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 35. Therefore, they need your help: can you construct a model that predicts which donors are most likely to donate more than EUR 35 for a re-activation campaign?

Datasets

DSC provides 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.

Project steps

- 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. Compare the performance of different models and check whether the models are interpretable.
- Evaluate the model using AUC, but also lift and cumulative gain curves.
- Try to make a business case. How many donors should be selected? What if a campaign costs EUR 0.5? Can DSC make profit?
- Try to give insights on the variables that are used in the model. E.g., if “gender” is in the model, are female donors more likely to respond?

Deliverables

- 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.
- Make a nice **presentation** (15 minutes) that you would use to convince DSC to use your model. Avoid technical details in the presentation. The presentation should at least include following topics:
    - Which models did you try
    - Which model had the best performance
    - Which feature selection algorithm did you use
    - Final model: which variables are included, what is the AUC on train and test data (is there overtraining?), show the cumulative gains and lift curve.
    - Final model: business case. Can DSC make profit using your model?

  However, if you feel you want to share more in the presentation, feel free to do so.
- **R code** - please comment and structure the code.

Deadlines

- Plan of action: November 12th (send to nele.verbiest@pythonpredictions.com topic: “IESEG - group X - plan of action”)
- R code: November 21st (send to nele.verbiest@pythonpredictions.com topic: “IESEG - group X - R code”)
- Presentation: November 21st (send to nele.verbiest@pythonpredictions.com topic: “IESEG - group X - presentation”)