

# Project: Predictive analytics in Fundraising

Geert Verstraeten

## INTRODUCTION



The objective of this project is to obtain hands-on experience with a concrete and real problem in predictive analytics. In this project, you receive real-life data of **campaigns and donors of an existing fundraising organization**. This organization possesses a database of a number of lapsed donors (**individuals and organizations who have donated in the past, but did not donate more than 5 EUR in the previous two years**). The organization has executed different **reactivation campaigns**, and aims to increase its efficiency on these campaigns by only targeting those **donors with the highest potential** to respond to the reactivation campaign.

You will receive data of historic reactivation campaigns, and the gifts and donors related to these campaigns.

**Goal of the predictive model is to predict whether a donor will respond to a future reactivation campaign with a gift of at least 30 EUR.**

More concretely, you receive data of:

- **Donors:** general data of donors that were at least once targeted in a reactivation campaign
- **Campaigns:** general information about the reactivation campaigns in scope
- **Selections:** which donors were targeted in which reactivation campaign. However, all selections for the last reactivation campaign (7541) are removed, as this campaign will serve as an out-of-sample, out-of-time validation
- **Gifts:** all historic gifts of donors who were targeted at least once in a reactivation campaign. This includes all previous gifts of donors (on other campaigns as well). However, all gifts for the last reactivation campaign (7541) are removed, as this campaign will serve for an out-of-sample, out-of-time validation
- **Score:** which donors were targeted in the last reactivation campaign. This campaign will be used as out-of-sample, out-of-time validation. This table is in reality an extract from the **Selections** table.

## GOAL

On April 4<sup>th</sup> 2016 at 1pm, we expect the following from each team:

- Create a predictive model based on historical data. Also create a 30 minute **presentation** about the project, as if presented towards the general management and the head of analytics at the fundraising organization.

**This presentation should contain at least a brief project description, the predictive timeline, a summary slide about the predictors created, the final model used, the variables selected, interpretation of the model and the main predictors, the predictive performance (AUC, cumulative response and cumulative gains of the final model on the validation set) and, if possible, the potential financial implications (profit). The goal of the presentation is to convince management to use this model for targeting in future reactivation campaigns.**

- Create a prediction (=ranking) for all donors in the **Score** data set. In other words, we expect a prediction of response for all selected donors in the last reactivation campaign (7541).

**On April 4<sup>th</sup> at 1pm, provide a tab-delimited .txt file containing as first variable donorID and as second variable a ranking, where the highest value represents the highest probability. This table should contain one line per selected donor in the Score data set (last reactivation campaign, 7541).**

- Provide the final **code** used (rather clean, if possible – we do not need code of explorations)

## SUGGESTED PROJECT APPROACH



### 1. Project definition

For the detailed definition of this project, we refer to the pdf "Project definition Fundraising Reactivation" of the first session (see Introduction Descriptive and Predictive Analytics on February 24th 2016).



### 2. Data Preparation

- **Draw the predictive timeline**
  - Start with an example of one reactivation campaign
  - Next, draw timelines for all reactivation campaigns
- **Start – as always – by exploring your data**
  - How many reactivation campaigns are available for analysis?
  - Do you have general donor information for all donors in the selections?
  - Does each campaign obtain at least one gift
  - Count the number of donations and average donation amount per campaign
  - etc
- **Create the population:** which observations should be in your analytical base table
- **Create the target:** how do you determine success (=a subsequent gift of at least 30 EUR)
- **Create the potential predictors:** create variables that are potentially useful in predicting the target
- **Use data partitioning** to create training, selection and validation sets
- **Perform data preprocessing:** consider at least the basics discussed in Session 2 of this class



### 3. Model Building

- **Create a logistic regression model:** every team should consider this as the benchmark
- **Optional:** feel free to use another approach



### 4. Model Validation

- **Calculate at least AUC, Cumulative Response and Cumulative Gains of your final model**



### 5. Model Usage

- **Perform a scoring of all selected donors for the last reactivation campaign (7541)**
- **Create a convincing presentation, proving your competence, understanding of the methodology, the problem, and the results**

## DATA DESCRIPTION

**TABLE 1: donors**

Table containing general information per donor (one line per donor)

Variable name	Variable Description	Values / Examples
donorID	Unique donor identification number	100001
gender	Gender; family type or company	M=Male, F=Female, C=Couple, S=Company, U=Unknown
language		N=Dutch, F=French
zipcode	4-digit postal code. The first digit represents a high-level grouping of region. The two first digits represent a lower-level grouping of region, etc.	1861
region	Detailed area code (if available)	13049A090

**TABLE 2: campaigns**

Table containing general information per campaign and communication (one line per communication)

Variable name	Variable Description	Values / Examples
campID	Campaign identification number	5233
date	Date on which the campaign was sent (= drop date)	26SEP2012
cost_unit	Cost per letter sent	0.8352535955
commID	Communication identification number. a communication represents a type of letter, for example a Dutch letter versus a French letter, a letter with incentive versus without incentive, etc.	12164
language	Language code of the communication	FR=French, NL=Dutch
sent	Number of letters sent in this communication	3046
description	Description of the communication	Inact FR

**TABLE 3: selections**

Table containing the selections for reactivation campaigns: which donor was selected for which campaign and communication.

Variable name	Variable Description	Values / Examples
campID	Campaign identification number	5233
commID	Communication identification number. a communication represents a type of letter, for example a Dutch letter versus a French letter, a letter with incentive versus without incentive, etc.	12164
donorID	Unique donor identification number	100001

**TABLE 4: gifts**

Table containing historic gifts for donors in reactivation campaign (one line per gift)

Variable name	Variable Description	Values / Examples
campID	Campaign identification number	5233
commID	Communication identification number. a communication represents a type of letter, for example a Dutch letter versus a French letter, a letter with incentive versus without incentive, etc.	12164
donorID	Unique donor identification number	100001
amount	Amount of the gift in EUR	12.39
date	Date of the gift	23/10/1998