# Homework 6

Your Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student ID:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**(100 points) Python practice for classification**

**Use the best setting of decision trees, from HW4**

**1). Run different ensemble methods with this best tree model**

**2). Use oversampling to deal with imbalance issues, though the imbalance issue may not be serious in this data. Run decision tree again, and observe the results**

**3). Use feature selection and reduction, and re-run decision tree models, and observe whether the results can be improved.**

**Note, for part 1) and 2), you do not need to change parameters for decision trees, just used the best setting from HW4.**

**For part 3), you need to tune up parameters to decision trees again**

* Use Loans\_20K.csv data by using 10-fold cross validation
* Use Loans\_200K.csv data by using 75% as training, 25% as testing
* Loan term as label

Note:

* You need to change different/multiple parameters to find the best model.
* You can find data sets from “slide & data” on blackboard system

Submission

* The ipynb and saved html files

A comparison of different parameters and metrics (accuracy, F1)