Company Bankruptcy Prediction – a Kaggle Competition

For the project, you are required to use the data provided to predict company bankruptcy. For each company, various features are provided in the training data. You can use methods that you learned in our class to try your hands on this real life problem.

Tasks:

1. Split the training data provided into training and validation. You can also use cross-validation.
2. You can choose to use techniques to deal with imbalanced data, but this is not required. When you use under-sampling or over-sampling, please remember that you can **only** do that on your training set, **not** your validation set, **not** your test set either.
3. Choose **three** classification methods introduced in the class to train on the model. Use either the validation set or the cross-validation to tune hyper-parameters of your models.
4. Report the performance (accuracy, F1-score for both classes, AUC) for each method in the validation set or cross-validation.
5. Choose the best model and make your prediction of the test set.

Marking:

This project will be 25% of the course grade.

Part of grade will be determined using the **F1-score of the default class**. So you want to optimize your model parameters using the F1-score of the default class in the (cross) validation set.

You can also go beyond of what we learned in the class. Reasonable additional contributions can be awarded with a small amount of additional points.

Report:

1. Submit a Python notebook or R script with sufficient comments.
2. Submit a separate PDF report on the methods you choose, how you tune the hyper-parameters, results you obtained for each method, and how you choose your best model. The report is limited to at most 8 pages long, with figures.
3. Submit these two files separately. Do not Zip them together.

Academic Integrity

You can use the codes provided in our course. Please use the appropriate referencing. However, you **cannot** directly copy and paste other codes found online. Your submission will be compared with online material to check for similarity using plagiarism software.

This is an **individual project**. You can discuss with other classmates. But you need to write all codes yourself.

In the end, the goal for this project is to sharpen your own skills!