



# Assignment 1: Predicting Customer Churn

The objective of this assignment is to do a quick refresh on machine learning, to set up your favorite development environment, to process a simple dataset, apply machine learning algorithms that were studied in CS5785, and write up the results in the form of a presentation.

This homework is due on **February 5th, at 11:59 PM EST**. Upload your homework to the CMS.

Please upload the submission as a single .zip file. It should include:

1. The slideshow in pdf format.
2. The source code for all your experiments (and figures), and the instructions on how to run your code.

You can perform the following tasks using Python and any additional libraries that you wish to use.

## Introduction

The file `telco_customer_churn.zip` contains information about customer churn for a large telecom provider. Customer churn signifies which customers have stopped using your company's product or service during a certain time frame. The file's fields include data on each customer, which includes demographic data, the services they purchased, their monthly and total charges, and whether they churned within a month when the data was captured. You can assume that the data was uniformly sampled from the full database.

## Analysis

Pick and apply two different predictive methods (of your choice) to predict whether the customers churned in the given period. You may need to convert the data to a format that is more compatible with the models you use, such as one-hot encoding, etc. Also, you may use any models that you think are appropriate. Examples include Logistic Regression, Decision Tree, etc.

# Presentation

Prepare a slideshow, as if to be delivered to the CEO, which answers the following questions/tasks:

1. Describe and summarize the data (one slide).
2. Produce two visualizations of the data that are meaningful to the presentation. Explain (in one sentence for each figure) why do you think them meaningful (two slides)
3. Explain your method (one slide). Which predictive methods does the best on predicting customer churn? Provide an explanation that is based on formal evaluation methods. Good samples are k-fold cross-validation of accuracy and ROC curve. You should evaluate your models on dedicated subsets of data that are not involved in training.
  - a. Is the data balanced? If not, how does this affect your evaluation? What have you done to minimize its impact? (Hint: what is the accuracy of a prediction method that always outputs "No"?)
4. Report the correlations between data entries and customer churn based on your predictive models (one slide).
5. Provide recommendations for reducing churn, based on the data (one slide).

# Appendix

Explanations of data fields in the provided data:

customerID	A unique ID for each customer
gender	Whether the customer is a male or a female
SeniorCitizen	Whether the customer is a senior citizen or not (1, 0)
Partner	Whether the customer has a partner or not (Yes, No)
Dependents	Whether the customer has dependents or not (Yes, No)
tenure	Number of months the customer has stayed with the company
PhoneService	Whether the customer has a phone service or not (Yes, No)
MultipleLines	Whether the customer has multiple lines or not (Yes, No, No phone service)
InternetService	Customer's internet service provider (DSL, Fiber optic, No)
OnlineSecurity	Whether the customer has online security or not (Yes, No, No internet service)
OnlineBackup	Whether the customer has online backup or not (Yes, No, No internet service)
DeviceProtection	Whether the customer has device protection or not (Yes, No, No internet service)
TechSupport	Whether the customer has tech support or not (Yes, No, No internet service)
StreamingTV	Whether the customer has streaming TV or not (Yes, No, No internet service)
StreamingMovies	Whether the customer has streaming movies or not (Yes, No, No internet service)
Contract	The contract term of the customer (Month-to-month, One year, Two year)
PaperlessBilling	Whether the customer has paperless billing or not (Yes, No)
PaymentMethod	The customer's payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))
MonthlyCharges	The amount charged to the customer monthly
TotalCharges	The total amount charged to the customer
Churn	Whether the customer churned or not (Yes or No)