Big Data and Finance Class Complementary Project June 22nd 2025

Problem : Credit Approval Model

**As a complementary project students who did not validate the class have to individually solve the following credit approval problem.**

The goal of this problem is to build models to predict the credit approval status (column TARGET) using the other variables (or a subset of them). The dataset can be downloaded from this url dataset.

Students have to implement two solutions:

- a first based on a random forest model **(5pt)**,
- and a second one based on a one hidden layer deep network model **(5pts)**.

Proposed models have to be properly evaluated through a randomized three fold cross validation on model parameters specified by students, with 20% of the dataset used as test set **(2.5pts)**.

It has to be noticed that the problem at hand is what is called an unbalanced problem with far more samples with negative labels than with positive labels. Students have to find a solution to address this issue (**2.5pts**).

For random forest solutions, for the best model, input variables importances have to be visualized and analyzed **(2.5pts)**.

For neural network solutions, students have to propose, implement, and analyze a method to assess input variables importances over final classification performances **(2.5pts)**.

Students have one week to work on their project and deliver their solution into a jupyter notebook that professors can replay if needed. **It has to be noticed that notebook clarity will be part of the mark. Points can be lost for unclear notebooks.**