



Université de
Sherbrooke

Département d'informatique IFT 599 / IFT 799 – Science des données

Plan d'activité pédagogique Automne 2026

Enseignant	Shengrui Wang
Courriel :	shengrui.wang@usherbrooke.ca
Local :	D4-1018-1
Téléphone :	+1 819 821-8000 x62022
Disponibilités :	Disponible les après-midi de mardi et de jeudi pour répondre aux questions en personne.

Site web du cours : MS Teams

Horaire	Exposé magistral :	Mardi	10 h 30 à 12 h 20	salle D7-2023
		Jeudi	10 h 30 à 11 h 20	salle D7-2023

Description officielle de l'activité pédagogique¹

Cibles de formation : Comprendre et maîtriser des théories et méthodes de base pour la science des données.

Contenu : Inférence statistique : procédures statistiques fondamentales, estimation des paramètres d'un modèle, tests des hypothèses liées aux caractéristiques structurelles d'un modèle, intervalle de confiance pour les paramètres de modèle. Techniques de forage de données : analyse exploratoire des données, prétraitement, visualisation, recherche et extraction des règles d'association, classification et prédiction, analyse de regroupement. Recherche d'information : principe, concepts de base, indexation, engins de recherche. Applications dans divers domaines tels que la santé, l'intelligence d'affaires, les réseaux sociaux et la finance.

Crédits	3
Organisation	3.0 heures d'exposé magistral par semaine 6.0 heures de travail personnel par semaine
Préalable	IFT339 et (STT290 ou STT418)
Particularités	Aucune

¹<https://www.usherbrooke.ca/admission/fiches-cours/ift599>

1 Présentation

Cette section présente les cibles de formation spécifiques et le contenu détaillé de l'activité pédagogique. Cette section, non modifiable sans l'approbation du comité de programme du Département d'informatique, constitue la version officielle.

1.1 Mise en contexte

De nos jours, le volume de plus en plus important de données disponibles (que l'on appelle données massives, mégadonnées ou bien « big data ») et leur démocratisation combinée avec les puissances de calcul et de stockage, nous amènent vers une nouvelle ère d'analyse des données qui comporte beaucoup de défis. En effet, les mégadonnées qui ne cessent d'augmenter jour après jour sont souvent peu exploitées, alors qu'elles cachent des connaissances décisives face au marché et à la concurrence. La plupart des techniques d'analyse de données traditionnelles échouent à extraire de l'information pertinente sur ce type de données. Pour combler ce besoin, la science des données s'est rapidement développée ces dernières années. Elle s'intéresse à tous les aspects liés aux données notamment la collecte, le nettoyage, le filtrage, l'anonymisation, le cryptage, la visualisation, le stockage, le transfert, l'analyse et la modélisation (classification et prédiction), etc. L'appellation « science des données » est issue de celle du forage de données (« Data Mining ») qui a été le porteur de la discipline avec l'apprentissage automatique (« machine learning ») depuis une vingtaine d'années. La science des données est donc l'ensemble des algorithmes et méthodes destinés à l'exploration et l'analyse de grandes bases de données informatiques en vue de détecter dans ces données des règles, des associations, des tendances inconnues (non fixées a priori), des structures particulières restituant de façon concise l'essentiel de l'information utile, et de prédire les résultats des processus sous jacents afin de générer des valeurs pour la société.

1.2 Cibles de formation spécifiques

Ce cours vise à initier l'étudiante et l'étudiant aux concepts fondamentaux de sciences de données et d'appliquer ces notions à des problèmes concrets. À la fin de cette activité pédagogique, l'étudiante ou l'étudiant sera capable de :

1. comprendre les principaux concepts de la science des données et le processus d'extraction des connaissances dans les bases de données ;
2. maîtriser des concepts et techniques d'inférence statistique ;
3. maîtriser les techniques d'analyse des données transactionnelles, plus particulièrement la recherche des règles d'association ;
4. maîtriser des techniques d'agrégation et de classification les plus utilisées dans la pratique ;
5. maîtriser des techniques de prétraitement des données et de réduction de la dimension (essentiellement la gestion de la redondance et la sélection des attributs) ;
6. comprendre la problématique des données de grande dimension et maîtriser des techniques d'agrégation utilisées sur ce type de données ;
7. se familiariser avec des algorithmes de classement des pages Web tels que HITS et PageRank ;
8. se familiariser avec des algorithmes pour les systèmes de recommandation.
9. comprendre le lien qu'il y a entre la sciences des données et divers applications en lien avec l'informatique de gestion.

1.3 Contenu détaillé

Thème	Contenu	Nbr. d'heures	Objectifs	Lectures ¹
1	Concepts de base : processus d'extraction des connaissances dans les bases de données; prétraitement des données; mesures de similarités; techniques de réduction de la dimension (ACP : analyse en composantes principales); techniques de sélection d'attributs; concepts de base pour la détection des anomalies; visualisation; Introduction de Weka (progiciel d'analyse des données).	8	1 et 5	[2] [3] chapitres 1 et 2
2	Analyse d'association et analyse des séquences : introduction aux données transactionnelles; principe de l'algorithme <i>A priori</i> et, si le temps permet, l'algorithme <i>FP-growth</i> ; recherche et extraction des règles d'associations; extraction et utilisation des patrons significatifs pour l'analyse des séquences.	8	3	[2] [3] chapitres 5 et 6
3	Inférence statistique : procédures statistiques fondamentales; estimation des paramètres d'un modèle; tests des hypothèses liées aux caractéristiques structurelles d'un modèle; Intervalle de confiance pour les paramètres de modèle.	4	1 et 2	[2]
4	Agrégation (méthodes de base) : partition et hiérarchie; agrégation basée sur la densité; données catégoriques et transactionnelles; Validation des résultats.	6	4 et 5	[2] [3] chapitre 7
5	Agrégation (méthodes avancées) : détection des anomalies; agrégation des données de grandes dimensions et des données complexes.	6	4, 5 et 6	[2] [3] chapitre 8 [4]
6	Systèmes de recommandation : principes et éléments de base; principe des algorithmes basés sur le contenu; approches basées sur le filtrage collaboratif; méthodes basées sur la ressemblance directe; méthodes basées sur la sémantique latente.	6	5 et 8	[2] [1]
7	Exploration du Web : principes et éléments de bases; algorithmes de classement des pages Web (HITS et <i>PageRank</i>); extraction des communautés dans le Web.	2	7	[2]

¹ Les lectures indiquées ne sont là qu'à titre indicatif. L'enseignant est libre de choisir un autre document de référence

2 Organisation

Cette section propre à l'approche pédagogique de chaque enseignante ou enseignant présente la méthode pédagogique, le calendrier, le barème et la procédure d'évaluation ainsi que l'échéancier des travaux. Cette section doit être cohérente avec le contenu de la section précédente.

2.1 Méthode pédagogique

La présentation de cours sera offerte de façon magistrale. Il y aura trois TPs. Des séances d'exercices et de révision seront intégrées dans le déroulement du cours. Les présentations PowerPoint, suffisamment détaillées, seront fournies aux étudiants au fur et à mesure de l'avancement du cours par documents partagés dans le Teams du cours.

2.2 Calendrier

Semaine	Commençant le	Thème	Travaux pratiques
1	2026-08-31	1	
2	2026-09-07	1	
3	2026-09-14	1 et 3	
4	2026-09-21	3	
5	2026-09-28	3 et 4	
6	2026-10-05	4	Remise TP1
7	2026-10-12	Révision et 4	
8	2026-10-19	Semaine des examens périodiques	
9	2026-10-26	Relâche	
10	2026-11-02	5	
11	2026-11-09	5	
12	2026-11-16	6	
13	2026-11-23	6	Remise TP2
14	2026-11-30	7	
15	2026-12-07	Révision et 2	Remise TP3
16	2026-12-14	Semaine des examens finals	
17	2026-12-21	Semaine des examens finals	

2.3 Évaluation

Les évaluations incluent trois TP (30 %), l'intra (30 %) et le final (40 %). Pour chaque TP, la remise inclut les programmes, les données et un rapport technique. Aucun retard n'est permis. Les TP doivent être faits par équipe de deux (2) ou trois (3) personnes.

2.3.1 Qualité de la langue et de la présentation

Conformément à l'article 17 du Règlement facultaire d'évaluations des apprentissages² l'enseignante ou l'enseignant peut retourner à l'étudiante ou à l'étudiant tout travail non conforme aux exigences quant à la qualité de la langue et aux normes de présentation.

2.3.2 Plagiat

Le plagiat consiste à utiliser des résultats obtenus par d'autres personnes afin de les faire passer pour sien et dans le dessein de tromper l'enseignante ou l'enseignant. Vous trouverez en annexe un document d'information relatif à l'intégrité intellectuelle qui fait état de l'article 9.4.1 du Règlement des études³. Lors de la correction de tout travail individuel ou de groupe une attention spéciale sera portée au plagiat. Si une preuve de plagiat est attestée, elle sera traitée en conformité, entre autres, avec l'article 9.4.1 du Règlement des études de l'Université de Sherbrooke. L'étudiante ou l'étudiant peut s'exposer à de graves sanctions qui peuvent être soit l'attribution de la note E ou de la note zéro (0) pour un travail, un examen ou une activité évaluée, soit de reprendre un travail, un examen ou une activité pédagogique. Tout travail suspecté de plagiat sera transmis au Secrétaire de la Faculté des sciences. Ceci n'indique pas que vous n'avez pas le droit de coopérer entre deux équipes, tant que la rédaction finale des documents et la création du programme restent le fait de votre équipe. En cas de doute de plagiat, l'enseignante ou l'enseignant peut demander à l'équipe d'expliquer les notions ou le fonctionnement du code qu'elle ou qu'il considère comme étant plagié. En cas d'incertitude, ne pas hésiter à demander conseil et assistance à l'enseignante ou l'enseignant afin d'éviter toute situation délicate par la suite.

2.4 Échéancier des travaux

Les dates de remise des travaux seront indiquées sur les énoncés.

2.5 Utilisation d'appareils électroniques et du courriel

Selon le règlement complémentaire des études, section 4.2.3⁴, l'utilisation d'ordinateurs, de cellulaires ou de tablettes pendant une prestation est interdite à condition que leur usage soit explicitement permise dans le plan de cours.

Dans ce cours, l'usage de téléphones cellulaires, de tablettes ou d'ordinateurs est autorisé. Cette permission peut être retirée en tout temps si leur usage entraîne des abus.

Tel qu'indiqué dans le règlement universitaire des études, section 4.2.3⁵, toute utilisation d'appareils de captation de la voix ou de l'image exige la permission de la personne enseignante.

Note : Je réponds aux questions posées par courriel à l'extérieur des périodes de cours.

Je ne suis pas rapide à répondre aux courriels.

²https://www.usherbrooke.ca/sciences/fileadmin/sites/sciences/documents/Etudiants_actuels/Informations_academiques_et_reglements/2017-10-27_Reglement_facultaire_-_evaluation_des_apprentissages.pdf

³<https://www.usherbrooke.ca/registraire/droits-et-responsabilites/reglement-des-etudes/>

⁴https://www.usherbrooke.ca/sciences/fileadmin/sites/sciences/documents/Etudiants_actuels/Informations_academiques_et_reglements/Sciences_Reglement_complementaire.pdf

⁵<https://www.usherbrooke.ca/registraire/droits-et-responsabilites/reglement-des-etudes/>

3 Matériel nécessaire pour l'activité pédagogique

Diapositifs du cours.

4 Références

- [1] DESROSIER, CHRISTIAN AND KARYPIS, GEORGE : *A Comprehensive Survey of Neighborhood-based Recommendation Methods*, pages 107–144. Springer US, Boston, MA, 2011.
- [2] SHENGRUI WANG : Acétates du cours et des documents supplémentaires .
- [3] TAN, PANG-NING AND STEINBACH, MICHAEL AND KUMAR, VIPIN : *Introduction to Data Mining, (Second Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2018. Ce livre contient plus que 65% de la matière du cours (thèmes : 1, 2, 3 et 4). Voir le site web du livre <http://www-users.cs.umn.edu/~kumar/dmbook/index.php>.
- [4] WU, SHU AND WANG, SHENGRUI : Information-theoretic outlier detection for large-scale categorical data. *IEEE transactions on knowledge and data engineering*, 25(3) :589–602, 2013.

Délits relatifs aux études

Extrait du règlement des études (Règlement 2575-009)

Sont notamment considérés comme un délit relatif aux études les faits suivants :

- a) commettre un plagiat, soit faire passer ou tenter de faire passer pour sien, dans une production évaluée, le travail d'une autre personne, des passages ou idées tirés de l'œuvre d'autrui ou du contenu, de toute forme, généré par un système d'intelligence artificielle (ce qui inclut notamment le fait de ne pas indiquer la source et la référence adéquate);
- b) commettre un autoplagiat, soit soumettre, sans autorisation préalable, une même production, en tout ou en partie, à plus d'une activité pédagogique ou dans une même activité pédagogique (notamment en cas de reprise);
- c) usurper l'identité d'une autre personne ou procéder à une substitution de personne lors d'une production évaluée ou de toute autre prestation obligatoire;
- d) fournir ou obtenir toute forme d'aide non autorisée, qu'elle soit collective ou individuelle (incluant l'assistance provenant d'un système d'intelligence artificielle), pour une production faisant l'objet d'une évaluation;
- e) obtenir par vol ou toute autre manœuvre frauduleuse, posséder ou utiliser du matériel non autorisé de toute forme (incluant le matériel numérique et celui généré par un système d'intelligence artificielle) avant ou pendant une production faisant l'objet d'une évaluation;
- f) copier, contrefaire ou falsifier un document pour l'évaluation d'une activité pédagogique;
- k) posséder ou avoir à sa portée un appareil électronique ou numérique interdit durant une activité d'évaluation;

[...]

Un [guide sur l'intégrité intellectuelle](#) vous est rendu disponible par le service des bibliothèques et des archives de l'Université de Sherbrooke, afin de bien comprendre les différents délits et ainsi éviter d'être aux prises avec un dossier disciplinaire et une ou des sanctions.

Les mesures pouvant être imposées à titre de sanctions disciplinaires sont les suivantes :

- a) la réprimande simple ou sévère consignée au dossier étudiant pour la période fixée par l'autorité disciplinaire ou à défaut, définitivement. En cas de réprimande fixée pour une période déterminée, la décision rendue demeure au dossier de la personne aux seules fins d'attester de l'existence du délit en cas de récidive;
- b) l'obligation de reprendre une production ou une activité pédagogique, dont la note pourra être établie en tenant compte du délit survenu antérieurement;
- c) la diminution de la note ou l'attribution de la note E ou 0;

[...]

Balises d'utilisation des outils d'intelligence artificielle générative

Autorisés ou pas dans les situations d'apprentissage et d'évaluation ?

NIVEAU 0

NIVEAU 1

NIVEAU 2

NIVEAU 3

NIVEAU 4

L'utilisation des outils d'intelligence artificielle générative (IAg) est limitée, voire complètement interdite parce que la personne enseignante considère que l'usage de ces outils nuit au développement de compétences essentielles. Ces compétences peuvent être disciplinaires, comme elles peuvent être d'ordre méthodologique, rédactionnel ou informationnel. Considérant que l'utilisation des IAg requiert un esprit critique, il peut s'agir d'une situation d'apprentissage ou d'évaluation sans IAg qui vise à développer celui-ci.

Dans ces situations, **la personne étudiante produit le travail.**

L'utilisation prononcée des IAg est permise parce que la personne enseignante considère que les personnes étudiantes sont en mesure d'exercer un esprit critique et sont capables de juger de la qualité des contenus produits par les IAg. Ou encore, l'utilisation est encouragée parce que la situation d'apprentissage ou d'évaluation proposée contribue à développer leur esprit critique.

Dans ces situations, l'IAg produit le travail préliminaire, alors que **la personne étudiante s'assure de sa qualité en l'améliorant.**



Utilisation interdite

Le **NIVEAU 0** signifie que l'**utilisation est interdite**.

Ceci signifie que si la personne enseignante a un motif de croire qu'il y a eu l'utilisation d'une IAg dans une situation d'évaluation, elle doit dénoncer les faits auprès de la personne responsable des dossiers disciplinaires universitaires. Il s'agit d'un délit relatif aux études tel que stipulé dans le [Règlement des études](#).



Utilisation limitée

Le **NIVEAU 1 D'UTILISATION** signifie que l'**utilisation est autorisée uniquement pour assister l'apprentissage dans le domaine disciplinaire ou des langues**.

Dans ce contexte, la personne étudiante **est tenue de déclarer l'utilisation qu'elle en a faite** selon les consignes fournies par la personne enseignante sans quoi l'utilisation peut être considérée comme un délit. Par exemple :

Domaine disciplinaire :

- S'inspirer
- Générer des idées
- Explorer un sujet pour mieux le comprendre
- Générer du matériel pour apprendre

Domaine des langues :

- Identifier ses erreurs et se les faire expliquer
- Reformuler un texte
- Générer un plan pour aider à structurer un texte
- Traduire un texte



Utilisation guidée

Le **NIVEAU 2 D'UTILISATION** signifie que l'**utilisation est autorisée pour améliorer un travail produit par la personne étudiante**.

Dans ce contexte, la personne étudiante **est tenue de déclarer l'utilisation qu'elle en a faite** selon les consignes fournies par la personne enseignante sans quoi l'utilisation est considérée comme un délit. Par exemple :

- Analyser des contenus
- Obtenir une rétroaction
- Évaluer la qualité de son travail à partir de critères
- Demander à être confronté relativement à ses idées, à sa démarche
- Diriger les processus de résolution de problèmes



Utilisation balisée

Le **NIVEAU 3 D'UTILISATION** signifie que l'**utilisation est autorisée pour produire un travail qui sera amélioré**.

Dans ce contexte, la personne étudiante **est tenue de citer selon les normes¹ le contenu généré par l'IAg ou de déclarer l'utilisation qu'elle en a faite** selon les consignes fournies par la personne enseignante sans quoi l'utilisation est considérée comme un délit. Par exemple :

- Résumer ou rédiger des parties d'un texte
- Générer un texte ou un modèle d'une production et l'adapter
- Réaliser des calculs mathématiques
- Produire du code informatique
- Résoudre des problèmes complexes
- Répondre à une question
- Générer des images, ou autres contenus multimédias



Utilisation libre

Le **NIVEAU 4 D'UTILISATION** signifie qu'**aucune restriction spécifique n'est imposée**.

Dans ce contexte, la personne étudiante **est tenue de citer selon les normes¹ le contenu généré par l'IAg ou de déclarer l'utilisation qu'elle en a faite** selon les consignes fournies par la personne enseignante sans quoi l'utilisation est considérée comme un délit.

Ce niveau inclut tout ce qui précède, de l'exploration à la production, ainsi que toute autre tâche particulière jugée complexe.

À considérer avant l'utilisation d'outils d'intelligence artificielles génératives

Si, en tant que personne étudiante envisagez d'utiliser un outil d'intelligence artificielle générative (IAG) lorsque l'évaluation autorise les niveaux 1 à 4 d'utilisation mentionnés précédemment.

Dans ce cas, gardez à l'esprit les éléments clés suivants.

- Vous assumez la responsabilité de tout le contenu produit, avec ou sans IAG, et intégré à votre production.
- Les produits des outils d'IAG peuvent très souvent comporter **des erreurs ou des faussetés** (hallucinations) : on doit donc impérativement valider tout contenu généré par ces outils.
- Dans l'état actuel de la Loi sur le droit d'auteur du Canada, les **productions faites par l'IAG sont du domaine public**, puisque les outils d'IAG ne sont pas reconnus comme des auteurs au sens de la Loi et que les contenus générés ne répondent pas aux critères d'une œuvre protégée, notamment aux critères d'originalité.
- L'entreprise qui fournit le service pourrait émettre certaines exigences dans ses conditions d'utilisation. Comme l'algorithme et le code informatique appartiennent à l'entreprise qui les a développés, nous devons tenir compte de ces conditions. Celles-ci pourraient également fournir des précisions relatives à la **réutilisation des données soumises (confidentialité)**.

Comment déclarer l'utilisation d'outils d'intelligence artificielle générative

Dans l'esprit d'une conduite intègre et responsable, vous devez TOUJOURS mentionner de façon explicite toute utilisation de l'intelligence artificielle, conformément au Règlement des études (9.4.1 Délits relatifs aux études). De plus, à des fins pédagogiques, il est recommandé de toujours intégrer à la production les requêtes, de même que les réponses intégrales générées par les outils d'IAG. Celles-ci pourront être intégrées directement dans le corps du texte ou en note de bas de page. Les réponses longues pourraient être insérées en annexe de votre document ou dans des documents supplémentaires, selon les directives de la personne enseignante.

L'utilisation de ces deux documents s'avèrera utile, ils se trouvent sous licence libre, donc vous pouvez utiliser les tableaux et les adapter selon votre besoin:

1. Modèle de citation : Ce formulaire, à remplir par l'enseignant, donne un exemple aux étudiants de citation de l'IAG dans la réalisation d'un travail évalué ou non.
2. Déclaration d'usage : Ce formulaire, à remplir par les étudiants, doit être remis avec une réalisation afin de déclarer l'usage de l'IAG dans la réalisation, qu'elle soit évaluée ou non.

Référence

La Faculté des sciences tient à remercier le SSF pour la production des documents.

- Cabana, M. et Côté, J.-A. (2024). Balises d'utilisation des outils d'intelligence artificielle générative. Service de soutien à la formation, Université de Sherbrooke. Sous licence [CC BY 4.0](#).
- Cabana, M. et Beaudet, M. (2024). Directives de déclaration de l'utilisation de l'intelligence artificielle générative dans une production étudiante. Service de soutien à la formation, Université de Sherbrooke. Sous licence [CC BY 4.0](#).
- Cabana, M. (2024). Formulaire de déclaration de l'utilisation de l'intelligence artificielle générative dans une production étudiante. Service de soutien à la formation, Université de Sherbrooke. Sous licence [CC BY 4.0](#).