

Intuit Redux



The purpose of this exercise is to gain experience modeling the response to an upsell campaign with Tree-based models and Neural Networks (or Deep Learning). You will use the same data (“intuit75k.pkl”) as in the previous assignment.

I have created assignments on DataCamp for you to work through to help prepare with your group to complete this assignment successfully.

Assignment guidelines:

1. As in the previous assignment, determine which of the 22,500 businesses in the test set (i.e., training == 0) to mail in wave-2 (i.e., generate a list of IDs). Your **full grade (10 points)** for this group assignment will be based on the **profit** (not ROME) achieved using this list of IDs. After you submit the list of IDs I will be able to determine the final outcome because all customers who did not respond in wave-1 were actually mailed in wave-2. You will **not** have access to this extra data set with information on response in wave-2.
2. Your selection of models should contain **only** Tree-based models (e.g., decision trees, random forests, gradient boosting from sklearn) or Neural Networks (MLP in sklearn or DL models in Keras).
3. **No case write-up is required** but the code you used to estimate different models must be reproducible and must be pushed to **GitLab** before the start of class. Also, make sure the code is well commented and formatted (e.g., the Black code formatting for python).
4. You can run python code on the server (<https://rsm-compute-01.ucsd.edu/>) in Jupyter notebooks and/or VS Code
5. I recommend you use Git and GitLab to move estimation code to and from the server. GitGadget is installed on the server and is available from Jupyter or the Addins Dropdown in Rstudio. You can also use the Git extension that is part of Jupyter. See the playlist below for more information

https://www.youtube.com/playlist?list=PLNhtaetb48EfTV5_ZW5X7GQdBo2-l26PP

6. As before, assume each mail piece costs \$1.41 and that the margin (or net revenue) from each responder, excluding the mailing cost, is \$60. Make sure to correct the (predicted) purchase probabilities appropriately (i.e., the 50% drop-off as discussed in class). There is no need to calculate how to “project” your profits to the remaining customers in this assignments.
7. Each model used should return a predicted probability of response so you can use it to calculate a “mailto” variable
8. With your team, create a video presentation that lasts no more than 10 minutes and that covers your approach to building and finding the best model for this problem. Upload your

video to Panopto and submit the link to the video to Canvas through “Group Assignment: Intuit Redux (Video submission)”.

9. **Before noon on the day of class** please post the list of IDs you want to mail to **Canvas**. The Python file(s) used for estimation should be submitted through **GitLab** before the start of class as usual. The reason that the list of IDs must be submitted earlier is that I need to compile all the results before class. Compiling the results takes some time so please help me out and stick to the following instructions **exactly!**

Please post a CSV file with **only 2 variables**

- a. The original **id** variable from the intuit75.pkl dataset (please do **not** rename the variable). The column should contain all IDs from the test set. Do **NOT** delete IDs you don’t want to target from the dataset, i.e., **the file must have 22,500 rows**.
- b. A variable named **“mailto_wave2”** (lower case) that is TRUE if you want to target a customer in wave-2, and FALSE if you do not.
- c. Please name the dataset using the **first** names of all group members separated by underscores “_”, plus your group name (e.g., Nancy_Yu_Manuel_MightyDucks.csv). If your group name is more than one word please leave out spaces or underscores (e.g., instead of “Mighty Ducks” use “MightyDucks”). Also, the name should not contain any symbols.
- d. An example submission file will be added to the assignment GitLab repo you can see the required format (i.e., Nancy_Yu_Manuel_MightyDucks.csv)
- e. **Please double and triple check that your targeting submission file is in the exact right format so you do not end up the group that crashed my evaluation code!**