PROJECT

DSC is an organization that helps non-profit organisations finding donors for their good causes. They have a huge database with candidate donors. DSC wants to launch a new re-activation campaign: they want to send letters to donors that have been inactive for a long time, hoping that they will donate again. They only want to send letters to candidate donors that are likely to donate more than EUR 30. Therefore, they need your help: can you construct a model that predicts which donors are most likely to donate more than EUR 30 for a re-activation campaign?

To this end, they provide 4 datasets:

- Donors.csv contains general information about the candidate donor database
- Gifts.csv contains information about the gifts that these candidate donors donated in the past for campaigns that are no re-activation campaigns
- Campaign20130411.csv: donors that were selected for a previous re-activation campaign on April 11th 2013, and the amount that they gave for this campaign
- Campaign20140115.csv: donors that were selected for a previous re-activation campaign on January 15th 2014, and the amount that they gave for this campaign.

Feel free to approach the project as you like. If you need more guidance, you can follow the steps below:

- You can use the donors in Campaign20130411.csv as train data and the donors in Campaign20140115.csv as test data.
- Calculate the target for train and test data
- Construct variables. You can use information in donors.csv, but also information in gifts.csv. For the latter, mind the timeline!
- Use a feature selection algorithm to select relevant variables
- Construct a model. You can use logistic regression, or any other method you like if the performance is better (SVM, NB, decision tree, KNN, …)
- Evaluate the model using AUC, but also lift and cumulative gain curves
- Select the best model you can get.
- Try to make a business case. How many donors should be selected? What if a campaign costs EUR 0.5? Can DSC make profit? These are only examples – be creative.
- Try to give insights on the variables. E.g., are female donors more likely to respond?
- Make a nice presentation (15 minutes) that you would use to convince DSC to use your model. Try to avoid technical details, focus on business insights and actionability.

Deliverables (send to nele.verbiest@pythonpredictions.com topic: IESEG):

- A plan of action (can be short):
  - How will you divide the work
  - When will you organize meetings to discuss the results
  - How are the deadlines organized in your group.
- A presentation of 15' that you will present with your group. Be prepared for Q & A
- R code - please comment and structure the code.

Deadlines:

- Plan of action: November 16th
- R code: December 9th
- Presentation: December 12th