

1SC4292 - Les données de la cohorte E3N/E4N pour l'identification de grande tendance

Responsables : **Laurent LE BRUSQUET , Arthur TENENHAUS**

Langues d'enseignement : **FRANCAIS**

Campus où le cours est proposé : **CAMPUS DE PARIS - SACLAY**

Nombre d'heures d'études élèves (HEE) : **40**

Nombre d'heures présentielles d'enseignement (HPE) : **24**

Année académique : **2024-2025**

Niveau avancé : **non**

Présentation, objectifs généraux du cours :

- Dans le cadre de cet enseignement d'intégration, nous proposons de mettre en œuvre des méthodes statistiques/machine learning pour répondre à des problématiques d'épidémiologie. En particulier, on s'intéressera aux données de la cohorte E3N/E4N (Etude Epidémiologique auprès de femmes de la MGEN), qui concernent donc des centaines de milliers de femmes (<https://www.e4n.fr>). Il s'agit donc ici d'un problème d'analyse statistique d'un jeu de données « Big Data » où l'on cherche à dégager des grandes tendances au niveau de la population à partir de données massives bruitées et incomplètes.
- Partenaire associé : INSERM
- Lieu : Campus Paris-Saclay

Période(s) du cours (n° de séquence ou hors séquence) :

ST4

Prérequis :

Cours de Statistique et Apprentissage

Déroulement, organisation du cours :

Les élèves impliqués dans ce projet seront répartis en sous-groupes qui analyseront les données aux travers de différentes techniques. Les données et le problème étant fournis par des membres de l'INSERM, les étudiants seront amenés à les rencontrer à plusieurs reprises. Enfin, ils auront à restituer leurs résultats devant l'ensemble des acteurs du projet (tous les élèves impliqués dans le projet, partenaires, encadrants).

Organisation de l'évaluation :

Soutenance orale et/ou rapport à l'issue de l'enseignement d'intégration.

Acquis d'apprentissage visés dans le cours :

À l'issue de ce cours, les élèves seront capables de définir, comprendre, choisir une méthode statistique/machine learning et de la mettre en œuvre en adéquation avec le problème posé.

Description des compétences acquises à l'issue du cours :

- 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.

Bibliographie :

The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition. February 2009. Springer.

<https://web.stanford.edu/~hastie/Papers/ESLII.pdf>