DIRO, Hiver 2022 **Chargé de cours : Mohamed Lokbani**

Secrétariat DIRO : 2194, Pavillon André-Aisenstadt
2920, chemin de la Tour
Université de Montréal

IFT 6501

Traitement informatique des mégadonnées

Modalités

Cours théorique et travaux pratiques durant un trimestre. Pour la session d'hiver 2022, il y aura 13 semaines de cours à raison de 4 heures de cours par semaine. Le cours est un mélange de théorie et de pratique.

Jour	Heure	Section	Lieu	Type
Mardi	12h30-14h30	A	Zoom (ou) 1411 Pav. André-Aisenstadt	Cours magistral
Vendredi	10h30-12h30	A	Zoom (ou) 1409 Pav. André-Aisenstadt	Cours magistral

La première séance de **cours** est prévue pour le mardi 11 janvier 2022.

Le cours aura lieu à distance via Zoom jusqu'à au moins le 31 janvier 2022.

Site web

Toutes les informations relatives au cours seront disponibles sur le site du cours sur Studium.

Présentation du cours

Ce cours a pour objectif de passer en revue les différents outils informatiques nécessaires pour traiter des données. Nous allons d'abord vous initier au langage Python. Par la suite, nous allons présenter les paquetages Python nécessaires pour préparer, traiter et visualiser les données. Nous allons rappeler des notions mathématiques et statistiques, nécessaires à la compréhension du cours. Nous allons étudier des modèles statistiques classiques et des techniques de l'apprentissage automatique. Nous allons examiner quelques cas reliés à la santé publique. Finalement, nous allons terminer le cours en présentant des moteurs de traitement de données rapides dédiés au « Big Data ».

Quelques références

« Data science par la pratique : [fondamentaux avec Python] », Joel Grus, Eyrolles (traduction de: « Data science from scratch : first principles with Python ») 2e édition.
https://www.eyrolles.com/Informatique/Livre/data-science-par-la-pratique-9782212679076/

« Python pour le Data Scientist », Emmanuel Jakobowicz, Dunod, 2e édition.
https://www.dunod.com/sciences-techniques/python-pour-data-scientist-bases-du-langage-au-machine-learning-1

« Intro to Python for Computer Science and Data Science », Deitel & Deitel, Pearson.

https://www.pearson.com/us/higher-education/program/Deitel-Intro-to-Python-for-Computer-Science-and-Data-Science-Learning-to-Program-with-AI-Big-Data-and-The-Cloud/PGM2392788.html

D'autres références seront communiquées durant la session.

Évaluation

Projet final: 30%

Le projet final consiste à examiner des données de A à Z en appliquant les différentes techniques étudiées durant le cours. Le projet est annoncé à la fin de la session et sera réalisé sur un long week-end.

Projet de session: 20%

Le projet de session consiste à examiner une thématique donnée. Une présentation est faite par la suite par chaque étudiant en classe (à distance si les mesures sanitaires ne le permettent pas).

Travaux pratiques: 30%

Une série de travaux pratiques à faire durant toute la session. La durée du travail peut varier en fonction des sujets traités.

Quiz hebdomadaires: 20%

Une série de questions sur les différentes notions du cours fraichement étudiée. Les quiz sont de courte durée à faire dans la semaine.

La note sur 100 est convertie en note littérale (A+, A, A-, etc.) à la fin du cours seulement, selon un barème qui dépendra à la fois de la moyenne du groupe et de la répartition des étudiants. De plus, la note globale doit satisfaire les exigences de la faculté où l'étudiant s'est inscrit.

Le plagiat à l'Université de Montréal est sanctionné par le Règlement disciplinaire sur la fraude et le plagiat concernant les étudiants. Pour plus de renseignements, consultez le site http://www.integrite.umontreal.ca

Dates importantes

Date limite d'annulation d'inscription sans frais, le 21 janvier 2022.

Date limite d'abandon avec frais, le 18 mars 2022.

Sujets traités

Introduction : Big Data et science des données pour la santé publique
Langage python pour la science des données : les bases de Python, les listes, opérateurs relationnels et logiques, les dictionnaires
Utilisation des librairies Numpy, Pandas, Matplotlib et Seaborn
Rappel probabiliste et analyse statistique
Apprentissage machine : apprentissage supervisé et non supervisé
Réseau de neurones et apprentissage profond
Écosystème BigData : Apache Spark et Hadoop