FTEC4003 Course Project Task 2: Startup Failure Prediction.

1. Background

- This data comes from an institution of venture capital. This venture capital institution wants us to help them predict whether a startup company will succeed or fail. The top priority of this investment is to avoid huge losses. So we need to find out which startups will fail. Specifically, the task is to train a binary classifier to do the binary classification based on given information. With the trained classifier, we could help venture capital predict whether a startup will fail or not.

2. Data Set Information

- The data are attributes of a company's basic information.
- train.csv
  - The training set with 21 input attributes and one output attribute (i.e., class attribute)
- assignment-test.csv
  - The testing set with 21 input attributes. You need to identify the class of each item.

This data set contains two files:

1. train.csv
   - The training set with known labels
2. assignment-test.csv
   - The testing set without labels (the "failure" Attribute).

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_2:
   - This is a command line tool to evaluate your result. We will use the F1-score to measure your result. Please note that it's the F1-score of the positive class "1". Details may refer to lecture notes.

- Usage: Press "command + space" to open spotlight search and type in "terminal", then type in the following command. You should replace `./submission_2.csv` with your own path to the submission_1.csv. Please note that `./` denote the current position of the command line and `submission_2.csv` denote your submission file name.

```
./evaluate_2 ./submission_2.csv
```

3. Ubuntu setting: evaluate_2_ubuntu:
   - This is a command line tool to evaluate your result. We will use the F1-score to measure your

result. Please note that it's the F1-score of the positive class "1". Details may refer to lecture notes.

- Usage: Press "ctrl + alt + t" to launch a terminal and input the following command.

```
./evaluate_2_ubuntu ./submission_2.csv
```

4. Windows setting: evaluate_2.exe:

- This is a command line tool to evaluate your result. We will use the F1-score to measure your result. Please note that it's the F1-score of the positive class "1". Details may refer to lecture notes.

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

```
./evaluate_2.exe ./submission_2.csv
```

3. Goal

- The classification goal is to predict whether the startup will fail in the future (i.e., Identify the value of feature 'failure', 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

startups' basic information

- index: unique ID of companies.
- attr1: relating to non-core business.
- attr2: relating to the overall business.
- attr3: relating to core business.
- attr4: relating to core business.
- attr5: relating to the cash flows.
- attr6: relating to liabilities and the cash flows.
- attr7: relating to liabilities and assets.
- attr8: relating to profit.
- attr9: relating to profit.
- attr10: relating to profit.
- attr11: relating to profit.
- attr12: relating to profit.
- attr13: relating to profit.
- attr14: relating to assets.
- attr15: relating to assets.
- attr16: relating to assets.
- attr17: relating to assets.
- attr18: relating to the investment.
- attr19: relating to cash flows.
- attr20: relating to cash flows.

b) Output variable

- failure: whether the startup will fail in the future.