TEST 1

Data Information: "default of credit card clients.csv"

Columns:

LIMIT\_BAL: amount of the given credit

GENDER: (1 = male; 2 = female).

EDUCATION: (1 = graduate school; 2 = university; 3 = high school; 4 = others).

MARRIGE: (1 = married; 2 = single; 3 = others).

AGE: (year).

PAY\_0 to PAY\_6:

Past monthly repayment status:

PAY\_0 = the repayment status in September, 2005; PAY\_1 = the repayment status in August, 2005; . . .; PAY\_6 = the repayment status in April, 2005.

The measurement scale is: -1 = pay duly; 1 = payment delay for one month; 2 = payment delay for two months; . . .; 8 = payment delay for eight months; 9 = payment delay for nine months and above.

BILL\_AMT1 to BILL\_AMT6:

Amount of bill statement.

PAY\_AMT1 to PAY\_AMT6:

Amount of previous payment.

default payment next month: (Yes = 1, No = 0)

TASK:

Use “default of credit card clients.csv” dataset to build a decision tree, target variable “default payment next month”. Use all other variables as predictors.

1. Present a histogram of “LIMIT\_BAL”

(submit the figure only)

2. In the variable "default payment next month", how many cases are there for 0 and 1?

(submit the answer only)

3. Build a decision tree with “default payment next month” as the target variable, and all other columns as the predictors.

Use 80% data in the training set, 20% in the testing set.

Set all random state as 100.

(Submit both the code and the tree plot)

4. How many rows of data are included in the training and testing data? (submit the answer only)

5. Prune the tree (set random state as 100), so that the minimal impurity decrease required to split a node is 0.01

Present the pruned tree

(Submit both the code and the tree plot)

6. What is the accuracy for the original tree and pruned tree?

(submit the answer only)

7. What are the sensitivity and specificity of the original tree and pruned tree?

(submit the confusion matrix and the calculation used)

8. What do sensitivity and specificity tell us in this context?

(explain in 1-2 sentences)

9. Write down the decision rules learned from the tree in plain English.