

To: Prof. John Sparks
From: 658795799
Date: April 27, 2015
Re: Model to Contact Sales Prospects to purchase home repairs

This memo is in response to your request for an analysis regarding the following three questions:

1. Given our marketing budget of \$25,000 which modeling technique should be used to maximize sales?
2. What is the overall performance of the models?
3. Provide a very brief description of the modeling techniques.

By using MARS and contacting the prospects most likely to purchase home repairs from Mr. Handyman within the marketing budget of \$25,000 we can convert 67 prospects into customers and yield \$33,500 in revenue. This represents an increase in savings of \$7,500 relative to using random selection to contact prospects. Overall MARS out-performs the other technique used (Random Forest, Logistic Regression and CHAID) for the bottom half as seen on the gains chart on page 2. Descriptions of the two techniques are provided on page 3 and 4.

Analysis of Models for Marketing Budget

The constraints of the budget to maximize sales are as follows: The overall budget is \$25,000 and the cost per contact is \$30. Therefore we can contact 833 prospects ($\$25,000/\$30=833$). An additional assumption of this analysis is that each sale of home repairs from Mr. Handyman will yield \$500 in revenue.

Within the constraint of the marketing budget and given the revenue assumption, which modeling technique should we use to contact the sales prospects and yield the maximum amount of revenue? Table 1 summarizes the performance of the four modeling techniques vs. random selection. An analysis follows that also explains how the figures in the table 1 were calculated.

Table 1: Revenue comparison for Modeling technique vs Random selection

Modeling Technique	Revenue at cut-point	Increase in Revenue vs Random Selection
Random Selection (Baseline)	\$26,000	
MARS	\$33,500	\$7,500
Logistic Regression	\$32,000	\$6,000
Random Forest	\$30,500	\$4,500
CHAID	\$29,500	\$3,500

Table 1: Repeated for convenience

Modeling Technique	Revenue at cut-point	Increase in Revenue vs Random Selection
Random Selection (Baseline)	\$26,000	
MARS	\$33,500	\$7,500
Logistic Regression	\$32,000	\$6,000
Random Forest	\$30,500	\$4,500
CHAID	\$29,500	\$3,500

Random Selection (or Baseline) If we select sales prospects randomly for the purchase of home repairs, then we would expect to account for the same proportion of sales as we do the proportion of prospects that we contact. For example, if we contact 10% of the prospects then it would account for 10% of the sales. If we contact 20% of the prospects then it would account for 20% of the sales, etc. The marketing budget allows us to contact 50% of our file. There were 103 total sales on the file. 50% of 103 is 52 total sales prospects with an associated revenue of \$26,000 (52 X \$500).

MARS By comparison, using MARS we reach 67 sales prospects among the 833 sales prospects for a maximum revenue of \$33,500 (67 X \$500). This yields an increase in revenue of \$7,500 relative to using the random selection (\$33,500 - \$26,000).

Logistic Regression If Logistic Regression is used, we would earn \$1,500 less in revenue relative to using the MARS technique. Specifically, this technique converts 64 prospects into customers with an associated revenue of \$32,000 (64 X \$500).

Random Forest If Random Forest is used, we would earn \$3,000 less in revenue relative to using the MARS technique. Specifically, this technique converts 61 prospects into customers with an associated revenue of \$30,500 (61 X \$500).

CHAID If CHAID is used, we would earn \$4,000 less in revenue relative to using the MARS technique. Specifically, this technique converts 59 prospects into customers with an associated revenue of \$29,500 (59 X \$500).

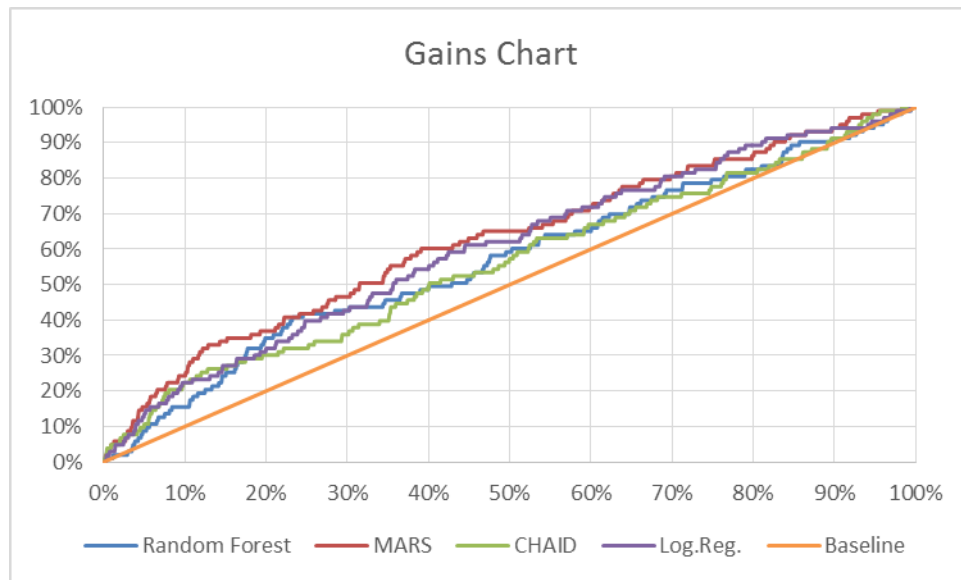
Based on these figures, MARS modeling technique should be used as it allows for maximizing the revenue with the best number of prospects contacted for the sales of home repairs.

Overall Performance Comparison

The gains chart shown on page 3 provides a method to summarize the overall performance of four models. Generally, a gains chart shows comparison of the four models versus the random selection (baseline).

GAINS CHART

Figure 1: Gains Chart for Predicting models vs Baseline



The lift curve for MARS is above the curve for Logistic Regression, Random Forest and CHAID for the range 3% to 52% of total prospects. The lift curve for MARS, Logistic Regression and CHAID have almost the same lift value from 0% and 3%. The lift curve for MARS and Logistic Regression have almost the same lift value from 52% and above. Therefore, MARS generally has a stronger performance in predicting the sales prospects across certain range of percentage of total prospects i.e. from 3% to 52%.

Modeling Techniques

The four techniques used in this analysis were MARS, Logistic Regression, Random Forest and CHAID. A description of each is provided below.

MARS. MARS is a form of regression analysis technique which recognizes the kinks or turning points that exist in the linear relationship. It identifies the kinks when the slope of the line changes remarkably.

Logistic Regression. Logistic regression is a regression technique which measures the relationship between the binary dependent variable and one or more independent variables. The log of odds of response is used as the parameter estimate value to model the linear combination between the independent variables and dependent variable.

Random Forest. Random Forest is an ensemble of trees. Random forest technique is a modified tree learning algorithm that randomly removes the small number of variables and builds the best tree without those variables. It uses the average of the predicted rates for all the different trees to find the best estimate for the probability of response for each prospect.

CHAID. CHAID, Chi-Square Automatic Interaction Detection, computes the chi-square statistic value between the dependent variable and all of the independent variables. The independent variable with the large chi-square statistic is selected. Data is split into groups based on the independent variable. Each of these sub-groups is examined and the process is repeated until a stopping rule is met.

Summary

The results of this analysis have shown that the maximum sales can be reached using MARS as opposed to Logistic Regression, Random Forest and CHAID and that regression clearly out-performs a random selection of prospects. The revenue from using MARS for the sales model to purchase home repairs is \$33,500. MARS also has superior performance relative to Logistic Regression, Random Forest and CHAID across certain range of percentage of total prospects i.e. from 3% to 52%. A description of the four techniques was provided on page 3 and 4.