# INTERPRETABILITY AND ALGORITHMIC FAIRNESS

| | |
|---|---|
| **Term: Fall 2024** | **Instructor: Prof. Christophe Pérignon** |
| **Location: HEC Campus** | **Email:** perignon@hec.fr |

## PROJECT DESCRIPTION:

**Goal:** The goal of this project is to apply most of the techniques presented during this course on a dataset provided by the instructor and available on Slack.

**Groups:** Students have to work by group of five people. The groups have to be set by the Student Representatives.

**Delivery:** Each group must deliver a computer code (Python, R, etc) and a slide presentation, which both have to be sent by email to the instructor before Monday September 30, 9:40AM.

**Presentation:** Each group has to make a presentation on campus to a jury on Wednesday November 30 between 9:40AM and 5:50PM (exact time for each team TBC). Presentation 15 minutes + Q&A 10 minutes.

**Dataset:** The dataset can be downloaded from Slack (dataproject2024.xlsx). The dataset must be deleted at the end of the project.

**Step 1:** Use the estimated default probability (PD) provided in the dataset. Implement one or two surrogate model(s) to interpret the unknown model used to generate PD.

**Step 2:** Estimate your own black-box machine learning model forecasting default. Each model is specific to a group of students and cannot be developed in collaboration with another group.

**Step 3:** Analyze the forecasting performance of your own model.

**Step 4:** Global interpretability: Implement one or two surrogate model(s) to interpret your own model. Compare the results provided in Steps 1 and 4.

**Step 5:** Global interpretability: Implement the PDP method to interpret your own model. Compare the results provided in Steps 4 and 5.

**Step 6:** Local interpretability: Implement the ICE method to interpret your own model.

**Step 7:** Local interpretability: Implement the SHAP method to interpret your own model. Compare the results provided in Steps 6 and 7.

**Step 8:** Performance interpretability: Implement the permutation importance method and/or the XPER method to identify the main drivers of the predictive performance of your model. Are the drivers of the performance metric (Step 8) similar to the drivers of the individual forecasts identified by SHAP (Step 7).

**Step 9:** Assess the fairness of your own model with respect to age (protected attribute). Use a statistical test for the following three fairness definitions: Statistical Parity and Conditional Statistical Parity (groups are given in the dataset). Discuss your results.

**Step 10:** Implement a FPDP using a fairness measure. Discuss your results.

## PROJECT EVALUATION:

| | |
|---|---|
| Technical skills | (0 to 10 points, 1 point per step) |
| Presentation skills | (0 to 5 points)* |
| Q&A management | (0 to 5 points)* |
| Slides & Code | (0 to 5 points) |
| Total | (0 to 25 points) |

* These parts of the grade can vary across team members.