HOMEWORK 2

(Participate in Kaggle competition **AND** upload your jupyter notebook or python script in NYU Classes)

Goal:

Given a training set of past churn data, your goal in this homework is to predict whether a person will leave the network (churn) or stay.

(Only) Packages to be used: Numpy , Pandas, SciPy, scikit-learn

Steps:

1. Train and build a model based using the training data: any model studies in class is acceptable, such as a decision tree, logistic regression, or support vector machine.
2. Use your model to predict the outputs for the test data.
3. Upload your prediction as a file to the kaggle competition to be evaluated and ranked.

Data Description:

Each row represents a customer of the network, with the parameters for each customer described below.

The data consists of 20,000 customers, split into 90% (18,000) for training data and the remaining 10% (2,000) as test data (holdout).

You can find the labeled training data in 'train.csv' and unlabeled test data in 'test.csv'.

Features of each customer:

Features of each customer:

COLLEGE	Is the customer college educated?
INCOME	Annual income
OVERAGE	Average overcharges per month
LEFTOVER month	Average number of leftover minutes per
HOUSE tract)	Estimated value of dwelling (from census
HANDSET_PRICE	Cost of phone
OVER_15MINS_CALLS_PER_MONTH over) per month	Average number of long calls (15 mins or

AVERAGE_CALL_DURATION	Average duration of a call
REPORTED_SATISFACTION	Reported level of satisfaction
REPORTED_USAGE_LEVEL	Self-reported usage level
CONSIDERING_CHANGE_OF_PLAN their plan	Whether the customer considered changing
LEAVE (Target variable)	Did the customer stay or leave (churn)?

Submission Format (in kaggle competition):

For every student in the competition, submission files should contain two columns: 'ID & 'LEAVE'. ID will have values from 0 to 1999 and 'LEAVE' column should have predicted outputs of 2,000 test samples of test.csv
Note: You can download 'test_submit.csv' to know how should your submission file be.

Submission Format (in NYU Classes):

You must submit a jupyter notebook or a python script which has the model which was used to submit the file in Kaggle competition.

Team Size: At most 2 students in one team.

Complete Homework Grading Criteria:

1. Based on your rank on the leaderboard in Kaggle Competition
2. Your uploaded jupyter notebook file in NYU Classes (any one of the students in the team can submit the jupyter notebook in NYU Classes)