# Homework 8

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

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

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

We want to use the loans data, and would like to use clustering to improve the task of classification.

1). Use KMeans clustering to build clusters, use elbow method to find the optimal K value

2). Evaluate KMeans outputs by the following two ways

a). compare the centroids to observe whether you can find significant patterns

b). use KMeans output as inputs to the classification task (by using decision trees as classifier), and observe whether you can improve the classification results

3). Build hierarchical clustering, and create K clusters. Use the same optimal value for K as the K value in step 1). Repeat step 2)b)

* Use Loans\_20K.csv data only
* Loan term as label in the classification task

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