1SC4292 – Data from the E3N / E4N cohort for trending identification

Instructors: Laurent Le Brusquet, Arthur Tenenhaus
Department: DOMINANTE - VIVANT, SANTÉ, ENVIRONNEMENT
Language of instruction: FRANCAIS
Campus: CAMPUS DE PARIS - SACLAY
Workload (HEE): 40
On-site hours (HPE): 27,00

Description

- As part of this integration teaching, we propose to implement statistical/machine learning methods to address epidemiological issues. In particular, we will focus on the data of the E3N/E4N cohort (Epidemiological Study of Women in the MGEN), which therefore concerns hundreds of thousands of women (https://www.e4n.fr). This is therefore a problem in the statistical analysis of a "Big Data" dataset in which we try to identify major population trends based on massive, noisy and incomplete data.

- Associated partner: INSERM

1. Location : Paris-Saclay Campus

Quarter number
ST4

Prerequisites (in terms of CS courses)
Statistics and Learning

Class components (lecture, labs, etc.)
The students involved in this project will be divided into sub-groups that will analyze the data using different techniques. As the data and the problem are provided by INSERM members, students will have to meet them several times. Finally, they will have to present their results to all the actors of the project (all the students involved in the project, partners, supervisors).

Grading
Oral defense and/or written report at the end of the integration course.

Course support, bibliography

The Elements of Statistical Learning : Data Mining, Inference, and Prediction. Second Edition. February 2009. Springer. https://web.stanford.edu/~hastie/Papers/ESLII.pdf

Learning outcomes covered on the course

At the end of this course, students will be able to define, understand, choose a statistical/machine learning method and implement it in accordance with the problem at hand.

Description of the skills acquired at the end of the course

C1 Analyse, design and implement complex systems made up of scientific, technological, social and economic dimensions.
C2.1 Thoroughly master a domain or discipline based on the fundamental sciences or the engineering sciences.
C3.1 Be proactive and involved, take initiatives
C3.5 Put forward new tools with either continual progress or disruptive solutions as the goal
C3.6 Evaluate the efficiency, feasibility and strength of the solutions offered. / proposed solutions
C3.7 Make pragmatic and informed choices with the aim of producing tangible results.
C6.1 Identify and use the necessary software for one's work (including collaborative tools) and adapt digital responses according to the context.