Influencing Elections with Statistics: Targeting Voters with Logistic Regression Trees

Rusch, Thomas; Lee, Ilro; Hornik, Kurt; Jank, Wolfgang; Zeileis, Achim

Published: 01/01/2012

Document Version
Publisher's PDF, also known as Version of record

Influencing Elections with Statistics: Targeting Voters with Logistic Regression Trees

Thomas Rusch, Ilro Lee, Kurt Hornik, Wolfgang Jank, Achim Zeileis

Research Report Series
Report 117, March 2012

Institute for Statistics and Mathematics
http://statmath.wu.ac.at/

Influencing Elections with Statistics: Targeting Voters with Logistic Regression Trees

Thomas Rusch
WU (Wirtschafts-universität Wien)

Ilro Lee
University of New South Wales

Kurt Hornik
WU (Wirtschafts-universität Wien)

Wolfgang Jank
University of South Florida

Achim Zeileis
Universität Innsbruck

Abstract

Political campaigning has become a multi-million dollar business. A substantial proportion of a campaign's budget is spent on voter mobilization, i.e., on identifying and influencing as many people as possible to vote. Based on data, campaigns use statistical tools to provide a basis for deciding who to target. While the data available is usually rich, campaigns have traditionally relied on a rather limited selection of information, often including only previous voting behavior and one or two demographical variables. Statistical procedures that are currently in use include logistic regression or standard classification tree methods like CHAID, but there is a growing interest in employing modern data mining approaches. Along the lines of this development, we propose a modern framework for voter targeting called LORET (for logistic regression trees) that employs trees (with possibly just a single root node) containing logistic regressions (with possibly just an intercept) in every leaf. Thus, they contain logistic regression and classification trees as special cases and allow for a synthesis of both techniques under one umbrella. We explore various flavors of LORET models that (a) compare the effect of using the full set of available variables against using only limited information and (b) investigate their varying effects either as regressors in the logistic model components or as partitioning variables in the tree components. To assess model performance and illustrate targeting, we apply LORET to a data set of 19,634 eligible voters from the 2004 US presidential election. We find that augmenting the standard set of variables (such as age and voting history) together with additional predictor variables (such as the household composition in terms of party affiliation and each individual's rank in the household) clearly improves predictive accuracy. We also find that LORET models based on tree induction outbeat the unpartitioned competitors. Additionally, LORET models using both partitioning variables and regressors in the resulting nodes can improve the efficiency of allocating campaign resources while still providing intelligible models.

Keywords: campaigning, classification tree, get-out-the-vote, logistic regression, model tree, model-based recursive partitioning, political marketing, voter identification, voter segmentation, voter targeting.

1. Introduction

"Decisions are made by those who show up", said President Bartlet, a character from a popular

TV show, *The West Wing*. The character in the show used the line to motivate a college audience to voice their opinion by showing up at the polls. Getting eligible voters to actually vote ("get-out-the-vote"; GOTV) is an important goal in countries with a democratic political system and a lot of resources are spent on achieving that goal. Take the 2008 US presidential race for example. In that year, the world witnessed the amount of money raised and spent reaching unprecedented heights. By spending over USD 1 billion, the Obama and McCain campaigns tried to persuade and mobilize voters to engage in the political process by casting their vote on November 4th. However, even with monumental campaign effort, and large out-laying of resources, only 61.7% of eligible voters did cast their ballot.

1.1. Campaigning, mobilization and turnout in the United States

The impact of partisan campaigning or nonpartisan get-out-the-vote efforts on mobilization and turnout has been subject to numerous scientific investigations over the last 20 years see e.g., Whitelock, Whitelock, and van Heerde (2010); Baek (2009); Karp and Banducci (2007); Steel, Pierce, and Lovrich (1998); Finkel (1993); Gelman and King (1993). Starting from an early 'minimal effect' hypothesis (i.e., the idea that political campaigns only marginally mobilize, persuade or convert voters), the general sentiment nowadays is that campaigning does indeed have measurable effects on mobilization of (core) supporters (Holbrook and McClurg 2005; Hillygus and Jackman 2003). This mobilization, in turn, has been shown to have an effect on increasing overall turnout and on getting additional votes for a specific candidate (Holbrook and McClurg 2005; Cox and Munger 1989).

As a result, campaigns are spending huge amounts of money on mobilizing voters. Despite this spending, campaigns often fail to mobilize voters for the campaign's cause. Take the United States for example, where the "professionalization" (Muller 1999) of campaigning has had its origin[1] (Plasser 2000). Arguably, nowhere else is political campaigning a bigger business then in the United States. However, despite increased political consultancy and the hundreds of millions of campaign spending, the average voter turnout since 1980 during the Presidential election years has only been 56%; see also Table 1.

Table 1 shows voter turnout, total spending of presidential candidates since 1980 as well as total spending adjusted for inflation at 2008 CPI (consumer price index) rates (i.e., real expenditures). Figure 1 shows the bivariate relationship between turnout and the logarithm of real total campaign expenditures per eligible voter along with a fitted linear regression line. While some caution is warranted when interpreting a linear regression fitted to just 8 observations, there is clearly a positive association. The 2004 and 2008 elections saw especially increased expenditures per voter accompanied by a noticable increase in voter turnout. Given the relationship between campaign spending and turnout, campaigns are well advised to spend money on mobilizing voters (Baek 2009; Hall and Bonneau 2008). However, as campaigns increasingly face limited resources and budget constraints (in addition to public sentiment against excessive spending during times of economic hardship), it is important to allocate resources as efficiently as possible.

From a marketing point of view, voter mobilization is a two-step process (cf. Goldstein and Ridout 2002). In the first step, campaigns need to craft measures that best motivate people

[1] The professionalization of politial campaigning spread form the US to many democratic countries all over the world (Sussman and Galizio 2003). Accordingly we will focus on the US system but the ideas are easily generalizable to other democratic countries as well.

Year	Turnout (in %)	Expenditures (in mill. USD)	Real expenditures (at 2008 rates)
2008	61.7	1,324.7	1,324.7
2004	60.1	717.9	818.2
2000	54.2	343.1	429.0
1996	51.7	239.9	329.2
1992	58.1	192.2	295.0
1988	52.8	210.7	383.5
1984	55.2	103.6	214.7
1980	54.2	92.3	241.2
Mean	56.0	403.1	504.4
Sd	3.6	422.0	381.6
Min	51.7	92.3	214.7
Max	61.7	1,324.7	1,324.7

Table 1: Individual and aggregated turnout rate (votes for highest office divided by the voting-eligible population) for presidential elections in the United States and the money spent by all candidates (in million USD). The fourth column lists the real expenditures (inflation-adjusted at 2008 rates). Source: McDonald (2012) and `http://www.opensecrets.org/` and `http://www.bls.gov/data/inflation_calculator.htm`.

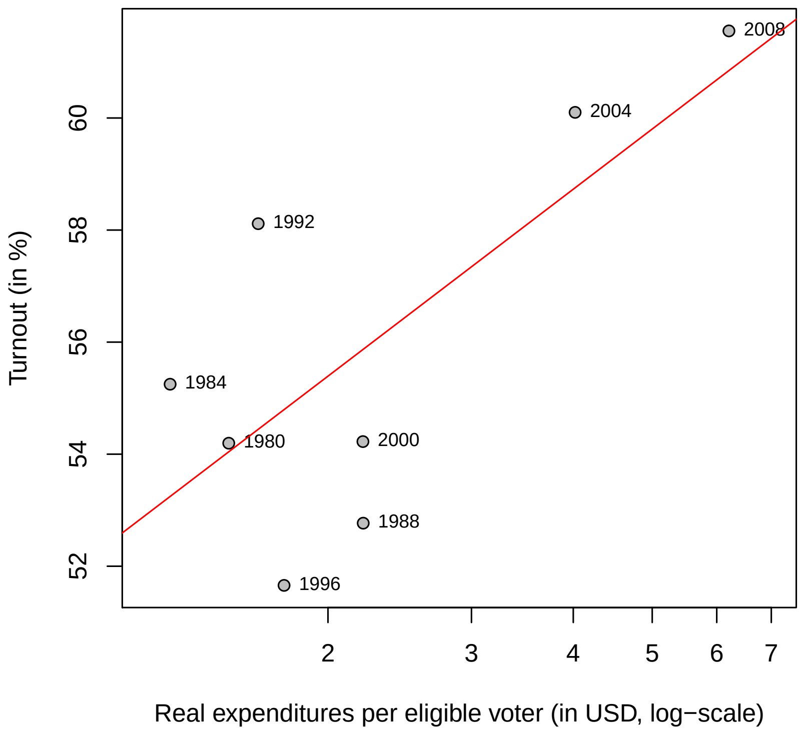


Figure 1: The relationship between real expenditures per eligible voter (in USD, log-scale) and turnout (in %) in the US since 1980 along with fitted linear regression (solid line). The model has $R^2 = 0.52$ and its slope implies an expected increase of turnout by 0.5 percentage points for a 10% increase of expenditures per eligible voter.

	2008	2006	2004
Expenditures (in USD)	2,501,605	2,231,941	1,848,822
Votes received	164,562	71,651	113,040
Cost per vote (in USD)	15.20	31.15	16.36

Table 2: Election costs for Congressman Barrow (Georgia's 12th congressional district). Source: Secretary of State, Georgia, http://www.opensecrets.org/.

to turn up at the polls, i.e., to assure the effectiveness of mobilization. In the second step, campaigns need to identify suitable people that should be subjected to these measures (also known as voter targeting).

The first step includes decisions on which marketing measures to use. Since many measures lack in effectiveness, numerous studies have been designed on this subject, investigating diverse measures such as TV ads, canvassing and face-to-face contacting, telephone calls or negative campaigning (e.g., Green and Gerber 2008; Hansen and Bowers 2009; Ridout 2009; Lau, Sigelman, and Rovner 2007; Gillespie 2010).

In the second step of the marketing process, campaigns need to *identify* the "right" recipients for these marketing messages. To our knowledge, there has only been little work on this topic in the literature, some notable exceptions including Wielhouwer (2003); Parry, Barth, Kropf, and Jones (2008) or Murray and Scime (2010). Identifying the right people to target is important because it reduces wasteful spending (e.g., targeting a person who is very unlikely, or not even eligible, to vote would be considered extremely wasteful) and allows campaigns to efficiently allocate their limited resources.

As point in case consider Table 2 which shows how much money has been spent by the campaign of Congressman Barrow of Georgia's 12th congressional district, and how many votes he received in three consecutive election years (in the US, members of the house of representatives get elected every two years). During each of the three elections, the campaign spent similar amounts of money, however, in 2006 it cost the campaign double the amount of money for each vote it received. Assuming that the campaign was targeting roughly the same voters in all three elections, one cannot help but wonder whether it would have been better to target a lower number of people in 2006 (a midterm election year where turnout is generally lower). The identification of people who only vote in general elections might have helped in order to spend the available resources more efficiently. For such a precise identification of voters, statistics offers a number of suitable tools.

1.2. Voter targeting in get-out-the-vote (GOTV) campaigns

In his standard source on political targeting, Malchow (2008, p. 1) defines voter targeting as "...the process by which a campaign predicts which voters it needs to persuade to win." These voters include those who are undecided as well as those who are in favor of the issue at hand, at least in principle, but who need some encouragement to turnout. Malchow (2008) opines that efficient identification and prediction of which voter should be targeted is going to be one of the future major issues in campaigning. This is also reflected in his alternate definition of targeting as being "the process of determining which voters you need for victory and identifying them as efficiently as possible" (Malchow 2008, p. 7).

This goal of voter targeting shares similar objectives with that of consumer targeting in marketing. However, there are several structural reasons as to why political campaign marketing differs from that of consumer marketing. Following Quelch (2008) these are: (i) the lower number of choices for voters in general elections than for consumers, (ii) that voters have to live with the majority's decision which might dampen their enthusiasm and (iii) that most of the voters only get to vote every couple of years on a fixed date while consumers can usually decide when to when and where purchase. Additionally, (iv), singling out a niche may work fine for marketing a product, but politicians cannot win by targeting just a single segment as they need to get the majority of votes. This may be the reason why political marketing is generally considered to be less successful than consumer marketing.

How is targeting carried out?

Campaigns basically try to mobilize voters who (however loosely) identify themselves with any party or candidate. They do not necessarily try to convince voters to cast their ballot for a specific candidate. Thus they may simply aim at increasing the number of people who show up at polls. Malchow (2008) describes targeting for turnout as a targeting procedure for which the campaign needs to know or predict the likelihood that a voter will actually vote, regardless of whether it is for persuasion or mobilization purposes, as well as making a strategic decision which range of prediction is of interest.

To make such predictions, campaigns are employing many different techniques, some of which are founded in statistical reasoning. This also pertains to the campaigns gearing up for the 2012 presidential election which are showing a strong interest in statistics for decision making. President Obama's campaign is actively seeking for data miners to join his campaign for reelection[2][3][4]. In addition, not only the incumbent is seeking help from statistics, but also some of his challengers such as the Texas Governor, Rick Perry[5][6]. The increased media coverage of the importance of statistics in election campaigns supports this effect.

When targeting voters, campaigns rely on data that are either public or proprietary. Public data offer a limited number of variables such as aggregate number of turnout, while data sets from proprietary sources often contain much richer information. Usually the most important variables that are collected are records of the individual voting history. The aptitude of voting history as a predictor for future election attendance has already been established (Denny and Doyle 2009) and consequently it is the gold standard in targeting (Malchow 2008). However there might be predictive power in additional variables that are often ignored.

Traditionally, campaigns have relied on simple deterministic rules for choosing who to target, e.g., using information from the last four comparable elections as the main predictors for future voting behavior. Intuitively, someone who voted in all four out of the last four elections is seen as a most likely voter whereas someone who did not vote in any of the four elections is very unlikely to vote now. However, forecasting the behavior of persons with other patterns (i.e., who voted sometimes but not always) is not clear in this very simple setup.

This has sparked interest in adopting probabilistic approaches in place of deterministic rules

[2] `http://www.cnn.com/2011/10/09/tech/innovation/obama-data-crunching-election/index.html`

[3] `http://andrewgelman.com/2011/10/data-mining-efforts-for-obamas-campaign/`

[4] `http://www.politico.com/blogs/bensmith/0711/Obama_campaign_hiring_data_mining_scientists.html`

[5] `http://thecaucus.blogs.nytimes.com/2011/08/22/rick-perrys-scientific-campaign-method/`

[6] `http://www.campaignsandelections.com/magazine/us-edition/267417/technology-bytes-perryand39s-social-scientists-and-obamaand39s-data-brigade.thtml`

based solely on the voting history. For instance, Malchow (2008) promotes a linear probability model as well as tree-like models such as CHAID (Kass 1980) for political microtargeting. Murray and Scime (2010) suggest decision trees as well. Other state-of-the-art approaches that are used include logistic or probit regression.

When using probabilistic models for GOTV targeting, campaigns are interested in assigning each voter an individual probability to show up at election day. Based on this estimated probability, it stands to reason that using targeting plans on people with a value around 0.5 is worthwhile (Malchow 2008), whereas targeting people with predicted probabilities near 0 or 1 is considered a waste. This is in accordance with results on how to best allocate campaign resources in general (Brams and Davis 1973; Snyder 1989), namely spending more resources on highly contested seats or states where the race is tight. In fact, a person with a predicted probability near zero is almost definitely not going to vote, regardless of how compelling the mobilization message is. A person with a predicted probability of one is going to turn out at the polls anyways, without the need for extra persuasion. In both cases, targeting those people would not lead to an increase in turnout, yet it would consume resources and hence be wasteful. However, voters with a predicted probability in a "targeting range" around 0.5 may be "convincable" to show up at the polls using the right incentive. Malchow (2008) suggests a targeting range of $[0.3, 0.7]$. Clearly, we can be hopeful to sway a person with a probability of voting of say, 0.35 as long as we get the right message to her. On the other hand, while a person with a probability of 0.68 might be going to vote without being targeted specifically, it should not hurt to encourage her a bit more.

1.3. A new unified framework for voter targeting

In this paper we propose a new and flexible statistical framework for voter segmentation that generalizes two standard models currently used in political targeting. In fact, our framework encompasses logistic regression as well as classification trees and allows for a combination of both within the same model. We refer to the resulting framework as <u>lo</u>gistic <u>re</u>gression <u>t</u>ree (LORET) models. LORET models are very flexible in that, in their simplest form, they reduce to a majority vote (May 1952) or naive Bayes (Hand and Yu 2001) model; on the other hand, they also allow regression-like modeling of predictors as well as tree-like partitioning of the sample space under the same umbrella. We investigate LORET models of varying degrees of flexibility, and compare them with a particular focus on the benefits that they provide for political decision makers. We apply LORET to a novel data set of Ohio voters and find that, depending on the nature of the race, different statistical methods lead to relevant differences in how campaign budgets are best allocated.

This paper is organized as follows. In Section 2, we present a statistical framework for voter targeting that combines logistic regression models with recursive partitioning. Section 3 describes a case study for which we apply the methods. There, we explain how we evalute the framework and investigate properties of our targeting approach from a campaign's point of view. The corresponding results can be found in Section 4. We finish with conclusions and some general remarks in Section 5.

2. LORET: Modeling and predicting voting behavior

Currently, campaigns employ methods like logistic regression or tree-based methods for voter prediction and targeting (Malchow 2008). Using this as a backdrop, we introduce a general framework, logistic regression trees (LORET), that encompasses and extends these approaches. Briefly, the idea is the following: Instead of fitting a global logistic regression model to the whole data, one might fit a collection of local regression models to subsets or segments of the data (i.e., a segmented logistic regression model) in order to obtain a better fit and higher predictive accuracy. Since usually the "correct" segmentation is not known, it needs to be learned from the data, for example by using recursive partitioning methods.

In what follows we start with the general formulation for logistic regression models for one or more segments and then show how for more than one segment, the segmentation can be learned with recursive partitioning.

2.1. Segmented logistic regression

Let $y_i \in \{0,1\}$ denote a Bernoulli random variable for the i-th observation, $i = 1, \dots, N$, and $\mathbf{x}_i$ denote a p-dimensional covariate vector $(x_{i1}, \dots, x_{ip})$. Let us assume there are r (known or estimated) disjoint segments in the data. For each segment $k = 1, \dots, r$, we can then specify a logistic regression model for the relationship between y and $x_1, \dots, x_p$ within that segment,

$$P(y_i = 1 | \mathbf{x}_i; \boldsymbol{\beta}^{(k)}) = \pi_i = \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta}^{(k)})}{1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta}^{(k)})}, \tag{1}$$

where $k = k(i)$ is the segment to which observation i belongs and π_i denotes the probability to belong to class "1". The segment-specific parameter vector is $\boldsymbol{\beta}^{(k)}$ and its estimates are referred to as $\hat{\boldsymbol{\beta}}^{(k)}$, which can be easily obtained (given the segmentation) via maximum likelihood (see e.g., McCullagh and Nelder 1989). Based on the associated estimated probabilities, classification can then be done by

$$\hat{y}_i(c_0) = \begin{cases} 1 & \text{if } \hat{\pi}_i \geq c_0 \\ 0 & \text{if } \hat{\pi}_i < c_0. \end{cases} \tag{2}$$

where $c_0 \in [0,1]$ is a specific cutoff value (but could, in principle, also be specified to be different for different segments).

If there is only a single segment (i.e., a root node and hence a known segmentation), LORET in (1) reduces to a logistic regression model. Here the conditional distribution of the response variable y is estimated given the status of p covariates. Evaluation of the logistic model at the estimated parameter vector $\hat{\boldsymbol{\beta}}$ yields the predicted probabilities, $\hat{\pi}_i$. If the model uses no covariates as regressors, it further reduces to a majority vote (May 1952) or naive Bayes model (Hand and Yu 2001), i.e., a logistic regression model with only an intercept or simply the relative frequency of class "1" transformed to the logit scale. The upper row in Figure 2 illustrates majority vote and logistic regression on an artificial set of data. The former fits a single constant, the latter a single logistic function of x to the entire data.

If there are more than one segment and the segmentation were known, then LORET can still be simply seen as estimating a maximum likelihood model from a binomial likelihood in each segment. This time however, one needs to specify interactions between factors corresponding to the segments and the cofficients of a logistic regression model to estimate the LORET.

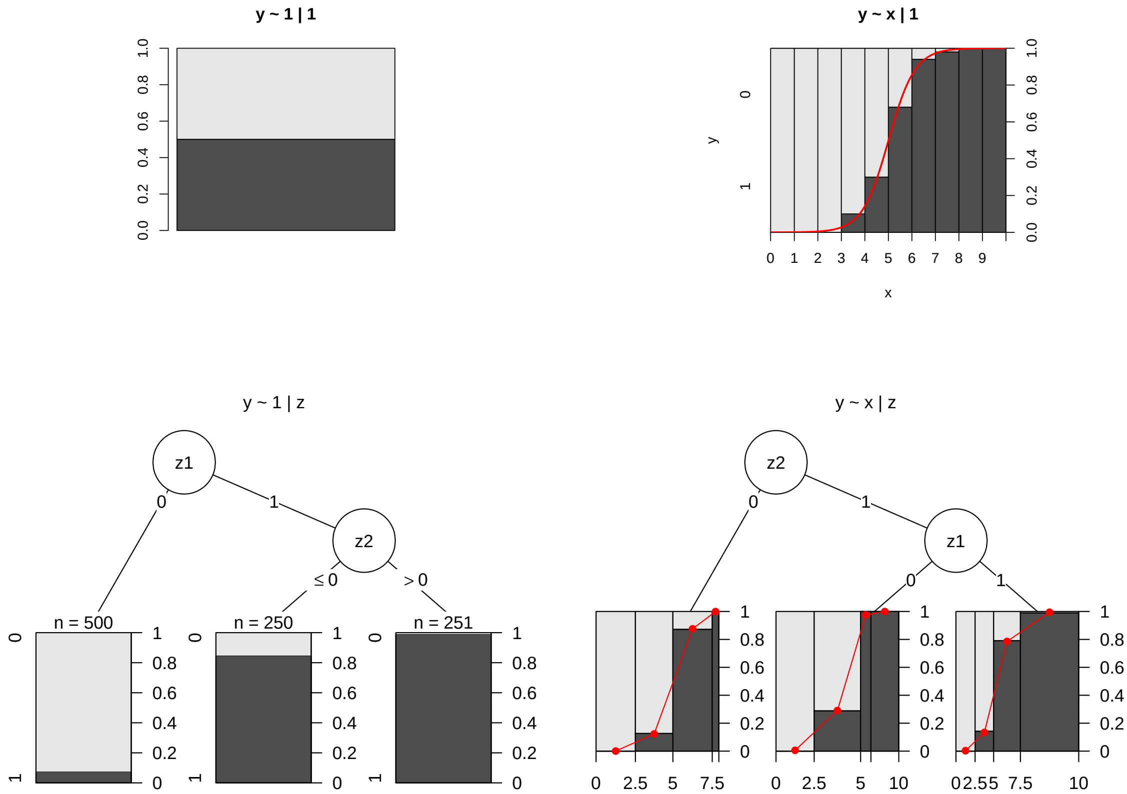


Figure 2: A visualization of the different cases of LORET. In the upper left panel there is the $y \sim 1 \,|\, 1$ LORET, fitting a constant. In the upper right the $y \sim x \,|\, 1$ LORET (logistic regression) is displayed, which is a single function of x for the whole data set. The lower left panel displays a $y \sim 1 \,|\, z$ LORET where the data set is partitioned based on the state of predictor variables z and in each partition a constant is fitted. In the lower right panel, the $y \sim x \,|\, z$ LORET can be found. Here the data set is again partitioned based on z but this time a logistic function of x is fitted in the partitions. Hence it combines the $y \sim 1 \,|\, z$ and $y \sim x \,|\, 1$ LORET.

If the segmentation is unknown, however, it needs to be learned from the data. Two popular approaches for this are mixture models (e.g., mixtures of experts or latent class regression) or some type of algorithmic search method. Recursive partitioning (often called a tree, Zhang and Singer 2010) is a popular example of the latter. Trees are usually induced by splitting the data set along a function of the predictor variables into a number of partitions or segments. The segments are usually chosen by minimizing an objective function (e.g., a heterogeneity measure or a negative log-likelihood) for each segment. The procedure is then repeated recursively for each resulting partition. This approach approximates real segments in the data and yields a segmentation for which maximum likelihood estimation of parameters in each segment can be carried out, as is done in LORET.

Method	Regressor variables	Partitioning variables	Schema
Majority vote	none	none	$y \sim 1 \mid 1$
Logistic regression	yes	none	$y \sim x \mid 1$
Classification tree	none	yes	$y \sim 1 \mid z$
Model tree	yes	yes	$y \sim x \mid z$

Table 3: Various instances of LORET.

2.2. Recursive partitioning

Let us assume we have an additional, ℓ-dimensional covariate vector $\mathbf{z} = (z_1, \dots, z_\ell)$. Based on these covariates we learn the segmentation, i.e., we search for r disjoint cells that partition the predictor subspace. Depending on whether the logistic model used for y in each segment has any covariates or just a constant, there are two algorithmic approaches we can use: classification trees and trees with logistic node models.

Classification trees

If the logistic model is an intercept-only model and we have a number of partitioning variables $z_1, \dots, z_\ell$, then LORET can be estimated as a classification tree. An illustration of a classification tree can be found in the lower left panel of Figure 2, where the data is first partitioned into three subsets and a intercept-only model is fitted to each subset separately. Hence in each terminal node the model is a constant. A wide variety of algorithms have been developed to fit classification trees; among them are: CHAID (Kass 1980), CART (Breiman, Friedman, Olsen, and Stone 1984), C4.5 (Quinlan 1993), QUEST (Loh and Shih 1997), CTree (Hothorn, Hornik, and Zeileis 2006) and many others. In this paper, we use CART and CTree, which are examples of a biased and an unbiased tree algorithm, respectively.

Trees with logistic node models

If there are partitioning variables $\mathbf{z} = (z_1, \dots, z_\ell)$ as well as regressor variables for the logistic node model $\mathbf{x} = (x_1, \dots, x_p)$, we get the most general type of LORET, which is a "model tree". The model is illustrated in the lower right panel in Figure 2. Like in a classification tree, the data is first partitioned into subsets. However, in contrast to a classification tree, separate logistic regressions with regressors are employed in each terminal node. Thus, the resulting model tree essentially combines data-driven partitioning like a classification tree with model-based prediction in a single approach. Different algorithms have been proposed to estimate model trees with logistic node models, including: SUPPORT (Chaudhuri, Lo, Loh, and Yang 1995), LOTUS (Chan and Loh 2004), LMT (Landwehr, Hall, and Eibe 2005), and MOB (Zeileis, Hothorn, and Hornik 2008). In what follows, we will use the MOB algorithm with a logistic node model for estimating the most general version of LORET, as it proved to have good properties for these kind of data (Rusch and Zeileis 2012).

To simplify notation and to stress the similarities, we will use a simple schema to refer to the different LORET types (cf. Table 3): The majority vote model will be refered to as $y \sim 1 \mid 1$, the global logistic regression model as $y \sim x \mid 1$, the classification tree model as $y \sim 1 \mid z$ and the full LORET model as $y \sim x \mid z$.

3. Case study: Ohio voters 2004

To illustrate our targeting framework, we use a unique set of data from the state Ohio in the US. Most US states collect and report voter registration information but the data is not readily and easily accessible (US Election Assistance Comission 2010). The collection of voter registration data is done at the county level and most of the states aggregate the data. However, due to technical and resource limitations, political campaigns often turn to political and marketing data providers who add value by collecting, maintaining, and updating the voter registration data. The voter registration data would typically include name, address, phone, gender, party affiliation, age, vote history (elections that each voter voted), and ethnicity (in many of the southern states). The data providers not only add value by standardizing the data that is collected from each state or county, they also add other potentially relevant behavioral information such as income, type of occupation, education, presence of children, property status (rental or owning), and charities that the person donated to. We use such a proprietary data set which was provided by Aristotle, Inc., one of the leading campaign application and data providers in the industry.

Our data set consists of $20,000$ eligible voters from Ohio. Ohio has proven to be an important state because in every election since 1964, the winner of that state has ultimately won the presidency. Also since 2000, the presidential vote difference between the Republican and Democratic candidates has been 4% or lower. Thus Ohio has been considered one of the top "battleground" states in every recent election.

3.1. Data description

Our set of data includes a total of 77 variables, many which are socio-demographic categorical variables like gender, job category or education level. The data set also contains records on past voting behavior from 1990 to 2004 in general elections, primary or presidential primary elections and other elections (all coded as binary variables – i.e., voted or not). We added three composite or aggregate variables: the raw count of elections a person attended, the number of elections a person attended since registering and the relative frequency of attended elections since registering. After removal of missing values and inconsistent entries, there are a total of $N = 19634$ records with 80 variables per record. Our target variable is the voting behavior ("yes"/"no") in the 2004 presidential election. This election is considered to be unusual in the campaign's high emphasis on face-to-face voter mobilization within neighborhoods and social networks (Middleton and Green 2008) as well as the sharp increase in turnout (see Table 1) and is therefore particularly well suited for illustration.

3.2. Two sets of predictors: Voting history only vs. kitchen sink data

As pointed out earlier, campaigns relied on very limited information when it comes to political targeting. While some of the literature on voter targeting also recommends taking into account a person's age (Malchow 2008; Karp, Banducci, and Bowler 2008), the most commonly and often solely used piece of information is the person's voting history (Malchow 2008). Thus, one of the goals of this study is to investigate the potential benefits of including additional information (besides a person's voting history) into the targeting model. To that end, we compare and contrast two sets of predictors:

- The first set employs the standard information used by many campaigns, which is also

LORET	Regressor variables	Partitioning variables	Partitioning algorithm
$y \sim 1 \mid 1$	none	none	–
$y \sim s \mid 1$	s	none	–
$y \sim s + e \mid 1$	$s + e$	none	–
$y \sim 1 \mid s$	none	s	CART, CTree
$y \sim 1 \mid s + e$	none	$s + e$	CART, CTree
$y \sim s \mid e$	s	e	MOB

Table 4: LORET versions combined with the two variable groups and the algorithms used to estimate the partition. The standard variable set of age and voting history is labeled "s" and the set of additional variables with "e" (hence all variables together are "$s + e$").

recommended in literature. The standard variables used by campaigns are a person's voting history, recorded over the the last four elections, and age. We call this set "s" for "standard".

- The second set contains all other variables available, i.e., 'the "kitchen sink". In our case this includes variables like gender, occupation, living situation, party affiliation, party makeup of the household ("partyMix"), position within the family ("hhRank" and "hhHead"), donations for various causes, education level, relative frequency of attended elections so far ("attendance") and many others. These variables constitute a set of additional variables, labeled with "e" for "extended".

3.3. Model specification

The combination of the two variable sets with the different LORET models leads to model specifications as displayed in Table 4. The models either employ only the standard set of variables or the combination of the standard and the extended set. For unpartitioned models, the parameters are estimated with maximum likelihood. If a partition is induced, we learn it with three different algorithms (CART, CTree and MOB) depending on the nature of the node model. Please note that if age is specified as a parameter in the logistic model part (i.e., for models $y \sim s \mid 1$, $y \sim s{+}e \mid 1$ and $y \sim s \mid e$), a quadratic effect will be used (cf. Rusch and Zeileis 2012). If age is included as a partitioning variable we use the untransformed variable since partitioning algorithms are invariant to monotone transformations such as taking squares.

All recursive partitioning algorithms that we employ allow for tuning with metaparameters. These tuning parameters can be used to avoid overfitting of the tree algorithms and control how branchy the tree becomes. Quite generally it can be said that the less branchy a tree is, the less prone it is to overfitting. In the algorithms we used, a higher number of observations per node, a lower tree depth, and a stricter split variable selection criterion all lead to smaller trees. At the same time our specification should grant enough flexibility for the algorithm to approximate a complex non-linear relationship in the data.

For CART the maximal depth of the tree and the minimum number of observation per node (minsplit) are available to control the tree appearance. We use a maximal tree depth of 7 and a minsplit of 100 (which corresponds to roughly 0.5% of the observations). For CTree and MOB the significance level of the association or stability tests respectively and the minimum number of observation per node can be used to tune the algorithm. We employ a global significance level of $\alpha = 1 \times 10^{-6}$. This appeared sensible since the high number of observations might easily lead to spurious significance that is mainly due to the sample size. Hence we reduce the probability of "false positive" selection of a split variable or split point by specifying a low significance level. For minsplit we use 100 for CTree (the same as for CART) and 1000 for MOB (which enables reliable estimation of the node model).[7]

3.4. Model evaluation

We compare the different LORET specifications in terms of their ability to predict potential voters accurately and to allow for efficient targeting. Of particular interest is how data-driven approaches like trees compare to the model-driven approach of logistic regression and whether the combination of the two can lead to substantial improvements. We measure the performance of all models with different learning and test sets using different data- and domain-driven criteria. These criteria include standard measures from the data mining literature (such as predictive accuracy and ROC curves), and measures that arise from an election campaign and voter targeting practitioner point of view. We elaborate on each of these in more detail below. Additionally, we put emphasis on the interpretability of the models and model parameters that result from applying the LORET framework.

Learning and test samples via bootstrapping

We employ the benchmarking framework of Hothorn, Leisch, Zeileis, and Hornik (2005) to evaluate and compare different models via bootstrapping (see also Efron and Tibshirani 1993; Hastie, Tibshirani, and Friedman 2009). That is, we fit a model based on a learning set of size N which is sampled randomly (with replacement) from the entire set of data. The fitted model is then used to predict the out-of-bag test set which consists of observations that were not part of the learning sample. Ten folds of learning and test samples, $f = 1, \ldots, 10$ were used. To provide a further benchmark, we also train and evaluate all models on the whole data set. This allows us to gauge the tendency of a model to overfit as well as how close out-of-bag and in-sample performance are.

Measuring predictive accuracy

For each method, we assess the classification accuracy (acc_f) on each test set f at a given cutoff value $c_0 = 0.5$[8]. To estimate overall predictive accuracy, we use the average over all bootstrap samples $\overline{acc}$. When using the full data set as training and test set (i.e., within-sample performance), we denote the accuracy by acc_0.

Furthermore, we use the ROC curve for model comparison. It displays the false positive rate vs. the true positive rate. For a given threshold value, we average the ROC curves across all bootstrap samples. The area under the ROC curve for bootstrap sample f, auc_f, serves as a cutoff-independent measure of classification accuracy and we calculate it via the Wilcoxon statistic (Wilcoxon 1945). Once again, we average it over all bootstrap samples, $\overline{auc}$ and use

[7] The results were not sensitive to the choice of metaparameters. For CART, we explored depths from 3 to 20. For the global significance levels of CTree and MOB, we explored values of 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05 and 0.1. For the minimum number of observations a node must contain we explored values of 20, 50, 100, 150, 200, 250 and 500 for all methods. For these choices of depth, number of observations per node and significance level, the results were very similar.

[8] For simplicity, we use the same cutoff value of 0.5 for all segments k.

auc_0 to denote the in-sample area under the curve. For all the classification measures above, higher values imply better predictive capability. By using simultaneous pairwise confidence intervals (using Tukey's all-pairwise comparison contrasts and controlling for the family-wise error rate, cf. Hothorn, Bretz, and Westfall 2008), we assess whether differences in predictive accuracy (between two models) are significant or not. To account for the dependence structure of bootstrap samples, we center the accuracies beforehand (see Hothorn *et al.* 2005).

Measuring targeting effectiveness

While the above measures are interesting from a statistical point of view, a campaign may also want to gauge the monetary gain from applying LORET for targeting. Hence we investigate the targeting effectiveness in a simulated targeting environment.

A targeting range (such as $[0.3, 0.7]$) will contain both voters and non-voters. That is, it will contain individuals who will vote regardless of whether we target them with a customized message or not – and, as we have argued earlier, spending resources on such individuals might be a waste. However, the targeting range will also contain individuals who would not have voted out of their own motivation, but who, with the help of the right targeting message at the right time, will change their mind and will go to the polling stations after all. We will refer to these latter individuals broadly as "non-voters." Spending resources on non-voters is not wasteful, especially if there is a chance of swaying them. Thus, a targeting method is most effective, if – for a given targeting range – it identifies the largest number of non-voters and at the same time the lowest number of voters. We therefore assess the cost-benefit of a targeting method in the following way:

Since we know the outcome for the data at hand, we can treat each training/test sample as a possible targeting situation and compare costs for the presented methodology. We assign a monetary value to convincing a real non-voter to attend an election and see how the different LORET models fare in terms of overall cost. To do this we use a linear cost-benefit function for every method m which can be set up for each test sample f, $f = 1, \dots, 10$.

Let o denote the number of individuals identified in the targeting range (i.e., with predicted probabilities within $[0.3, 0.7]$). We target each of these individuals (e.g., by mail, telephone, email, etc) which incurs a cost of c per individuum. Thus, the overall cost of targeting all o individuals equals $o \times c$. Let us assume that out of these o targeted individuals, n were non-voters. Let us assume further that our targeting efforts are effective in the sense that they turn a fraction v of all non-voters into a voter. In other words, while there are n non-voters, our targeting actions turns $n \times v$ of them into voters. Turning non-voters into voters can be assumed to carry a monetary benefit and we denote that benefit by b. Thus, the overall benefit of targeting equals $n \times v \times b$. This leads to a cost-benefit equation of the form

$$s = (n \times v \times b) - (o \times c) \tag{3}$$

Here, s stands for either the loss (if s is negative and hence $o \times c$ bigger than $n \times v \times b$) or gain (if s is positive and hence $o \times c$ smaller than $n \times v \times b$) of targeting. Notice that o and n depend on the chosen LORET version, so we index it with the superscript m. In addition, each test sample is different hence they also depend on the bootstrap sample f. Thus, let $o_f^{(m)}$ denote the number of individuals which model m applied to bootstrap sample f predicts to be in the targeting range. Similarly, let $n_f^{(m)}$ denote all the non-voters contained in $o_f^{(m)}$. We compute the cost-benefit of model m for our hypothetical targeting situation by computing the average over all bootstrap samples.

$$\bar{s}^{(m)} = \frac{1}{F}\sum_{f=1}^{F}(n_f^{(m)} \times v \times b) - (o_f^{(m)} \times c) \tag{4}$$

For each model m, we explore $\bar{s}^{(m)}$ over a range of plausible values for v, b and c.

We also investigate the break-even point, b_0, i.e., the minimum benefit value of turning a non-voter into a voter for which, at a given targeting cost per person and a given effectiveness, the overall cost-benefit equals zero. We calculate it via the identity

$$b_{0f}^{(m)} = \frac{o_f^{(m)} \times c}{n_f^{(m)} \times v}, \tag{5}$$

which is proportional to the ratio of people in the targeting range and the number of real non-voters for a given ratio of targeting cost and targeting effectiveness $\left(b_{0f}^{(m)} \propto \frac{o_f^{(m)}}{n_f^{(m)}}\right)$. Again we average over all bootstrap samples f to get $\bar{b}_0^{(m)}$.

4. Results

In this section, we compare the methods from Section 2 using the measures described in Section 3.

4.1. Predictive accuracy

Looking at Figure 3 which shows boxplots of the predictive accuracy for the bootstrap samples as well as the within-sample accuracy (denoted by a cross) at a cutoff value of 0.5, one can see quite clearly how the different models from Table 4 behave. First, using both variable sets (the standard set and the extended set together) leads to a huge improvement in predictive accuracy as compared to just using the standard set. Interestingly, the improvement of using both the "s" and "e" variables over using only "s" is bigger than the improvement of using only "s" over using no covariates at all (cf. Figure 3). Second, LORET versions that employ recursive partitioning feature a better performance than global regression models alone. This holds for either using only the standard variable set as well as the combination of extended and standard set. This can also be seen in Figure 4 which displays the average classification accuracies as a function of different cutoff values in the upper panel and the mean ROC curves in the lower panel (averaged over the $F = 10$ out-of-bag samples).

Table 5 gives a detailed summary of the different performance measures for all models. The benchmark of the naive model $y \sim 1 \,|\, 1$ is an average prediction accuracy of $\overline{acc} = 70.36\%$ and an average AUC of $\overline{auc} = 0.5$, averaged over all test sets.

Global logistic regression models $y \sim s \,|\, 1$ and $y \sim s + e \,|\, 1$ display improved performance ($\overline{acc} = 74.97\%$ and $\overline{auc} = 0.740$ for the standard set and $\overline{acc} = 84.57\%$ and $\overline{auc} = 0.886$ for the combined set) with a huge improvement of the model that uses both variable sets.

Both classification tree algorithms, CART and CTree, used to estimate $y \sim 1 \,|\, s$ and $y \sim 1 \,|\, s + e$ result in a generally better performance compared to logistic regressions, both on the

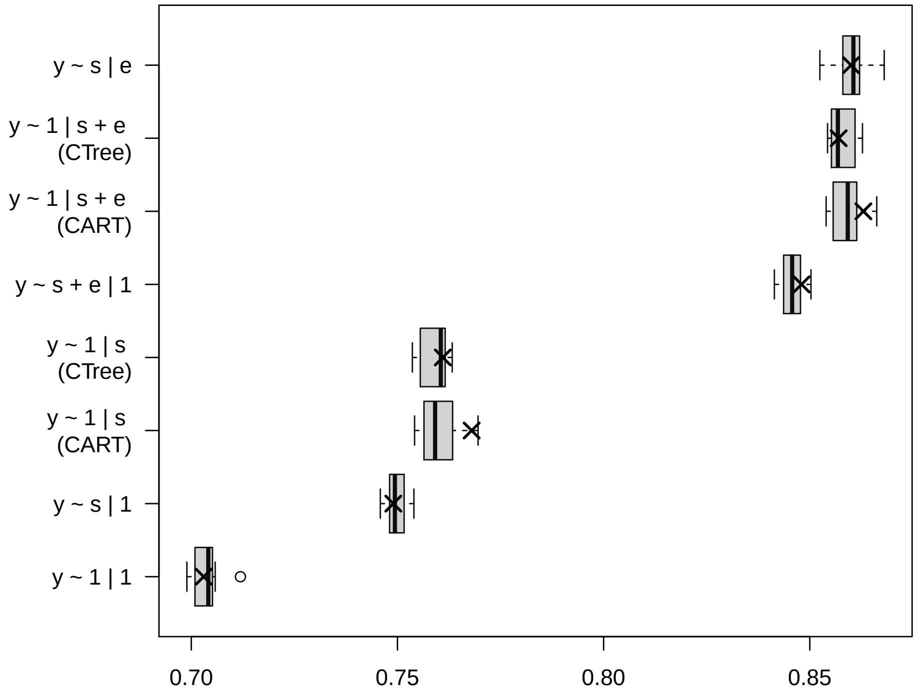


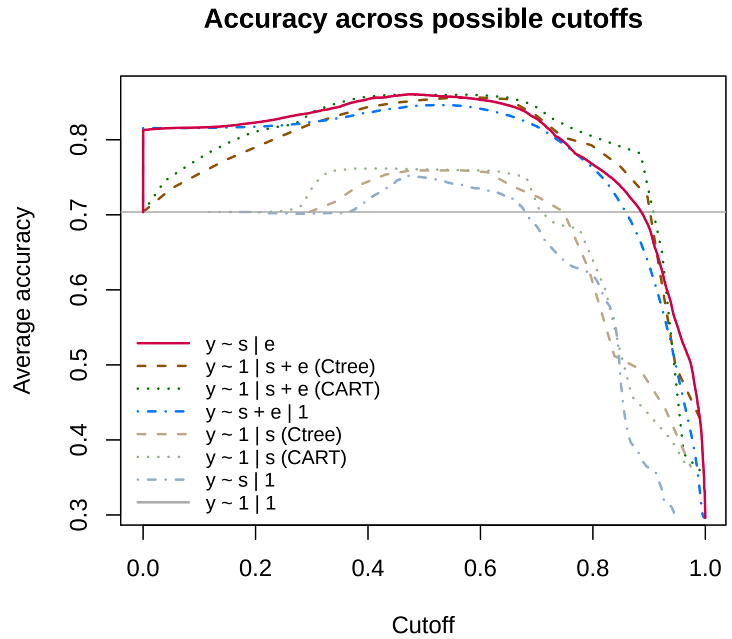
Figure 3: Boxplots of accuracies for all 10 out-of-bag samples for each LORET instances. The cross denotes the within-sample prediction accuracies of each model (acc_0).

standard set of predictors as well as for combining the standard and the extended set. Their performance peaks for the combined set with values of $\overline{acc} = 85.96\%$ and $\overline{auc} = 0.878$ for $y \sim 1 \,|\, s + e$ (CART) and $\overline{acc} = 85.78\%$ and $\overline{auc} = 0.898$ for $y \sim 1 \,|\, s + e$ (CTree).

For the LORET that uses the standard set of predictors as the model in the terminal nodes of the tree and the extended set of predictors for partitioning, i.e., $y \sim s \,|\, e$ result values of $\overline{acc} = 85.98\%$ and $\overline{auc} = 0.906$, respectively. Notice that this model yields the best mean AUC and, at this cutoff, the highest mean accuracy.

The performance differences of models using only standard variables and models employing both the standard and the extended variable sets are evident (see Table 5 and Figure 3). Makin use of the additional variables leads to highly improved performance.

However, the differences among the models employing the combined set themselves (especially between global logistic regression model and partitioned models) are not that strong. Therefore, to establish that these performance differences are not just due to chance, we calculated simultaneous 95%-confidence intervals of all pairwise performance differences between the models that use the combined set of variables based on their accuracy as well as AUC. The former can be found in the upper panel of Figure 5, the latter in the lower panel. We can see that the global logistic regression model performs significantly worse compared to the partitioned models. The tree methods perform best in terms of the accuracy and there are no significant differences amongst them. In contrast, in terms of the cutoff free measure AUC the $y \sim s \,|\, e$ LORET significantly outperforms all other methods.



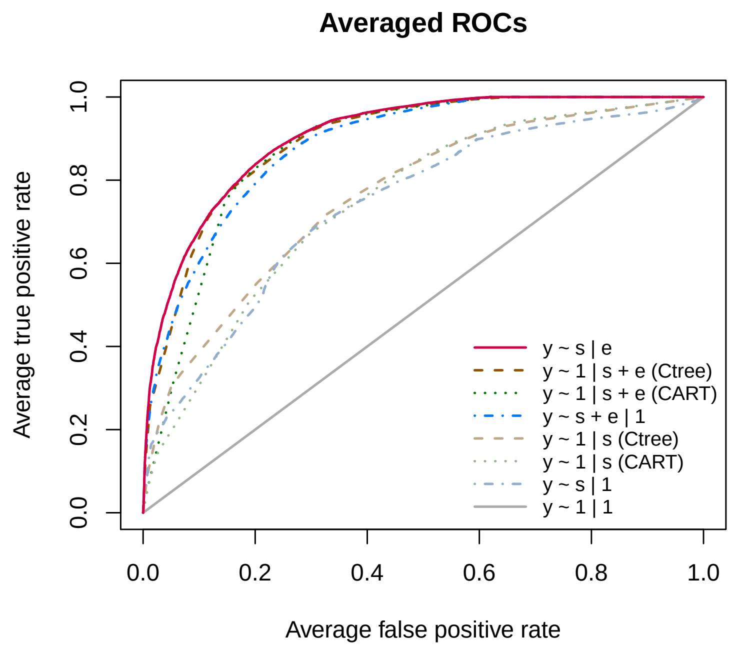


Figure 4: Performance indicators for different models. The upper panel displays features the average accuracies for the range of different cutoffs for the various LORET instances (for majority vote the average accuracy is displayed as a constant). The lower panel features the averaged receiver operating characteristic (ROC) curve for the different models. Threshold averaging has been used for all methods except majority vote.

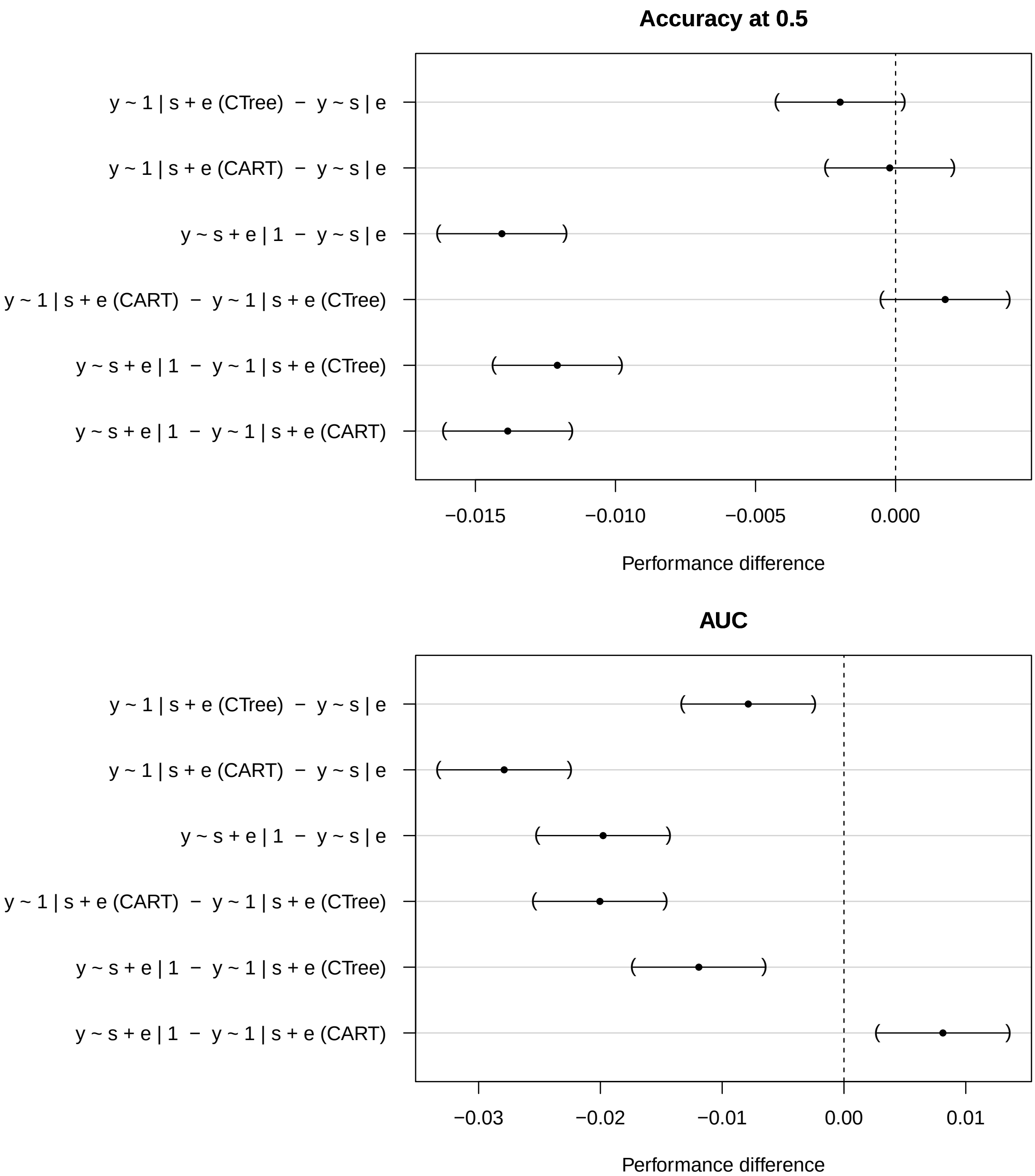


Figure 5: Simultaneous pairwise confidence intervals of the differences of mean accuracies at a cutoff 0.5 over all 10 out-of-bag samples (upper panel) and differences of the average area under the ROC curve (AUC) over all 10 out-of-bag samples (lower panel) for all methods employing the combination of the standard and extended variable set.

Method	Bootstrap samples					Full sample			
	$\overline{acc}$	$se(acc)$	$\overline{auc}$	p	$\tilde{r}$	acc_0	auc_0	p_0	r_0
$y \sim 1 \mid 1$	0.704	0.004	0.500	1	1.0	0.703	0.500	1	1
$y \sim s \mid 1$	0.750	0.002	0.740	8	1.0	0.749	0.739	8	1
$y \sim 1 \mid s$ (CTree)	0.759	0.004	0.765	1	15.0	0.761	0.762	1	14
$y \sim 1 \mid s$ (CART)	0.760	0.005	0.745	1	28.5	0.768	0.746	1	27
$y \sim s + e \mid 1$	0.846	0.003	0.886	57	1.0	0.848	0.888	57	1
$y \sim 1 \mid s + e$ (CTree)	0.858	0.003	0.898	1	18.0	0.857	0.898	1	18
$y \sim 1 \mid s + e$ (CART)	0.860	0.004	0.878	1	23.5	0.863	0.886	1	23
$y \sim s \mid e$	0.860	0.004	0.906	8	9.5	0.860	0.909	8	8

Table 5: Summary of performance indicators for each LORET instance. For the bootstrap samples, $\overline{auc}$ means the area under the ROC curve averaged over all 10 out-of-bag test sets. $\overline{acc}$ is the overall classification accuracy averaged over all test sets and $se(acc)$ its standard error. Complexity is given as the number of estimated parameters per segment (terminal node) p and the median number of segments $\tilde{r}$. For the full sample models (fitted and evaluated on all observations), the accuracy is given by acc_0, the AUC by auc_0 and the number of terminal nodes and cofficients in each node by r_0 and p_0, respectively.

4.2. Cost-benefit analysis

We evaluate the cost-benefit equation in (4) for each of our candidate models[9]. To that end, we investigate a range of scenarios for c (the cost of targeting a single person), b (the monetary benefit of turning a non-voter into a voter) and v (the effectiveness of a targeting message, that is, the proportion of non-voters that it will convert to voters). In fact, for c we investigate values of USD 5 and 15 as examples of low and high targeting costs. This is reasonable, since the 2008 Obama campaign spent roughly USD 8 on each vote President Obama got. Furthermore, we assume that the effectiveness v of a campaign can be either 0.3 or 0.1. While 0.3 is probably quite optimistic, a value of 0.1 would only require mobilizing every 10th non-voter to go to the polls. Putting a number on the monetary benefit b of turning a non-voter into a voter is the biggest challenge. In fact, b might be very small for campaigns that are expected to win in a landslide (i.e., for campaigns where one or two extra voters do not make any difference). However, for campaigns that expect a very close race, b might be extremely large. One example from recent history is the 2000 presidential election. In that election, George W. Bush won the State of Florida (and subsequently the presidency) from Al Gore by a margin of about only 500 votes (see, e.g., Agresti and Presnell 2002). Clearly, in such tight races, campaigns would put an extremely large value on b. In our analysis, we investigate values of b ranging between USD 0 and USD 500.

Figure 6 shows the results. The abscissa refers to different values of b; on the ordinate we find $\bar{s}^{(m)}$ as defined in (4). Notice that positive values of $\bar{s}^{(m)}$ correspond to a monetary gain; negative values indicate losses. Figure 6 displays scenarios for the four different combinations of c and v, starting with $c = 5$ and $v = 0.3$ (top left panel) and ending with $c = 15$ and $v = 0.1$ (bottom right panel).

[9] We only evaluate it for models based on the complete set of predictors since we have found in the previous section that using the standard set of predictors only leads to suboptimal performance.

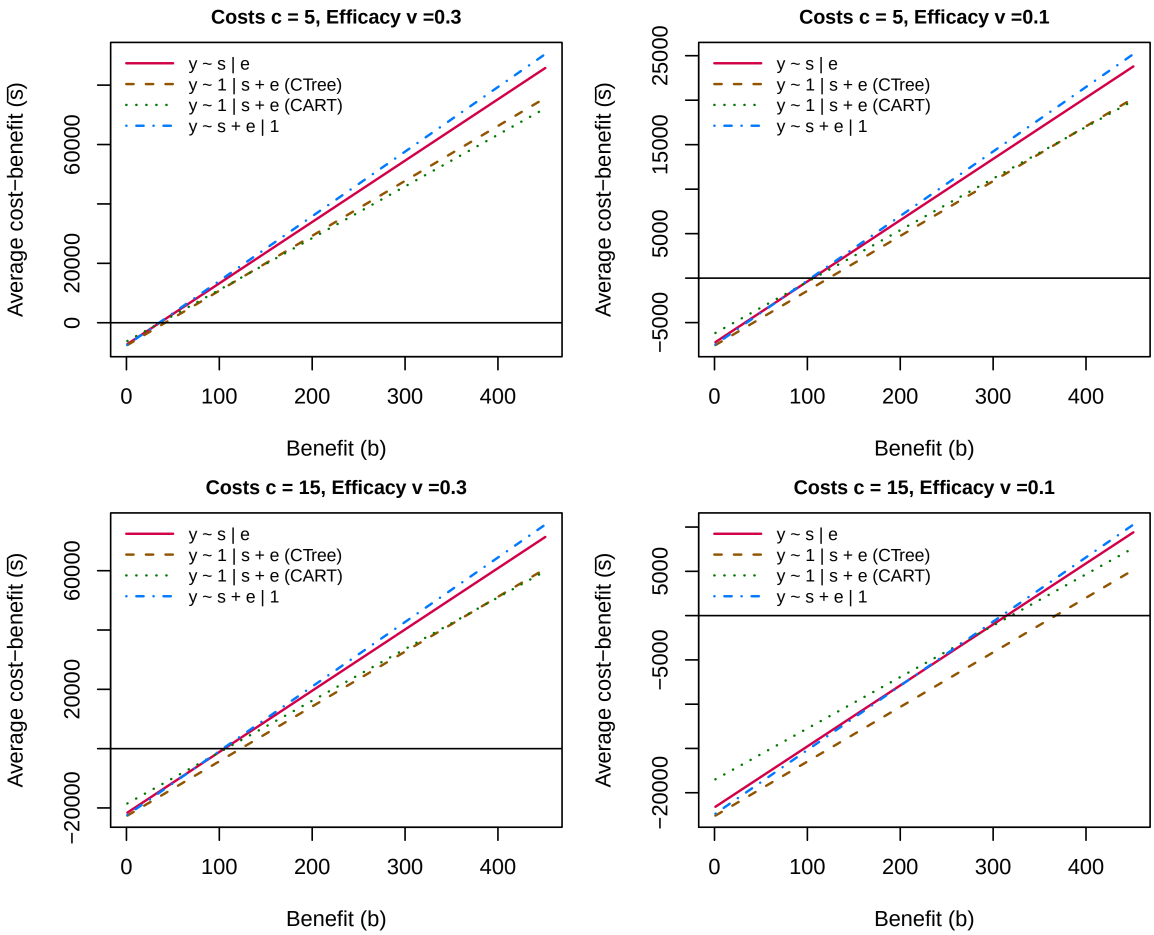


Figure 6: Average linear cost-benefit functions for different versions of LORET. The assumed costs c were USD 5 and USD 15 and the assumed efficacy v of the targeting measure was 0.3 and 0.1. The monetary benefit for turning a non-voter into a voter is depicted on the x-axis, the overall loss ($\bar{s} < 0$) or gain ($\bar{s} > 0$) of the targeting is displayed on the y-axis. The targeting range was $[0.3, 0.7]$.

We can see that the slopes of the cost-benefit function is lowest for classification trees $y \sim 1 \,|\, s + e$. For the CART-based classification tree however, the intercept is highest. This means that for a small benefit b of turning a non-voter into a voter and for a high cost c of targeting, a CART-based classification tree will perform best (i.e., leads to the lowest cost), but only in the loss region (i.e., the lowest loss occurs). With increasing values of b, $y \sim s + e \,|\, 1$ and $y \sim s \,|\, e$ both perform increasingly better – notice the much larger slope which suggests that both methods are especially valuable when there is a large benefit in turning a non-voter into a voter (such as in a tight races). Here, the global logistic regression model eventually performs best for high values of b. The implication for election campaigns is that the LORET framework can be used as a toolbox to increase monetary efficiency of voter targeting, tailor-made for different circumstances. Exactly how it should be used depends on the nature of the race.

Regarding the break even point (which is proportional to $\frac{o_f}{n_f}$, the ratio of people in targeting range to the number of real non-voters for a given ratio of cost and effectiveness), we calculate the mean of the $\frac{o_f}{n_f}$ as $\bar{b}_0 = 2.06$ for $y \sim s{+}e \,|\, 1$, $\bar{b}_0 = 2.09$ for $y \sim s \,|\, e$, $\bar{b}_0 = 2.15$ for $y \sim 1 \,|\, s{+}e$ (CART) and $\bar{b}_0 = 2.46$ for $y \sim 1 \,|\, s + e$ (CTree) under the assumption of USD 1 targeting cost and perfect effectiveness of the targeting measures (i.e., $v = 1$). Hence, *ceteris paribus*, targeting with $y \sim s + e \,|\, 1$ amortizes targeting costs fastest, closely followed by $y \sim s \,|\, e$.

4.3. Interpretability of LORET models

Apart from being able to provide a high classification accuracy, the LORET framework allows to fit interpretable and easily intelligible models that provide further insight into the dynamics of voting behavior relevant for voter targeting. This is one of the major strengths of this approach as compared to "black-box" methods with high predictive capabilities. As point in case, consider the most general LORET, $y \sim s \,|\, e$. Since it has the highest accuracy and AUC and enables efficient targeting for a high benefit of turning non-voters around, we fit it to the whole data set to shed more light on its performance and the turnout of our sample. A table of the decision rules and the coefficients for the logistic regression model in each terminal node can be found in Table 6.

We can see that the segmentation is driven by only four variables, the party composition of the household for each voter ("partyMix"), the relative frequency of attended elections ("attendance"), the rank of the individual in the household ("hhRank", with "1" being highest and "3+" being lowest) and if the person is the head ("H") or a member ("M") of the household ("hhHead"). Hence most partitioning variables are concerned with the household structure rather than with individual-level variables. This is in accordance with literature on the importance of the household for voting behavior (e.g., Cutts and Fieldhouse 2009). We can further see that for all of those individuals for whom "partyMix" is unknown, the probability to vote is zero (actually a case of quasi-complete separation, Albert and Anderson 1984).

The segmentation gives rise to different logistic models that provide additional targeting suggestions for a campaign. We find substantial heterogeneity in the data set as to how voting history influences the outcome. For instance, in node 7 (people who attended elections quite often so far) we see that a higher turnout in earlier elections is associated with a relatively low probability to vote in 2004. Hence these people usually cast their ballot, but for some reason they did less so in 2004. This appears to be a segment that would have been ready for targeting.

The influence of age is also interesting. We specified a quadratic effect and see that, apart from node 10, the estimated probability increases with increasing age just to slow down and reverse. This turning point is rather high for nodes 7, 8 and 10 (70 to more than a 100 years) but substantial in nodes 12 (53.5 years) and 13 (51.1 years). For node 10 it is even at an age of 42. Node 10 is special insofar as it contains young people that have a low rank in the household.

5. Conclusions

In this paper a framework for voter targeting has been proposed, that combines ideas of trees with the idea of logistic regression, coined LORET. The performance of different specifications of LORET with different algorithms in terms of predictive accuracy as well as intelligibility of the models for an exemplary data set has been investigated. Furthermore, a simple linear

Node	Partitioning variables				Regressor variables							
	partyMix	attend.	hhRank	hhHead	const.	gen00	gen01	gen02	gen03	ppp04	age	age^2
2	unknown	–	–	–	$-\infty$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
					(–.–)	(–.–)	(–.–)	(–.–)	(–.–)	(–.–)	(–.–)	(–.–)
6	allD	≤ 0.48	–	–	0.508	0.840	−1.474	0.287	−0.750	0.442	0.054	−0.038
					(0.623)	(0.269)	(0.212)	(0.212)	(0.212)	(0.231)	(0.024)	(0.022)
7	allR, onlyRorD	≤ 0.48	–	–	0.427	0.740	−0.465	0.756	−0.075	0.708	0.011	−0.004
					(0.660)	(0.239)	(0.174)	(0.185)	(0.177)	(0.169)	(0.028)	(0.027)
8	allR, allD,	> 0.48	–	–	2.760	0.277	−1.164	0.352	−1.890	−0.952	0.035	−0.017
	onlyRorD				(0.948)	(0.339)	(0.352)	(0.379)	(0.604)	(0.354)	(0.025)	(0.021)
10	noneRorD, noneD,	–	–	3+	4.057	0.781	0.591	1.249	1.520	0.677	−0.250	0.272
	noneR, legal				(0.797)	(0.128)	(0.203)	(0.165)	(0.214)	(0.212)	(0.052)	(0.076)
12	noneRorD, noneD,	–	$< 3+$	H	−3.630	1.415	−0.010	1.521	2.218	1.694	0.116	−0.108
	noneR, legal				(0.339)	(0.079)	(0.111)	(0.105)	(0.167)	(0.223)	(0.013)	(0.012)
13	noneRorD, noneD,	–	$< 3+$	M	−1.868	1.217	0.086	1.081	1.700	1.603	0.079	−0.078
	noneR, legal				(0.428)	(0.113)	(0.148)	(0.133)	(0.193)	(0.262)	(0.019)	(0.021)

Table 6: A tabular representation of the terminal nodes for the $y \sim s \,|\, e$ LORET for the whole Ohio voter data set. The first column lists the terminal node numbers. The next four columns list the partitioning variables (party mix, attendance, household rank, and household head) and the split point (if any). The last eight columns list the coefficients (upper row) and standard errors (lower row) for the fitted logistic models in the nodes. Please note that the values for the quadratic effect of age have been multiplied by 100 for readability.

cost-benefit analysis of targeting within this framework has been illustrated.

We find that the flexibility introduced by the tree structure leads to more accurate predictions. Furthermore, the framework enables the use of different targeting strategies for different situations. It is easy to understand or communicate to people who are familiar with logistic regression and/or trees and as such the framework is well suited for the purpose of voter targeting.

Regarding the special cases of LORET, a tree with a logistic node model (estimated with the MOB algorithm) may be the most useful default version. For our data, it has the best cutoff-independent predictive accuracy (measured by AUC) and the highest predictive accuracy (at a cutoff of 0.5). Additionally it has the advantage of being easily intelligible and of providing insight for refined targeting. As a result, decisions based on the $y \sim s \,|\, e$ LORET are easy to communicate to campaigns that already use logistic regression. Furthermore it has good potential for cost-efficient targeting, at least based on our sample.

The other instances of LORET, however, are not without merit either. Specifically, a LORET of the $y \sim 1 \,|\, s + e$ type is a good choice if it is not clear how the functional form in the nodes should look like or if there is no standard set of variables to be used in the terminal nodes. Here the nonparametric nature of classification trees show their advantage. If the focus of targeting lies in reducing targeting costs alone, logistic regression and model trees allow most flexible resource allocation and hence may lead to most efficent targeting. For our data set, targeting based on the $y \sim s + e \,|\, 1$ LORET performed best in the cost-benefit analysis. Therefore, even a LORET with just a root node can come in handy.

With the benefits analysed above, one would consider how to incorporate this technique into the overall campaign strategy. Although it is outside of the scope of this study, it needs to be pointed out that it is important for the campaigns to implement any GOTV programs on the likely supporters of the campaign if the intention is to increase the turnout of the supporters. There are three ways that campaigns target likely supporters. First, campaigns use voting results data per precinct from previous elections and gather a general understanding of the demographic and geographic profile of potential supporters. Second, more commonly, they conduct polls with representative samples. The additional benefit of running the polls is that the campaign can be more specific in profiling potential supporters and issues that would motivate them to turn out to vote. Third, campaigns use short surveys over the phone or go door to door interviewing voters to identify individuals who are supporters as well as potential supporters. The primary benefit of using this method is that campaigns can have specific individual level identification of potential supporters. This would also give campaigns the ability to customize communications to each individual. Once the campaigns have better knowledge of the potential voters profile and the likelihood of them voting, campaigns can maximize the return for each dollar spent targeting potential voters by communicating on issues that matter to them and only target voters who are likely to turn out to vote.

Another use of this modelling technique would be to suppress potential supporters of the opposition. This is often called negative campaigning or using “dirty tricks”, but it is logical for campaigns to use this method to target voters who might fit into a profile that classify them as potential but not strong supporters of the campaign’s opponent. Common ways the campaigns often incorporate this strategy would be to send negative attack message about the opponent to discredit the opponent’s character or even distort facts to create confusion. Another tactic that a campaign could use is to assist or send anonymous support for another candidate that shares the similar political philosophy. For example, for the 1992 presidential election, Ross Perot was an independent candidate; however he had a great amount of support from mostly republican party supporters. The democratic candidate Bill Clinton benefited from Perot dividing the republican electorate. In 2000, the democratic campaign was faced with the similar problem. Ralph Nader was an independent presidential candidate that attracted support from primarily democrats. The republican candidate George W. Bush benefited from it as his opponents had to campaign for the same pot of voters.

The bottom line is that this framework does not change the commonly used campaign tactics but it would influence campaign strategy because it is a more precise tool that would allow campaigns to target the recipients of positive or negative messages more accurately and efficiently which would give more options. With the LORET framework, campaigns have a flexible and versatile toolbox for GOTV targeting that can be customized to meet the requirements at hand.

For further research and practical application, it is possible to improve aspects of interest in GOTV campaigns. For example, it might be fruitful to use techniques such as artificial neural networks or ensembles of tree methods to improve predictive accuracy[10]. Regularized logistic regression models might prove to be a sensible alternative to the tree approach, especially in terms of interpretability and variable selection. It could also be interesting to improve the cost-benefit aspect by defining an appropriate objective function that explicitly incorporates the targeting costs which can then be minimized to yield LORET models that use these specific loss functions rather than the standard ones.

Computational details

All calculations have been carried out with the statistical software R 2.12.0–2.14.1 (R Development Core Team 2011). Logistic regression was fitted with the `glm()` function. Recursive partitioning infrastructure was provided by the packages **party** for `mob()` (Zeileis *et al.* 2008) and `ctree()` (Hothorn *et al.* 2006), as well as **rpart** (Therneau and Atkinson 1997; Therneau, Atkinson, and Ripley 2011) for CART. We used the **ROCR** package (Sing, Sander, Beerenwinkel, and Lengauer 2005, 2009) for calculating and plotting performance measures and ROC curves and **multcomp** (Hothorn *et al.* 2008) for the simultaenous confidence intervals.

References

Agresti A, Presnell B (2002). "The 2000 Presidential election in Florida: Misvotes, undervotes, overvotes." *Statistical Science*, **17**(4), 436–440.

Albert A, Anderson JA (1984). "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models." *Biometrika*, **71**(1), 1–10.

Baek M (2009). "A Comparative Analysis of Political Communication Systems and Voter Turnout." *American Journal of Political Science*, **53**(2), 376–393.

[10]We used random forests and logistic model trees with boosting in the nodes during the course of the study. On this data set their performance was not significantly better than the performance of the LORET models.

Brams SJ, Davis MD (1973). "Resource-Allocation Models in Presidential Campaigning: Implications for Democratic Representation." *Annals of the New York Academy of Sciences*, **219**, 105–123.

Breiman L, Friedman JH, Olsen RA, Stone CJ (1984). *Classification and Regression Trees.* Wadsworth, Pacific Grove, CA.

Chan KY, Loh WY (2004). "LOTUS: An Algorithm for Building Accurate and Comprehensible Logistic Regression Trees." *Journal of Computational and Graphical Statistics*, **13**, 826–852.

Chaudhuri P, Lo WD, Loh WY, Yang CC (1995). "Generalized Regression Trees." *Statistica Sinica*, **5**, 641–666.

Cox G, Munger M (1989). "Closeness, Expenditures, and Turnout in the 1982 U.S. House Elections." *American Political Science Review*, **83**, 217–231.

Cutts D, Fieldhouse E (2009). "What Small Spatial Scales Are Relevant as Electoral Contexts for Individual Voters? The Importance of the Household on Turnout at the 2001 General Election." *American Journal of Political Science*, **53**, 726–739.

Denny K, Doyle O (2009). "Does Voting History Matter? Analysing Persistence in Turnout." *American Journal of Political Science*, **53**(1), 17–35.

Efron B, Tibshirani R (1993). *An Introduction to the Bootstrap.* Chapman & Hall/CRC, Boca Raton, FL.

Finkel S (1993). "Reexamining the 'Minimal Effects' Model in Recent Presidential Elections." *Journal of Politics*, **55**, 1–21.

Gelman A, King G (1993). "Why Are American Presidential Election Polls so Variable when Votes Are so Predictable?" *British Journal of Political Science*, **23**, 409–519.

Gillespie A (2010). "Canvasser Affect and Voter Response: Results From National Focus Groups." *American Politics Research*, **38**(4), 718–758.

Goldstein K, Ridout TN (2002). "The Politics of Participation: Mobilization and Turnout over Time." *Political Behavior*, **24**(1), 3–29.

Green DP, Gerber AS (2008). *Get out the Vote: How to Increase Voter Turnout.* 2nd edition. Brookings Institution, Washington DC.

Hall MG, Bonneau CW (2008). "Mobilizing Interest: The Effects of Money on Citizen Participation in State Supreme Court Elections." *American Journal of Political Science*, **52**(3), 457–470.

Hand DJ, Yu K (2001). "Idiot's Bayes – Not so Stupid after All?" *International Statistical Review*, **69**, 385–399.

Hansen BB, Bowers J (2009). "Attributing Effects to a Cluster-Randomized Get-out-the-Vote Campaign." *Journal of the American Statistical Association*, **104**(487), 873–885.

Hastie T, Tibshirani R, Friedman JH (2009). *Elements of Statistical Learning*. 2nd edition. Springer-Verlag, New York.

Hillygus DS, Jackman S (2003). "Voter Decision Making in Election 2000: Campaign Effects, Partisan Activation, and the Clinton Legacy." *American Journal of Political Science*, **47**(4), 583–596.

Holbrook TM, McClurg SD (2005). "The Mobilization of Core Supporters: Campaigns, Turnout and Electoral Composition in United States Presidential Elections." *American Journal of Political Science*, **49**, 689–703.

Hothorn T, Bretz F, Westfall P (2008). "Simultaneous Inference in General Parametric Models." *Biometrical Journal*, **50**(3), 346–363.

Hothorn T, Hornik K, Zeileis A (2006). "Unbiased Recursive Partitioning: A Conditional Inference Framework." *Journal of Computational and Graphical Statistics*, **15**(3), 651–674.

Hothorn T, Leisch F, Zeileis A, Hornik K (2005). "The Design and Analysis of Benchmark Experiments." *Journal of Computational and Graphical Statistics*, **14**, 675–699.

Karp JA, Banducci SA (2007). "Party Mobilization and Political Participation in New and Old Democracies." *Party Politics*, **13**(2), 217–234.

Karp JA, Banducci SA, Bowler S (2008). "Getting out the Vote: Party Mobilization in a Comparative Perspective." *British Journal of Political Science*, **38**, 91–112.

Kass GV (1980). "An Exploratory Technique for Investigating Large Quantities of Categorical Data." *Journal of the Royal Statistical Society C*, **29**(2), pp. 119–127.

Landwehr N, Hall M, Eibe F (2005). "Logistic Model Trees." *Machine Learning*, **59**, 161–205.

Lau RR, Sigelman L, Rovner IB (2007). "The Effects of Negative Political Campaigns: A Meta-Analytic Reassessment." *American Political Science Review*, **69**(4), 1176–1209.

Loh WY, Shih YS (1997). "Split Selection Methods for Classification Trees." *Statistica Sinica*, **7**, 815–840.

Malchow H (2008). *Political Targeting*. 2nd edition. Predicted Lists, LLC, Sacramento, CA.

May KO (1952). "A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decisions." *Econometrica*, **20**, 680–684.

McCullagh P, Nelder JA (1989). *Generalized Linear Models*. 2nd edition. Chapman and Hall, New York.

McDonald M (2012). "Turnout Rates, 1980–2010." United States Election Project. http://elections.gmu.edu/ [accessed 2012-02-16].

Middleton JA, Green DP (2008). "Do Community-Based Voter Mobilization Campaigns Work even in Battleground States? Evaluating the Effectiveness of MoveOn's 2004 Outreach Campaign." *Quarterly Journal of Political Science*, **3**(1), 63–82.

Muller MG (1999). "Electoral Campaigning as an Occupation – The Professionalization of Political Consultants in the United States." *Politische Vierteljahresschrift*, **40**(1), 198–199.

Murray GR, Scime A (2010). "Microtargeting and Electorate Segmentation: Data Mining the American National Election Studies." *Journal of Political Marketing*, **9**(3), 143–166.

Parry J, Barth J, Kropf M, Jones ET (2008). "Mobilizing the Seldom Voter: Campaign Contact and Effects in High-Profile Elections." *Political Behavior*, **30**(1), 97–113.

Plasser F (2000). "American Campaign Techniques Worldwide." *Harvard International Journal of Press-Politics*, **5**(4), 33–54.

Quelch J (2008). "How Political Marketing Can Learn from Consumer Marketing." http://blogs.hbr.org/quelch/2008/01/how_political_marketing_can_le.html.

Quinlan JR (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishing, San Mateo, California.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Ridout TN (2009). "Campaign Microtargeting and the Relevance of the Televised Political Ad." *Forum – A Journal of Applied Research in Contemporary Politics*, **7**(2), 1–13.

Rusch T, Zeileis A (2012). "Ganining Insight with Recursive Partitioning of Generalized Linear Models." *Journal of Statistical Computation and Simulation*. Forthcoming.

Sing T, Sander O, Beerenwinkel N, Lengauer T (2005). "ROCR: Visualizing Classifier Performance in R." *Bioinformatics*, **21**(20), 3940–3941.

Sing T, Sander O, Beerenwinkel N, Lengauer T (2009). *ROCR: Visualizing the Performance of Scoring Classifiers*. R package version 1.0-4, URL http://CRAN.R-project.org/package=ROCR.

Snyder JM (1989). "Goals and the Allocation of Campaign Resources." *Econometrica*, **57**(3), 637–660.

Steel BS, Pierce JC, Lovrich NP (1998). "Public Information Campaigns and 'At-Risk' Voters." *Political Communication*, **15**(1), 117–133.

Sussman G, Galizio L (2003). "The Global Reproduction of American Politics." *Political Comunication*, **20**(3), 309–328.

Therneau TM, Atkinson EJ (1997). "An Introduction to Recursive Partitioning Using the rpart Routine." *Technical Report 61*, Section of Biostatistics, Mayo Clinic, Rochester. URL http://www.mayo.edu/hsr/techrpt/61.pdf.

Therneau TM, Atkinson EJ, Ripley BD (2011). *rpart: Recursive Partitioning*. R package version 3.1-50, URL http://CRAN.R-project.org/package=rpart.

US Election Assistance Comission (2010). "The Impact of the National Voter Registration Act of 1993 on the Administration of Elections for Federal Office 2009–2010."

Whitelock A, Whitelock J, van Heerde J (2010). "The Influence of Promotional Activity and Different Electoral Systems on Voter Turnout: A Study of the UK and German Euro Elections." *European Journal of Marketing*, **44**(3-4), 401–420.

Wielhouwer PW (2003). "In Search of Lincoln's Perfect List – Targeting in Grassroots Campaigns." *American Politics Research*, **31**(6), 632–669.

Wilcoxon F (1945). "Individual Comparisons by Ranking Methods." *Biometrics Bulletin*, **1**, 80–83.

Zeileis A, Hothorn T, Hornik K (2008). "Model-Based Recursive Partitioning." *Journal of Computational and Graphical Statistics*, **17**(2), 492–514.

Zhang H, Singer B (2010). *Recursive Partitioning and Applications.* 2nd edition. Springer-Verlag, New York.

Affiliation:

Thomas Rusch
Institute for Statistics and Mathematics
WU (Wirtschaftsuniversität Wien)
Augasse 2–6
1090 Wien, Austria
E-mail: Thomas.Rusch@wu.ac.at