

La science des données en santé : applications pratiques

- Cours (CM) 16h
- Travaux dirigés (TD) 32h
- Travaux personnels 16h

Description du contenu de l'enseignement

Ce cours offre une approche d'apprentissage par la pratique dans le domaine des sciences des données en santé. L'objectif pour les étudiants est d'être capable d'identifier et de mettre en œuvre les méthodologies adaptées pour répondre à des problématiques polyvalentes en lien avec les données de la recherche biomédicale.

Ce cours est organisé autour de la réalisation de quatre travaux personnels réalisés en petits groupes de travail. Le cours magistral sert à présenter les notions indispensables à la réalisation des travaux personnels. Les TD sont des séances de travail collectif visant à accompagner l'avancée des travaux personnels.

Ce cours constitue une étape importante dans le développement de compétences techniques (capacité à préparer les données, capacité à tirer parti du machine learning et de l'intelligence artificielle,...) et personnelles (approche proactive de la résolution des problèmes, communication efficace, sensibilité métier, curiosité intellectuelle,...) en science des données de santé.

Compétences à acquérir

À l'issue de ce cours l'étudiant est capable de :

- maîtriser les concepts-clés et le vocabulaire de base de la science des données en santé
- être en capacité de préparer les données, les analyser, les présenter
- savoir tirer parti du machine learning et de l'intelligence artificielle
- avoir une approche proactive de la résolution des problèmes liés aux données de santé
- appréhender l'existence de difficultés nécessitant le recours à un expert
- se familiariser avec les problématiques de la recherche clinique

Bibliographie, lectures recommandées

- <https://scikit-learn.org/stable/>
- <https://keras.io>
- Müller A.C. & Guido S., *Introduction to machine learning with Python*, O'Reilly , 2017.
- Chollet F. with Allaire J.J., *Deep learning with R*, Manning, 2018.

Modalités d'organisation et de suivi

Méthodes pédagogiques :

Pédagogie basée sur l'apprentissage par la pratique. Les étudiants sont accompagnés dans la réalisation d'analyses de données de santé en condition simulant les attentes professionnelles au sein de start-up ou de départements orientés « data analyse ».

L'UE consiste en la réalisation de 4 travaux. Chacun consiste en 2 h de CM et 2h de TD (introduction et conceptualisation du travail), 2h CM et 2h TD (réalisation du travail), 4h TD (restitution) soit 16hCM, 32h TD et le reste (4h par module) de travail perso.

Une partie de l'accompagnement et des TD est dématérialisée à l'aide d'outils de travail à distance.

Le travail en équipe entre étudiants est favorisé et indispensable.

- Contrôle continu - pas d'examen final

Travail personnel indispensable :

Une implication forte est demandée aux étudiants pour se familiariser avec les outils et langages de programmation (R et python) utilisés dans ce cours.

Plan :

Travail 1 : *Visualization challenge*

Introduction de la visualisation de données - travail en petits groupes. Tous les groupes d'étudiants cherchent à créer un *dashboard* pour représenter et rendre intelligible les données d'une même source (type opendata) de données de santé régulièrement réactualisées. Pour la motivation et la cohésion de la promotion, les groupes sont mis en concurrence dans un challenge avec un vote final pour classer les tableaux de bord créés.

Travail 2 : *Prediction challenge*

Travail d'introduction à la pratique du *machine learning* et du *deep learning*. Tous les groupes d'étudiants disposent d'un même jeu de donnée en lien avec des problématiques de classification ou de régression sur des données de santé. Les groupes sont mis en concurrence dans un challenge de type *Kaggle*

Travail 3 : Analyse des données d'une étude clinique

Les groupes travaillent à l'analyse et la présentation de données réelles issues d'essais thérapeutiques. Ils disposent de l'ensemble des documents habituels d'une étude clinique (protocole, eCRF commenté, données, ...). Les étudiants sont amenés à se pencher sur toutes les étapes allant de la mise en forme des données transmises par des équipes de data management jusqu'à la mise en forme synthétique des données dans le cadre d'une présentation de résultats.

Travail 4 : Analyse de données observationnelles issues d'un entrepôt de données de santé ou d'un infocentre.

Les groupes travaillent à l'analyse et la présentation de données réelles issues d'un entrepôt de donnée santé ou d'un infocentre biomédical. Les problématiques de nettoyage, visualisation,

analyse de données massives sont abordées. L'objectif est de fournir un compte rendu de type synthèse d'analyse pour répondre à une question à l'aide d'un *data-mart*.

Évaluation : contrôle continu (oral en session 2)

Responsable :

Faculté de Pharmacie
74 ROUTE DU RHIN
67400 ILLKIRCH

Dr JULIEN GODET
74 ROUTE DU RHIN
67400 ILLKIRCH
03.688.54183
julien.godet@unistra.fr