# National Veterans Organization

## Overview:

The data sets are in comma delimited format. The learning dataset <cup98LRN.txt> contains 95412 records and 481 fields. The first/header row of the data set contains the field names.

The validation dataset <cup98VAL.txt> contains 96367 records and 479 variables. The first/header row of the data set contains the field names.

THE RECORDS IN THE VALIDATION DATASET ARE IDENTICAL TO THE RECORDS IN THE LEARNING DATASET EXCEPT THAT THE VALUES FOR THE TARGET/DEPENDENT VARIABLES ARE MISSING (i.e., the fields TARGET\_B and TARGET\_D are not included in the validation data set.)

The data dictionary (for both the learning and the validation data set) are included in the file <cup98DIC.txt>. The fields in the data dictionary are ordered by the position of the fields in the learning data set. The dictionary for the validation data set is identical to the dictionary for the learning data set except the two target fields (target\_B and target\_D) are missing in the validation data set. Blanks in the string (or character) variables/fields and periods in the numeric variables correspond to missing values.

Each record has a unique record identifier or index (field name: CONTROLN.) For each record, there are two target/dependent variables (field names: TARGET\_B and TARGET\_D). TARGET\_B is a binary variable indicating whether or not the record responded to the promotion of interest ("97NK" mailing) while TARGET\_D contains the donation amount (dollar) and is only observed for those that responded to the promotion.

## Task:

Review the learning data set to identify which members should receive a request for fundraising. Each request for fundraising costs $1. Your stakeholder only has a $40,000 budget to run this campaign. Your assignment is to maximize the potential fundraising for this campaign.

## Submission Requirements:

A outline of the steps taken including:

* assumptions made
* how you turned assumptions into facts
* summary of different experiments conducted
* all code used
* provide conclusions from your findings from each phase/step taken

Overall conclusion of results including:

* Total funding raised
* Total costs
* Total disposable funding
* List of donors to be targeted
* Brief summary on why the chosen model is the best fit model

## Grading:

* Data health check – 10%
* Feature management – 10%
* Experiment approach – 10%
* Management of assumptions – 20%
* Logical thought process of conclusion and summaries – 10%
* Coding efficiency – 10%
* Overall results interpretation (models, campaign recommendations, alignment of results to conclusion) – 30%