

SDS3786

Laboratoire en sciences des données

Détails du cours

Description du cours

La gestion et l'analyse des données. Un cours de laboratoire et projet incluant: la manipulation et le nettoyage de données; l'analyse exploratoire et la visualisation de données; méthodes de régression et de classification; évaluation de résultats et écriture de rapports.

Objectifs:

- apprendre à travailler en équipe de façon efficace
- apprendre à gérer le temps et les projets
- marier le travail quantitatif aux aspects non-quantitatifs de la science des données
- travailler avec des données réalistes
- commencer à élaborer un portefeuille de projets de science des données

Mon approche pédagogique

1. Il faut FAIRE qqch (écrire, travailler avec les données, exercices, preuves, etc.).
2. Il faut NE PAS RÉUSSIRE (nous n'apprenons pas du succès, mais des échecs).
3. Il faut être AUTO-DIDACTE (fouiller pour trouver l'information requise, en ligne, dans des livres, sur YouTube, etc.).
4. Il faut utiliser un procédé ITÉRATIF (c'est en forgeant que l'on devient forgeron).
5. Il faut REMETTRE qqch (le perfectionisme n'est pas nécessairement un atout).

Détails du cours

Instructeur: Patrick BOILY

Horaire:

- LUN 14:30-17:20 (LMX 241)

Disponibilité:

- en classe
- sur Slack
- sur Zoom

Référence:

- [*Data Understanding, Data Analysis, & Data Science*](#), P.Boily, DAL/QED

Site web:

- sds3786-a24.slack.com
- data-action-lab.com/sds

Évaluation

- 8 laboratoires – 60% [cf. horaire détaillé]
- Projet (proposition, rapport de progrès, rapport final, présentation) – 30% [cf. détails]
- Participation – 10%

Je ne compte que vos 7 meilleurs laboratoires.

Mais soyez prudents: c'est une police d'assurance, pas une carte « sortie de prison ».

Semaine

1	<i>Cours 1: Introduction + Présentation des données + RStudio/Jupyter</i>	
2	<i>Cours 2: Programmation 101 + « Reporting » + écriture technique</i>	Choix d'équipe (20 sept)
3	Labo 1: Les aspects non-techniques de la science des données (9-23 sept)	
4	Labo 2: La visualisation des données (24-30 sept)	
5	Labo 3: Le traitement des données (1-7 oct)	
6	<i>Semaine de lecture</i>	
7	<i>Cours 3: Projet final, proposition du projet final à remettre</i>	Proposition
8	Labo 4: La grammaire des graphiques (22-28 oct)	
9	Labo 5: La mise en récit de données (29 oct-4 nov)	
10	<i>Cours 4: projet final, rapport de progrès à remettre</i>	Rapport de progrès
11	Labo 6: L'apprentissage automatique et les règles d'association (5-18 nov)	
12	Labo 7: La classification et la régression (19 nov-2 déc)	
13	Labo 8: Le regroupement (19 nov-2 déc)	

Présentation
Rapport final
(à déterminer)

Laboratoires

Chaque rapport de laboratoire doit comporter au **maximum 6 pages**.

La contribution de chacun des membres doit être donnée **de manière explicite** pour chaque rapport de laboratoire.

Pour chaque laboratoire, il n'y a que 5 notes possibles:

- F (0%), C (62.5%), B (72.5%), A- (82.5%), A+ (95%)

Vous n'aurez pas le temps de réaliser les laboratoires si vous ne vous êtes pas préparé en regardant les vidéos et/ou en lisant les notes de cours au préalable.

Vous devez vous présenter à chaque laboratoire avec un plan d'action.

Projet

Tout au long du semestre, vous travaillerez (en groupe) sur un projet de science des données de votre choix (vous devez d'abord obtenir mon approbation).

- Choix d'équipe: 20 sept
- Proposition: 21 octobre [5]
- Rapport de progrès: 11 nov [5]
- Présentation finale: à déterminer [10]
- Rapport final et portefeuille: à déterminer [10]

Le rapport final doit comporter au **maximum 10 pages**. La présentation finale dure **15 minutes**.

La contribution de chacun des membres doit être donnée **de manière explicite**.

Un mot sur les pré-requis et le codage

Il n'y a pas d'autre solution - pour tirer le meilleur parti du cours, vous devez **en maîtriser les pré-requis**: calculus, algèbre linéaire, probabilités et statistiques. Il serait utile d'avoir fait des études plus avancées (régression, séries chronologiques, etc.), mais ce n'est pas nécessaire.

Une chose dont vous aurez du mal à vous passer est la connaissance des environnements de programmation. Je vous suggère R et Rstudio. Mais R n'est pas un « vrai » langage de programmation et la plupart des emplois requièrent Python et/ou SQL, alors pourquoi R ?

Franchement, **ce n'est pas la bonne question à poser**: tous les langages sont plus ou moins équivalents et une fois que vous savez comment fonctionne l'un d'entre eux, vous avez la possibilité de comprendre les autres.

Autre raison: il ne s'agit pas seulement d'un cours sur les « résultats/l'efficacité », mais aussi sur les concepts et les aspects non-techniques (présentation, travail d'équipe).

Attentes

Engagement: l'université s'attend à ce que vous consacriez 135 à 150 heures à chaque cours [8 à 10 heures par semaine], en classe, à la lecture, aux projets, etc.

Je m'attends aussi à ce que vous participiez aux discussions (en classe, sur Slack) et que vous soyez un membre actif de vos équipes de travail (vous pourrez choisir vos propres groupes).

Mais il ne s'agit que d'un cours, d'un trimestre, d'une année, d'un diplôme, d'une vie. Ne perdez pas cela de vue. Je vous encourage à prendre toutes les précautions raisonnables pour optimiser votre santé physique et mentale. Prenez soin de vous et de vos proches. Aucun problème n'est trop petit (ou trop grand) pour être ignoré.

Le travail d'équipe

Question: Pourquoi confier un travail de groupe alors que l'on sait qu'il en résulte toujours une répartition inégale du travail entre les membres du groupe ?

Réponse (selon K. Wheeler): Je vais commencer par dire que je détestais les projets de groupe lorsque j'étais étudiant. Il y avait toujours au moins un fainéant qui ne faisait pas sa part du travail. Il y en avait toujours un ou deux qui finissaient par faire deux fois plus de travail que les autres.

Cependant, je continue à confier de tels projets, non pas parce qu'ils sont particulièrement justes, mais parce qu'ils permettent d'acquérir des **compétences subtiles mais vitales**.

Dans le monde professionnel, vous aurez des employeurs qui chargeront des groupes de travailleurs de terminer des projets collectivement. **Cette situation ne peut être évitée.** Les responsables ne vont pas vous superviser en permanence pour s'assurer que vous faites votre part du travail. Ce qu'ils attendent, c'est que le groupe réalise le projet - ils ne se soucient absolument pas de savoir qui a fait 47% du travail et qui en a fait 8%, DU MOMENT QUE LE GROUPE L'ACHEVE.

Le travail d'équipe

La vie est ainsi faite que vous ne pouvez pas toujours être à la hauteur, pour diverses raisons (certaines meilleures que d'autres). Parfois, vos coéquipiers prennent le relais pour vous, parfois vous prenez le relais pour eux. Mais que se passe-t-il lorsque vous avez un fainéant chronique ?

L'apprentissage ne porte pas seulement sur le sujet du projet. Il s'agit aussi de savoir comment vous allez, **EN TANT QUE GROUPE**, gérer le problème de l'inégalité de la charge de travail et de la qualité.

Supposons que Nick soit le paresseux de l'équipe. Allez-vous :

- m'approcher et vous plaindre que Nick ne fait pas sa part du travail, en supposant que je vais régler le problème pour vous ?
- dissimuler le fait que Nick ne fait aucun travail et que les autres vont le remplacer (et s'attribuer le mérite du travail qu'il n'a pas fait) en prenant le relais ?
- avoir une conversation franche en groupe sur les personnes qui ne font pas leur part de travail (y compris en regardant personnellement Nick dans les yeux lorsque vous le critiquez) ?
- diviser le projet en sections et demander à chaque membre du groupe de "signer" sa section (documentant l'effort individuel), afin qu'il soit clair que Nick n'a pas fait grand-chose ?
- inclure une brève réflexion narrative dans laquelle chaque membre de l'équipe s'arrête et discute de ce qu'il a ajouté au projet, ce qui montre clairement que Nick n'a pas fait sa part ?
- demander à chaque membre de votre équipe de m'envoyer un rapport d'avancement, documentant que Nick ne fait pas sa part ?
- laisser le projet dans son ensemble s'effondrer parce que Nick est un fainéant ?
- développer des compétences de leadership et de persuasion pour trouver un moyen de motiver Nick à faire sa part ?

Le travail d'équipe

En tant qu'étudiant effectuant un travail d'équipe, tout cela semble **injuste, artificiel, et exaspérant**. Mais il s'agit là d'une réaction qui repose sur une idée fausse, à savoir que ce sont les notes, et uniquement les notes, qui comptent.

Pour les projets de groupe dans le monde réel, vous devez rapidement réaliser à quel point il est important que les individus trouvent la manière de gérer ce problème **en tant que membres d'un groupe plus large** et ne comptent pas uniquement sur le patron pour intervenir en permanence.

La politique du "**Cover Your Ass**" (qui consiste à documenter le travail par étapes) est une bonne compétence à développer. Il en va de même de la capacité à **confronter les tire-au-flanc** et à les aborder **honnêtement** et **directement**. Il en va de même pour la **prise de décision en groupe** sur la manière de traiter les membres défaillants du groupe. Il en va de même pour les **compétences interpersonnelles** permettant de travailler ensemble et de **motiver** les paresseux ou fainéants. Il en va de même pour le **leadership**. Vous ne maîtriserez peut-être pas toutes ces compétences, mais si vous en **maîtrisez au moins une**, vous aurez un réel avantage dans le monde du travail.

Le travail d'équipe

Je sais que le travail en équipe n'est pas toujours parfaitement équitable, mais il est **parfaitement réaliste**. Il est utile d'apprendre à connaître nos propres attitudes et comportements au sein des groupes et à résoudre (en groupe) les situations dans lesquelles un membre du groupe est dysfonctionnel.

Je peux énumérer des façons de gérer le problème en classe, mais **tant que vous n'êtes pas confronté à la réalité, tout cela demeure abstrait**. Le plus important est d'essayer (et d'échouer) de trouver des moyens de faire fonctionner les choses, puis de répéter avec d'autres moyens, jusqu'à ce que cela fonctionne et devienne une seconde nature. Si je vous dis comment vous y prendre pour travailler en groupe, c'est comme si, en tant que nageur expérimenté, je décrivais verbalement comment nager à des personnes qui n'ont jamais nagé - **comprendre la théorie ne signifie pas qu'on puisse l'appliquer sans pratique**.

Il ne s'agit pas seulement de savoir quelles stratégies nous pouvons utiliser pour surmonter les dysfonctionnements de la dynamique de groupe - il s'agit d'avoir le courage et la motivation d'en choisir une, de la mettre en pratique et de l'appliquer. C'est ce que j'espère que les étudiants retireront du travail de groupe, et **il est presque impossible de le quantifier ou de le mesurer de manière significative**.

Si vous abordez tout cela sérieusement, les compétences acquises grâce au travail en équipe sont **bien plus utiles** que n'importe quelle autre leçon que vous apprendrez en réalisant le projet lui-même.

Foire aux questions

Question: 6 pages??? Il nous faut plus de temps !! Et il y a trop de laboratoires !!

Réponse: Je crains que ce ne soit vrai. 6 pages, **ce n'est pas beaucoup**... et c'est une période **très, très courte** pour travailler sur un laboratoire. C'est là tout l'intérêt.

Je ne vous demande PAS la perfection. Je vous demande au contraire de:

- réfléchir à la problématique;
- élaborer un plan d'analyse et décider de ce qui doit être conservé ou supprimé;
- le mettre en œuvre; et
- rendre compte des résultats de votre analyse.

Foire aux questions

Question: On peut utiliser ChatGPT?

Réponse: Même si cela va à l'encontre de toutes mes convictions, **je l'autorise, dans la limite du raisonnable**. Mais méfiez-vous des incohérences de style et... eh bien, je ne devrais pas pouvoir dire que vous l'avez utilisé.

L'outil pourrait s'avérer utile **pour nettoyer votre écriture** et/ou **pour traduire du texte/code**. Mais vous devez rester conscient que la rédaction technique générée par l'I.A. n'est impressionnante que pour les personnes qui ne comprennent pas les questions techniques en premier lieu.

Si vous l'utilisez pour effectuer des analyses, vous vous retrouverez avec des œufs sur la figure.

Ce n'est pas une bonne idée.

Foire aux questions

Question: Je ne suis pas doué pour les relations humaines et je déteste le travail en équipe. Puis-je travailler seul ?

Réponse: Non. La science des données est un sport d'équipe. Je ne vous demande pas de devenir extravertis ou de changer votre attitude envers les gens : je vous demande de faire en sorte que le travail d'équipe... fonctionne. Si ce n'est pas votre tasse de thé, ce cours n'est pas pour vous. Sans rancune.

MAIS... je m'attends à ce que vous respectiez vos coéquipiers et vous-même (tant sur le plan académique que social). Les règles universitaires régissant les interactions s'appliquent (en classe, sur Slack, etc.).

Foire aux questions

Question: D'accord, d'accord. Si je m'associe avec certains élèves pour un projet, dois-je m'associer avec eux pour les autres projets ?

Réponse: Non. Si quelqu'un ne fait pas ce qu'il faut, ou si travailler avec quelqu'un vous stresse, vous n'êtes pas obligé de travailler à nouveau avec cette personne.

Parfois, vos coéquipiers devront vous remplacer, et vice-versa. Mais la **communication est essentielle:** si vous attendez la veille de la date limite pour leur dire que vous ne pouvez pas faire votre part, la note de votre équipe en sera affectée et vos coéquipiers ne voudront plus travailler avec vous, ce qui est tout à fait compréhensible. Ce n'est pas le lieu pour les secrets.

Je vais vous traiter tous comme des adultes. Comportez-vous en conséquence.

Liens importants

Intégrité académique et inconduite académique

- <https://www.uottawa.ca/notre-universite/politiques-reglements/reglements-academiques/a-4-integrite-academique-inconduite-academique>

Accommodements religieux

- <https://www.uottawa.ca/notre-universite/politiques-reglements/reglements-academiques/a-5-accommodements-religieux>

Accommodements académiques

- <https://www.uottawa.ca/notre-universite/politiques-reglements/reglements-academiques/a-6-accommodements-academiques>

Liens importants

Dates importantes et échéances scolaires

- <https://www.uottawa.ca/etudiants-actuels/dates-importantes-echeances-scolaires>

Centre de santé et mieux-être étudiant

- <https://www.uottawa.ca/vie-campus/sante-mieux-etre/centre-sante-mieux-etre-etudiant>

Ne vous laissez intimider ni par la matière, ni par le professeur !

Profitez de ma disponibilité, retrousses vos manches et mettez le paquet ! (c'est plus facile que vous ne le pensez, je vous le promets).