

General Rules:

1. Please DO NOT make your solution publicly available. For example by pushing your code to public repository.
2. Your submission should be in standard archive format (zip, tar.gz, tar), please don't use proprietary archive.
3. Please do not include any binary executable as part of your submission (jar files, class files, header files, pyc files). If your solution needs a binary executable to run, you could tell us on how to obtain them inside your README.md file.

Problem Statement:

Customer attrition is the state in customer lifecycle where they stopped using a service. Most of the time, the cost of reviving customer who is in attrition cycle is greater than preventing them from attrition in the first place.

We provide you with user transaction data for two months (Dec-Jan) as a training set, and their last transaction date after one-month interval as a testing set.

The training data consists of order-related information such as merchant identifier, amount paid in IDR, promotion type, and amount of promotion in IDR, and time of transaction. We also provide you with information of whether or not the user is a premium user.

Supposed that our attrition cutoff is 20 days, we want to predict among our users which ones are going to churn on 1st of march. We give you the test set of last transaction date of users prior to 1st of march, as a basis of what could be used on your prediction.

On this assignment, we are going to look on an optimized negative log-likelihood (NLL for short) as our evaluation.

Deliverables:

You are given 168 hour to work on this assignment (24hour x 7). Your submission should contain:

1. Code
 - a. You can submit the code a IPython notebook or any other notebook kernel.
 - b. We expect to see at least two unit test cases. One for feature engineering (what the data should look like prior to model), and one for prediction (what

comes out from the model). Please think on how your model would behave on an unseen data and develop your test case on it.

- c. Please write a short README.md file on how to run your code.
- d. Please use git version control to show us the evaluation of your work. Your submission therefore should contain the .git metadata directory. You should put the dataset to .gitignore as it will inflate the size of your submission.

2. Report

- a. Write a description of your approach, it should be no longer than 2 A4 pages in pdf format.
- b. If you tried multiple models, please show us their NLL score.
- c. Include a visualization to get your point across.
- d. Try to be clear and concise.
- e. Include a possible improvement you might try if you are given longer time to work on this problem.