

Université de
Sherbrooke

Département d'informatique
IFT 870 / BIN 710 – Forage de données

Plan d'activité pédagogique
Hiver 2023

Enseignant

Aïda Ouangraoua

Courriel : Aida.Ouangraoua@USherbrooke.ca

Local : D4-1010-7

Téléphone : +1 819 821-8000 x62014

Disponibilités : Le vendredi de 13 h 30 à 16 h

Responsable(s) : Direction du département

Site web du cours : <https://cobius.usherbrooke.ca/ift870/>

Horaire

Exposé magistral :	Lundi	9h30 à 10h20	salle D3-2040
	Jeudi	8h30 à 10h20	salle D7-2021

Description officielle de l'activité pédagogique¹

Cibles de formation :	Apprendre les concepts et les techniques de base pour l'entrepôt et le forage de données. Comprendre le rôle du forage dans la réalisation d'un entrepôt de données. Acquérir des habiletés pour développer des systèmes de forage. Appliquer ces concepts et techniques pour la résolution de problèmes.
Contenu :	Entrepôts de données et technologie pour le forage : caractéristiques et architecture, métadonnées, modèles multidimensionnels, démarche de construction d'un entrepôt. Processus de forage. Techniques de forage : caractérisation du forage descriptif versus prédictif, prétraitement de données, recherche et extraction des règles d'association, méthodes pour la classification et la prédiction, analyse de faisceaux, évaluation de résultats.
Crédits	3
Organisation	3 heures d'exposé magistral par semaine 6 heures de travail personnel par semaine
Particularités	Aucune

¹<https://www.usherbrooke.ca/admission/fiches-cours/ift870>

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

Le cours IFT870 - Forage de données (BIN 710 : pour la bio-informatique) est un cours à option du 2e et du 3e cycle en informatique. Nous y introduisons les principaux aspects du forage de données et nous nous focalisons sur l'utilisation du langage Python et de la bibliothèque scikit-learn pour le développement d'applications de forage de données. Il ne présuppose aucune connaissance préalable en forage de données ou intelligence artificielle. Les principales méthodes et les aspects pratiques du forage de données seront introduites au fur et à mesure. Des bases en programmation avec le langage Python et les bibliothèques Numpy et Matplotlib sont souhaitables.

1.2 Cibles de formation spécifiques

Ce cours vise des étudiantes et des étudiants qui souhaitent découvrir les principales méthodes de forage de données et approfondir leurs connaissances des aspects pratiques pour le développement d'applications dédiées à des problèmes réels de forage de données. Il ne se focalise pas sur les concepts mathématiques fondamentaux du forage de données, mais plutôt sur les aspects pratiques pour l'utilisation d'algorithmes avec le langage Python et la bibliothèque scikit-learn. À la fin de cette activité pédagogique, l'étudiante ou l'étudiant sera capable :

1. De distinguer et comprendre les algorithmes de forage de données les plus utilisés en pratique ;
2. De maîtriser les étapes de l'implémentation de solutions pour des problèmes de forage de données ;
3. De maîtriser l'utilisation de Python, scikit-learn, pandas, matplotlib et les bibliothèques scientifique de Python pour la mise en oeuvre de solutions.

1.3 Contenu détaillé

Thème	Contenu	Nbr. d'heures	Objectifs	Travaux	Lectures
1	Introduction : Concepts de base, types de données, classification générale des problèmes et algorithmes usuels de forage de données.	6	1		Chap. 1 de [2] , [1] et [3]
2	Exploration des données : Types de données et d'attributs, descriptions statistiques, mesures de similarité et dissimilarité, visualisation.	6	2, 3	✓	Chap. 2 de [2] et [3]
3	Représentation et prétraitement des données : Nettoyage, intégration, réduction, transformation et discrétisation des données.	6	2, 3	✓	Chap. 3 de [2] et [3] , et 4 de [1]
4	Fonctions prédictives : Classification, régression, algorithmes usuels, évaluation, sélection et amélioration de modèles.	6	1, 2, 3	✓	Chap. 8,9 de [2] , 2,5 de [1] et 4,5 de [3]
5	Fonctions descriptives : Association, corrélations, clustering, détection d'anomalies, comparaison et évaluation de clusterings.	6	1, 2, 3	✓	Chap 10-12 de [2] , 3 de [1] , et 8,10 de [3]
6	Forage de données complexes : Données séquentielles, graphes et réseaux, applications du forage de données.	6	1, 2		Chap. 13 de [2] , et 7 de [1]

1. Le cours doit comprendre au moins quatre travaux pratiques couvrant tous les sujets marqués «✓» dans le tableau.
2. 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

Le cours sera donné sous forme d'exposés magistraux. Les exposés magistraux auront pour objectif d'introduire les concepts de base, les algorithmes et les aspects pratiques correspondant avec Python et scikit-learn pour chaque thème, ainsi que faire des exercices. Le cours comporte quatre travaux pratiques. Le but de ces travaux est d'implémenter les concepts et techniques vus en cours.

Compte tenu du contexte actuel (pandémie due au COVID-19), il se peut que le cours ait lieu en totalité ou en partie à distance d'une façon différente de ce qui est énoncé ci-dessus. Notez que vous en serez informés rapidement si tel est le cas.

2.2 Calendrier

Semaine	Date	Thème
1	2023-01-02	1
2	2023-01-09	1
3	2023-01-16	2
4	2023-01-23	2
5	2023-01-30	3
6	2023-02-06	3
7	2023-02-13	4
8	2023-02-20	Examen périodique
9	2023-02-27	Relâche
10	2023-03-06	4
11	2023-03-13	5
12	2023-03-20	5
13	2023-03-27	6
14	2023-04-03	6
15	2023-04-10	Révision
16	2023-04-17	
17	2023-04-24	Examen final

2.3 Évaluation

Devoir	30 %
Examen intra	30 %
Examen final	40 %

2.3.1 Qualité de la langue et de la présentation

Conformément à l'article 17 du règlement facultaire d'évaluation 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.

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

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 autorisées. 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 ne réponds à aucun courriel.

L'utilisation de la période de disponibilité du vendredi après-midi est recommandée pour poser vos questions en personne. L'utilisation de Teams est recommandée pour poser vos questions en tout temps. Les réponses seront envoyées dans les 48 h. Sauf pour les demandes personnelles, les réponses seront envoyées à toute la classe.

3 Matériel nécessaire pour l'activité pédagogique

La matière du cours provient essentiellement des manuels listés dans la section des références. Le Web est également une source de références dans le domaine. Les acétates du cours seront disponibles sur Teams et sur le site Web du cours. Les travaux pratiques seront réalisés dans le langage Python avec les bibliothèques et outils : Numpy, Scipy, pandas, matplotlib, et Jupyter Notebook. Les devoirs du cours doivent être remis sous forme de fichiers notebook .ipynb interprétables à l'aide de Jupyter Notebook.

4 Références

- [1] ADREAS C. MÜLLER ET SARAH GUIDO : *Introduction to Machine Learning with Python*. O'Reilly Media Inc, 2017.
- [2] JIAWEI HAN, MICHELINE KAMBER ET JIAN PEI : *Data Mining : Concepts and Techniques*. Morgan Kaufmann Publishers Inc, 3rd édition, 2011.
- [3] PANG-NING TAN, MICHAEL STEINBACH ET VIPIN KUMAR : *Introduction to Data Mining*. Addison-Wesley Longman Publishing Co., Inc, 2018.
- [4] PEDREGOSA, FABIAN ET AL. : Scikit-learn : Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 01 2012.

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

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

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

L'intégrité intellectuelle passe, notamment, par la reconnaissance des sources utilisées. À l'Université de Sherbrooke, on y veille!

Extrait du Règlement des études (Règlement 2575-009)

9.4.1 DÉLITS RELATIFS AUX ÉTUDES

Un délit relatif aux études désigne tout acte trompeur ou toute tentative de commettre un tel acte, quant au rendement scolaire ou une exigence relative à une activité pédagogique, à un programme ou à un parcours libre.

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 ou des passages ou des idées tirés de l'œuvre d'autrui (ce qui inclut notamment le fait de ne pas indiquer la source d'une production, d'un passage ou d'une idée tirée de l'œuvre d'autrui);
 - 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 aide non autorisée, qu'elle soit collective ou individuelle, pour une production faisant l'objet d'une évaluation;
 - e) obtenir par vol ou toute autre manœuvre frauduleuse, posséder ou utiliser du matériel de toute forme (incluant le numérique) non autorisé 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;
- [...]

Par plagiat, on entend notamment :

- Copier intégralement une phrase ou un passage d'un livre, d'un article de journal ou de revue, d'une page Web ou de tout autre document en omettant d'en mentionner la source ou de le mettre entre guillemets;
- reproduire des présentations, des dessins, des photographies, des graphiques, des données... sans en préciser la provenance et, dans certains cas, sans en avoir obtenu la permission de reproduire;
- utiliser, en tout ou en partie, du matériel sonore, graphique ou visuel, des pages Internet, du code de programme informatique ou des éléments de logiciel, des données ou résultats d'expérimentation ou toute autre information en provenance d'autrui en le faisant passer pour sien ou sans en citer les sources;
- résumer ou paraphraser l'idée d'un auteur sans en indiquer la source;
- traduire en partie ou en totalité un texte en omettant d'en mentionner la source ou de le mettre entre guillemets ;
- utiliser le travail d'un autre et le présenter comme sien (et ce, même si cette personne a donné son accord);
- acheter un travail sur le Web ou ailleurs et le faire passer pour sien;
- utiliser sans autorisation le même travail pour deux activités différentes (autoplagiat).

Autrement dit : mentionnez vos sources
