# IEE 520/BMI 555 Fall 2018 Project

**Due by 11:59 PM (MST) on December 5, 2017. Please submit to Blackboard.**

**INDIVIDUAL PROJECT—YOUR WORK SHOULD BE YOUR OWN.**

## Objective

Use the data provided with this project. One file contains the **training data** that you should use for model building. A second file contains **testing data** that will be used to evaluate your model.
The training and testing data were selected randomly from the original data set.

Build a classification model for this data based on the methods described in the course. You may choose any classifier and use any options covered in the course. It is often better to work to tune a particular classifier than to superficially jump to many classifiers.

## Evaluation

You will be primarily evaluated by the performance of your model, but a written report is also required. Your model will be scored as the **balanced error rate**. This is the average of the error rate on each class. This is **different** from the overall (weighted) error rate. This balanced error rate encourages models to predict each class equally well. Because the classes are unbalanced, you should work to minimize the error rate on each class. For example, consider the following confusion matrix:

| | Prediction | | | |
|---|---|---|---|---|
| Actual | Class 0 | Class 1 | Class 2 | Error rate |
| Class 0 | 900 | 50 | 50 | 100/1000 = 0.1 |
| Class 1 | 50 | 100 | 50 | 100/200 = 0.5 |
| Class 2 | 100 | 100 | 200 | 200/400 = 0.5 |

Overall error rate = 400/1600 = 0.25
**Balanced error rate = (0.1 + 0.5 + 0.5)/3 = 0.37**

A simple adjustment is to sample from the training data to create a new data set with equal rows from each class that is used to train your model. As mentioned in class, you might *upsample* (select the same instance more than once) to create equal instance for the majority class or *downsample* to create equal instances for the minority class. This extra step might not even be needed in some examples.

## Predictions and Report

Upload two files.
1) Submit a written report with a brief description of what methods you tried, how you evaluated your methods, your final model, and why you choose your final model and parameter settings. The description of your final model and parameters needs to provide sufficient detail to be reproduced. Probably about five pages are sufficient. Name your file BMI555IEE520_Report2018_*yourfullname*.

2) Submit predictions for the test data in a comma separated values (CSV) file with two columns: the first column is the row number, the second column is the predicted class. No headings. Name your file BMI555IEE520_Results2018_*yourfullname*.csv.