Individual Mini Project
Due: 24:00, June 26, 2018

Student ID: _____ Name: _____

Loan Default Prediction

This project asks you to build a prediction model which can determine whether a loan will default. Please notice that this is in fact a classification problem. The `bank.csv` file from a real bank contains historical information of 9,099 customers. This dataset has one class variable and 31 predictor variables. The variable description is given below.

Table. Variable description

Type	Variable Name	Description
Continuous	age	Age of a customer
	amt.bad.credit	Amount of bad credit that a customer has
	dep.by.guar	Amount of deposit by guarantors
	act.loan	Actual amount of loan taken by a customer
	appr.loan	Approved amount of loan
	num.bad	Number of bad credits that a customer has
	num.loan	Number of loans guaranteed by a borrower (customer)
	num.default	Number of a customer's loan default in past
	days.default	Number of default days
	rev.facility	Number of revolving facility
	gen.loan	Number of general loans
	days.curr.default	Number of default days for current loan
	aver.days.default	Average number of default days in past
	amt.default.cred.3	Amount of credit card defaults within 3 months
	amt.default.cred.6	Amount of credit card defaults within 6 months
	num.default.cred.3	Number of credit card defaults within 3 months
	num.default.cred.6	Number of credit card defaults within 6 months
	val.realestate.3rd	Value of real estate as security by the 3rd party
	saving.3rd	Saving as security by the 3rd party
	val.realestate	Value of real estate for loan
	saving	Amount of saving of a client as security
	total.security	Total amount of security
	total.3rd.security	Total amount of security by the 3rd party of all security
	aver.payback	Average payback period (days) per loan
	sum.payback	Sum of payback periods (days)
	spread.rate	Average spread interest rate
	fixed.rate	Average fixed interest rate
Categorical	gender	gender of a customer (1: male, 2: female)
	loan.status	Loan status of a customer (1: general loan only, 2: revolving facility only, 3: both)
	curr.credit	Current credit status of a customer (0: good, 1: bad)
	guarantor	Is a client a guarantor for others' loans? (0: no, 1:yes)
Class	default	Loan default (0: no, 1: yes)

Your project result will be evaluated by the following criteria.

- Model goodness in terms of AUC (Area under ROC curve)
- Originality of your method to solve the class imbalance problem

- Use of data mining techniques that you have learned in the class
- Your efforts put into the project

Please keep in mind that you must report all details to arrive at your final model. Your project score will be determined by the details. Please turn in your project as a hard copy by the due (Do not upload on the icampus.). There will a drop-box in front of the door of office 27409. Please also make sure that NO late project is accepted.