**Programming Assessment**

**Bud Lab - Internship**

1. Raw dataset: <https://archive.ics.uci.edu/ml/datasets/Adult>

2. Classification Algorithms (pick 3 out of 13)

* **Logistic regression**
* Linear discriminant analysis
* Mixture discriminant analysis
* Quadratic Discriminant Analysis
* Flexible Discriminant Analysis
* Support Vector Machine
* k-Nearest Neighbors
* Naive Bayes
* Classification and Regression Trees (CART)
* C4.5
* PART
* Bagging CART
* **Random Forest**
* **Gradient Boosted Machine**

3. Procedures (include full scripts with necessary comments and citations)

* Data cleaning and validation
  + Please include a data description (e.g. samples/instance, features/attributes and etc.)
* Outlier and missing value detection (if possible)
  + Please explain what outlier detection methods/approaches you used
  + Please explain how you handle missing values
* Model fitting
* Classification prediction
* Results comparisons and visualization