Homework #4
(Ensemble Learning)
Due: Sunday, September 30, 2018

Note: Submit your homework by pushing your code to your Gitlab account

A. REQUIRED PART

Goal

Another area where data science and machine learning play a huge role is in choosing if granting a loan. This is a particularly hot field as many start-ups feel that bank loan models can be improved. Therefore, there is space to come up with better loan strategies that can benefit both the lender and the borrower.

In this challenge, you will have access to loan data from a bank and will have to improve their model.

Description

We have access to a specific bank loan data. We have data about all loans asked to the bank, whether the bank decided to grant it and, finally, whether the borrower managed to repay it. We also have info about the borrower at the moment she is asking for the loan.

You have to come up with a better strategy to grant loans. Specifically you should:
* Build a model which is better than the bank model. For simplicity, assume that:
  - If you grant the loan and the it doesn't get repaid, you lose 1
  - If you grant the loan and the it does get repaid, you gain 1
  - If you don't grant the loan, you gain 0
* Using the rules above, compare bank profitability vs your model profitability
* Describe the impact of the most important variables on the prediction. Also, focus on the variable "is_employed", which describes whether the borrower is employed when she asks for the loan. How does this variable impact the model? Explain why
* Are there any other variables, not in the data provided, that you'd have liked to include in the model?

Dataset link:

https://drive.google.com/file/d/1Rj9Vto1ZnP6WM_qZsfWBmlxNtgf-GCZz/view?usp=sharing

https://drive.google.com/file/d/18z0B70G378FGxJ2i7DFJXK0tC2XdVuQU/view?usp=sharing

B. OPTIONAL PART

Apply ‘Ensemble Learning’ if not done in A)