Assignment - Acquisition Analytics

**Problem Statement**

By now, you would have understood the process for building a response model based on the bank marketing data set. This assignment is similar, with a slight twist.

To reiterate the response model problem discussed in the lectures: You wanted to predict the probability of a response from each prospect and target the ones most likely to respond to the next telemarketing campaign. The steps were as follows:

1. Identify relevant predictor variables for a response using EDA.
2. Build predictive models and choose the best one.
3. Sort the prospects in order of decreasing probability of response (predicted by the best model) and target the top X% (or top Y deciles), where X would be determined by your business objective (e.g., maximising the overall response rate/number of responders at a fixed marketing cost).

If you look at the important variables included in the final model, you will see that one of them is  ‘duration’, which refers to the duration of the phone call in seconds. When the model is presented to the Chief Marketing Officer (CMO), she notices one glaring error in the model: It contains ‘duration’ as a variable. However, when an executive picks up the phone and calls a prospective client, he/she is unsure of the duration of the call.

Now there are two problems with including the variable ‘duration’ in the model:

* The prospect data procured by the marketing team does not contain ‘duration’, since the call has not been made yet.
* In your analysis of marketing cost and response, you assumed that the cost of a phone call is independent of its duration (₹1 per call), which is not true.

**Tasks**

To solve these problems, you should resolve to build another model without including the variable ‘duration’. This will help you understand the relationship of the other variables with the response.

Also, set the business objective to achieving 80% of total responders at the minimum possible cost. The total number of responders is the total number of prospects who responded, from the available data of about 45,000 prospects.

Based on this information, calculate the X in the top X%, i.e., how many prospects should be called to meet the business objective. Further, create a presentation for the CMO highlighting the results and the methodology employed.

**Checkpoints**

Note: Before starting the assignment, you may find it helpful to create a unique ID for each prospect. Also, you only have to submit a Jupyter Notebook for this assignment, and so, you have to report some of the metrics as comments in the Jupyter Notebook.

The checkpoints for the assignment are as follows:

1. Perform data preparation (no marks are awarded for this step)
   * You can use the code provided in the lectures to complete all the data preparation steps
2. Build a logistic regression model without using the variable 'duration'
   * Select variables using the usual methods
   * Sort the data points in decreasing order of the probability of response
   * Find the optimal probability cut-off and report the relevant evaluation metrics
3. Create a data frame with the variables prospect ID, actual response, predicted response, predicted probability of response, duration of the call in seconds and cost of the call
   * While creating the data frame, list the cost of call for each prospect in a new column
4. Find the number of top X% prospects you should target to meet the business objective
   * Report the average call duration for targeting the top X% prospects to the CMO
5. Create a lift chart
   * The x-axis should show the number of prospects contacted; the y-axis should show the ratio of the response rate using the model and the response rate without using the model
6. Determine the cost of acquisition
   * Consider cost = 1\*number of contacts made in the current campaign; determine the cost incurred for acquiring 80% of customers using the predictive model
7. Create a small presentation for the CMO highlighting your findings and the methodology used.

Note: For this assignment, use the same bank marketing data set that you saw in the Acquisition Analytics module.

feedback

**Report an error**