.05	0.39	0.26	-0.13
outdoor_1	0.62	0.47	0.33	-0.14
homeowner	0.90	0.53	0.36	-0.17
conservative_donor	0.01	0.38	0.12	-0.27

2. Baseline model with all cleaned predictors

Table 10: Baseline Model (All Predictors)

Predictor	Est	Std Err	t-Stat	P-Value	Sig?
(Intercept)	0.83	0.76	1.08	0.28	
age	0.00	0.00	-0.66	0.51	
party_reg_state	-0.02	0.06	-0.44	0.66	
party_primary_state	-0.02	0.05	-0.46	0.64	
sexM	-0.42	0.05	-8.21	0.00	***
sexU	-0.30	0.27	-1.12	0.26	
combined_ethnicity_4wayB	0.88	0.22	4.07	0.00	***
combined_ethnicity_4wayH	-0.10	0.21	-0.45	0.65	
combined_ethnicity_4wayW	-0.75	0.20	-3.67	0.00	***

Predictor	Est	Std Err	t-Stat	P-Value	Sig?
children_3plus	-0.30	0.13	-2.38	0.02	*
homeowner	0.02	0.15	0.11	0.91	
renter	0.12	0.17	0.72	0.47	
educationhigh school	-0.03	0.08	-0.42	0.68	
educationno hs degree	0.01	0.12	0.07	0.95	
educationpost graduate	0.36	0.09	4.03	0.00	***
educationsome college	0.01	0.07	0.11	0.91	
educationunknown	-0.03	0.11	-0.29	0.77	
hasreligion	-0.18	0.08	-2.18	0.03	*
catholic	0.26	0.08	3.24	0.00	**
christian	0.28	0.12	2.38	0.02	*
bible_reader	-0.04	0.14	-0.30	0.76	
interest_in_religion	-0.42	0.13	-3.21	0.00	**
donrever_1	0.17	0.07	2.41	0.02	*
liberal_donor	0.87	0.09	9.19	0.00	***
conservative_donor	-1.54	0.31	-4.95	0.00	***
contbrel_1	-0.61	0.14	-4.54	0.00	***
contbpol_1	0.11	0.10	1.13	0.26	
contbhlt_1	-0.05	0.08	-0.69	0.49	
blue_collar	0.09	0.07	1.34	0.18	
farmer	0.32	0.39	0.80	0.42	
professional_technical	-0.08	0.07	-1.10	0.27	
retired	0.13	0.08	1.64	0.10	
apparel_1	0.15	0.06	2.65	0.01	**
bookmusc_1	0.05	0.07	0.74	0.46	
electrnc_1	-0.02	0.06	-0.34	0.74	
boatownr_1	-0.03	0.08	-0.42	0.68	
cat_1	-0.20	0.07	-2.88	0.00	**
environm_1	0.08	0.05	1.52	0.13	
outdgrdn_1	-0.10	0.07	-1.39	0.16	
outdoor_1	-0.02	0.08	-0.28	0.78	
guns_1	-0.24	0.07	-3.64	0.00	***
golf_1	-0.05	0.06	-0.78	0.43	
investor_1	0.03	0.07	0.42	0.67	
veteran_1	-0.07	0.09	-0.79	0.43	
expensive_items_1	-0.02	0.06	-0.38	0.70	
cnty_pct_religious	0.37	0.21	1.82	0.07	.
cnty_pct_evangelical	-1.36	0.27	-5.10	0.00	***
density_cleansuburban	0.48	0.07	7.21	0.00	***
density_cleanurban	0.67	0.08	7.93	0.00	***
marital_status_cleanother	-1.73	1.15	-1.50	0.13	
marital_status_cleansingle	0.23	0.07	3.19	0.00	**
log_ppi	0.02	0.05	0.53	0.59	
log_income	-0.11	0.07	-1.57	0.12	
log_nchildren	-0.01	0.01	-1.09	0.28	
age.na	0.78	0.59	1.32	0.19	
cnty_pct_religious.na	-12.40	124.55	-0.10	0.92	
cnty_pct_evangelical.na	0.61	1.26	0.48	0.63	