Project

Overview

DSC is an organization that helps non-profit organizations to find donors for their good causes. They have a huge database with candidate donors. DSC now wants to launch a new re-activation campaign where they want to send letters to donors that have been inactive for a long time, hoping that they will donate again. The CEO & Head of Data Analytics (conveniently, this is the same person for this exercise) have to be convinced that your model is better than just randomly selecting donors.

DSC is mostly interested in customers that have donated €30 or more, but you can be flexible in this decision, depending on your business case.

DEADLINES :

Q&A session (optional) : 16 november 2022

Scored customer dataset : 29 november 2022

Presentation & technical handover : 30 november 2022

Datasets

DSC provides 5 datasets:

- **donors.csv** contains general information about the candidate donor database.
  - *donorID:* The unique ID of the candidate donor.
  - *zipcode:* The postal code of the candidate donor's main address.
  - *province:* The province of the candidate donor's main address.
  - *region:* The region of the candidate donor's main address.
  - *gender:* The gender of the candidate donor (M = male, F = female).
  - *language:* The candidate donor's preferred communication language (FR = French, NL = Dutch, EN = English).
  - *dateOfBirth*: The date of birth of the candidate donor.

- **gifts.csv** contains information about the gifts that these candidate donors donated in the past for campaigns that are no re-activation campaigns
  - *donorID*: The unique ID of the candidate donor.
  - *campaignID*: The unique ID of the campaign (donation was made outside of a campaign if missing).
  - *amount*: The amount of the donation (in EUR)
  - *date*: The date of the donation payment.
- **selection campaign 6169.csv:** donors that were selected for a previous re-activation campaign on 04/09/2018.

- **selection campaign 7244.csv:** donors that were selected for a previous re-activation campaign on 18/06/2019.
- **campaigns:** contains information about the different campaigns organized by DSC since December 2014.
  - *campaignID*: The unique ID of the campaign.
  - *date*: The date of the campaign (when the letters were sent).
  - *lettersSent*: The number of letters sent.
  - *CostUnit*: The unitary cost of the campaign (Cost of the campaign divided by the number of letters sent).

Project steps

Use the information of the campaign 6169 as train data and the information of the campaign 7244 as test data.

1. Make sure all data is clean and describe your data to get an idea of what's in there.
2. Calculate the target for train and test data: All candidates donors from the campaign selections who actually made a donation (information found in the gifts table). Start off with donations larger or equal to €30.
3. Construct features. You can use information available in the donors table, but also information in the gift table. For the latter, mind the timeline! Only information known before the event should be used to train a model to make prediction.
4. Use a feature selection algorithm to select relevant variables (Optional).
5. Construct a model. You can use any type of algorithm. Compare the performance of different models and check whether the models are interpretable.
6. Evaluate the model using AUC, but also lift and cumulative gain curves.
7. Try to make a business case using the campaign information. How much your model would have increased the performance of the campaigns?
8. 5 days before the deadline, you will get an additional dataset with donorID's that you have to score with your model by the deadline.

Tip: A lot of interesting information can be derived from the gifts table. Organize a good brainstorm in your team to come up with interesting features.

Final results

Presentation

Convince the CEO & Head of Data Analytics of DSC (fictively) that your model has added value and that it should be used in future campaign. Do this in a 10 min presentation.

These people would like to know:

- What information do you use to make your predictions?
- Who are your donors? What is their profile?
- How good does your model perform? (lift, cumulative gains and cumulative response)
- How much they would gain from using your model and how many people should be contacted, assuming a campaign cost of €0.80 per letter?

Technical handover

- A well-documented notebook or set of notebooks that you used for this project.
- Scored set of DonorID's with the **probability** that they will make a donation.