

Information générale

Cours	
Titre	BIO3033/6033 - Science des données en biodiversité
Nombre de crédits	3
Sigle	BIO3033 et BIO6033
Site StudiUM	BIO3033/6033 - Science des données en biodiversité
Faculté / École / Département	Arts et sciences / Département de sciences biologiques
Trimestre	Automne
Année	2024
Mode d'enseignement	En présentiel
Déroulement du cours	Le cours se déroule le vendredi de 12h30 à 15h30, en présentiel; chaque séance se compose de trois modules de 50 minutes (et 10 minutes de pause entre chaque)
Charge de travail hebdomadaire	Environ deux heures de lecture par semaine, et environ deux heures de travail pour les évaluations

Enseignant		
Timothée Poisot	Titre	Professeur agrégé
	Coordonnées	timothee.poisot@umontreal.ca
	Disponibilités	Sur demande: https://usemotion.com/meet/timotheepoisot/officehours

Description du cours	
Description simple	Introduction aux concepts de science des données, utilisant des exemples des sciences de la biodiversité; gestion et nettoyage des données, classification et régression, modèles de distribution d'espèces, changements climatiques.
Description détaillée	Les objectifs de ce cours sont <ol style="list-style-type: none"> 1. présenter les bonnes pratiques de gestion et de nettoyage des données biologiques 2. se familiariser avec les concepts principaux en science des données 3. illustrer les concepts théoriques via une étude de cas 4. comprendre comment les modèles et les scénarios peuvent participer au processus de prise de décision
Place du cours dans le programme	Ce cours est offert en troisième année et aux cycles supérieurs, et est optionnel

Apprentissages visés

Objectifs généraux	<ol style="list-style-type: none"> 1. Initier aux concepts de cycle de vies des données et des modèles, notamment par la production d'une "persona" 2. Développer la capacité à automatiser le contrôle de qualité et le nettoyage des données 3. Comprendre comment on passe des données à la prédiction, via une étude de cas qui s'étend sur la majorité de la session 4. Contextualiser les prédictions, via des ateliers participatifs qui simulent le processus de prise de décision
Objectifs d'apprentissage	<ol style="list-style-type: none"> 1. Évaluer la qualité d'un jeu de données pour une question biologique 2. Recommander l'utilisation de méthodes de machine learning pour

- la prédiction
3. Évaluer la qualité d'un modèle prédictif
 4. Identifier les sources possibles de biais dans un jeu de données ou un modèle

Calendrier des séances

6 septembre 2024	Titre	Introduction
	Contenus	Présentation du cours, méthodes pédagogiques, étude de cas, discussion sur l'IA générative
	Activités	Atelier - ça veut dire quoi, "des données"? Atelier - comment un algorithme peut "apprendre"?
13 septembre 2024	Titre	Le cycle de vie des données
	Contenus	Comprendre la place des données dans le processus de recherche
	Activités	Atelier - collecte de données
	Lectures et travaux	Lectures à faire avant la séance (les fichiers seront sur StudiUM)
	Évaluation	Rédiger sa "persona" (version 1), à faire après la séance
20 septembre 2024	Titre	Réconciliation, nettoyage, et contrôle de qualité
	Contenus	Utilisation de OpenRefine pour le nettoyage des données
	Activités	Atelier - données de prévalence de mpox Démonstration - nettoyage de données d'inventaire biologique avec OpenRefine
	Lectures et travaux	Lectures à faire avant la séance
	Évaluation	Texte court - biais dans les données de mpox
4 octobre 2024	Titre	Méthodes de regroupement
	Contenus	Données LANDSAT, algorithme de k-means
	Activités	Théorie - le regroupement Démonstration - classification du paysage du sud de la Corse
	Lectures et travaux	Lectures du chapitre sur le regroupement; sauf mention contraire, les séances suivantes s'accompagnent d'un chapitre à lire après la séance
11 octobre 2024	Titre	Descente de gradient
	Contenus	Prédiction de variables quantitatives, rappel sur la régression linéaire, notion d'optimalité
	Activités	Théorie - la descente de gradient Démonstration - structure des réseaux trophiques Atelier - comment parler de la performance d'un modèle?
18 octobre 2024	Titre	Validation des modèles
	Contenus	Validation croisée, fonctions de perte, choix d'une mesure de performance
	Activités	Théorie - la validation croisée Démonstration - floraison des cerisiers et changements

		climatiques Atelier - formuler des recommandations sur l'utilisation de ce modèle
	Évaluation	Texte court - correction d'un texte décrivant l'utilisabilité du modèle
1 novembre 2024	Titre	Classification binaire
	Contenus	Introduction de l'étude de cas: Sitta whiteheadi
	Activités	Théorie - la classification binaire Démonstration - premier brouillon de modèle Débat - que vaut ce modèle?
	Lectures et travaux	En plus des notes de cours, un article qui sera communiqué sur StudiUM, à lire après la séance
	Évaluation	Rédiger sa "persona" (version 2), à faire après la séance, en s'imaginant comme utilisateur de modèles de distribution d'espèce
8 novembre 2024	Titre	Sélection et préparation des variables
	Contenus	Notions de dimensionalité, de transformation, et de data leakage
	Activités	Théorie - ACP, Whitening, sélection stepwise, IVF Démonstration - amélioration du modèle de la semaine précédente Débat - qui faut-il croire, les biologistes, ou les données?
	Évaluation	Texte - identification de variables en lien avec la biologie de l'espèce
15 novembre 2024	Titre	Hyper-paramètres
	Contenus	Modification des hyper-paramètres du modèle, courbes d'apprentissage, incertitude, imbalanced classification
	Activités	Théorie - moving threshold classification Démonstration - optimisation du modèle de la semaine précédente Atelier - comment peut-on parler d'incertitude?
	Évaluation	Texte court - est-ce bien utile de vouloir répondre à une question par vrai ou faux?
22 novembre 2024	Titre	Interprétabilité des modèles
	Contenus	Réponses partielles, valeurs de Shapley et algorithme de Shapley Monte-Carlo
	Activités	Théorie - comprendre l'effet des variables sur la prédiction Démonstration - expliquer pourquoi le modèle fait certaines prédictions Atelier - comprendre, décrire, expliquer, prédire... que font vraiment ces modèles?
	Lectures et travaux	Plusieurs articles sur StudiUM à lire après la séance
	Évaluation	Texte - utiliser les modèles pour guider l'action?
29 novembre 2024	Titre	Modèles d'ensemble
	Contenus	Comparaison des modèles, sélection de modèles,

	Activités	utilisation pour guider la pratique Théorie - bagging, ensembles homogènes et hétérogènes Discussion - identification de nouveaux hôtes de beta-coronavirus
	Lectures et travaux	Un article sur StudiUM à lire avant la séance
	Évaluation	Rédiger sa "persona" (version 3), à faire après la séance, en se mettant dans le rôle de formuler des recommandations aux modélisateurs

6 décembre 2024	Titre	Conclusion et perspectives
	Contenus	Discussion sur l'utilisation de ces méthodes pour la gouvernance globale de la biodiversité; présentation du monitoring framework du KM-GBF; projet BON-in-a-Box pour l'accès équitable à la modélisation
	Activités	Théorie - comment les sciences des données sont utilisées pour guider les politiques en biodiversité Atelier - qui a le droit de faire des prédictions? Discussion - retour sur le cours
	Lectures et travaux	Article sur StudiUM à lire avant la séance
	Évaluation	Texte - la place du machine learning dans le travail des biologistes

Attention ! Exceptionnellement, l'enseignant peut apporter des modifications aux dates des évaluations. Le cas échéant, l'enseignant doit obtenir l'appui de la majorité des étudiants de sa classe. Veuillez vous référer à l'[article 4.8 du Règlement des études de premier cycle](#) et à l'[article 28 du Règlement pédagogique de la Faculté des études supérieures et postdoctorales](#).

Évaluations

Calendrier des évaluations

13 septembre 2024	Activité	Persona (version 1)
	Objectifs d'apprentissage visés	Comprendre son positionnement par rapport aux sciences des données avant le début du cours
	Critères d'évaluation	Succès/échec
	Pondération	5
20 septembre 2024	Activité	Identification des biais dans les données de mpox
	Objectifs d'apprentissage visés	Identifier les forces et faiblesses d'un jeu de données dans un contexte spécifique; formuler des recommandations; argumenter via un texte
	Critères d'évaluation	Grille d'évaluation quantitative avec échelle
	Pondération	5
18 octobre 2024	Activité	Avis sur l'utilisabilité du modèle
	Objectifs d'apprentissage visés	Identifier les forces et faiblesses dans la description d'un modèle; fournir un argumentaire encadrant l'utilisation d'un modèle dans un contexte spécifique
	Critères d'évaluation	Grille d'évaluation quantitative avec échelle
	Pondération	10
1 novembre 2024	Activité	Persona (version 2)

	Objectifs d'apprentissage visés	Se projeter dans un rôle d'utilisateur des résultats de la modélisation
	Critères d'évaluation	Succès/échec
	Pondération	8
8 novembre 2024	Activité	Identification de variables expliquant la distribution d'une espèce
	Objectifs d'apprentissage visés	Faire un lien entre les connaissances de biologiste, et les compétences techniques en modélisation
	Critères d'évaluation	Grille d'évaluation quantitative avec échelle
	Pondération	12
15 novembre 2024	Activité	Est-ce qu'une réponse vrai/faux est toujours souhaitable?
	Objectifs d'apprentissage visés	Comprendre comment représenter l'information d'une prédiction; contextualisation de l'incertitude du modèle
	Critères d'évaluation	Grille d'évaluation quantitative avec échelle
	Pondération	5
22 novembre 2024	Activité	Utiliser les modèles pour guider des actions de conservation
	Objectifs d'apprentissage visés	Se projeter dans un rôle d'utilisateur de modèles, dans un contexte de prise de décision; argumenter en faveur / contre l'utilisation des modèles dans le processus de prise de décision
	Critères d'évaluation	Grille d'évaluation quantitative avec échelle
	Pondération	20
29 novembre 2024	Activité	Persona (version 3)
	Objectifs d'apprentissage visés	Se projeter dans un rôle de commanditaire de modèles prédictifs, qui guide le travail de scientifiques des données
	Critères d'évaluation	Succès/échec
	Pondération	10
6 décembre 2024	Activité	Quelle est la place du machine learning dans la biologie contemporaine?
	Objectifs d'apprentissage visés	Écrire un essai argumenté qui contextualise l'ensemble du matériel du cours dans la pratique de la biologie
	Critères d'évaluation	Grille d'évaluation quantitative avec échelle
	Pondération	30

Attention ! Exceptionnellement, l'enseignant peut apporter des modifications aux dates des évaluations. Le cas échéant, l'enseignant doit obtenir l'appui de la majorité des étudiants de sa classe. Veuillez vous référer à l'[article 4.8 du Règlement des études de premier cycle](#) et à l'[article 28 du Règlement pédagogique de la Faculté des études supérieures et postdoctorales](#).

Consignes et règles pour les évaluations

Absence à une évaluation	Tous les travaux sont remis via StudiUM, il n'y a pas d'examen final (évaluation continue)
Dépôts des travaux	Les travaux sont à remettre sur StudiUM, au format PDF, et doivent obligatoirement comporter le sigle du cours, le matricule, et le nom de l'étudiant-e.

Matériel autorisé

Tout matériel est autorisé, à l'exception du recours à l'IA générative, que ce soit pour générer du texte ou des images. Toute utilisation de l'IA générative, y compris pendant les ateliers en classe, sera considérée comme du plagiat.

Rappels

Utilisation des technologies en classe

Enregistrement des cours

L'enregistrement des cours n'est généralement pas autorisé. Si, pour des raisons valables, vous désirez enregistrer une ou plusieurs séance(s) de cours, vous devez préalablement obtenir l'autorisation écrite de votre enseignant au moyen du formulaire prévu à cet effet [Demande d'autorisation pour l'enregistrement d'un cours](#).

Notez que la permission d'enregistrer NE donne PAS la permission de diffuser l'enregistrement.

Ressources

Ressources obligatoires

Documents

Ouvrages en réserve à la bibliothèque

Soutien à la réussite

De nombreuses activités et ressources sont offertes à l'Université de Montréal pour faire de votre vie étudiante une expérience enrichissante et agréable. La plupart d'entre elles sont gratuites. Explorez les liens ci-dessous pour en savoir plus.

[Centre de communication écrite](#)

[Centre étudiant de soutien à la réussite](#)

[Services des bibliothèques UdeM](#)

[Soutien aux étudiants en situation de handicap](#)

Cadres réglementaires et politiques institutionnelles

Règlements et politiques

Apprenez à connaître les règlements et les politiques qui encadrent la vie universitaire.

Règlement des études

[Règlement des études de premier cycle](#)

Que vous soyez étudiant régulier, étudiant libre ou étudiant visiteur, connaître le règlement qui encadre les études est tout à votre avantage. Consultez-le !

[Règlement pédagogique des études supérieures et postdoctorales](#)

Politique-cadre sur l'intégration des étudiants en situation de handicap

[Politique-cadre sur l'intégration des étudiants en situation de handicap](#)

[Demande d'accommodement et responsabilités](#)

Renseignez-vous sur les ressources disponibles les mieux adaptées à votre situation auprès du Bureau de soutien aux étudiants en situation de handicap (BSESH). Le deuxième lien ci-contre

présente les accommodements
aux examens spécifiques à
chaque faculté ou école

Intégrité, fraude et plagiat

Problèmes liés à la gestion du temps, ignorance des droits d'auteurs, crainte de l'échec, désir d'égaliser les chances de réussite des autres – aucune de ces raisons n'est suffisante pour justifier la fraude ou le plagiat. Qu'il soit pratiqué intentionnellement, par insouciance ou par négligence, le plagiat peut entraîner un échec, la suspension, l'exclusion du programme, voire même un renvoi de l'université. Il peut aussi avoir des conséquences directes sur la vie professionnelle future. Plagier ne vaut donc pas la peine !

Le plagiat ne se limite pas à faire passer un texte d'autrui pour sien. Il existe diverses formes de manquement à l'intégrité, de fraude et de plagiat. En voici quelques exemples :

- Dans les travaux : Copier un texte trouvé sur Internet sans le mettre entre guillemets et sans citer sa source ; Soumettre le même travail dans deux cours (autoplégat) ; Inventer des faits ou des sources d'information ; Obtenir de l'aide non autorisée pour réaliser un travail.
- Durant les évaluations : Utiliser des sources d'information non autorisées ; Obtenir des réponses de façon illicite ; S'identifier faussement comme un étudiant du cours.

[Site Intégrité](#)

[Les règlements expliqués](#)