PROJECT – IMAFA 2024-2025
ActuarAI SOFTWARE

1. LIFE INSURANCE PROJECT: LIFE ANNUITY TEMPORARY AND DEFERRED

STUDY: Generate a dataset and train some Machine Learning (ML) models to price **Life Annuities due at the end of the year temporary (n) and deferred (m').**
PRICING TOOL (code and working executable): User Interface from where we get Actuarial & Machine Learning prices.

    a. Actuarial theoretical formula
- Write single and annual premium formulae in the report depending on:
  - **x** Age
  - **m** Number of premium payments (payable at the beginning of each year, from the issue date)
  - **n** Term
  - **i** Technical Rate
  - **A** Annuity amount due at the end of the year
  - **m'** Annuity deferment

    b. Dataset
- Mortality table to use: **TGF05**
  https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000241832
- Randomly generated from all parameters (x, m, n, i, A, m', **generation**) in order to get a set (N = size of the dataset: from 1 to 1 000) of premiums with:

  x in: 20, 30, 40, 50, 60
  m in: 1, 5, 10, 20, 30, 40
  m' in: 0, 1, 5, 10, 20, 30, 40
  n in: 1, 5, 10, 20, 30, 40, 50, 60
  i in: 0, 0.5%, 1%, 1.5%, 2%, 2.5%
  A in: 50, 100, 200, 400, 800, 1000, 2000
  Generation is the year of birth of the insured person.
- Dataset study: provide in the report some analysis on the generated dataset (statistics, graphics and comments).

    c. Machine Learning models study
- Train and analyze some (at least two) ML models (scores, adjustment parameter, learning curve…) in order to get models providing good estimation of the insurance product prices.
- Do some dataset pre-processing if it is relevant.
- Describe in the report how you proceeded to train each model, choose the final parameter(s) and dataset size.

    d. Pricing tool with **Python**
       Provide the end user with a pricing tool where he can:
- Choose the size (N) of the ML dataset (will be generated on fly then),
- specify parameters of a life annuity due at the end of the year temporary and deferred (x, m, n, i, A, m'),
- get the Actuarial and ML prices,
- see the difference between both prices,
- see the parameter used and scores of each ML model.

2. NON-LIFE INSURANCE PROJECT: Vehicle Insurance Fraud Detection

Vehicle insurance fraud involves conspiring to make false or exaggerated claims involving property damage or personal injuries following an accident. Some common examples include staged accidents where fraudsters deliberately "arrange" for accidents to occur; the use of phantom passengers where people who were not even at the scene of the accident claim to have suffered grievous injury, and make false personal injury claims where personal injuries are grossly exaggerated.

The provided dataset contains vehicle dataset - attribute, model, accident details, etc along with policy details - policy type, tenure etc.

The goals are:

- to study and analyze the dataset,
- to do some pre-processing if needed,
- to analyze/train some (at least two) Machine Learning models,
- to provide the end user with a python tool so that he can detect if a claim application is fraudulent or not when he submits the required data.

3. INFO

Backend: python
Deadline: 12th March 2025
Group: 3 per group but, up to two groups of 2 persons is possible.