Successfully Implementing Predictive Analytics in Direct Marketing
John Blackwell and Tracy DeCanio, The Nature Conservancy, Arlington, VA

ABSTRACT
Successfully Implementing Predictive Analytics in Direct Marketing

An analysis of the benefits of modeling in direct marketing as well as a review of the dangers associated with common data mining mistakes. The presentation will review specific examples of how SAS® data mining products have been successfully utilized at the one of the world's largest non-profit organizations to dramatically improve ROI.

Specifically, this session will focus on 1) measuring the true value of predictive models and 2) preparing data for modeling and avoiding pitfalls and 3) building models and selecting the best for deployment and 4) the importance of industry expertise and questioning what the data are "saying" and 5) assessing the readiness of your organization in building analytics capacity in-house.

INTRODUCTION
As organizations scale back marketing budgets many are foregoing the expense associated with predictive analytics. However, the importance of analytics must be viewed in the context of being ready to bounce back when the economy rebounds as opposed to stressing that there can also be a short term benefit despite the current climate. While it is true that data-mining might not be able to completely reverse decreases in revenue, it can have immediate positive return on investment. This paper presents a case study for one of world's largest fundraising organizations who used analytics several years ago to mitigate the impact of a negative economic environment and continues to do so today in order to efficiently raise money.

BACKGROUND
The Nature Conservancy ("TNC") is the world’s largest environmental non-profit with a mission to “preserve the plants, animals and natural communities that represent the diversity of life on Earth by protecting the lands and waters they need to survive.” TNC is able to fund its mission primarily through fundraising from which it generates about five-hundred million dollars a year from its 3 million members. While data mining is used throughout the conservancy this paper will focus on its application in direct-mail campaigns.

BEFORE ANALYTICS
Prior to 2003, TNC’s decisions about who to mail and how much to ask for were mostly done through “gut-feel” and rules of thumb that had often not been tested empirically. While the organization was mostly successful, events in 2003 including a weaker economy and negative press put downward pressure on TNC’s ability to fundraise. Indeed, a membership base which had previously been growing steadily came to an abrupt halt. To maintain its reputation and continue its mission The Nature Conservancy needed to increase it revenues through increased efficiencies. The organization has met the challenge through making fact-based decisions with an emphasis on analytics. Today when TNC plans a direct mail solicitation they determine who to mail based on which donors are the most likely to respond with the largest gifts.

The following SAS Enterprise Miner diagram shows a typical modeling development flow for a response:
Details of the value behind the ensemble model are discussed in more detail later in the paper.

The score from the above model (using a binary target for response) is then multiplied by a predicted gift amount generated through a separate model. The resulting product is a prediction of gross revenue per mailing. The marketing team will then decide who to mail based on a chart summarizing the model results such as the sample below. The chart shows the donors ranked in deciles according to their model score and shows the predicted response, revenue per mailing sent, profit and the cost associated with raising each dollar.

<table>
<thead>
<tr>
<th>Model Rank</th>
<th>Size of Group</th>
<th>Resp. Rate</th>
<th>Rev. Per Piece</th>
<th>Profit</th>
<th>Cost Per $ Raised</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>96,715</td>
<td>6.9%</td>
<td>$7.46</td>
<td>$668,961</td>
<td>$0.07</td>
</tr>
<tr>
<td>2</td>
<td>96,740</td>
<td>5.4%</td>
<td>$2.67</td>
<td>$206,185</td>
<td>$0.20</td>
</tr>
<tr>
<td>3</td>
<td>96,294</td>
<td>4.4%</td>
<td>$1.70</td>
<td>$111,441</td>
<td>$0.32</td>
</tr>
<tr>
<td>4</td>
<td>96,778</td>
<td>3.8%</td>
<td>$1.24</td>
<td>$68,075</td>
<td>$0.43</td>
</tr>
<tr>
<td>5</td>
<td>97,917</td>
<td>3.6%</td>
<td>$0.98</td>
<td>$42,595</td>
<td>$0.55</td>
</tr>
<tr>
<td>6</td>
<td>95,817</td>
<td>2.9%</td>
<td>$0.75</td>
<td>$19,859</td>
<td>$0.72</td>
</tr>
<tr>
<td>7</td>
<td>104,412</td>
<td>2.4%</td>
<td>$0.55</td>
<td>$594</td>
<td>$0.99</td>
</tr>
<tr>
<td>8</td>
<td>89,794</td>
<td>1.8%</td>
<td>$0.41</td>
<td>($11,857)</td>
<td>$1.32</td>
</tr>
<tr>
<td>9</td>
<td>95,245</td>
<td>1.5%</td>
<td>$0.30</td>
<td>($22,643)</td>
<td>$1.79</td>
</tr>
<tr>
<td>10</td>
<td>97,424</td>
<td>1.3%</td>
<td>$0.20</td>
<td>($32,891)</td>
<td>$2.67</td>
</tr>
</tbody>
</table>
As illustrated above, the model is able to identify a significant portion of the available audience for any given mailing that is not profitable to mail (wherever the cost per dollar raised is greater than one dollar) in addition to identifying which donors will be the most profitable.

**MAKING THE CASE**

An initial hurdle in making the case for predictive modeling was to convince management that it would be beneficial not only to the organization but also to the membership population. Charitable giving is highly seasonal and while many people may prefer to give at the end of the year some are also more likely to give around tax time and others over the summer months. It was clear that we had to show that modeling could help to better understand the seasonality of the subsets of the membership population and could actually result in a decrease in the amount of direct mail sent without a corresponding decrease in revenue. For example a model could allow us to target our audience when they were ready to give as opposed to sending an expensive, steady stream of solicitations throughout the year.

**PROVING THE VALUE**

In their book *Competing on Analytics*, Davenport and Harris refer to the “detour” that some organizations need in order to prove the value of fact-based decisions before analytics can be fully implemented. At The Nature Conservancy there was much skepticism about the notion that a model could make better predictions than current business practices. For the first few campaigns that were modeled we randomly split the available audience into two segments and allowed the Marketing team to apply their usual selection criteria to the first half. For the second half, we used a model select the accounts to be mailed. At the completion of the campaign the results were analyzed and shown to the decision-makers. In all cases the modeled select significantly out-performed the “standard” select. The chart below shows how the modeling boosted revenue for monthly solicitation campaigns in the first year after the “detour”. The bars show dollar amount per piece sent and the line represents the percentage increase in total revenue per piece mailed associated with the model select.

While needing to prove the value of modeling took more time and resulted in foregone revenue, the results were definitive and allowed us to get the buy-in from management that we needed to expand the scope of analytics within the organization.
AVOIDING PITFALLS

TOO MUCH DATA?
The massive amounts of data that are now available about an organization's customers are one reason predictive modeling has been so effective. However, if not treated with caution, in some cases it can lead to a weaker model. Vendors that deal in customer data append will often tout the hundreds of demographic and psychographic variables that can be appended to an individual's database record. In addition, the low cost of storage allows most companies to retain hundreds of data points associated with a customer relationship over time, which can be used in model development.

However there is always the danger that correlations of marketing data in a development data set are merely noise and not predictors of future behavior. There are some well-known examples of chance correlations associated with political events. For example, from 1952 to 1976 when an American League team won the World Series a Republican would win the U.S. Presidency. A similar more recent example involving the Washington Redskins was true until the election of 2004. Examples such as the above are often cited as curiosities and are not necessarily taken seriously. However, because a correlation is plausible does not necessarily mean that it will provide any more value than the above. While data is the backbone of a strong predictive model in direct-marketing, it can also weaken models when the apparent correlation is not one that holds up. An article in the Economist on August 18, 2007 effectively sums up the problem: "As the old saw has it, garbage in, garbage out. The difficulty comes when you do not know what garbage looks like."

How do you know what variables are truly predictive and which are just correlated by chance in a development data set? While you can never say what is a true predictor with absolute certainty, there are several approaches that you can use to avoid using noise variables and to lessen their impact if they are selected by a modeling algorithm. The following approach can help to identify which variables are more likely than others to be meaningless and also to lessen the impact that the "garbage" has on a model.

A CASE STUDY
As illustration I created a dataset that contained a binary dependant variable and had about 50 potential independent variables taken from the marketing data base. For the purpose of providing an example, 100 random variables were appended to the actual data. The data were split into development, validation and test samples and an initial model was built only allowing for the random variables to be used as input variables. When run through a stepwise logistic regression and entropy-reduction decision tree model, each model selected between four and six variables. This example is merely to illustrate that even with high significance levels for variable selection, when the number of candidate variables is large enough some will be selected even in the event when there is nothing beyond noise.

In the next scenario the same models were run using the random and the real input variables. In this case both models continue to select two of the random variables. However, the overall fit statistics on the test partition (as shown below) are strong and do not suggest any cause for concern.

<table>
<thead>
<tr>
<th>Model</th>
<th>Test: Roc Index</th>
<th>Test: Gini Coefficient</th>
<th>Test: Kolmogorov-Smirnov Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Tree</td>
<td>0.709</td>
<td>0.419</td>
<td>0.289</td>
</tr>
<tr>
<td>Regression</td>
<td>0.715</td>
<td>0.429</td>
<td>0.329</td>
</tr>
</tbody>
</table>

When re-running the model without the random variables we can see that the fit statistics on the test partition of the data set improve. In the case of the logistic regression the same non-random variables are selected, however, given the higher ROC index on the test data this suggests stronger coefficients. In the case of the decision tree, the random variables that had been selected were selected as the second to last split and removing them allowed the tree to grow differently, resulting in a stronger model. Since at any depth a decision tree will always look for the next most significant split without regard to the best possible combination of splits, if it is confused at one level by noise, predictive variables might never be selected.
Model | Test: Roc Index | Test: Gini Coefficient | Test: Kolmogorov-Smirnov Statistic
--- | --- | --- | ---
Decision Tree | 0.738 | 0.477 | 0.356
Regression | 0.742 | 0.483 | 0.360

The above examples serve to show that it is easy for noise variables to be selected by models, and the emphasis on any candidate predictor variable should be less about deciding whether the split is plausible and more about questioning whether the relationship is real. While it may be impractical to fully analyze every candidate variable for every model and still meet deadlines, one way to guard against noise is to perform in-depth analysis on new variables appended to the file as well as variables that are only selected by one particular modeling technique.

Similar to the above, sometimes merely questioning the data and treating all relationships with scientific skepticism can be useful. A fairly well-known example of the pitfalls associated with not questioning what the data are saying is a look at the apparent correlation between SAT scores by state and the amount of money spent on education:

![Average SAT Score vs Average Expense per Pupil](chart.png)

The above chart shows that on the balance, the more spent on education the lower the test scores are. However, someone familiar with the subject would be able to explain that the correlation is not meaningful since the states that spend less on education are over-represented in states where students take the ACT as opposed to the SAT. In these states the students that take the SAT are often those that are applying to competitive east-coast schools that require the test, making the data useless as a display of the effect of money on education.

The more familiar an analyst is with the industry and the data (s)he is analyzing, the less likely (s)he is to be fooled by an apparent variable that looks to be predictive as in the case above. In the SAT example perhaps it is counter-intuitive enough that most people would question it. However, if the data showed the trend reversed most people would probably never question the apparent relationship.

Another effective approach to building strong models in spite of the preponderance of data is to lessen the impact of any one variable selected by a model by averaging or “ensembling” the scores of several models. Within any
individual model the emphasis is to make it satisfy a certain threshold using as few inputs as possible. However we have found that most of the strongest models that we have created use a large number of input variables and lessen the impact any one of them can have through averaging. For example, a simplified, yet illustrative model flow is shown below:

This flow shows the development data being partitioned and fed through a neural network, regression and a decision tree node before the ensemble node averages the scores of all three individual models to create a new model. Since none of the models select the same set of independent variables the total number is larger, however, the averaging helps to ensure that no one variable has a particularly large impact on the final score. This is particularly true for variables with weaker correlations to the target since these are less likely to be selected by all three modeling techniques. And it is within these weaker yet important inputs where the noise data are most likely to be found.

When the data set that combined the random and non-random variables was run through the model diagramed above the power of the ensemble model is significantly higher than any of the individual models. Random variables were still selected as predictors; however their influence was reduced. There are several reasons this could be the case including the fact that the different models treatment of the random variables largely cancel each other out when combined. The strength of ensembling is also true of the data set without the random variables. As displayed in the ROC chart below the green line represents the averaged model and the others show the individual models.
CONCLUSIONS
While there are many pitfalls associated with modeling in direct marketing they can be avoided through a careful analysis of the data combined with modeling techniques that do not place too much significance on variables that may in fact be noise. Predictive analytics will almost always enable better business decisions. While customer or donor behavior may have changed due to the current economic climate, the value provided by a model will always be better than that driven by instinct or gut-feel. When organizations try and cut the expense associated with analytics it might be incumbent on the analysts to once again take the “detour” as described by Davenport and Harris to prove the value of analytics (even if this has been done previously) by setting up regular experiments to show the difference between the performance of modeled selections for mailings versus a non-modeled select.

REFERENCES:


ACKNOWLEDGMENTS
SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the authors at:

John Blackwell  
The Nature Conservancy  
4245 N. Fairfax Drive  
Arlington, VA  22310  
Work Phone: 703-841-8786  
Email: jblackwell@tnc.org  
Web: www.nature.org

Tracy DeCanio  
The Nature Conservancy  
4245 N. Fairfax Drive  
Arlington, VA  22310  
Work Phone: 703-841-4581  
Email: tdecanio@tnc.org  
Web: www.nature.org