Instructions

- This is a practice problem as a warm-up for the MiTH
- The data for the problem is uploaded in SCT
    - You would be given two data sets, train and test in csv formats
    - Train data will have the target column while for the test, you would find only the independent attributes but not target.
    - You need to make predictions on this test data.
    - You need to develop and tweak the model based on the error metric given to you.
- You need to submit the following
    - A submission csv file which contains your predictions on the test data set needs to be uploaded. You would have 5 chances to upload your test predictions.
    - Based on your predictions, the system would give you the score. Based on the score, if you can tweak the model and upload the test predictions again to check if the score improves. You have five such chances to upload your predictions

Problem Statement:

The given data contains financial ratios of various firms and the target attribute tells whether the firm has got bankrupt in the subsequent years or not. Your objective to come-up with a predictive model which helps to predict whether given firm going to be bankrupt or not.

Data set and Description:

Data: Train.csv
Description: Desc.txt

Expectations:

- ✓ Data preprocessing
    - Read the data; Deal the column names and headers properly
    - Check for highly correlated, and zero-variance attributes
    - Decide on imputation and standardization of data. Justify
    - Tool (Use either R or Python, based on your level of comfort)

- ✓ Model Building
    - Client is more interested in identifying the bankruptcy so true positives are more important
    - Apply ML algorithms to predict defaulters
    - Use all the relevant techniques which you have learnt so far
    - Analyze the results of each model implemented
    - Tune the models by feature selection, and hyper parameter tuning;
    - Analyze the results again; Get the best model

- Tool: At least 4 models should implement (2 in R and 2 in Python)

✓ Challenges/Further scope

- Report challenges faced if any, and scope to improve further

✓ Deliverables

- Entire R and Python code(s)
- Predictions csv submissions