

DATA MINING ET BIG DATA

Semestre 4

Code Apogée : **DBH2071M**

3 ECTS

11 heures de CM et 10 heures de TD

Master Santé Publique
Parcours M2 B3S Biostatistique, Biomathématique
Bioinformatique et Santé.

Composante de gestion de l'UE :

Responsables de l'UE :

DOMAINE SCIENCES TECHNOLOGIE SANTE

Département de Biologie Humaine et ECL

Céline HELBERT, Delphine MAUCORT-BOULCH

Programme de l'enseignement :

L'objectif du cours est de donner aux étudiants les outils pour aborder des problèmes statistiques du Big Data sous l'angle des problèmes d'apprentissage en grande dimension. La grande dimension pourra être vue sous l'angle d'un grand nombre de variables et/ou d'un échantillon de grande taille (nombre d'individus important).

Une conférence introductive permettra de donner les contours de ce qui relève du Big Data.

Ensuite, les grandes classes de méthodes d'apprentissage seront étudiées : modèles linéaires pénalisés, modèles additifs généralisés, modèles à base d'arbres (CART, random forests, etc).

Cette UE est mutualisée avec la filière MIR (« Mathématiques et Ingénierie du Risque ») de l'option Mathématiques et Décision de la troisième année de l'École Centrale.

Modalités d'évaluations de l'unité d'enseignement :

Session 1 : 2 rapports.

Session 2 : Pas de session 2.

(Modalités données à titre indicatif, les modalités précises sont votées tous les ans par le CA sur proposition du CFVU)

Vos enseignants dans cette UE : D. MAUCORT-BOULCH, C. HELBERT.