FTEC4003 Course Project Task 1: Debt Default Prediction.

1. Background

- This data comes from clients of a loan company. Through a thorough survey, these clients have already had some records about their credits. The company wants to determine which potential clients will most likely be debt defaulters through these records.

2. Data Set Information

- The data is related to a loan problem. The clients' information is about their basic information and credit situations.

This data set contains two files:

1. train.csv
    - The training set with seven input attributes and one output attribute (i.e., class attribute).
2. assignment-test.csv
    - The testing set with seven input attributes. You need to identify the class of each item.

Other files

1. samplesubmission.csv:
    - This is a sample file to show the output format. The wrong format will lead to an unknown result.
2. macOS setting: evaluate_1:
    - This is a command line tool to evaluate your result. We will use the F1-score of the positive class "1" to measure your result.
    - Usage: Press "command + space" to open spotlight search and type in "terminal," then type in the following command. You should replace `./submission_1.csv` with your own path to the submission_1.csv.

```
./evaluate_1 ./submission_1.csv
```

3. Ubuntu setting: evaluate_1_ubuntu:
    - This is a command line tool to evaluate your result. We will use the F1-score of the positive class "1" to measure your result.
    - Usage: Press "ctrl + alt + t" to launch a terminal and input the following command.

```
./evaluate_1_ubuntu ./submission_1.csv
```

4. Windows setting: evaluate_1.exe:
    - This is a command line tool to evaluate your result. We will use the F1-score of the positive class "1" to measure your result.

- Usage: Press "command + r" and then type "cmd" in the dialog box to launch a terminal. Then type in the command:

```
./evaluate_1.exe ./submission_1.csv
```

3. Goal

- The classification goal is to predict if the client will be a debt defaulter (i.e., Identify the value of feature 'Class,' 1 for yes and 0 otherwise). We evaluate the prediction by the F1-score of the positive class "1".

4. Attribute Information

a) Input variables

clients' basic information

1. index: Unique ID of clients.
2. EA1: encrypted attribute 1.
3. EA2: encrypted attribute 2.
4. EA3: encrypted attribute 3.
5. EA4: encrypted attribute 4.
6. EA5: encrypted attribute 5.
7. EA6: encrypted attribute 6.

b) Output variable

1. Class: whether the client is a defaulter (category: '0': not a defaulter, '1': a defaulter).