

**Stat 6910, Section 003**  
**Statistical Learning and Data Mining II**  
**Fall 2018**

**Final Project Proposal**

**Matt Isaac**  
**Colby Wight**  
**Jared Hansen**

(i) Goal: Data and Problem

Our data set can be found at: <https://www.kaggle.com/sonujha090/bank-marketing>.

This data set comes from the phone call marketing campaign of a Portuguese bank. The goal of this marketing campaign was to get their customers to use a term deposit product (“CD”). The dataset contains a binary “yes/no” response, with 16 other features that can be used for prediction. There are roughly 45,000 observations in the data set. Some of the features given include: age, occupation, marital status, education level, and account balance.

Luckily the data is already very clean. There may be a few missing values present, but otherwise the data is in a format that is ready for predictive modeling.

The goal of our analysis is to help the bank identify (classify) customers that are likely to respond positively to the campaign and agree to a term deposit (“1’s”). One of the challenges that we anticipate having to deal with is imbalanced class sizes: there are far more 0’s (“no’s”) than 1’s (“yes’s”), so we’ll try some sampling methods to ameliorate this issue. Additionally, we’ll have to make some decisions about splitting up the data into training, test, and validation sets, as well as utilizing cross-validation or other methods for obtaining accuracy metrics.

(ii) Methods

This problem lends itself well to several methods, a few of which we’ve discussed in class. The methods we’re planning to use include:

- Logistic regression
- Support vector machines
- Neural networks
- Random forests and/or decision trees
- QDA (maybe, but probably not)

(This selection of methods satisfies the requirement of “Apply(ing) at least 4 ML methods to the problem, 1-3 methods from those covered in class, at least one method not covered in class”.)

(iii) Contributions of Each Member

Our plan is to assign each member one specific machine learning method (TBD), and split the work for the fourth method. We’ll work together collaboratively to aggregate our findings and write a report detailing the results our analyses.